Journal of Safety Research 82 (2022) 38–47

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Adapting, adopting, and advancing change: A framework for future research in the psychology of occupational safety

Steve Granger [*,1], Nick Turner [2]

Haskayne School of Business, University of Calgary, Canada

A R T I C L E I N F O

Article history:
Received 7 December 2020
Received in revised form 29 October 2021
Accepted 19 April 2022
Available online 28 April 2022

Keywords:
Appraisals
Occupational safety
Safety climate
Safety performance
Stress

A B S T R A C T

Introduction: While there are numerous reviews of the research on the psychology of occupational safety, these studies provide weak guidance on where the research should go next. Accordingly, we introduce a simple framework for thinking about future research in this area: the *adapting, adopting, and advancing change* framework. This framework summarizes how external, technological, and theoretical developments have driven research in the psychology of occupational safety and uses these observations as evidence to imagine ways in which they may continue to do so. *Method:* We critically reviewed seminal research in the psychology of occupational safety using the *adapting, adopting, and advancing change* framework. Adapting to change means considering external changes such as the fluctuating nature of work and the labor market. Adopting change refers to incorporating the latest technological and technical advances to facilitate more robust research methods and analyses. Finally, advancing change refers to theoretical advances and how they will push psychology of occupational safety research forward. *Results:* We highlight several avenues for future research that emerge at the convergence of the framework's three themes, including developing the safety skill construct, assessing variation in demand appraisals on safety outcomes, distinguishing safety climate from related constructs, and examining safety constructs that are usually considered as outcomes (e.g., injuries) as predictors instead. *Conclusions:* In doing so, we provide a clear structure to help researchers better identify the most effective directions for future research on the psychology of occupational safety.

1. Introduction

Overall, workplaces in developed nations have become remarkably safer over the past several decades as indicated by declining occupational injury and fatality rates (Association of Workers' Compensation Boards of Canada, 2020; Bureau of Labor Statistics, 2020; Health and Safety Executive, 2020). Some posit that physically dangerous jobs have become rarer and safer in themselves (e.g., Aldrich, 1997; Fishback & Kantor, 2000), while others argue that improved efforts to measure work injuries and fatalities systematically have enabled better safety management (e.g., Pfeffer, 2018). Despite these improvements, progress in safety still evades large segments of the working population (e.g., gig workers; Tran & Sokas, 2017) and has yet to realize the potential of contemporary methods and statistics in providing more nuanced insights into why work and workplaces have become safer in developed nations (e.g., the role of change and variation over time; Gabriel et al., 2019). Further, psychological theory stemming from studying behaviors in other domains (e.g., appraisal and dual process models; Ellsworth, 2013; Evans, 2008) offers many untapped opportunities to advance psychology of occupational safety research.

Having a framework to organize and imagine such issues will yield valuable insights into the field of occupational safety research, guiding scholars toward the most effective and promising avenues of study. Therefore, the key contribution of this paper is to provide an organizing framework for charting trends in the psychology of occupational safety research and structuring where the field may be headed. Many excellent conceptual reviews (e.g., Barling & Frone, 2004; Beus, McCord, & Zohar, 2016; Granger, Turner, & Grocutt, 2021; Hofmann, Burke, & Zohar, 2017; Tetrick, 2017) and empirical reviews (e.g., Burke et al., 2011; Christian, Bradley, Wallace, & Burke, 2009; Cornelissen, Van Hoof, & De Jong, 2017; Feltner et al., 2016; Nahrgang, Morgeson, & Hofmann, 2011) have summarized the field of psychology of occupational safety, but they typically consider future directions as an

* Corresponding author at: Haskayne School of Business, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada.

E-mail addresses: nicholas.turner@ucalgary.ca, steven.granger@ucalgary.ca (S. Granger).

[1] Steve Granger, https://orcid.org/0000-0003-2924-356X.
[2] Nick Turner, https://orcid.org/0000-0001-8369-931X.

https://doi.org/10.1016/j.jsr.2022.04.004
0022-4375/

afterthought, if at all. In contrast, this paper brings possible future avenues to the fore by organizing important trends and highlighting gaps in knowledge. More generally, we suggest that our framework can be applied to practically any area of social scientific research in that we offer explicit guidance for organizing future research along the lines of adapting, adopting, and advancing change.

These three central themes of our framework—adapting to change, adopting change, and advancing change (see Fig. 1)—are applied as follows. First, we contend that psychology of occupational safety research must continue to *adapt* to external changes (e.g., shifts in the nature of work, work environments, and worker demographics) to ensure that the evidence coming from and feeding the research accurately reflects the lived experiences of working populations. Thus, adaptation involves changes in what, who, when, and where researchers focus their attention. Second, *adopting* the latest technology and technical advances in occupational safety research has implications for methodology, research design, and data analysis. This adoption may shift how we measure what we measure, what kind of analyses we can conduct, and, in turn, what inferences we can draw from the data. Third, we argue that safety research must *advance* better explanations, which involves changes in theorizing. This advancement has the potential to alter the types of questions we ask (Kuhn, 1977)—questions about why constructs of interest are conceptualized the way they are and how those constructs tie together—which will subsequently influence the adaptations and adoptions outlined in the first two themes.

We ground our framework in a critical review of the existing research on the psychology of occupational safety by first tracing the field's history and organizing these developments through the lens of the framework's three themes to highlight potential empirical, methodological, and theoretical challenges. We then outline specific future directions that sit at the convergence of the three themes. Ultimately, this framework has the potential to enrich and reshape the lenses through which psychology of occupational safety is researched and practiced.

2. The past, present, and future of psychology of occupational safety research

2.1. Adapting to change

The psychology of occupational safety has been a topic of study since the late 19th and early 20th centuries, when data on accidents and injuries became more easily available to researchers (for examples of this early work, see Greenwood & Woods, 1919; Hoffman, 1909; Myers, 1915). However, occupational safety was generally considered as peripheral to the central topics of productivity and efficiency in early organizational sciences (e.g., efficiency and fatigue; Goldmark, 1912; Münsterberg, 1913). It was not until the late 1920s and early 1930s that safety became a central topic per se (Brakeman & Slocombe, 1929). The interest in occupational safety, both in the public and among researchers, continued steadily throughout the mid-20th century as more people joined the workforce and measurement of accidents and injuries accumulated to the point that the rising costs of poor safety could no longer be ignored (Swuste, van Gulijk, & Zwaard, 2010).

The mid- and late 20th century saw a remarkable diversification in the type of work that people did (Quinlan, 1999), as well as changes in worker demographics (e.g., diversity in age, gender, and ethnic background of the labor force; Toossi, 2002). As this diversification continued, so did the diversity of occupational safety research. While most of the early research on occupational safety was located within specific organizations or industries (Brundage, 1927), later efforts sought to generalize findings beyond the study context (e.g., Kerr, 1957) and to test frameworks across diverse occupations and industries (e.g., Zohar, 1980).

Contextualizing research within a particular occupation or industry allows researchers to enhance the degree to which measures accurately capture phenomena in a specific workplace (i.e., construct validity; Shadish, Cook, & Campbell, 2002) and expands our understanding of specific phenomena (e.g., trends and reasons for injuries among Alaskan loggers; Springer, Lucas, Castrodale, & McLaughlin, 2018). At the same time, adopting a wider approach by testing frameworks across organizations and industries helps to assess the generalizability of constructs and theories (Shadish et al., 2002). The concept of safety climate—shared perceptions and expectations about the relative importance of safety in the workplace—is a prime example of a highly studied topic in the psychology of occupational safety that has been adapted to specific occupations or industries and has also been examined in a generalized way across occupations and industries (e.g., lone worker and aviation safety climate vs. general safety climate; Huang et al., 2013; O'Connor, O'Dea, Kennedy, & Buttrey, 2011; Zohar, 1980, 2010). As a result, our understanding of the nuances of safety climate, such as its key elements, antecedents, and outcomes, has grown substantially (e.g., Beus, Payne, Arthur, & Muñoz, 2019; Jiang, Lavaysse, & Probst, 2019).

Occupational safety research should continue this beneficial approach of simultaneous specificity and generalizability. However, we argue that to ensure the most effective application of the empirical evidence, researchers conducting systematic and meta-analytic reviews on occupational safety must incorporate these types of distinctions more clearly. While meta-analyses are increasingly possible and valuable in occupational safety research, there are potential issues involved in broadly integrating all existing research without carefully considering the contexts in which it was conducted. Inappropriately translating evidence from one context to another can yield results that range from innocuous or unsuitable to seriously harmful in their consequences for decisions regarding occupational safety.

While current occupational safety research has generally adapted well to external changes in the nature of work, numerous potential challenges remain. Certain external changes, such as the growing trends toward temporary, contract, and gig work (Howard, 2017; Quinlan, Mayhew, & Bohle, 2001), are proving particularly important for safety in a number of ways. First, temporary work typically refers to short-term employment, and research shows that workers are most at risk of injury within their first year of employment, especially the first three months (Burt, 2016; Van Zelst, 1954). Second, temporary work often involves little training (Hopkins, 2017), which is a robust predictor of safety (Burke et al., 2011). Third, job security ranges from limited to non-existent in temporary work and is another important predictor of work-related injuries (Probst & Brubaker, 2001). Fourth, the highly contingent and precarious nature of temporary work may motivate employees to underreport injuries, based on the fear that doing so would reduce their chances of being rehired or gaining permanent status (Collinson, 1999). Fifth, contingent and precarious work is typically undertaken by vulnerable populations, including those who are younger and older, less educated, and of racial or ethnic minority (Herbert & Landrigan, 2000; Hopkins, 2017; Howard, 2017).

Relatedly, another adaptation gap in the occupational safety research is a significant shift in worker demographics. For instance, people in developed countries are now retiring later than they used to (Rudolph, Marcus, & Zacher, 2018). While continued employment can have many benefits (e.g., contributions to the economy and psychological well-being; Herzog, House, & Morgan, 1991; Kowalski-Trakofler, Steiner, & Schwerha, 2005), older populations are particularly at risk for severe injuries, generally require more

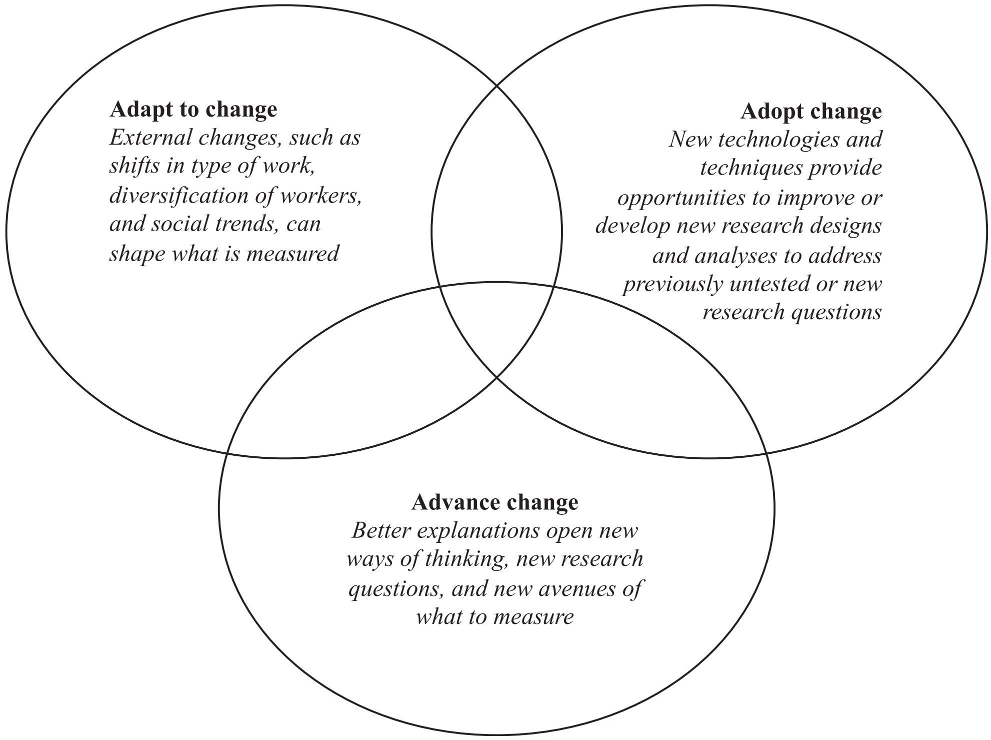


Fig. 1. General outline of organizing themes for review.

time to recover (Stoesz et al., 2020), and have greater vulnerabilities to other illnesses due to the effects of aging on immune function (Castelo-Branco & Soveral, 2014). Occupational safety researchers need to adapt to these changing demographics; this may involve determining whether interventions and other preventative steps mitigate the likelihood of harm toward specific subpopulations.

While there is a strong need for research addressing such external changes, there is also a need for research that controls context and randomizes the study population. Although considerable advances in field interventions have helped researchers draw causal inferences (such as the work of Total Worker Health® interventions; Anger et al., 2018; Feltner et al., 2016), more experimental research on occupational safety is required. Indeed, there are few studies that attempt to examine occupational safety within experimental settings (cf. Beck, Scholer, & Schmidt, 2017; Noort, Reader, & Gillespie, 2019; Probst, 2002). One reason for this gap may be that the independent variables in safety research are challenging to manipulate ethically and validly in an experimental context. However, there are some creative approaches to studying the effects of external changes on occupational safety. For example, inspired by rising job insecurity and increasing use of strategic downsizing, Probst (2002) created a manual labor work simulation in which participants either received or did not receive an emergency memorandum warning them about performance-related layoffs, and then examined the effect this had on safety compliance. The results from this and other safety-related experiments enable sorely needed cause-and-effect inferences.

2.2. Adopting change

Early research on occupational safety adopted a wide range of tools for measuring safety-related variables, such as physiological and psychomotor tests (e.g., Slocombe & Brakeman, 1930), interviews and observations (e.g., Hersey, 1936), organizational records (e.g., Hill & Trist, 1953), and informant- and self-reports (e.g., Davids & Mahoney, 1957; Kerr, 1950). These tools have not expanded significantly in decades—perhaps because the ideals of research design (i.e., randomization of participants, control of independent variable[s], multiple measurements pre-post manipulation) have changed very little since the early 20th century (Fisher, 1935). However, the technology for measurement has improved greatly. Today it includes advances in validating measurements (e.g., advanced software for testing psychometrics), more reliable instruments (e.g., recorders for interviews instead of memory and notes), better access to participants (e.g., devices with remote Internet access), and the ability to store and navigate big data (e.g., increasingly powerful computers with specialized software).

Take the measurement of work-related injuries as an example. Researchers still rely heavily on physical and physiological measures (e.g., severity of body deformity or functional limitations resulting from work injury; Chin et al., 2018), interviews and observations (e.g., thematic analysis on the consequences of compensable injuries; Allen et al., 2016), archival and organizational records (e.g., administrative data on injuries; Barnes & Wagner, 2009), and informant- and self-reports of injuries (e.g., children's ratings of their parents' lost-time work injuries; Turner, Granger, Tucker, Deng, & Kelloway, 2021). However, the validity and reliability of the measurement tools, as well as the technology for gathering and handling this data has improved markedly. For instance, current best practices for measuring self-reported injury frequency have moved away from single-item count measures with long recall periods (e.g., 12 months; Bamberger, 2005) to multi-item indices of generalizable and occupation-specific minor- (e.g., superficial wounds such as scratches or abrasions) to major-injuries (e.g., concussion or fractures) incurred within shorter recall windows (e.g., 3 months; Tucker, Ogunfowora, & Ehr, 2016). These indices provide richer insight into the types of injuries incurred, allow for differentiation of severity, and balance the lapse of time needed to accumulate relatively infrequently occurring injuries with minimizing the underestimation of injuries found in longer recall periods (Andersen & Mikkelsen, 2008).

Similarly, analytical approaches have advanced tremendously. Early occupational safety research typically relied on displaying descriptive and frequency statistics (e.g., Hoffman, 1909) and correlations coefficients (e.g., Slocombe & Brakeman, 1930) before it slowly adopted the statistical analyses and software used in related fields of social science, especially psychology and economics. This adoption of technologically advanced measurement and statistical analyses allowed occupational safety researchers to address unanswered or poorly answered research questions in new ways and to ask new research questions to advance thinking.[3] Here we consider the methodological and technological advances that are likely to be more heavily adopted in future research.

There is growing interest in applying multilevel research designs and data analyses to occupational safety research. Multilevel approaches enable researchers to address questions related to variation at different levels of analyses, providing an opportunity to examine the effect of variation within and between these levels. Examples of a multilevel approach include examining individuals' safety motivation across time (e.g., Beus & Taylor, 2018) and between levels within organizations (e.g., systems-focused models; Hofmann et al., 2017). Considerable occupational safety research has involved rich hierarchical data (e.g., Boyle, 1980; Goodman & Garber, 1988). Had this research been conducted today, it would certainly have benefited from multilevel analyses (e.g., accounting for variance within and between work teams or dyads of workers, respectively, over time).

Moreover, technological advances in research tools have made multilevel research easier to conduct and analyze. Consider, for instance, study participants providing data at several time points throughout each day for two to three weeks, otherwise known as experience sampling (Beal, 2015). Safety research is likely to see a significant rise in the use of this method to improve several lines of research, including variation in safety behaviors over time (Olsen et al., 2020). Further improvements are also likely as smart wearables and miniaturized physiological devices advance in capability (e.g., enabling the gathering of psychological data in addition to physical measures), duration (e.g., battery power), accuracy (e.g., physiological measures), comfort, and affordability. Sending brief questions to participants through, for example, a smart watch could reduce the gap between experiences and responses to those experiences, altogether minimizing recall bias (Scollon, Prieto, & Diener, 2009).

Finally, streams of research in occupational safety have matured to an extent that large meta-analytic reviews have been and will increasingly be conducted (e.g., Christian et al., 2009; Nahrgang et al., 2011). These are helpful in considering the magnitude and variation of relationships as well as the relative importance of factors influencing and influenced by safety; aggregating tests of theory; examining the external validity of relationships; examining the conditional role of contexts, sample characteristics such as age and relative minority status, and other design features of primary studies (e.g., comparing longitudinal studies); and identifying gaps for future research attention. The future of meta-analytic reviews in occupational safety, as well as other methods and analyses outlined in this section, will be one that involves addressing the ecological changes highlighted in the previous section on adapting to change, distinguishing between levels of analysis as multilevel approaches are increasingly adopted, and testing theoretical and conceptual advances—as detailed in the following section.

[3] Despite advances in technology that have improved research measurement and analyses, it is crucial to emphasize that ideals of research design (i.e., random sampling and control of independent variables) are considerably more important. There is likely some wisdom in considering Stone-Romero's (2011) words—"data that are not worth analyzing are not worth analyzing well" (p. 42) —when re-evaluating previously collected data with new analytical strategies.

2.3. Advancing change

Research in occupational safety advanced by measuring the costs of and predicting accidents and injuries. It is only recently that the concept of safety has expanded in scope (i.e., as "an attribute of work systems reflecting the [low] likelihood of physical harm—whether immediate or delayed—to persons, property, or the environment during the performance of work;" Beus et al., 2016, p. 353). Rather, earlier researchers viewed safety as essentially synonymous with harmful outcomes and were primarily interested in identifying individual differences related to the occurrence of accidents and injuries (e.g., Slocombe & Brakeman, 1930). A leading idea at the time—accident proneness—was driven by an observation that a minority of individuals were liable to experience the majority of accidents and injuries (Farmer & Chambers, 1929). More recently, interest in dispositional inclination toward accidents and injuries has resurfaced (Clarke, 2016), with the latest meta-analysis on the topic suggesting that there may be some empirical basis to accident proneness (Visser, Pijl, Stolk, Neeleman, & Rosmalen, 2007).

However, there are multiple reasons to be skeptical about accident proneness and how it is understood. Theoretically, accident proneness does not adequately provide nor account for alternative arguments as to why previous accidents would be related to future accidents. Research that has found supporting evidence for accident proneness rarely accounts for other plausible reasons causing both prior and future injuries (Mohr & Clemmer, 1988)—otherwise known as the third variable problem of causation. This research also relies largely on frequency of previous injuries as the measure of accident proneness, which fails to capture disposition or liability and could just as well be a product of failures of work safety regulations to protect certain minority populations, or physical and psychological hindrances that arise from the previous injury. Further, accident proneness could in part be a reliable measurement artifact capturing the proneness to report accidents and injuries, rather than differentiating the actual frequency of accidents and injuries.

The next advance in occupational safety research was growing evidence for and theory around work design (Kerr, 1957) in explaining why certain individuals are more likely to be involved in accidents and injuries—particularly in terms of how work design shapes the relationships that individuals had with their work and work environment. Kerr (1957) presciently suggested that accidents and injuries should be considered low-quality performance behavior that has similar antecedents to work performance. He posited that jobs that were designed to provide workers with sufficient autonomy and empowerment and were not subject to excessive and unnecessary stress would have fewer accidents and injuries (foreshadowing general balance theories, such as job demands-control-support theory and job demands-resources theory, that would later gain dominance in organizational sciences; Bakker & Demerouti, 2017; Karasek, 1979; Van der Doef & Maes, 1999).

Although Kerr (1957) was prophetic about safety as work performance, his research was limited by the conceptual definition of safety as accidents and injuries. More recent expansion of the scope of safety separates outcomes from behaviors (Burke, Sarpy, Tesluk, & Smith-Crowe, 2002; Griffin & Neal, 2000). Safety-related behaviors refer to "any workplace behaviors that affect the likelihood of physical harm to persons" (Beus et al., 2016, p. 3). Research in this vein encompasses the presence of safety (safe work behaviors) and the absence of safety (unsafe work behaviors), as well as the relationship between behavior and the probability of risk (Beus et al., 2016). Safety behaviors can be much more informative than accidents in advancing safety research, as the latter only communicate the absence of safety—and only if or after harm is done (Beus et al., 2016). However, the major challenge of focusing on safety behaviors is that they need to be captured accurately and over time. We still know little about the variation in safety behavior within individuals, the factors that contribute to this variation, and the consequences of it for accidents and injuries.

The idea of viewing safety as performance-related lent itself to a number of expansions, such as borrowing the tenets of job performance theory (Borman & Motowidlo, 1993; Campbell, McCloy, Oppler, & Sager, 1993) to develop the idea of safety performance theory (Griffin & Neal, 2000). Safety performance involves safety compliance (which is equivalent to task performance and involves following safety rules) and safety participation (which is equivalent to contextual performance and involves voluntary behaviors to improve safety). Both aspects are shaped by the proximal factors of safety knowledge, skills, and motivation (Griffin & Neal, 2000; Neal & Griffin, 2004). Collectively referred to as the safety triad, these three factors drive safety performance and are mainly determined by individual differences (e.g., personality) and contextual factors (e.g., leadership and training), as well as history and experience (Neal & Griffin, 2004).

The literature largely supports the proposed links between a host of individual and contextual factors on safety knowledge and motivation, which are then related to safety behaviors (Christian et al., 2009). However, research has yet to test the notion of specific safety skills beyond the broad idea of non-technical skills related to safety (e.g., Flin & O'Connor, 2016; Yule, Flin, Paterson-Brown, & Maran, 2006)—despite the inclusion of safety skills in models that feature safety knowledge and safety motivation (e.g., Beus et al., 2016; Griffin & Neal, 2000). One study purporting to measure safety skills (Eklöf & Törner, 2002) only measured technical (i.e., procedural) safety knowledge and did not actually operationalize any skill measure. Developing and testing the concept of safety skills as a relevant antecedent to safety performance seems a relevant and timely parallel, given the important role of abilities in understanding and predicting general performance.

Another advancement in the safety literature was the adoption of balance theories to predict safety performance and outcomes. The two balance theories that have received the most empirical attention in occupational safety research are the job demands-control-support theory (e.g., Snyder, Krauss, Chen, Finlinson, & Huang, 2008; Turner, Stride, Carter, McCaughey, & Carroll, 2012) and the job demands-resources theory (Nahrgang et al., 2011). These theories suggest that an imbalance in demands and resources shapes the degree to which individuals can perform well. Demands refer to job aspects that call for physical and psychological effort of workers, while resources are job aspects that help achieve work goals, mitigate demands, or lead to employees' growth, learning, or development (Bakker & Demerouti, 2007). Balance theories have considerable generalizability, in that many job aspects can be classified as demands and resources and help to shape performance (Bakker & Demerouti, 2017). From a safety performance lens, job demands and resources influence safety motivation, which in turn influences safe and unsafe behaviors.

The literature generally reinforces the role of balance theories in predicting safety outcomes (Nahrgang et al., 2011). Not only do demands and resources rooted in job characteristics and contextual factors associate with safety outcomes directly, but they also indirectly shape outcomes through safety behaviors and psychological strain (Nahrgang et al., 2011). Yet, while the generalizability of the balance theories is their strength, it is also their weakness. Only recently has substantive evidence emerged for how demands and resources interact (e.g., Gonzalez-Mulé, Kim, & Ryu, 2020), and there is scant consensus about the most important types of demands and resources for predicting safety behaviors and outcomes. The job demands-control-support theory hints at which resources are the most critical (i.e., control and support), but neither of the balance theories is effective in clearly distinguishing among demands. Research distinguishing challenge and hindrance demands has proven useful (Clarke, 2012), but a recent meta-analysis highlights that assuming specific demands are challenge or hindrance stressors is misguided, and the distinction should instead be measured through appraisals (Mazzola & Disselhorst, 2019). These appraisals, in turn, are subject to variation within individuals, as little is known about the effects of this variation on safety.

The final and most prominent theory in workplace safety research is the application of organizational climate theory to the field in the form of safety climate (Beus et al., 2019; Kerr, 1950; Zohar, 1980). The broader climate theory suggests that an organization's collective expectations of how people behave will shape individual- and group-level behaviors (Zohar, 2010). These expectations typically represent the belief that certain behaviors will be reinforced or punished, and thus motivate people to behave accordingly. Applied to safety, organizational climate reflects the shared perceptions about the value of safety in the workplace (Zohar, 1980).

This sense of safety climate is one of the most robust predictors of safety-related behavior (Beus, Payne, Bergman, & Arthur, 2010). However, it is often unclear how safety climate differs from leadership (Clarke, 2013), safety culture (Guldenmund, 2000), and organizational goal setting and feedback (Reber & Wallin, 1984), particularly as many measures of safety climate refer to leaders' behaviors and the extent to which safety is a salient goal. While the evidence for these features as predictors is also strong, there is disagreement about the overlap between the behavior-outcome expectancy (i.e., results from safety climate) and safety motivation. Beus et al. (2016) argue that safety motivation is value-laden, while behavior-outcome expectancy is instrumental. This distinction makes sense theoretically but remains empirically untested.

Despite the gaps that the theories presented above have addressed, we still know little about causal direction in safety. For instance, accidents have consequences for perceptions of organizational safety policies and safety climate (Bergman, Payne, Taylor, & Beus, 2014; Beus et al., 2010; Madsen, 2009); yet factors such as these are often considered predictors of accidents and injuries. A high percentage of the workplace safety literature has examined the behavior-to-outcome relationship but has typically measured outcomes prior to behavior (e.g., Frone, 1998). This sequence is not only a limitation, but also an opportunity: future research can revisit prior research in meta-analytic reviews to test why the relationship may be the other way around, offering a more complete narrative of occupational safety.

2.4. Future research in psychology of occupational safety

The sections above provide an abridged review of the historical progression of the psychology of occupational safety research using our *adapting, adopting, and advancing change* framework. In doing so, they identify untested research questions as examples of how these three themes can play out. In this next section, we propose future research that blends the organizing themes of external changes, technological and technical developments, and theoretical advances (see Table 1) and includes (1) building a more refined understanding of safety skill and its connection to safety performance; (2) testing how the variation in the appraisals of

Table 1
Example future directions at the convergence of adapting, adopting, and advancing change.

Research topic	Adapting to change	Adopting change	Advancing change
Safety skill	Comparing specific vs generalized contexts to assess whether the etiology of safety skill varies by context and determine the extent to which it can be transferred across contexts	Intensive longitudinal designs, such as experience sampling, in conjunction with improved wearable technology for data collection to assess how safety skill develops and varies over time	Advancing safety performance theories and the emergence of safety skills as an independent concept grounded in cognitive psychology
Demand appraisals and safety	Comparing the nature of demands and their consequences on safety between gig workers and organizational workers	The measurement of acute versus chronic demands through latent growth curve modeling, physiological measures to assess impact of differentially appraised stressors	Advance balance theories and the role of variation in stress appraisals
Distinguishing safety motivation and safety climate	Experimental and controlled conditions to isolate the valence-instrumentality distinction	Incorporate measures of appraisals to test whether the argued distinction is valid	Advance safety performance and safety climate theories by testing important theoretical distinctions
Consequences of safety outcomes	Observed and expected demographic shifts, such as aging, and their impact on work safety associations	Meta-analytical reviews	Kindle theoretical development in research that has increasingly become atheoretical

demands and their interaction with resources influence safety behavior within balance theories; (3) empirically differentiating safety motivation and safety-related behavior-outcome expectancy to distinguish safety climate consequences; and (4) examining the relationship between safety outcomes and broader well-being phenomena.

2.5. Safety skill

First, as noted, safety performance theory suggests that the proximal determinants of safety knowledge, skill, and motivation collectively shape safety behavior. Yet, the idea of safety skill remains underdeveloped in the literature, with its absence possibly explaining considerable variance in safety performance beyond safety knowledge and motivation. Thus, the first question that arises is how safety skill may be effectively conceptualized to advance change. We propose that it should focus on attention and decision making, as under ideal conditions, individuals allocate an appropriate amount of attention to hazards in their work environment and decide on a course of action that avoids or mitigates those hazards for themselves and others.

The next advancing change question pertains to how safety skill develops and manifests. Theory from cognitive psychology, particularly dual-process models and naturalistic decision making, may be informative in this regard. Dual-process models suggest that there are two independent and relatively distinct modes through which thoughts arise and decisions are made (Evans, 2008): one is quick, efficient, and unconscious (i.e., System 1) and the other is slow, deliberate, and calculating (i.e., System 2). Relatedly, the naturalistic decision making framework posits that individuals realistically rely heavily on System 1, rapidly examining their surroundings and using experience to categorize a situation to determine a course of action (Lipshitz, Klein, Orasanu, & Salas, 2001). Together, these theories suggest that safety skill develops through experience and helps individuals quickly perceive and avoid hazards and threats they would otherwise be unaware of, even if they were fully attentive. In summary, we conceive of safety skill as a malleable and adaptable form of hazard detection activated in the presence of recognized hazards.

The approaches for empirically investigating safety skill can be informed by considerations around research design in the form of context (adapting change) and considerations around methodology in the form of time and relative importance (adopting change). Given that the ideas and research questions posed above focus on the "how" of safety skill, drawing on a body of existing research (e.g., Gherardi & Nicolini, 2002) that describes how safety is both embodied across various contexts and socially constructed becomes a vital first step. This will be important for hypothesizing about the etiology of safety skill and the extent to which it can be generalized across contexts and is specific to others. Further, factoring time into the approach will enable an understanding of how safety skill develops, how and for how long it activates, and how much it varies among individuals. Studies involving intensive longitudinal designs in conjunction with advanced wearable technology for data collection to assess how safety skill develops and varies will be key. Finally, considering relative importance will help determine whether safety skill contributes to safety performance beyond safety knowledge and safety motivation. These insights will be key for determining the extent to which safety skill advances change in safety performance theory.

2.6. Demand appraisals and safety

Second, extensive research is still needed to incorporate appraisals of demands in testing and advancing change in balance theories applied to occupational safety. Numerous insights could be gained from research that enhances the correspondence between demands and perceptions of demands on safety by revisiting transactional theory of stress (Ellsworth, 2013; Lazarus, 1991) and adopting the conceptual distinctions of challenge and hindrance stressors (Cavanaugh, Boswell, Roehling, & Boudreau, 2000). For example, it is possible that demands that are actually appraised as challenges (rather than just assumed to be) improve safety performance—similar to what they have been found to do for job performance (Podsakoff, LePine, & LePine, 2007). Considering the opposite direction, are there aspects of safety that may be better conceptualized as demands rather than outcomes—such as acute injuries appraised as hindrances? Further, expanding occupational safety research that applies appraisals within balance theories would also advance our understanding of how appraisals of demands vary within and between people over time, giving rise to new research questions focused on the chronic or acute nature of specific demands in shaping safety behavior and outcomes.

Future research that examines the role of demands in occupational safety can adopt change through various quantitative approaches. Once again, the role of time may be a prime consideration, both methodologically and in further distinguishing the type of demands to incorporate. Indeed, little research has differentiated between acute and chronic demands, as well as their interaction. Moreover, methodological considerations of time would allow researchers to test diverse questions about the change in appraisals and the effect that both the appraisals themselves and any changes in appraisals play on safety. Analytically, these questions can be examined through techniques that are growing in popularity based on the type of questions they allow researchers to answer. In particular, questions on the nature of demand appraisals (similar to those outlined above) could be addressed through latent growth curve modeling (Bollen & Curran, 2006) to determine the role that certain trends play on safety behavior and safety outcomes.

Variation in the appraisal of demands and the effect this variation has on safety may also change as a function of the nature of work being conducted. There is growing interest in adapting research to understand the challenges of gig work (Caza, Reid, Ashford, & Granger, 2021), how they differ from those experienced by traditional organizational workers, and their unique consequences for health and safety (Gregory, 2021). Examining or comparing the effect that variation of demand appraisals has on safety within this population (relative to traditional organizational workers) could not only expand the extent to which demands can be both challenges and hindrances, but also uncover substantive differences within this growing workforce.

2.7. Distinguishing safety motivation and safety climate

Third, safety climate communicates the behavior-outcome expectancies that employees have, which subsequently motivate (in the presence of positive outcomes, such as organizational rewards) safety behavior or demotivate (in the presence of negative outcomes, such as performance setbacks) safe behavior (Zohar, 2010). This behavior-outcome expectancy is cognitively distinct from safety motivation as outlined in safety performance theory (Griffin & Neal, 2000). In particular, Beus et al. (2016) argue from an expectancy theory lens (Vroom, 1964) that safety motivation is a function of the perceived value that individuals place on safety (i.e., valence), while behavior-outcome expectancy is a function of the correspondence between behavior and outcomes (i.e., instrumental). This conceptual advance in distinguishing safety climate may be useful for explaining safety climate's importance among other contextual factors, but still requires empirical testing.

Researchers can move this debate forward by taking approaches from the theme of adapting to change by controlling the context through experimental designs. Specifically, occupational safety researchers can create simulations in which values or instrumentalities of safe conduct become manipulated and examine the effect this has on safety behavior. Doing so would enable a comparison of the extent to which behavior-outcome expectancy is uniquely related to safety climate or safety motivation while holding the context constant. An important element to this experimental approach will be to adopt changes mentioned for advancing balance theories, such as incorporating appraisals as manipulation checks. By doing so, researchers can effectively demonstrate that safety motivation is driven by the values that individuals have and that the behavior-outcome expectancy is driven by instrumental means, which in turn would advance distinctions in safety-related motivations and their relative importance.

2.8. Consequences of safety outcomes

Finally, occupational safety researchers have traditionally considered accidents and injuries as outcomes. However, there is a call for more research on the consequences of these safety outcomes (Beus et al., 2016), with recent research starting to examine their bi-directional effects. For instance, Beus et al. (2010) have shown that injuries are associated with declines in future measures of safety climate. Further research in this domain may offer valuable insights for both theory and practice.

Theoretically, little is known about why previous injuries may have an association with various consequential phenomena. Consider, for example, the potential relationship between work-related injuries and mental health. Scattered evidence has accumulated to suggest an association (e.g., Kim & Choi, 2016), but little else can be said about the nature of the relationship. Does mental health predict work-related injuries or do work-related injuries predict mental health? We argue that both directions are plausible but that different theoretical rationales will need to be tested to advance our understanding of why.

The most productive methodological adoption that researchers can take to begin dissecting the relationships between traditional safety outcomes (e.g., work-related injuries) and other important phenomena would be to synthesize the existing literature through meta-analysis. Fortunately, most research on work-related injuries has either measured prior or subsequent injuries, allowing for some inferences about time. Having strong empirical support for the consequences of injuries will advance current thinking in injury prevention, injury reporting, and return to work. Further, researchers could carefully consider coding important demographic differences within samples included in a meta-analysis, thus testing the association between work injuries and mental health problems. Considering potential demographics mentioned in the adapting change theme, it is possible that age could be an important condition shaping the relationship given the differential consequences of injuries for older workers. Further, research has indicated that gender, education, and income, as well as marital and relative minority status may be useful to test through meta-analyses (Burke-Miller et al., 2006; Kim & Choi, 2016; Nelson, 2002; Salminen, 2013; Smith & Mustard, 2010; Williams & Mohammed, 2009).

3. Conclusion

We introduce an organizing framework of adapting, adopting, and advancing change to describe the historical progression of research in the psychology of occupational safety. This process highlights numerous external shifts in the nature of work, work environments, and worker demographics (i.e., adapting to change); technological and technical implications for methodology, research design, and analysis (i.e., adopting change); and advances in explanations and theory (i.e., advancing change). We then discuss problems with leading theories in this domain and potential future directions. Finally, we take these potential future directions and draw out future research questions and approaches that intersect the adapting, adopting, and advancing change themes. Overall, this paper provides a structure to summarize current research challenges and to imagine new ways of advancing research in the psychology of occupational safety.

4. Author note

We are grateful to Sandy Hershcovis, Eric Kearney, Zhanna Lyubykh, Catie Phares, Tim Vahle-Hinz, Ivana Vranjes, Julie Weatherhead, and Justin Weinhardt for their constructive feedback on earlier versions of this manuscript. The Canadian Centre of Advanced Leadership in Business and the Social Sciences and Humanities Research Council of Canada enabled collaboration on this paper.

References

Aldrich, M. (1997). *Safety first: Technology, labor, and business in the building of American work safety, 1870–1939*. The Johns Hopkins University Press.

Andersen, L. P., & Mikkelsen, K. L. (2008). Recall of occupational injuries: A comparison of questionnaire and diary data. *Safety Science, 46*(2), 255–260. https://doi.org/10.1016/j.ssci.2007.06.014.

Anger, W. K., Kyler-Yano, J., Vaughn, K., Wipfli, B., Olson, R., & Blanco, M. (2018). Total Worker Health® intervention for construction workers alters safety, health, well-being measures. *Journal of Occupational and Environmental Medicine, 60*(8), 700–709. https://doi.org/10.1097/JOM.0000000000001290.

Association of Workers' Compensation Boards of Canada (2020). *Statistics*. National Work Injury/Disease Statistics Program (NWISP). Retrieved from https://awcbc.org/en/statistics/.

Bakker, A. B., & Demerouti, E. (2007). The job demands-resources model: State of the art. *Journal of Managerial Psychology, 22*(3), 309–328. https://doi.org/10.1108/02683940710733115.

Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology, 22*(3), 273–285. https://doi.org/10.1037/ocp0000056.

Bamberger, P. A. (2005). Work-based critical incidents and problem drinking: Taking intrusive reactions, traumatic distress, and the kindling effect into account. *Work and Occupations, 32*(3), 257–289. https://doi.org/10.1177/0730888405277900.

Barling, J. E., & Frone, M. R. (2004). *The psychology of workplace safety*. American Psychological Association.

Barnes, C. M., & Wagner, D. T. (2009). Changing to daylight saving time cuts into sleep and increases workplace injuries. *Journal of Applied Psychology, 94*(5), 1305–1317. https://doi.org/10.1037/a0015320.

Beal, D. J. (2015). ESM 2.0: State of the art and future potential of experience sampling methods in organizational research. *Annual Review of Organizational Psychology and Organizational Behavior, 2*(1), 383–407. https://doi.org/10.1146/annurev-orgpsych-032414-111335.

Beck, J. W., Scholer, A. A., & Schmidt, A. M. (2017). Workload, risks, and goal framing as antecedents of shortcut behaviors. *Journal of Business and Psychology, 32*(4), 421–440. https://doi.org/10.1007/s10869-016-9450-0.

Bergman, M. E., Payne, S. C., Taylor, A. B., & Beus, J. M. (2014). The shelf life of a safety climate assessment: How long until the relationship with safety–critical incidents expires? *Journal of Business and Psychology, 29*(4), 519–540. https://doi.org/10.1007/s10869-013-9337-2.

Beus, J. M., McCord, M. A., & Zohar, D. (2016). Workplace safety: A review and research synthesis. *Organizational Psychology Review, 6*(4), 352–381. https://doi.org/10.1177/2041386615626243.

Beus, J. M., Payne, S. C., Arthur, W., Jr, & Muñoz, G. J. (2019). The development and validation of a cross-industry safety climate measure: Resolving conceptual and operational issues. *Journal of Management, 45*(5), 1987–2013. https://doi.org/10.1177/0149206317745596.

Beus, J. M., Payne, S. C., Bergman, M. E., & Arthur, W. (2010). Safety climate and injuries: An examination of theoretical and empirical relationships. *Journal of Applied Psychology, 95*(4), 713. https://doi.org/10.1037/a0019164.

Beus, J. M., & Taylor, W. D. (2018). Working safely at some times and unsafely at others: A typology and within-person process model of safety-related work behaviors. *Journal of Occupational Health Psychology, 23*(3), 402–416. https://doi.org/10.1037/ocp0000092.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: John Wiley & Sons Ltd.

Borman, W. C., & Motowidlo, S. M. (1993). Expanding the criterion domain to include elements of contextual performance: The meaning for personal selection research. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71–98). Jossey-Bass.

Boyle, A. J. (1980). 'Found experiments' in accident research: Report of a study of accident rates and implications for future research. *Journal of Occupational Psychology, 53*(1), 53–64. https://doi.org/10.1111/j.2044-8325.1980.tb00006.x.

Brakeman, E. E., & Slocombe, C. S. (1929). A review of recent experimental results relevant to the study of individual accident susceptibility. *Psychological Bulletin, 26*(1), 15–38. https://doi.org/10.1037/h0071735.

Brundage, D. K. (1927). A 10-year record of absences from work on account of sickness and accidents: Experience of employees of the Edison Electric Illuminating Co. of Boston, 1915 to 1924, Inclusive. *Public Health Reports, 1896–1970*, 529–550. https://doi.org/10.2307/4578201.

Bureau of Labor Statistics (2020). *Injuries, illness, and fatalities*. U.S. Department of Labor. Retrieved from https://www.bls.gov/iif/.

Burke, M. J., Salvador, R. O., Smith-Crowe, K., Chan-Serafin, S., Smith, A., & Sonesh, S. (2011). The dread factor: How hazards and safety training influence learning and performance. *Journal of Applied Psychology, 96*(1), 46–70. https://doi.org/10.1037/a0021838.

Burke, M. J., Sarpy, S. A., Tesluk, P. E., & Smith-Crowe, K. (2002). General safety performance: A test of a grounded theoretical model. *Personnel Psychology, 55*(2), 429–457. https://doi.org/10.1111/j.1744-6570.2002.tb00116.x.

Burke-Miller, J. K., Cook, J. A., Grey, D. D., Razzano, L. A., Blyler, C. R., Leff, H. S., ... Cook, W. L. (2006). Demographic characteristics and employment among people with severe mental illness in a multisite study. *Community Mental Health Journal, 42*(2), 143–159. https://doi.org/10.1007/s10597-005-9017-4.

Burt, C. D. B. (2016). *New employee safety*. Springer International Publishing.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.). *Personnel selection in organizations* (Vol. 3570, pp. 35–70). Jossey-Bass.

Castelo-Branco, C., & Soveral, I. (2014). The immune system and aging: A review. *Gynecological Endocrinology, 30*(1), 16–22. https://doi.org/10.3109/09513590.2013.852531.

Cavanaugh, M. A., Boswell, W. R., Roehling, M. V., & Boudreau, J. W. (2000). An empirical examination of self-reported work stress among U.S. managers. *Journal of Applied Psychology, 85*(1), 65–74. https://doi.org/10.1037/0021-9010.85.1.65.

Caza, B. B., Reid, E. M., Ashford, S. J., & Granger, S. (2021). Working on my own: Measuring the challenges of gig work. *Human Relations*. https://doi.org/10.1177/00187267211030098.

Chin, W. S., Guo, Y. L., Liao, S. C., Lin, K. H., Kuo, C. Y., Chen, C. C., & Shiao, J. S. C. (2018). Suicidality 6 years after occupational injury. *Journal of Clinical Psychiatry, 79*(5). https://doi.org/10.4088/JCP.17m11729.

Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. (2009). Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology, 94*(5), 1103–1127. https://doi.org/10.1037/a0016172.

Clarke, S. (2012). The effect of challenge and hindrance stressors on safety behavior and safety outcomes: A meta-analysis. *Journal of Occupational Health Psychology, 17*(4), 387–397. https://doi.org/10.1037/a0029817.

Clarke, S. (2013). Safety leadership: A meta-analytic review of transformational and transactional leadership styles as antecedents of safety behaviours. *Journal of Occupational and Organizational Psychology, 86*(1), 22–49. https://doi.org/10.1111/j.2044-8325.2012.02064.x.

Clarke, S. (2016). Accident proneness: Back in vogue? In *Occupational health and safety* (pp. 119–142). Routledge.

Collinson, D. L. (1999). Surviving the rigs': Safety and surveillance on North Sea oil installations. *Organization Studies, 20*(4), 579–600. https://doi.org/10.1177/0170840699204003.

Cornelissen, P. A., Van Hoof, J. J., & De Jong, M. D. (2017). Determinants of safety outcomes and performance: A systematic literature review of research in four high-risk industries. *Journal of Safety Research, 62*, 127–141. https://doi.org/10.1016/j.jsr.2017.06.009.

Davids, A., & Mahoney, J. T. (1957). Personality dynamics and accident proneness in an industrial setting. *Journal of Applied Psychology, 41*(5), 303–306. https://doi.org/10.1037/h0047608.

Ellsworth, P. C. (2013). Appraisal theory: Old and new questions. *Emotion Review, 5*(2), 125–131. https://doi.org/10.1177/1754073912463617.

Eklöf, M., & Törner, M. (2002). Perception and control of occupational injury risks in fishery–A pilot study. *Work & Stress, 16*(1), 58–69. https://doi.org/10.1080/02678370110114694.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629.

Farmer, E., & Chambers, E. G. (1929). A study of personal qualities in accident proneness and proficiency. *Industrial health research board report*. London: HMSO.

Feltner, C., Peterson, K., Palmieri Weber, R., Cluff, L., Coker-Schwimmer, E., Viswanathan, M., & Lohr, K. N. (2016). The effectiveness of Total worker health interventions: A systematic review for a National Institutes of Health pathways to prevention workshop. *Annals of Internal Medicine, 165*(4), 262–269. https://doi.org/10.7326/M16-0626.

Fishback, P. V., & Kantor, S. E. (2000). *A prelude to the welfare state: The origins of workers' compensation*. University of Chicago Press.

Fisher, R. A. (1935). *The design of experiments*. Oliver & Boyd.

Flin, R., & O'Connor, P. (2016). *Safety at the sharp end: A guide to non-technical skills*. CRC Press.

Frone, M. R. (1998). Predictors of work injuries among employed adolescents. *Journal of Applied Psychology, 83*(4), 565–576. https://doi.org/10.1037/0021-9010.83.4.565.

Gabriel, A. S., Podsakoff, N. P., Beal, D. J., Scott, B. A., Sonnentag, S., Trougakos, J. P., & Butts, M. M. (2019). Experience sampling methods: A discussion of critical trends and considerations for scholarly advancement. *Organizational Research Methods, 22*(4), 969–1006. https://doi.org/10.1177/1094428118802626.

Gherardi, S., & Nicolini, D. (2002). Learning the trade: A culture of safety in practice. *Organization, 9*(2), 191–223. https://doi.org/10.1177/1350508402009002264.

Goldmark, J. C. (1912). *Fatigue and efficiency: A study in industry*. Charities Publication Committee.

Gonzalez-Mulé, E., Kim, M. M., & Ryu, J. W. (2020). A meta-analytic test of multiplicative and additive models of job demands, resources, and stress. *Journal of Applied Psychology*. https://doi.org/10.1037/apl0000840.

Goodman, P. S., & Garber, S. (1988). Absenteeism and accidents in a dangerous environment: Empirical analysis of underground coal mines. *Journal of Applied Psychology, 73*(1), 81–86. https://doi.org/10.1037/0021-9010.73.1.81.

Granger, S., Turner, N., & Grocutt, A. (2021). The next best safety dollar: Using evidence to decide how to invest in workplace safety. *Organizational Dynamics, 50*(2). https://doi.org/10.1016/j.orgdyn.2020.100774 100774.

Gregory, K. (2021). 'My life is more valuable than this': Understanding risk among on-demand food couriers in Edinburgh. *Work, Employment, & Society, 35*, 316–331. https://doi.org/10.1177/0950017020969593.

Greenwood, M., & Woods, H. M. (1919). *The incidence of industrial accidents upon individuals: With special reference to multiple accidents* (Issue 4). HM Stationery Office [Darling and son, Limited, printers].

Griffin, M. A., & Neal, A. (2000). Perceptions of safety at work: A framework for linking safety climate to safety performance, knowledge, and motivation. *Journal of Occupational Health Psychology, 5*(3), 347–358. https://doi.org/10.1037/1076-8998.5.3.347.

Guldenmund, F. W. (2000). The nature of safety culture: A review of theory and research. *Safety Science, 34*(1–3), 215–257. https://doi.org/10.1016/S0925-7535(00)00014-X.

Health and Safety Executive (2020). Non-fatal injuries at work in Great Britain Retrieved from https://www.hse.gov.uk/statistics/causinj/index.htm.

Herbert, R., & Landrigan, P. J. (2000). Work-related death: A continuing epidemic. *American Journal of Public Health, 90*(4), 541. https://doi.org/10.2105/ajph.90.4.541.

Hersey, R. B. (1936). Emotional factors in accidents. *Personnel Journal, 15*, 59–65.

Herzog, A., House, J. S., & Morgan, J. N. (1991). Relation of work and retirement to health and well-being in older age. *Psychology and Aging, 6*(2), 202–211. https://doi.org/10.1037/0882-7974.6.2.202.

Hill, J. M. M., & Trist, E. L. (1953). A consideration of industrial accidents as a means of withdrawal from the work situation. *Human Relations, 6*(4), 357–380. https://doi.org/10.1177/001872675300600404.

Hoffman, F. L. (1909). Industrial accidents and industrial diseases. *Publications of the American Statistical Association, 11*(88), 567–603.

Hofmann, D. A., Burke, M. J., & Zohar, D. (2017). 100 years of occupational safety research: From basic protections and work analysis to a multilevel view of workplace safety and risk. *Journal of Applied Psychology, 102*(3), 375–388. https://doi.org/10.1037/apl0000114.

Hopkins, B. (2017). Occupational health and safety of temporary and agency workers. *Economic and Industrial Democracy, 38*(4), 609–628. https://doi.org/10.1177/0143831X15581424.

Howard, J. (2017). Nonstandard work arrangements and worker health and safety. *American Journal of Industrial Medicine, 60*(1), 1–10. https://doi.org/10.1002/ajim.22669.

Huang, Y., Zohar, D., Robertson, M. M., Garabet, A., Lee, J., & Murphy, L. A. (2013). Development and validation of safety climate scales for lone workers using truck drivers as exemplar. *Transportation Research Part F: Traffic Psychology and Behaviour, 17*, 5–19. https://doi.org/10.1016/j.trf.2012.08.011.

Jiang, L., Lavaysse, L. M., & Probst, T. M. (2019). Safety climate and safety outcomes: A meta-analytic comparison of universal vs. industry-specific safety climate predictive validity. *Work & Stress, 33*(1), 41–57. https://doi.org/10.1080/02678373.2018.1457737.

Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly, 24*(2), 285–308. https://doi.org/10.2307/2392498.

Kerr, W. (1950). Accident proneness of factory departments. *Journal of Applied Psychology, 34*(3), 167–170. https://doi.org/10.1037/h0061694.

Kerr, W. (1957). Complementary theories of safety psychology. *The Journal of Social Psychology, 45*(1), 3–9. https://doi.org/10.1080/00224545.1957.9714280.

Kim, J., & Choi, Y. (2016). Gender differences in the longitudinal association between work-related injury and depression. *International Journal of Environmental Research and Public Health, 13*(11), 1077. https://doi.org/10.3390/ijerph13111077.

Kowalski-Trakofler, K. M., Steiner, L. J., & Schwerha, D. J. (2005). Safety considerations for the aging workforce. *Safety Science, 43*(10), 779–793. https://doi.org/10.1016/j.ssci.2005.08.014.

Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice. In *The essential tension* (pp. 320–339). University of Chicago Press.

Lazarus, R. S. (1991). Psychological stress in the workplace. *Journal of Social Behavior & Personality*, *6*(7), 1–13. http://doi.org/1992-03575-001.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making, 14*(5), 331–352. https://doi.org/10.1002/bdm.381.

Madsen, P. M. (2009). These lives will not be lost in vain: Organizational learning from disaster in US coal mining. *Organization Science, 20*(5), 861–875. https://doi.org/10.1287/orsc.1080.0396.

Mazzola, J. J., & Disselhorst, R. (2019). Should we be "challenging" employees?: A critical review and meta-analysis of the challenge-hindrance model of stress. *Journal of Organizational Behavior, 40*(8), 949–961. https://doi.org/10.1002/job.2412.

Mohr, D. L., & Clemmer, D. I. (1988). The "accident prone" worker: An example from heavy industry. *Accident Analysis & Prevention, 20*(2), 123–127. https://doi.org/10.1016/0001-4575(88)90028-0.

Münsterberg, H. (1913). *Psychology and industrial efficiency*. Houghton, Mifflin and Company.

Myers, G. (1915). A study of the causes of industrial accidents. *Quarterly Publications of the American Statistical Association, 14*(111), 672–694.

Nahrgang, J. D., Morgeson, F. P., & Hofmann, D. A. (2011). Safety at work: A meta-analytic investigation of the link between job demands, job resources, burnout, engagement, and safety outcomes. *Journal of Applied Psychology, 96*(1), 71–94. https://doi.org/10.1037/a0021484.

Neal, A., & Griffin, M. A. (2004). Safety climate and safety at work. In J. Barling & M. R. Frone (Eds.), *The psychology of workplace safety* (pp. 15–34). American Psychological Association.

Nelson, A. (2002). Unequal treatment: Confronting racial and ethnic disparities in health care. *Journal of the National Medical Association, 94* (8), 666–668 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2594273/pdf/jnma00325-0024.pdf.

Noort, M. C., Reader, T. W., & Gillespie, A. (2019). Walking the plank: An experimental paradigm to investigate safety voice. *Frontiers in Psychology, 10*, 668. https://doi.org/10.3389/fpsyg.2019.00668.

O'Connor, P., O'Dea, A., Kennedy, Q., & Buttrey, S. E. (2011). Measuring safety climate in aviation: A review and recommendations for the future. *Safety Science, 49*(2), 128–138. https://doi.org/10.1016/j.ssci.2010.10.001.

Olsen, O. K., Hetland, J., Matthiesen, S. B., Hoprekstad, Ø. L., Espevik, R., & Bakker, A. B. (2020). Passive avoidant leadership and safety non-compliance: A 30 days diary study among naval cadets. *Safety Science, 1–8*. https://doi.org/10.1016/j.ssci.2020.105100.

Pfeffer, J. (2018). *Dying for a paycheck: How modern management harms employee health and company performance—and what we can do about it*. HarperCollins.

Podsakoff, N. P., LePine, J. A., & LePine, M. A. (2007). Differential challenge stressor-hindrance stressor relationships with job attitudes, turnover intentions, turnover, and withdrawal behavior: A meta-analysis. *Journal of Applied Psychology, 92*(2), 438–454. https://doi.org/10.1037/0021-9010.92.2.438.

Probst, T. M. (2002). Layoffs and tradeoffs: Production, quality, and safety demands under the threat of job loss. *Journal of Occupational Health Psychology, 7*(3), 211–220. https://doi.org/10.1037/1076-8998.7.3.211.

Probst, T. M., & Brubaker, T. L. (2001). The effects of job insecurity on employee safety outcomes: Cross-sectional and longitudinal explorations. *Journal of Occupational Health Psychology, 6*(2), 139–159. https://doi.org/10.1037/1076-8998.6.2.139.

Quinlan, M. (1999). The implications of labour market restructuring in industrialized societies for occupational health and safety. *Economic and Industrial Democracy, 20*(3), 427–460. https://doi.org/10.1177/0143831X99203005.

Quinlan, M., Mayhew, C., & Bohle, P. (2001). The global expansion of precarious employment, work disorganization, and consequences for occupational health: A review of recent research. *International Journal of Health Services, 31*(2), 335–414. https://doi.org/10.2190/607H-TTV0-QCN6-YLT4.

Reber, R. A., & Wallin, J. A. (1984). The effects of training, goal setting, and knowledge of results on safe behavior: A component analysis. *Academy of Management Journal, 27*(3), 544–560. https://doi.org/10.2307/256044.

Rudolph, C. W., Marcus, J., & Zacher, H. (2018). Global Issues in Work, Aging, and Retirement. In K. S. Shultz & G. A. Adams (Eds.), *Aging and work in the 21st century* ((2nd ed.,, pp. 292–324). Taylor & Francis.

Salminen, S. (2013). Occupational accidents: Prevalence, risk factors and health outcomes. *Journal of Risk and Governance*, *4*(3), 219–234. https://doi.org/1938929676?accountid=9838.

Scollon, C. N., Prieto, C.-K., & Diener, E. (2009). Experience sampling: Promises and pitfalls, strength and weaknesses. In E. Diener (Ed.), *Assessing well-being* (pp. 157–180). Springer.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Boston.

Slocombe, C. S., & Brakeman, E. E. (1930). Psychological tests and accident proneness. *British Journal of Psychology. General Section, 21*(1), 29–38.

Smith, P. M., & Mustard, C. A. (2010). The unequal distribution of occupational health and safety risks among immigrants to Canada compared to Canadian-born labour market participants: 1993–2005. *Safety Science, 48*(10), 1296–1303. https://doi.org/10.1016/j.ssci.2010.03.020.

Snyder, L. A., Krauss, A. D., Chen, P. Y., Finlinson, S., & Huang, Y.-H. (2008). Occupational safety: Application of the job demand–control-support model. *Accident Analysis & Prevention, 40*(5), 1713–1723. https://doi.org/10.1016/j.aap.2008.06.008.

Springer, Y., Lucas, D. L., Castrodale, L. J., & McLaughlin, J. B. (2018). Work-related injuries in the Alaska logging industry, 1991–2014. *American Journal of Industrial Medicine, 61*(1), 32–41. https://doi.org/10.1002/ajim.22784.

Stoesz, B., Chimney, K., Deng, C., Grogan, H., Menec, V., Piotrowski, C., ... Turner, N. (2020). Incidence, risk factors, and outcomes of non-fatal work-related injuries among older workers: A review of research from 2010 to 2019. *Safety Science, 126*. https://doi.org/10.1016/j.ssci.2020.104668 104668.

Swuste, P., van Gulijk, C., & Zwaard, W. (2010). Safety metaphors and theories, a review of the occupational safety literature of the US, UK and The Netherlands, till the first part of the 20th century. *Safety Science, 48*(8), 1000–1018. https://doi.org/10.1016/j.ssci.2010.01.020.

Tetrick, L. E. (2017). Trends in measurement models and methods in understanding occupational health psychology. *Journal of Occupational Health Psychology, 22*(3), 337–340. https://doi.org/10.1037/ocp0000076.

Toossi, M. (2002). A century of change: The US labor force, 1950–2050. *Monthly Labour Review, 125*(5), 15–28.

Tran, M., & Sokas, R. K. (2017). The gig economy and contingent work: An occupational health assessment. *Journal of Occupational and Environmental Medicine*, *59*(4), e63-e66. http://doi.org/JOM.0000000000000977.

Tucker, S., Ogunfowora, B., & Ehr, D. (2016). Safety in the c-suite: How chief executive officers influence organizational safety climate and employee injuries. *Journal of Applied Psychology, 101*(9), 1228–1239. https://doi.org/10.1037/apl0000116.

Turner, N., Granger, S., Tucker, S., Deng, C., & Kelloway, E. K. (2021). Parents' work injuries and children's mental health: The moderating role of children's work centrality. *Journal of Safety Research, 77*, 61–66. https://doi.org/10.1016/j.jsr.2021.02.002.

Turner, N., Stride, C. B., Carter, A. J., McCaughey, D., & Carroll, A. E. (2012). Job demands–control–support model and employee safety performance. *Accident Analysis & Prevention, 45*, 811–817. https://doi.org/10.1016/j.aap.2011.07.005.

Van der Doef, M., & Maes, S. (1999). The job demand-control (-support) model and psychological well-being: A review of 20 years of empirical research. *Work & Stress, 13*(2), 87–114. https://doi.org/10.1080/026783799296084.
Van Zelst, R. H. (1954). The effect of age and experience upon accident rate. *Journal of Applied Psychology, 38*(5), 313–317. https://doi.org/10.1037/h0062548.
Visser, E., Pijl, Y. J., Stolk, R. P., Neeleman, J., & Rosmalen, J. G. M. (2007). Accident proneness, does it exist? A review and meta-analysis. *Accident Analysis & Prevention, 39*(3), 556–564. https://doi.org/10.1016/j.aap.2006.09.012.
Vroom, V. (1964). *Motivation and work*. Wiley.
Williams, D. R., & Mohammed, S. A. (2009). Discrimination and racial disparities in health: Evidence and needed research. *Journal of Behavioral Medicine, 32*(1), 20–47. https://doi.org/10.1007/s10865-008-9185-0.
Yule, S., Flin, R., Paterson-Brown, S., & Maran, N. (2006). Non-technical skills for surgeons in the operating room: A review of the literature. *Surgery, 139*(2), 140–149. https://doi.org/10.1016/j.surg.2005.06.017.
Zohar, D. (1980). Safety climate in industrial organizations: Theoretical and applied implications. *Journal of Applied Psychology, 65*(1), 96–102. https://doi.org/10.1037/0021-9010.65.1.96.
Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. *Accident Analysis & Prevention, 42*(5), 1517–1522. https://doi.org/10.1016/j.aap.2009.12.019.

Steve Granger, MSc (Manitoba) is a PhD candidate in the Haskayne School of Business, University of Calgary, Canada. His research focuses on workplace safety, proactivity, and resilience among gig workers.

Nick Turner, PhD (Sheffield) is Professor and Distinguished Research Chair at the Haskayne School of Business, University of Calgary, Canada. His research focuses on work design, transformational leadership, and occupational health and safety.

Journal of Safety Research 82 (2022) 1–12

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

An assessment of the effect of green signal countdown timers on drivers' behavior and on road safety at intersections, based on driving simulator experiments and naturalistic observation studies

Wei Yan[a], S.C. Wong[a,c], Becky P.Y. Loo[b,c], Connor Y.H. Wu[d], Helai Huang[e], Xin Pei[f], Fanyu Meng[g,h,*]

[a] Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China
[b] Department of Geography, The University of Hong Kong, Pokfulam Road, Hong Kong, China
[c] Guangdong - Hong Kong - Macau Joint Laboratory for Smart Cities, China
[d] Department of Geospatial Informatics, College of Arts and Sciences, Troy University, Troy, AL, USA
[e] School of Traffic and Transportation Engineering, Central South University, Changsha, China
[f] Department of Automation, Tsinghua University, Beijing, China
[g] Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen, China
[h] Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, China

ARTICLE INFO

Article history:
Received 4 June 2021
Received in revised form 10 October 2021
Accepted 15 April 2022
Available online 27 April 2022

Keywords:
Signalized intersection safety
Green signal countdown timer (GSCT)
Driving behavior

ABSTRACT

Introduction: Motor-vehicle crashes at signalized intersections are a significant traffic safety problem. To address this problem, many Asian cities have installed signal countdown displays at signalized intersections, aiming to assist drivers to make correct decisions in response to traffic signals. *Method:* In this study, we assessed the short-term and long-term effects of green signal countdown timers (GSCTs) on road safety, using a combination of driving simulator experiments and naturalistic observations. *Results:* In our driving simulator experiments, 80 participants drove at 50 km/h in scenarios in which a car either approached a signalized intersection alone or following another car. In naturalistic observations, short-term (1-week) and long-term (1-year) intersection safety in the presence and absence of GSCTs were compared. These observations revealed that GSCTs reduced the number of red-light-running violations over the short term, but not over the long term. In fact, GSCTs appeared to lead to an overall increase in rear-end crash risk at intersections, as their presence resulted in drivers exhibiting more sudden acceleration and braking, and altered intersection-crossing speeds and patterns. *Conclusions*: The results suggest that GSCTs worsen safety at signalized intersections, and thus their removal should be considered.

1. Introduction

Traffic signals are commonly used at intersections to control traffic flow, minimize the number of conflicts, and improve traffic safety. However, traffic crashes at signalized intersections remain a significant problem. Approximately 40% of all traffic crashes in the United States occurred at intersections (NHTSA's National Center for Statistics and Analysis, 2010). The corresponding share was about 26.5% in Hong Kong (Road Traffic Accident Statistics, 2016). In an attempt to increase safety at signalized intersections, countdown display signals were installed alongside conventional traffic signals (Chen et al., 2009). For instance, green signal countdown timers (GSCTs) are, like the yellow signal phase, intended to provide drivers with additional information that allows them to respond to signals at intersections during the last few seconds of a traffic cycle (Huang et al., 2014).

However, whether GSCTs improve safety at intersections is a matter of debate. To illustrate, some studies have suggested that GSCTs negatively affect road safety (Mussa et al., 1996; Chen et al., 2015; Devalla et al. 2015; Lum & Halim, 2016). Lum and Halim (2006) evaluated the long-term effects of GSCTs based on before-and-after study data and found that GSCTs increased the rate of rear-end collisions. Mussa et al. (1996) found that drivers used the information from GSCTs to "race the clock." Devalla et al. (2015) have also been skeptical about the positive value of GSCTs. In contrast, Yu and Shi (2015) found that GSCTs had a positive effect on road safety by reducing the number of cars that crossed a signalized intersection. Similarly, Limanond et al.

* Corresponding author at: Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen, China.

E-mail address: fymeng91@gmail.com (F. Meng).

https://doi.org/10.1016/j.jsr.2022.04.001
0022-4375/

(2010) and Factor et al. (2012) have found that GSCTs were favored by a majority of local drivers in a public opinion survey. However, the lack of comprehensive and systematic analysis on GSCTs means that the net effects of their installation on intersection safety remain uncertain (Fu et al., 2016).

Regarding other benefits, GSCTs have been found to increase the number of early stops (Papaioannou & Politis, 2014), to reduce the numbers of red-light-running violations (Sharma et al., 2011) and right-angle collisions (Köll et al., 2004; Chiou & Chang, 2010), and to increase safety for left-turning drivers (Liu et al., 2012). Moreover, GSCTs may reduce start-up time (Ma, 2010) and improve traffic flow and through-movement (Limanond et al., 2009; Ni & Li, 2013; Sun et al., 2013), although these effects appear to be limited in magnitude (Liu et al., 2012). Regarding other disadvantages, in larger dilemma zones, GSCTs may increase the risk of rear-end collisions (Köll et al., 2004; Chiou & Chang, 2010; Ni & Li, 2013, Huang et al., 2014), and may lead more drivers to enter intersections during the later portions of a yellow signal or during a red signal (Long et al., 2011), with such entry involving acceleration (Ma et al., 2010). Similarly, GSCTs have been found to reduce saturation headway (Limanond et al., 2009; Limanond et al., 2010; Liu et al., 2012; Sun et al., 2013; Fu et al., 2016), which is another safety concern.

The above findings mostly came from field/observational studies of intersections before and after GSCT installation (Lum & Halim, 2006; Chiou & Chang, 2010; Sharma et al., 2011; Sun et al., 2013), or with and without GSCTs (Köll et al., 2004; Limanond et al., 2009; Chiou & Chang, 2010; Limanond et al., 2010; Long et al., 2011, Long et al., 2013; Ni & Li, 2013; Ma et al., 2010; Papaioannou & Politis, 2014). Video cameras were used to record naturalistic observational data of drivers' behavior at signalized intersections (Lum & Halim, 2006; Chiou & Chang, 2010; Ma et al., 2010; Long et al., 2013; Papaioannou & Politis, 2014; Yu & Shi, 2015). Furthermore, driving simulator–based experimental studies have also been conducted to study drivers' behaviors at signal intersections with and without GSCTs (Islam et al., 2016; Islam et al., 2017). For example, Islam et al. (2017) examined the driving behaviors of 55 drivers in the United States in a simulator experiment of multiple signal intersections with GSCTs, and concluded that GSCTs increased intersection safety by increasing drivers' propensity to stop and decreasing their average deceleration rates. In comparison with naturalistic observation studies, driving simulator–based experimental studies allow the capture of detailed information on each driver's driving behavior under specific circumstances, such as signalized intersections with GSCTs (Lee et al., 2002; Yan et al., 2007; Yan et al., 2018; Meng et al., 2019). Overall, the driving simulator–based experiments and naturalistic observation are complementary. The former is a safe and efficient method for simulating test scenarios within controlled parameters and simplified settings, whereas the latter involve realistic conditions and consider the persistence of an effect.

In this study, we combined the driver-simulator experiments with naturistic observations to examine the effects of GSCTs. We examined the short-term and long-term effects of GSCTs on several key indicators, including the number of red-light-running violations (Lum & Halim, 2006; Chiou & Chang, 2010; Limanond et al., 2010) and stop-decisions (Mussa et al., 1996; Lum & Halim, 2006; Long et al., 2011; Huang et al., 2014; Papaioannou & Politis, 2014). We also determined the effects of GSCTs on a range of variables, namely on the distance of a drivers' decision point from the stop line (Köll et al., 2004; Chiou & Chang, 2010; Long et al., 2011; Huang et al., 2014); on drivers' speed at the decision point or their approach speed, and on average traffic-flow speeds (Köll et al., 2004; Chiou & Chang, 2010; Ma et al., 2010; Ni & Li, 2013; Huang et al., 2014; Wu, 2014; Devalla et al., 2015); on time spent by drivers' at the decision point, and on drivers' reaction time (Köll et al., 2004; Chiou & Chang, 2010; Ni & Li, 2013; Huang et al., 2014); and on drivers' acceleration or deceleration behavior (Wu et al., 2009; Ni & Li, 2013) and ability to maintain speed (Wu et al., 2009). We analyzed the data using statistical tests of proportion (Lum & Halim, 2006, Ni & Li, 2013), fixed-effects logistic regression models (Köll et al., 2004; Wu et al., 2009; Chiou & Chang, 2010; Long et al., 2011; Wu, 2014), and analyses of variance (Limanond et al., 2009; Liu et al., 2012; Huang et al., 2014). The combined research methodology enables us to better determine the effects of GSCTs on driver behavior at intersections.

Following this introduction on literature review and the research gap, the next section describes the methods used in this study. It is followed by the results of the data analysis. Finally, we discuss the road safety implications and make recommendations to transport authorities about the GSCTs.

2. Methods

In this study, driving simulator experiments and naturalistic observation studies were combined to comprehensively assess the effects of GSCTs on traffic safety at intersections.

2.1. Driving simulator experiments

2.1.1. Recruitment and randomization

The recruitment and experimental phase lasted from August to December 2014. Permanent Hong Kong residents with valid full drivers licenses were contacted via online advertising, leaflet recruiting, a network of driver associations, and the online Research Participants Registration System, and invited to perform a driving simulator trial. Applicants who could not operate the driving simulator or experienced simulator sickness were excluded at the trial stage. The 80 drivers who were ultimately recruited consisted of 74 males and 6 females, were aged from 20 to 64 years (mean, 39.3 years; SD, 11.9 years), and comprised an even distribution of age groups, as shown in Table 1. All participants had a clear concept of the GSCT but had never driven on roads on which GSCTs were installed.

Most participants were frequent drivers who drove for an average of 11.19 h per week (SD, 18.08 h), and 12 participants were full-/part-time professional drivers. Each participant was invited

Table 1
Summary of participants' demographic data (sample size = 80).

Factor	Attribute	Frequency	Proportion (%)
Age (y)	18–24	9	11.25
	25–29	10	12.50
	30–34	12	15.00
	35–39	11	13.75
	40–44	10	12.50
	45–49	10	12.50
	50–54	7	8.75
	55 or older	11	13.75
Sex	Female	6	7.50
	Male	74	92.50
Marital status	Single	34	42.50
	Married	46	57.50
Car owner	Yes	39	48.75
	No	41	51.24
Education level	Primary	1	1.25
	Secondary	30	37.50
	Tertiary	49	61.25
Years license held	Less than 3 years	12	15
	3 to 10 years	19	23.75
	More than 10 years	49	61.25
Occupation as driver	Full time	6	7.5
	Part time	6	7.5
	No	68	85

to attend two experimental sessions, which took place over 2 days to minimize the effects of driver fatigue on test credibility. Table 1 summarizes the participants' demographic information.

The experimental procedures were explained to all the participants before the simulation study, and they were also informed that the experiment lasted approximately 10 h. The drivers received HK$50 for each hour of their participation. In each session, the participants had breaks between every two test scenarios, during which they were encouraged to rest, nap, and engage in stretching exercises in the laboratory waiting area, to minimize driving fatigue. The experiment was approved by the Human Research Ethics Committee for Non-Clinical Faculties of the University of Hong Kong.

2.1.2. Experimental design

A desktop-based driving simulator (XPDS 300 Driving Simulator, Version 1.6) with XPDS 2.3.1 software (especially designed for signal countdown scenarios) was used in the GSCT experiments. The simulator system consisted of a driving kit (a steering wheel, a pedal kit, and a driver's seat) and three 27-inch high-definition monitors. Two scenarios were used: a test vehicle traveling alone, and a test vehicle following another vehicle. Fig. 1 shows an example of the former scenario.

Fig. 1. Example of the "vehicle traveling alone" simulated scenario.

In the four-phase driving simulator test, we set the speed limit to 50 km/h to replicate urban road conditions in Hong Kong. Phase 1 was an introduction, in which the purposes and risks of the study were explained to the participants; phase 2 was a 10-min warm-up session, in which participants practiced driving on the simulator to familiarize the participants with the driving environments on the driving simulator; phase 3 was a pretest session, in which driving performance data were collected to record general driving behavior; and phase 4 was a signal display test, which comprised the main body of the study and in which participants' driving performance in scenarios with GSCTs and conventional signal displays were recorded. Fig. 2 illustrates the procedures. Each of the warmup, pretest, and test scenarios were repeated 10 times by each driver and were completed by all participants in a random counter-balanced order.

In Scenario 1, on each test run, the participants were asked to accelerate to reach the required speed and to then maintain that speed along the road. When the test vehicle passed the trigger point (see Fig. 3), the signal display appeared on the screen. As shown in Fig. 3 and Fig. 4(a), the participants were then required to maintain the same speed along a street segment with a visibility length of 200 m and then respond to the signal by either decelerating to a halt or passing through the signalized intersection.

Similarly, on each test run in Scenario 2, the participants were asked to accelerate to reach the required speed and to then follow the lead car to a signalized intersection. The lead car accelerated to 50 km/h and then either maintained this speed until the end of the run (a "go" decision) or halted at the stop line of the signalized

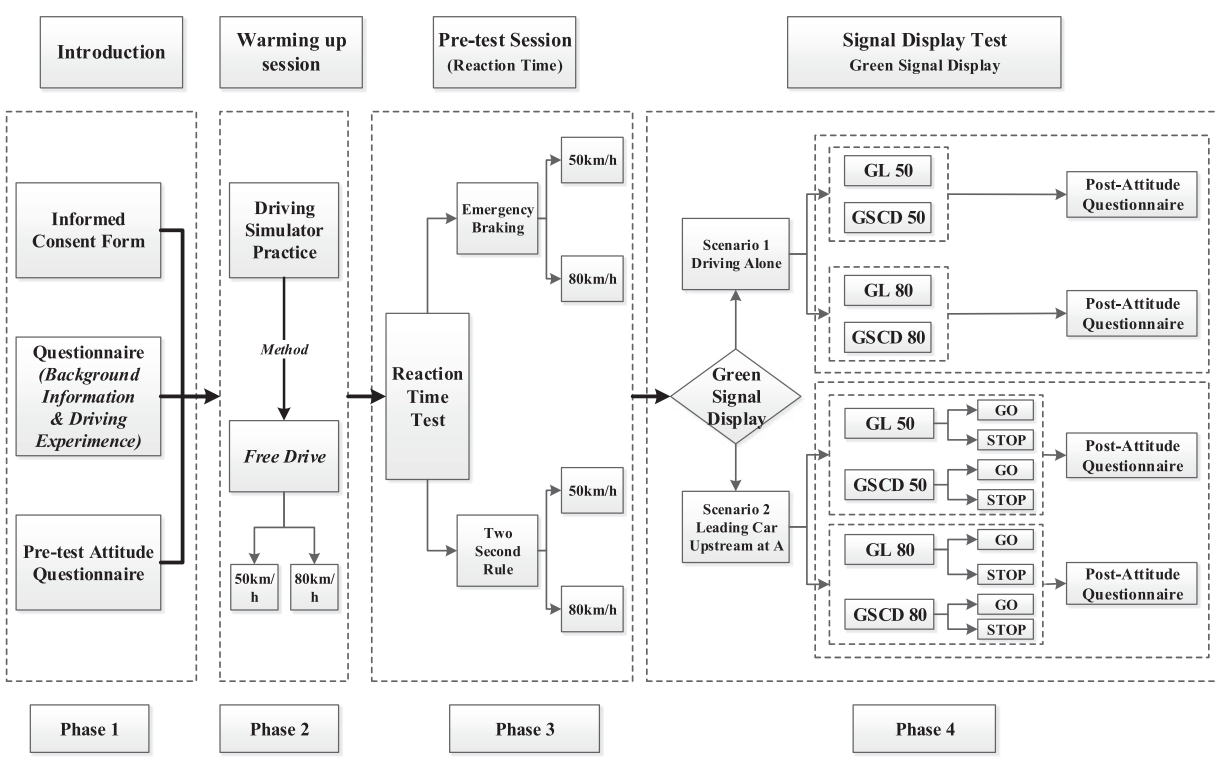


Fig. 2. Green signal countdown display experimental procedure (To focus the analyses and discussion, the comparison of different driving speeds is not discussed in this paper.).

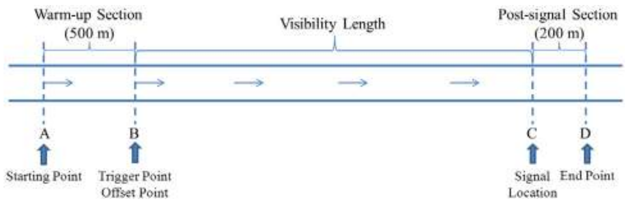


Fig. 3. Geometric design of the signal countdown display scenarios.

intersection (a "stop" decision), where this stop/go decision was set randomly with an equal chance. If "stop" was selected, the lead car began to decelerate at a specified rate and at an appropriate location to achieve a smooth stop at the stop line. Each participant thus made decisions in response to the traffic signals and the

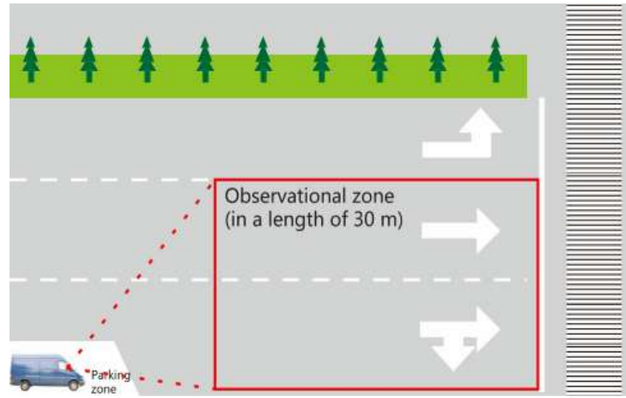


Fig. 5. Observational zone in the naturalistic survey.

GL at 50 km/h

GSCT at 50 km/h

(a) Scenario 1: driving alone

GL at 50 km/h

GSCT at 50 km/h

(b) Scenario 2: following a car

Fig. 4. Driving scenarios in the driving simulator (50 km/h).

movement of the lead car and either decelerated to a stop at the stop line (behind the lead car, or alone) or passed through the intersection (behind the lead car). Note that the test car has the opportunity to decide to stop or go only when the leading car decides to go, and the analyses of Scenario 2 are thus based only on the tests in which the lead car decides to pass through the intersection.

2.2. Naturalistic observations

2.2.1. Location selection

Maoming is one of the most populous cities in western Guangdong province. GSCTs were first installed in Maoming in 1997, and by 2008 the number of intersections with GSCTs had increased to 20. Thus, we assumed that drivers in Maoming were familiar with GSCTs, and performed a repeated cross-sectional observational study at an intersection in Maoming that was equipped with a GSCT, with permission from the local traffic security bureau. A video camera was installed to capture the traffic flow moving in three lanes, from east to west, on You Cheng Si Road. The average traffic volume of this main road was more than 2,000 vehicles per hour, with a peak of 3,000 vehicles per hour. The speed limit was 40 km/h, as set by the local transport department.

2.2.2. Observational survey

The observation zone in the survey extended 30 m from the stop line (Fig. 5). Two through-lanes were observed (the left-turn lane was excluded). Videos were recorded during three observational periods: a GSCT-installed period from December 21 to 22, 2009; a short-term GSCT-uninstalled period (during an uninstallation period that lasted for 1 week) from December 28 to 29, 2009; and a long-term uninstalled period (during an uninstallation period that lasted for 1 year) from December 27 to 28, 2010. None of these periods coincided with public holidays in China. The videos were recorded during morning and afternoon non-peak periods, from 10:00 am to 11:00 am and from 3:00 pm to 4:30 pm, and captured the performance of drivers of coaches, tractors, medium-sized buses, trucks, motorcycles, mopeds, tricycles, bicycles, and cars. The drivers of vehicles passing through this study area are considered to be randomly sampled, and their knowledge and experience with GSCTs thus represents the average level in Maoming. Assuming this average level remains constant during the study period, the variable (i.e., the drivers' level of familiarization with the GSCTs) is well controlled in this case.

2.3. Correlation and comparisons

2.3.1. Correlations and complementarity

The effects of GSCTs on intersection safety were examined using a driving simulator experiment and a naturalistic observation study, which are complementary approaches whereby one approach compensates for the weakness of the other. Various aspects of drivers' responses were scrutinized, particularly during the transition of signals from green to yellow and from yellow to red. Data on drivers' performance related to road safety, red-light-running violations, intersection-crossing decisions, and intersection-crossing speed were obtained using both approaches and compared. The complementarity of these two approaches helps to ensure the reliability of the findings.

2.3.2. Comparisons

Multiple scenarios of driving alone or following a vehicle were used in the driving simulator to simulate the behavior of a lead vehicle and a following vehicle on a road. All possible conditions at the intersections were included in a random condition sequence. Thus, this controlled experiment collected accurate data on drivers' behavior and on their pre-and post-study attitudes toward GSCTs. The naturalistic observations revealed the real decisions made by drivers at the intersection, in contrast to the hypothetical scenarios used in the simulation. These naturalistic data were used to analyze the short- and long-term effects on drivers' behavior as a function of time since the removal of GSCTs.

To account for the absence of GSCTs in Hong Kong, the sequence of study conditions in Maoming City was counter-balanced to eliminate the possible effects of consequence. Thus, the short-term effects and the relatively long-term effects of the uninstallation of GSCTs were assessed by comparison with a period in which GSCTs were installed. Table 2 presents the parameters examined in the driving simulator experiments and naturalistic observations study.

3. Data

Two sources of data were used in our analysis. Source 1 was the driving simulator experiments. The simulation data were recorded in a 30-Hz sampling frame in each run, which comprised four stages: acceleration, speed maintenance, response to signal displays, and crossing the intersection to finish the test. The data on safe-driving behavior was sampled and analyzed for each of the 5,760 test runs and comprised data driving speed, deceleration, driving stability, crossing speed, red-light-running violations, and drivers' stop/go decisions at a signalized intersection. We compared the drivers' response at signalized intersections with either GSCT or conventional traffic light (CTL) scenarios, when driving at 50 km/h alone or following another vehicle.

In addition, naturalistic observation data were collected during three periods at a signalized intersection. The driving behaviors were recorded: (a) when the GSCTs were in operation, and (b) 1 week after their uninstallation by the local traffic security bureau, as a short-term effect study. The same observational survey with the same settings at the same intersection and the same time of the day was repeated 1 year after the GSCTs' uninstallation, as a long-term effect study. The indicators of driving speed, intersection-crossing speed, red-light-running violations, and driving decisions sampled every 1.5 h were analyzed. In all, 131, 157, and 145 runs (only vehicles included) were observed with the GSCTs installed, after the GSCTs had been uninstalled for 1 week, and after the GSCTs had been uninstalled for 1 year, respectively. As in the driving simulator experiments, we observed drivers' decision-making behaviors at a signalized intersection in terms of their stop/go decisions and red-light-running violations. The effects of GSCTs were assessed in both datasets and compared.

4. Results

In this section, we present and compare the data on red-light-running violations, intersection-crossing speed, and stop/go deci-

Table 2
Comparison of driving simulator experiments and naturalistic observation study.

	Driving Simulator	Naturalistic Observation
Red-light-running violation	√	√
Intersection-crossing speed	√	√
Acceleration/deceleration pattern	√	√
Heterogeneity	√	×
Controlled variable study (scenario)	√	×
Counter-balanced sequence	√	×
Short-/long-term study	×	√
Realistic	semi-realistic	√
Drivers' (un)familiarity with GSCTs	unfamiliar	familiar

sion making at signalized intersections in all scenarios in the driving simulation experiments and in the naturalized observation study.

4.1. Decision zone

The participants' average reaction time was measured in the simulated driving test of an emergency braking scenario. The maximum deceleration set in the driving simulation system was $a = 3.886\ \text{m/s}^2$. The yellow-signal display phase was set at $t_a = 3$ s, according to the timing situation in Hong Kong. Calculation of the dilemma zone was applied to Scenarios 1 and 2 at the target speed of 50 km/h; the results are given in Table 3. Both scenarios fell within the option zone, a zone in which the participants decide whether to stop or go and typically exhibit indecision upon seeing the yellow signal displays (Parsonson, 1992; Gates et al., 2007; Zhang et al., 2014).

To measure the effects of GSCTs on participants' decision-making in the option zone, statistical tests of the proportions of vehicles in the option zone were performed to analyze the differences between the effects of CTLs and GSCTs. Table 4 shows that when the participants were following another vehicle, the number of vehicles in the option zone approaching an GSCT intersection was significantly greater (at the 5% significance level) than that in the option zone approaching a CTL intersection. Thus, the installation of a GSCT resulted in unsafe conditions for all road users. In Scenario 1, the corresponding difference was not significant.

Table 3
Calculated zone areas and definitions for the option zone (driving simulator experiment).

Safety parameters	Scenario 1: Driving alone	Scenario 2: Following a vehicle
Required speed (km/h)	50	50
Average reaction time (s)	0.63	0.76
Minimum reacting and safe-stopping distance (m)	33.63	35.41
Maximum yellow-light-passing distance (m)	41.67	41.67
Calculated dilemma zone definition	Option	Option

4.2. Red-light-running violations

The data were analyzed to evaluate the effects of GSCT installation on drivers' decision-making compared with CTLs. The observed data for the proportions of red-light-running violations in both the simulation and naturalistic observation were analyzed to determine whether there are statistically significant differences between these data. Table 5 gives the number of cases and the statistical significances for the CTL and GSCT conditions.

Analysis of the numbers of red-light-running violations indicates that GSCT installation decreases the overall number of violations (at the 1% significance level) relative to CTL installation in both simulation scenarios, implying that a GSCT makes the road safer for participants who are not previously familiar with these devices. However, our on-road naturalistic observation study of long-term effects found no statistically significant difference between the number of red-light-running violations when the intersection was installed with a GSCT or a CTL.

The explanatory variables that could influence the tendency for drivers to run red lights were explored using standard general, fixed-effects, and random-effects logistic regressions (Washington et al., 2020). Collinearity was identified among the explanatory variables in the dataset, and an optimal set of explanatory variables was explored for the regression. The log–likelihood and Akaike information criterion values indicated that the fixed-effects logistic model was optimal for both scenarios. The results in Table 6 suggest that the variables of yellow distance (a vehicle's distance from an intersection when the signal changes to yellow; 1% significance level), yellow speed (the instantaneous speed of a vehicle when the signal changes to yellow; 1% significance level), and treatment of GSCT installation (installed GSCT is 1, uninstalled is 0; 1% significance level) all significantly affect the tendency of participants to run red lights.

In summary, because the GSCT installation can be regarded as a short-term installation in the driving simulator test where intermediate responses of the drivers to the installation are collected, the results show that installation of a GSCT would reduce the number of red-light-running violations in the short term, but its long-term effects on safety remain questionable. Moreover, even in the short-term observation, when the traffic signal turned yellow, a greater distance to the stop line or a greater yellow speed can increase the probability of a driver running a red light.

Table 4
Number and statistical tests of proportion of vehicles in the dilemma/option zone (driving simulator experiment).

In option/dilemma zone			Scenario 1: Driving alone		Scenario 2: Following a vehicle	
			No.	Z_{stat}	No.	Z_{stat}
50 km/h	CTL	Option	179	−0.58	255	−2.24*
	GSCT		189		304	

*Significant at the 5% level.
Z-critical at 95% confidence level ($\alpha = 0.05$, two-tailed) is ± 1.96, and at 99% confidence level ($\alpha = 0.01$, two-tailed) is ± 2.58.

Table 5
Analysis of proportions of red-light-running violations for CTL and GSCT conditions.

Description of tests		Test statistics (Z_{Stat}) and significance at a 95% level		Red-light-running violation (n)	
		Z_{Stat}	SIG	CTL	GSCT
Driving simulator	Scenario 1: Driving alone	2.78	**	40	19
	Scenario 2: Following a vehicle	3.71	**	51	20
Naturalistic observation	Short term	0.94	-	4	6
	Long term	0.81	-	4	6
	Uninstalled in short vs. long term	−0.11	-	4	6

**Significant at the 1% level.
Z-critical is ±2.58 at the 99% confidence level ($\alpha = 0.01$, two-tailed).

Table 6
Explanatory variables of a fixed-effect logistic regression model of the tendency of participants to run red lights at intersections with GSCTs (driving simulator experiment).

Red-light-running violation variables	Scenario 1: Driving Alone			Scenario 2: Following a car		
	Coeff.	SE	Sig	Coeff.	SE	Sig
Yellow distance	+0.104	0.007	**	+0.095	0.007	**
Yellow speed	+0.045	0.012	**	+0.101	0.013	**
Treatment (GSCT)	−0.780	0.172	**	−0.438	0.144	**
Log likelihood	−471.505			1500.490		
AIC	953.010			1053.841		

**Significant at the 1% level. Coeff: coefficient. SE: standard error.

Table 7
Analysis of numbers of stop/go decisions in GL- and GSCT-installed scenarios.

Description of test		Test statistics (Z_{Stat})		Stop Cases (n)	
		Z_{Stat}	SIG	CTLs	GSCTs
Driving simulator	Scenario 1: Driving alone	6.71	**	417	564
	Scenario 2: Following a vehicle	5.39	**	1,190	1,348
Naturalistic observation	Short term	3.12	**	101	106
	Long term	1.28	-	108	106
	Uninstalled in short vs. long term	−1.91	-	101	108

**Significant at the 1% level.
Z-critical is ± 2.58 at the 99% confidence level ($\alpha = 0.01$, two-tailed).

4.3. Stop/go decisions at signalized intersections

The numbers of stop/go decisions were analyzed in both the simulation experiments and naturalistic observations; and the variables that influenced the drivers' decisions in the GSCT-installed case and the CTL-installed case were compared. Table 7 gives the number of stop/go decisions and statistical significance of these results.

Analysis of the numbers of stop/go decisions indicates that GSCT installation increases the likelihood of stop decisions at the 1% significance level in all scenarios of the simulation experiments, implying that the installation of GSCTs can guide participants who are not familiar with them to drive more conservatively in the short term. In our naturalistic observation study, we found that the short-term removal of GSCTs encouraged drivers to make more go than stop decisions, but this effect disappeared over the long-term (1 year), as the number of stop decisions returned to the previous level.

The explanatory variables that may affect a driver's decision to go were explored using the same three statistical methods (i.e., general, fixed-effects and random-effects logistic regression models) used for exploring those that affected a driver's decision to run a red light. A fixed-effects logistic regression model was found to be superior to other models. The results reveal that the variables of yellow distance, yellow speed, and the treatment (GSCT installation) all significantly affect a participant's stop/go decision (all at the 1% significance level). When a participant was following another vehicle, the factors of average speed (maintained speed; 1% significance level) and driving-maneuver stability (lateral-lane deviation; 5% significance level) also increased the probability of a participant making a go decision at an intersection with a GSCT installed, as shown in Table 8.

In summary, our results indicate that GSCTs have significantly positive short-term effects on a drivers' tendency to make stop decisions, but no significant long-term effects were observed.

4.4. Intersection-crossing pattern and intersection-crossing speed

Intersection-crossing pattern and speed are indicators of road safety at intersections. In our driving simulator experiment, we analyzed the overall intersection-crossing speed, as shown in Table 9, and found a statistically significant reduction in the intersection-crossing speed for cases in which participants decided to go (1% significance level when driving alone; 5% significance level when following a vehicle). The installation of a GSCT at an intersection led to significant decelerations in intersection-crossing speeds in all runs in both scenarios, compared with a CTL.

The results show that the average intersection-crossing speed reduced with the installation of a GSCT. However, the relatively high standard deviation of the GSCT condition indicates that intersection-crossing speeds varied widely and that the average speed for all runs was not representative of on-road conditions. A detailed investigation of the dataset revealed that the intersection-crossing speed could be compared more specifically based on intersection-crossing patterns, including acceleration and deceleration, in the driving simulator experiments, which pro-

Table 8
Explanatory variables that affected the participants' stop/go decisions at an intersection with a GSCT installed (driving simulator experiment).

Stop/go decision	Scenario 1: Driving alone			Scenario 2: Following a car		
	Coeff	SE	Sig	Coeff	SE	Sig
Average speed	+0.052	0.010	**			
Driving lane undulation	+0.198	0.080	*			
Yellow distance	−0.104	0.007	**	−0.095	0.007	**
Yellow speed	+0.045	0.012	**	+0.101	0.013	**
Treatment (GSCT)	−0.780	0.172	**	−0.438	0.144	**

**Significant at the 1% level; *Significant at the 5% level. Coeff: coefficient.

Table 9
Overall intersection-crossing speed in "go" cases in the CTL and GSCT conditions (driving simulator experiment).

Scenario		95% CI for Mean		Mean	SD	Abs. diff.	*F*-statistic
		Lower	Upper			(% diff.)	
1: Driving alone	CTL	50.92	52.17	51.54	7.40	−2.89	17.540**
	GSCT	47.31	49.99	48.65	13.56	(-7.08%)	
2: Following a vehicle	CTL	50.8	51.5	51.15	4.83	−0.85	5.817*
	GSCT	49.66	50.94	50.3	7.84	(-2.25%)	

**Significant at the 1% level; *Significant at the 5% level.
SD: standard deviation; CI: confidence interval.

vided accurate driving-performance data. Similar intersection-crossing patterns were found in the naturalistic observation data, as shown in Tables 10, 11, and 12.

In Table 10, the intersection-crossing speeds in the driving simulator experiment with CTL and GSCT conditions are grouped in terms of their acceleration and deceleration values as vehicles approached designated intersections. There was a more noticeable bimodal distribution of intersection-crossing speeds with GSCT installation than with CTL installation. The presence of GSCTs increased the differences between intersection-crossing speeds in each condition, and increased the variability in participants' driving behavior. In the acceleration runs, the average intersection-crossing speed was higher for intersections installed with GSCTs than CTLs, although the difference was not statistically significant. In the deceleration runs, the average intersection-crossing speed decreased significantly (at the 1% statistical significance level) for GSCT intersections. We also analyzed the intersection-crossing patterns in these runs, as shown in Table 11, but found no statistically significant differences between the patterns in the two conditions in either scenario.

Our driving simulator results indicate that the installation of GSCTs does not affect participants' intersection-crossing patterns when drivers are following other vehicles, but significantly affects (1% significance level) their intersection-crossing speed and decel-

Table 10
Comparison of intersection-crossing speeds between CTL and GSCT conditions (driving simulator experiment).

Scenario	Intersection Decision	Signal	95% CI for Mean		Mean	SD	Abs. diff.	*F*-statistic	GSCT/ CTL
			Lower	Upper			(% diff.)		
1: Driving alone	Acceleration	CTL	54.05	55.32	54.68	5.89	0.98	2.632	↑
		GSCT	54.57	56.75	55.66	8.30	(−1.76%)		
	Deceleration	CTL	45.67	47.52	46.60	6.81	−7.26	45.910**	↓
		GSCT	37.28	41.39	39.33	13.6	(−18.46%)		
2: Following another vehicle	Acceleration	CTL	52.84	53.58	53.21	4.04	0.43	2.005	↑
		GSCT	53.15	54.12	53.64	4.63	−0.80%		
	Deceleration	CTL	47.17	48.12	47.65	3.97	−2.69	19.889**	↓
		GSCT	43.77	46.14	44.96	8.90	(−5.99%)		

**Significant at the 1% level.
SD: standard deviation; CI: confidence interval.
↑ Speed increased with GSCT; ↓ Speed decreased with GSCT.

Table 11
Comparison of intersection-crossing patterns between GL and GSCT conditions (driving simulator experiment).

Scenario	*a*	CTL		GSCT		Z_{stat}
		No.	%	No.	%	
1: Driving alone	Acceleration (>5%)	93	17.13	99	25.00	−1.35
	Maintain (±5%)	149	27.44	173	43.69	−3.06**
	Deceleration (<−5%)	301	55.43	124	31.31	4.54**
2: Following another vehicle	Acceleration (>5%)	147	20.14	146	25.52	−1.10
	Maintain (±5%)	272	37.26	185	32.34	1.08
	Deceleration (<−5%)	311	42.60	241	42.13	0.11

**Significant at the 1% level.

Table 12
Comparison of intersection-crossing patterns between GSCT and CTL conditions (naturalistic observation).

Crossing Pattern (observation)		GSCT		GSCT replaced by CTL over short term		GSCT replaced by CTL over long term	
		Specific	Overall	Specific	Overall	Specific	Overall
Acceleration	Sudden acceleration	9.23%	40.86%	4.36%	32.67%	2.44%	30.64%
	Mild acceleration	31.63%	28.31%	28.20%			
Maintain	Constant speed	13.75%	13.75%	11.39%	11.39%	14.85%	14.85%
Deceleration	Mild deceleration	39.88%	45.38%	52.76%	55.94%	54.14%	54.52%
	Sudden braking	5.50%	3.18%	0.38%			

eration when they are driving alone. However, when driving alone, more participants overall decided to cross an intersection by accelerating or maintaining a steady speed. In contrast, the proportion of participants who decided to decelerate while crossing an intersection decreased from approximately 55% to less than 33%.

Table 12 shows that the naturalistic observation results were similar. When GSCTs were uninstalled, a much smaller proportion of drivers decided to accelerate to cross the intersection. In particular, the proportion that accelerated suddenly decreased by almost 7%. Meanwhile, most drivers behaved more conservatively and made better evaluations, such that cases of sudden braking decreased by 5%.

Fig. 6 shows the intersection-crossing speeds in the driving simulator experiment and the average intersection-approaching speeds in the naturalistic observation. Both factors were distributed similarly in bimodal patterns when GSCTs were installed in the intersections. (Note that although the distributions for acceleration when driving alone and deceleration following the lead vehicle with GSCTs are long-tail single-modal distributions, there is a tendency toward a typical bimodal distribution). These results show that the additional information provided by the GSCTs influenced the drivers' responses.

In our driving simulator experiments and naturalistic observation studies, drivers consistently showed different responses to GSCTs than to CTLs. Risk-taking drivers were more aggressive, as they tended to accelerate when approaching signalized intersections with GSCTs, and were more willing to take the risk of running a red light. However, risk-averse drivers behaved more conservatively, as they were inclined to decelerate when approaching signalized intersections installed with GSCTs. Such change of behavior among risk-taking drivers can lead to a greater incidence of sudden braking. Moreover, risk-averse drivers tended to slow down while crossing an intersection, which led to a bimodal distribution of intersection-crossing speeds. The bimodal intersection-crossing pattern was also found in the naturalistic observations with GSCTs. The intersection-crossing speed distributions in the experiments and the naturalistic observations were similar: most drivers were relatively risk-averse. Yet, some risk-taking drivers accelerated to cross an intersection when a GSCT was installed.

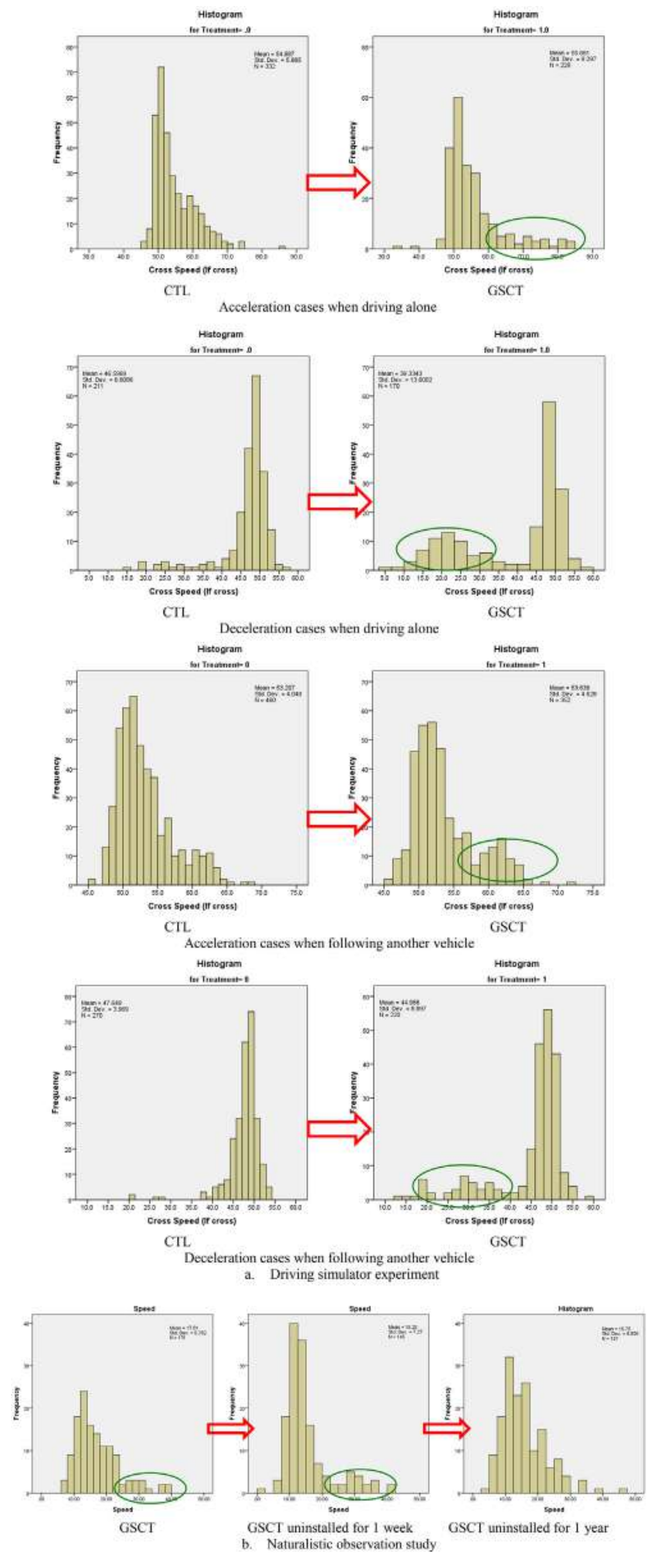


Fig. 6. Comparison of intersection-crossing speeds.

5. Discussion

This study incorporated results from both driving simulator tests and naturalistic observations. Results from these complementary approaches are integrated to examine the effects of GSCTs on red-light-running violations, stop/go decisions, and crossing speed; the factors that contributed to these specific driving behaviors; and participants' attitudes toward GSCTs.

5.1. Decision zone safety

Because driving through an option zone may lead to more rear-end collisions, and driving through a dilemma zone will result in more red-light-running violations, an increase in the proportion of cars in these decision zones is unsafe. In the simulations, GSCTs did not appear to contribute significantly to the increase in the number of option-zone runs when participants were driving alone at 50 km/h. These may have been due to the low speed and the lack of an effect on the participants (from another vehicle). However, the results indicate that GSCTs may cause more driving runs through option zones when drivers are following a vehicle, which represents real road conditions, and that this may lead to collisions and safety problems (Köll et al., 2004; Chiou & Chang, 2010; Ni & Li, 2013; Huang et al., 2014). In other words, the effects of GSCTs can be complex and that their positive effects may not persist over time (Lum & Halim, 2006; Chiou & Chang, 2010).

5.2. Tendency for red-running violations

The tests of red-light-running violations in the driving simulator experiments indicate that the installation of GSCTs decreased the overall number of red-light-running violations. This is consistent with Lum and Halim (2006) and Sharma et al. (2011). However, on-road naturalistic observation studies generally reported that the long-term percentage of red-light-running violations by pedestrians and motorcycles at intersections with GSCTs was higher than that at intersections without CTLs (Chen et al., 2015; Wu, 2014; Long et al., 2013). Moreover, Lum and Halim (2006) found an increase in rear-end collisions based on 4.5- and 7.5-month observational data. Our naturalistic observation also revealed that there was not a statistically significant increase in the proportion of red-light-running violations when GSCTs were uninstalled. These results suggest that the additional information provided by GSCTs does not improve road safety over the long run. In addition, the vehicle speed vehicle distance from the stop line when the signal lights changed to yellow were adversely affected by the presence of GSCTs.

5.3. Decision to go

The tests of the proportion of stop/go decisions indicate that the installation of GSCTs decreased the number of go decisions and increased the number of stop decisions in the driving simulator experiment. This is consistent with the findings of other studies (Chiou & Chang, 2010; Yu & Shi, 2015). In our naturalistic observation, we found that short-term removal of GSCTs encouraged more drivers to decide to go instead of to stop in the reverse-sequence experiment; however, these effects of GSCTs disappeared after 1 year, as the proportion of stop decisions returned to the previous level. This is consistent with the naturalistic observation findings reported in a study in Singapore (Lum & Halim, 2006).

5.4. Crossing pattern and crossing speed

As drivers make decisions on whether to go or stop at signalized intersections, the speed at which they cross an intersection is an important measure of intersection safety (Mussa et al., 1996; Ma et al., 2010). In the short term, the installation of GSCTs led to significant decreases in intersection-crossing speeds for all runs in both scenarios when the cars were traveling at 50 km/h. These results are similar to those of an on-road study of motorcycle traffic conducted by Wu (2014). However, the high standard deviation of our results indicate that the intersection-crossing speeds varied greatly and that the average intersection-crossing speeds for all runs does not represent the reality of road conditions.

Intersection-crossing speeds were grouped in terms of their acceleration and deceleration values as participants approached the intersections. A bimodal distribution of crossing speeds was found to be more prominent when GSCTs were installed. The installation of GSCTs increased the differences between participants' intersection-crossing speeds and increased the variability of participants' driving behavior. Our results also suggest that GSCT installation worsened the bimodality of intersection-crossing speed and increased the difference between the speeds of the acceleration and deceleration patterns. Specifically, risk-averse participants decelerated to stop in response to countdown timing information. However, this well-intentioned action might have led to violations if they cross an intersection at a reduced speed. In contrast, risk-taking participants tended to accelerate and race the countdown timing. This is consistent with the findings of Mussa et al. (1996), Ma et al. (2010), and Devalla et al. (2015), who have all shown that the installation of GSCTs tended to increase intersection-crossing speeds.

The distribution of intersection-crossing speeds at intersections where GSCTs were installed therefore shows that GSCTs increased the differences between the intersection-crossing speeds of participants approaching the signalized intersections (Long et al., 2013). The differences between participants' minimum and maximum speeds, and between participants' rates of sudden acceleration and sudden braking, were greater when intersections had GSCTs installed than when they had CTLs. The GSCTs thus made participants' driving behavior less predictable and their responses more variable: risk-taking participants were more prone to accelerate, whereas risk-averse participants were more likely to decelerate. Participants who were following other vehicles could not predict the behavior of the lead vehicle, and this uncertainty will increase the probability of rear-end collisions when the lead and following vehicles have different responses to the GSCT. The higher proportion of cases with sudden acceleration and sudden braking confirmed that GSCTs amplified the unsafe effects of participants' driving unpredictability during intersection crossings. Thus, although the lower speeds we observed suggest that GSCTs encouraged participants to be more conservative in their driving, this may not have positive effects on safety, as the increased rate of sudden braking observed in the naturalistic observations and the bimodality of driving speeds observed in the driving simulation mean that GSCTs widen speed variability and increase the risk of rear-end collisions at intersections.

6. Conclusions

This study examined the overall effects of GSCTs on driving behavior and road safety by using complementary experimental and observational methods. In the driving simulator study, the controlled conditions compared the influence of GSCTs versus CTLs on driving behaviors, in terms of their influence on the decision zone, the numbers of red-light-running violations, the number of stop/go decisions, intersection-crossing patterns, and intersection-crossing speeds. Eighty participants participated in the simulated driving experiment in scenarios of driving alone and following a vehicle at 50 km/h. In the naturalistic study, on-road observation data were analyzed by the numbers of red-light-running violations and stop/go decisions, and on intersection-crossing patterns and intersection-crossing speeds. Three periods of driver behavior were recorded: at the beginning of the observation, when the GSCTs were in operation; 1 week after GSCT uninstallation by the local traffic security bureau, as a short-term effect study; and one year after GSCT uninstallation. The behavioral data of driving speed, intersection-crossing speed, number of red-light-running violations, and number of driving decisions were sampled every 1.5 h in each period.

The results of the study are as follows.

(1) GSCT installation increased the number of runs that passed through decision zones, which might have led to safety problems over the short- and long-term.
(2) GSCT installation did not reduce the number of red-light-running violations over the long term. Specifically, drivers who were not previously familiar with GSCTs committed fewer red-light-running violations over the short-term, but this behavior did not persist over the long term. The driving performance factors of longer yellow distance and higher approaching speed may have increased the probability of drivers committing red-light-running violations in the short term.
(3) GSCT installation did not reduce the number of go decisions over the long term; only the short-term removal of GSCTs encouraged more drivers to decide to go instead of to stop. The driving-performance factors of shorter yellow distance and higher approaching speed may have also increased the probability of go decisions at intersections in the short term.

(4) GSCT installation worsened the bimodality of intersection-crossing speed, increased the speed differences between two acceleration and deceleration modes, encouraged risk-taking drivers to cross intersections at higher speeds, and caused indecision in risk-averse drivers that resulted in their decelerating during intersection-crossing. These increased rates of accelerated crossing and sudden braking could lead to increases in rear-end collisions at intersections.
(5) Drivers' vague and varied attitudes toward GSCTs could not be interpreted as robust support for GSCT installation, and its implications.

This study reveals that the evidence for positive effects of GSCTs on intersection safety is weak, at best. In particular, although GSCTs have some positive short-term effects, these effects vanish over the long term, and drivers exhibit unpredictable behaviors in response to GSCTs. These include high-risk intersection-crossing decisions and similar misjudgments based on overestimated self-evaluations of driving abilities, which may result in rear-end collisions and other preventable and potentially serious traffic accidents. Thus, GSCTs appear to decrease safety at intersections, especially when drivers have only a vague understanding of the effects of GSCTs, and their use should be reviewed. Therefore, to generally improve intersection safety, it is suggested that GSCTs be installed only at intersections where there are many red-light-running violations. Additionally, a reduced speed limit can be imposed at signalized intersections equipped with GSCTs to prevent an appreciable speed difference that results in conflicting driving behaviors. It is also suggested that drivers be educated and guided in their training to make more conservative decisions when passing through intersections with GSCTs to eliminate the second peak in the overall speed distribution.

It is suggested that future studies acquire a more balanced gender ratio of participants that matches the statistics of the local population.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This research was jointly supported by a Research Postgraduate Studentship and by grants from the University Research Committee of the University of Hong Kong (201309176024), Joint Research Scheme of National Natural Science Foundation of China/Research Grants Council of Hong Kong (Project Nos. 71561167001 and N_HKU707/15), and National Natural Science Foundation of China (Project No. 71671100). The second author was supported by the Francis S Y Bong Professorship in Engineering, and the second and third authors were also supported by the Guangdong - Hong Kong - Macau Joint Laboratory Program of the 2020 Guangdong New Innovative Strategic Research Fund, Guangdong Science and Technology Department (Project No.: 2020B1212030009).

References

Chen, H. Y., Zhou, H. G., & Hsu, P. (2009). What do we know about signal countdown timers? *ITE Journal on the Web, 7*, 72–76.
Chen, P. L., Pai, C. W., Jou, R. C., Saleh, W., & Kuo, M. S. (2015). Exploring motorcycle red-light violation in response to pedestrian green signal countdown device. *Accident Analysis and Prevention, 75*, 128–136.
Chiou, Y. C., & Chang, C. H. (2010). Driver responses to green and red vehicular signal countdown displays: Safety and efficiency aspects. *Accident Analysis and Prevention, 42*, 1057–1065.
Devalla, J., Biswas, S., & Ghosh, I. (2015). The effect of countdown timer on the approach speed at signalised intersections. *Procedia Computer Science, 52*, 920–925.
Factor, R., Prashker, J. N., & Mahalel, D. (2012). The flashing green light paradox. *Transportation Research Part F: Traffic Psychology and Behaviour, 15*, 279–288.
Fu, C., Zhang, Y., Qi, W., & Cheng, S. (2016). Effects of digital countdown timer on intersection safety and efficiency: A systematic review. *Traffic Injury Prevention, 17*, 98–103.
Gates, T. J., Noyce, D. A., Laracuente, L., & Nordheim, E. (2007). Analysis of driver behavior in dilemma zones at signalized intersections. *Transportation Research Record, 2030*, 29–39.
Huang, H., Wand, D., Zheng, L., & Li, X. (2014). Evaluating time-reminder strategies before amber: Common signal, green flashing and green countdown. *Accident Analysis and Prevention, 71*, 248–260.
Islam, M. R., Hurwitz, D. S., & Macuga, K. L. (2016). Improved driver responses at intersections with red signal countdown timers. *Transportation Research Part C: Emerging Technologies, 63*, 207–221.
Islam, M. R., Wyman, A. A., & Hurwitz, D. S. (2017). Safer driver responses at intersections with green signal countdown timers. *Transportation Research Part F: Traffic Psychology and Behaviour, 51*, 1–13.
Köll, H., Bader, M., & Axhausen, K. W. (2004). Driver behaviour during flashing green before amber: A comparative study. *Accident Analysis and Prevention, 36*, 273–280.
Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002). Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors, 44*, 314–334.
Limanond, T., Chookerd, S., & Roubtonglang, N. (2009). Effects of countdown timers on queue discharge characteristics of through movement at a signalized intersection. *Transportation Research Part C: Emerging Technologies, 17*, 662–671.
Limanond, T., Prabjabok, P., & Tippayawong, K. (2010). Exploring impacts of countdown timers on traffic operations and driver behavior at a signalized intersection in Bangkok. *Transport Policy, 17*, 420–427.
Liu, P., Yu, H., Wang, W., Ma, J., & Wang, S. (2012). Evaluating the effects of signal countdown timers on queue discharge characteristics at signalized intersections in China. *Transportation Research Record: Journal of the Transportation Research Board, 2286*, 39–48.
Long, K., Han, L. D., & Yang, Q. (2011). Effects of countdown timers on driver behavior after the yellow onset at Chinese intersections. *Traffic Injury Prevention, 12*, 538–544.
Long, K., Liu, Y., & Han, L. D. (2013). Impact of countdown timer on driving maneuvers after the yellow onset at signalized intersections: An empirical study in Changsha, China. *Safety Science, 54*, 8–16.
Lum, K. M., & Halim, H. (2006). A before-and-after study on green signal countdown device installation. *Transportation Research Part F: Traffic Psychology and Behaviour, 9*, 29–41.
Ma, W. Y., Liu, Y., & Yang, X. G. (2010). Investigating the impacts of green signal countdown devices: Empirical approach and case study in China. *Journal of Transportation Engineering, 136*, 1049–1055.
Meng, F., Wong, S. C., Yan, W., Li, Y. C., & Yang, L. (2019). Temporal patterns of driving fatigue and driving performance among male taxi drivers in Hong Kong: A driving simulator approach. *Accident Analysis & Prevention, 125*, 7–13.
Mussa, R., Newton, C., Matthias, J., Sadalla, E., & Burns, E. (1996). Simulator evaluation of green and flashing amber signal phasing. *Transportation Research Record: Journal of the Transportation Research Board, 1550*, 23–29.
NHTSA's National Center for Statistics and Analysis. (2010). Crash factors in intersection-related crashes: An on-scene perspective. 1200 New Jersey Avenue S.E., Washington, DC 20590.
Ni, Y., & Li, K. (2013). Estimating rear-end accident probabilities at signalized intersections: A comparison study of intersections with and without green signal countdown devices. *Traffic Injury Prevention, 15*, 583–590.
Papaioannou, P., & Politis, I. (2014). Preliminary impact analysis of countdown signal timer installations at two intersections in Greece. *Procedia Engineering, 84*, 634–647.
Parsonson, P. S. (1992). *NCHRP synthesis of highway practice: Signal timing improvement practices* (p. 172). Washington DC: TRB National Research Council.
Road Traffic Accident Statistics. (2016). *Transport Department of the Government of the Hong Kong Special Administrative Region*. Figure 2.12 Road traffic accidents at junction by junction type, junction control type and severity 2016. Available at: URL [http://www.td.gov.hk/filemanager/en/content_4866/f2.12.pdf].
Sharma, A., Vanajakshi, L., Girish, V., & Harshitha, M. (2011). Impact of signal timing information on safety and efficiency of signalized intersections. *Journal of Transportation Engineering, 138*, 467–478.
Sun, W., He, Z., Xia, X., & Xu, F. (2013). Exploring impacts of countdown timers on queue discharge characteristics of through movement at signalized intersections. *Procedia - Social and Behavioral Sciences, 96*, 255–264.
Washington, S., Karlaftis, M. G., Mannering, F., & Anastasopoulos, P. (2020). *Statistical and econometric methods for transportation data analysis*. CRC Press.
Wu, Y.C. (2014). *Motorcycle safety in developing countries: A case study of Maoming, China*. Doctoral dissertation. The University of Hong Kong. Retrieved from http://hdl.handle.net/10722/206703.
Wu, W. J., Juan, Z. C., & Jia, H. F. (2009). Drivers' behavioral decision-making at signalized intersection with countdown display unit. *Systems Engineering - Theory and Practice, 29*, 160–165.
Yan, W., Xiang, W., Wong, S. C., Yan, X., Li, Y. C., & Hao, W. (2018). Effects of hands-free cellular phone conversational cognitive tasks on driving stability based on driving simulation experiment. *Transportation Research Part F: Traffic Psychology and Behaviour, 58*, 264–281.

Yan, X., Radwan, E., & Guo, D. (2007). Effects of major-road vehicle speed and driver age and gender on left-turn gap acceptance. *Accident Analysis & Prevention, 39* (4), 843–852.
Yu, S., & Shi, Z. (2015). Analysis of car-following behaviors considering the green signal countdown device. *Nonlinear Dynamics, 82*, 731–740.
Zhang, Y., Fu, C., & Hu, L. (2014). Yellow light dilemma zone researches: A review. *Journal of Traffic and Transportation Engineering, 1*, 338–352.

Wei Yan received her Ph.D. from the University of Hong Kong, and is currently working as a Transport Specialist in the World Bank. Her research interests include human factors in road safety, drivers' behaviors and sustainable transport.

S.C. Wong is the Francis S Y Bong Professor in Engineering and the Chair of Transportation Engineering in The University of Hong Kong. His research interests include road safety, traffic signals, continuum modeling of traffic equilibrium problems and theory of traffic flow.

Becky Y.P. Loo is a full Professor in the Department of Geography and the Director of Institute of Transport Studies in The University of Hong Kong. Prof. Loo's core research interests are transportation, smart technologies and cities. In particular, she is interested in applying spatial analysis, surveys and statistical methods in analysing pertinent issues related to smart cities, sustainability, regional transport infrastructure and development, transit-oriented development, walkable communities, and road safety.

Connor Wu joined Troy University in August 2018. He was a Postdoctoral Fellow in the Department of Population Health Sciences at Virginia Tech between 2016 to 2018. Before that, he was a Research Associate at the University of Hong Kong (HKU). He earned a Ph.D. degree in Geography from HKU, an MSc in Geomatics (GIS) from the Hong Kong Polytechnic University, and a BEng in Computer Science from the Beijing Normal University. He has experience in data collection and mining, statistics, and spatiotemporal analysis with GIS and remote sensing skills, and his research interests include public health, epidemiology, climate change, and transportation safety.

Helai Huang is a professor of transportation engineering in the School of Traffic and Transportation Engineering, Central South University (CSU). He serves as associate dean of the School, the founding director of CSU Urban Transport Research Centre and Key Laboratory of Smart Transportation of Hunan Province. His research interests include traffic safety, transportation planning and ITS. Dr. Huang is the Editor-in-chief of Accident Analysis and Prevention (Elsevier), an editorial board member of Analytic Method in Accident Research (Elsevier) and Journal of Geography and Regional Planning.

Xin Pei received the B.S. and M.S. degrees from Tsinghua University, China, in 2005 and 2007, respectively, and the Ph.D. degree from The University of Hong Kong in 2011. She is currently a Research Associate Professor with the Department of Automation, Tsinghua University. She has published more than 70 SCI/SSCI/EI indexed articles, and serves as a reviewer for several international journals and conferences, including IEEE TRANSACTION ON INTELLIGENT TRANSPORTATION SYSTEMS, Accident Analysis and Prevention, Transportmetrica, ITSC Conference, and TRB Annual Meeting. Her current research interests include road safety evaluation and driving behavior analysis.

Fanyu Meng is a Research Assistant Professor in Academy of Advanced Interdisciplinary Studies in Southern University of Science and Technology. His Research interests are road safety, intelligent transportation systems and public safety in smart cities.

Journal of Safety Research 82 (2022) 241–250

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

An exploratory study of drivers' EEG response during emergent collision avoidance

Xiaomeng Li [a,b,*], Liu Yang [c], Xuedong Yan [a]

[a] MOT Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044, China
[b] Queensland University of Technology (QUT), Centre for Accident Research and Road Safety-Queensland (CARRS-Q), Kelvin Grove, Queensland, 4059, Australia
[c] School of Transportation, Wuhan University of Technology, Wuhan 430063, China

ARTICLE INFO

Article history:
Received 22 July 2020
Received in revised form 11 May 2021
Accepted 27 May 2022
Available online 7 June 2022

Keywords:
EEG
Collision avoidance
Driving simulator
Driver gender
Driver profession

ABSTRACT

Introduction: EEG (electroencephalogram) has been applied as a valuable measure to estimate drivers' mental status and cognitive workload during driving tasks. However, most previous studies have focused on the EEG features at particular driver status, such as fatigue or distraction, with less attention paid to EEG response in emergent and safety–critical situations. This study aims to investigate the underlying patterns of different EEG components during an emergent collision avoidance process. *Method:* A driving simulator experiment was conducted with 38 participants (19 females and 19 males). The scenario included a roadside pedestrian who suddenly crossed the road when the driver approached. The participants' EEG data were collected during the pedestrian-collision avoidance process. The log-transformed power and power ratio of four typical EEG components (i.e., delta, theta, alpha and beta) were extracted from four collision avoidance stages: Stage 1-normal driving stage, Stage 2-hazard perception stage, Stage 3-evasive action stage, and Stage 4-post-hazard stage. *Results:* The activities of all four EEG bands changed consistently during the collision avoidance process, with the power increased significantly from Stage 1 to Stage 4. Drivers who collided with the pedestrian and drivers who avoided the collision successfully did not show a significant difference in EEG activity across the stages. Male drivers had a higher delta power ratio and lower alpha power ratio than females in both hazard perception and evasive action stages. *Conclusions:* Enhanced activities of different EEG bands could be concurrent at emergent and safety–critical situations. Female drivers were more mentally aroused than male drivers during the collision avoidance process. *Practical Applications:* The study generates more understanding of drivers' neurophysiological response in an emergent and safety–critical collision avoidance event. Driver state monitoring and warning systems that aim to assist drivers in impending collisions may utilize the patterns of EEG activity identified in the collision avoidance process.

1. Introduction

Road crashes lead to a large number of casualties every year, remaining a serious public safety concern worldwide. According to the statistics of World Health Organization (WHO, 2018), road crash is the eighth leading cause of deaths globally, claiming over 1.5 million fatalities and 50 million injuries each year. Given that most casualties due to road crashes are preventable, reducing or even eliminating crashes on roads and creating a safer road environment has been a primary mission of the transportation authorities in many countries. For example, the Vision Zero (or Towards Zero) road safety principles targeting at zero severe injuries on the road have been implemented in many countries such as Sweden, Australia, Canada, Norway, and France (Johansson, 2009). It has been agreed that efforts to improve road safety should be approached from all components of the road transport system, including road users, transportation tools, and road infrastructures.

Human factors have been deemed responsible for the occurrence of a significant proportion of crashes on roads (Petridou & Moustaki, 2000), among which driver awareness of safety and hazards while driving is critical. The reaction and response of drivers in a few seconds before an impending collision play a decisive role in the outcomes of the crash (Harb et al., 2009). Generally, drivers' collision evasive process includes risk perception, decision-making, and action response (Li et al., 2016, 2018). Many previous studies

* Corresponding author at: Queensland University of Technology (QUT), Centre for Accident Research and Road Safety-Queensland (CARRS-Q), Kelvin Grove, Queensland, 4059, Australia.

E-mail addresses: xiaomeng.li@qut.edu.au (X. Li), yang.liu@whut.edu.cn (L. Yang), xdyan@bjtu.edu.cnxdya, n@bjtu.edu.cnxdya, n@bjtu.edu.cn (X. Yan).

https://doi.org/10.1016/j.jsr.2022.05.015
0022-4375/

have focused on drivers' behavioral performance in this process and identified that the action types (i.e., deceleration, acceleration, or turning the steering wheel) and the extent to which the action is mainly applied depend on the situation features and driver characteristics (Li et al., 2019; Markkula et al., 2012). Compared with behavioral performance, the responses in physiological and psychological levels in critical situations are more implicit, and thus harder to be detected. However, according to the risk perception theory, psychological and physiological measures (e.g., eye-movement, EEG, heart rate) show an early indication of an individual's mental status regarding whether and how h/she is going to respond toward an emergent situation (Crundall et al., 2012; Guo et al., 2018; Jones et al., 2014; Li et al., 2018). Up to date, little attention has been paid to drivers' psychological and/or physiological changes in an impending collision situation. A deep understanding of the patterns in drivers' psychological and physiological activities before an emergent collision could help predict the collision outcomes and develop advanced monitoring and warning technologies to assist in collision avoidance.

EEG (electroencephalogram) is the measurement of the brain's spontaneous electrical activity over a period of time by a net of electrodes covered on the scalp surface (Schomer & Da Silva, 2012). It has been used as a valid tool to estimate the subject's mental status and workload while performing specific tasks (e.g., cognitive, visual, or motor tasks; Galán & Beal, 2012). EEG signals are basically classified into four bands according to the frequency, namely, delta (0.5–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), and beta (14–30 Hz) (Yang et al., 2019). Prior research suggested that changes in the activities of different EEG bands represent the different mental states of the subject. For example, the delta band has the highest amplitude and lowest wave frequency among the four bands, and it is frequently observed in adults during sleeping (Ako et al., 2003). Theta band could be observed in drowsy or arousal states (Stenberg, 1992). Alpha and beta bands have been reported to have a reverse relationship with alpha activity decreased and beta activity increased on tasks that required attentional demands (Jap et al., 2011; Schier 2000; Schwartz et al., 1989). Specifically, Ray and Cole (1985) found that increased alpha activity was related to tasks that required little or no attention to the environment, while increased beta activity was associated with emotional and cognitive tasks such as active thinking or concentration.

The road safety research applying EEG measures has mainly focused on two areas. The first area is fatigued driving or driver sleep deprivation (Otmani et al., 2005; Perrier et al., 2016). Many studies have shown that changes in slow-wave activity such as delta and theta waves were related to the transition to fatigue state (Borghini et al., 2014; Xiaoli et al., 2009). These studies demonstrated the potential of using EEG as a neurophysiological indicator of driver fatigue. The second area is distracted driving or driver inattention (Lin et al. 2011; Sonnleitner et al. 2014). Activities in theta, alpha, and beta have been widely used as indicators of cognitive distraction, and it is suggested that changes in the activities of these bands are often a result of brain engagement in multiple tasks (Lin et al., 2011; Sonnleitner et al., 2014). Some other studies have also applied machine learning techniques such as SVM (Support Vector Machine) and CNN (Convolutional Neural Networks) with EEG as one of the inputs to classify specific driving style (e.g., aggressive driving) or driving behavior (e.g., car-following) (Hernández et al., 2018; Yang et al., 2018).

Overall, most previous studies involving driver EEG have focused on driver-status-related safety research, and the status has been primarily limited to driver fatigue or distraction. These studies provide essential support for identifying and detecting hazardous driver status and promoting advanced driver monitoring/warning devices to improve driver safety on roads. In contrast, few studies have investigated drivers' EEG responses in specific driving situations, especially the safety–critical situation when the driver encounters an impending collision. The situation occurs regardless of whether the driver is in a hazardous status (e.g., fatigue or distraction) or not, and the driver response in the situation is directly correlated to the occurrence and severity of a collision. To bridge the gap, this study focuses on drivers' EEG response in an emergent collision situation that is created using an advanced driving simulator, and a detailed examination of the EEG activity within each step of the collision avoidance is conducted.

In collision events, driver characteristics such as gender and profession have been identified as factors associated with driving performance and collision risks. Male and female drivers perceive and assess risks differently (Harris & Jenkins, 2006), and a large number of studies reported that males engaged in more risk-taking behaviors and had a higher crash involvement rate per driving mileage than females (Massie et al., 1997; Regev et al., 2018). Professional drivers typically have high exposure to a great variety of traffic conditions, and thus they are deemed to be more skillful and capable of coping with hazards on roads (Chen et al., 2021). In the same scenario condition, research has reported that professional drivers have a lower crash rate than non-professional drivers, while non-professional drivers are more likely to take abrupt brake action to avoid collisions (Wu et al., 2016). Therefore, the study considers the role of driver characteristics (e.g., driver gender and profession) in the EEG response during the collision event. Specifically, the study has the following three research objectives and corresponding hypotheses:

(1) The study explores the patterns of different EEG bands (i.e., delta, theta, alpha, and beta) in the collision avoidance process. It is hypothesized that the EEG bands show significant changes as the collision event develops.
(2) The study identifies the difference in EEG response between the collision and non-collision groups across the collision avoidance process. It is hypothesized that the collision group has more slow-wave activities (e.g., delta, theta bands) and fewer fast-wave activities (e.g., alpha, beta bands) than the non-collision group.
(3) The study investigates the effect of driver characteristics (i.e., driver gender and driver profession) on the EEG response in the collision avoidance process. It is hypothesized that if driver gender and profession are associated with the collision outcome, they may play a significant role in EEG response during the collision avoidance process.

2. Methodology

2.1. Apparatus

The BJTU Driving Simulator was used to conduct the experiment (see Fig. 1). It is a high-fidelity, high-performance driving simulator with a linear motion base capable of operating with 1 degree of freedom (pitching). It has a full-size vehicle cabin (Ford Focus) with a genuine operational interface, environmental noise and movement simulation system, digital video replay system, and vehicle dynamic simulation system. The simulated environment is projected with a front/peripheral field of view of 300 degrees at a resolution of 1400×1050 pixels and left, middle, and right rear-view mirrors. The Simvista and Simcreator software is used for driving scenario design, virtual traffic environment simulation, and virtual road modelling. The Neuroscan system is used to collect the participants' EEG data (see Fig. 1). It is composed of a SynAmps2TM amplifier and an electrode cap with 64 channels. The distribution of 64 channels follows the 10–20 standard channel system defined by the International EEG Association. The EEG

Fig. 1. The BJTU Driving simulator and Neuroscan EEG system.

recording is entirely non-invasive, with high temporal resolution (sampling rate 1000 Hz) that allows for assessing instantaneous brain activity.

2.2. Participants

A total of 45 participants were recruited for the experiment. Due to motion sickness or equipment problems, 38 participants finished the experiment with complete data collected. The participants consisted of 19 female drivers and 19 male drivers. According to their profession, the participants could be divided into 20 professional taxi drivers (full-time) and 18 non-professional drivers who only used their cars for daily commuting. The average driving mileage was 67.5 thousand km per year for professional drivers and 20.1 thousand km per year for non-professional drivers. The average age of all participants was 34.5 years, ranging from 31 to 40 years, with a standard deviation of 3.1 years. Prior to the study, each participant was required to hold a valid driver's license and have a minimum driving mileage of 20 thousand kilometers.

2.3. Scenario design

The collision event was designed on a suburban road of 5 km long with a speed limit of 80 km/h. It was a two-way, two-lane road segment without a guardrail or medial strip. A pedestrian stood on the roadside 15 m away from the road center on a straight road segment of the driving route. The standing position of the pedestrian was not visible to the drivers as the buildings blocked their view. As the driver approached the pedestrian and the time to arrive at the pedestrian position was reduced to 3.5 s, the pedestrian was triggered to run across the road in front of the vehicle at a speed of 15 km/h (see Fig. 2). A collision would occur between the vehicle and the pedestrian if the driver did not take any actions such as braking or turning the steering wheel. The entire driving route consisted of signalized and non-signalized intersections, curves, and straight segments. Some ambient vehicles were arranged on the opposite lane except the straight road segment where the pedestrian-crossing event occurred. There were no other pedestrians or unexpected events on the route to avoid participants' speculation or any interference on their performance.

2.4. Experiment procedure

Upon arrival, each participant read and signed an informed consent form (per IRB). A pre-drive questionnaire was used to collect the participants' demographic information, such as age, gender, profession, and driving mileage. The participants were advised to drive as they normally would and to comply with traffic laws as they did in real-world driving. They were also notified that they could quit the experiment at any time in case of motion sickness or any discomfort. Before the experiment, each participant had 5–10 minutes practice driving session to get familiar with the vehicle operation, such as acceleration, deceleration, and left/right turn. After completing the experimental drive, each participant finished a short survey to collect their ratings on the sense of reality regarding the simulator control and driving scenario.

Fig. 2. Collision scenario.

2.5. Data processing and analysis

Participants' EEG data were collected during the whole drive. EEG toolbox in Matlab 2017a was used to process the raw data, and a series of data pre-processing procedures were conducted. First, a band-pass FIR filter (0.5–30 Hz) was used to remove undesirable signals and noise (Kar et al., 2010). Second, unqualified channels were removed with standards mentioned in Delorme et al. (2006), and ICA (Independent Component Analysis) was used to decompose the EEG signal. Third, ADJUST function tool was applied to identify the artifacts in EEG signals automatically (Mognon et al., 2011). The ADJUST could automatically detect and remove several types of artifacts, including the ocular artifacts caused by blinks, vertical, and horizontal eye movements and the generic artefacts caused by high impedance conditions or electrical instabilities in the recording device (Mognon et al., 2011). Lastly, the EEG data were re-referenced to the average reference and baseline correction was performed. This EEG pre-processing strategy has been commonly applied in many previous studies, and more detailed processing steps could refer to Yang et al. (2018) and Mognon et al. (2011).

After pre-processing, the EEG data within specific time windows were extracted. The time windows were determined based on the drivers' collision avoidance process. According to previous studies (Li et al., 2016, 2018), drivers' collision avoidance process could be divided into four stages, including normal driving stage (Stage 1), hazard perception stage (Stage 2), evasive action stage (Stage 3), and post-hazard stage (Stage 4). As shown in Fig. 3, the normal driving stage (Stage 1) refers to the situation when the hazard has not emerged. Stage 2 involves hazard perception and decision-making. It starts from when the pedestrian started to run and ends as the driver began to brake. The evasive action stage (Stage 3) starts as the driver began to press the brake pedal and ends as the driver began to release the brake pedal (note that all drivers took brake action to avoid the collision during the experiment). The post-hazard stage (Stage 4) indicates the end of the emergent collision event as the driver started to release the brake pedal and then press the accelerator pedal. According to the vehicle control data from the experiment, the average time duration for Stage 2 to Stage 4 was 2.04 s (SD = 0.32), 1.30 s (SD = 0.44), 1.36 s (SD = 0.56), respectively. Stage 1 was a stable driving period during which no abrupt event occurred. The study set 5 s before the pedestrian started to run as Stage 1 and divided it into two sub-stages equally (Stage 1–1 and Stage 1–2) with each sub-stage 2.5 s so that the duration was comparable with the subsequent stages. The purpose of dividing two sub-stages was to examine whether drivers' EEG activity remained stable without the pedestrian collision task. In addition, Stage 1–2 was used as a baseline stage to identify changes in the EEG bands in the subsequent collision avoidance stages.

Regarding each stage, Fast Fourier Transformation (FFT) was applied to transform the EEG data from a time domain into a frequency domain. Using the frequency domain data, the log-transformed power (LTP) and the power ratio (PR, i.e., the proportion of a certain band power to the sum of all bands power) of four EEG bands (i.e., delta (0.5–3 Hz), theta (4–7 Hz), alpha (8–13 Hz) and delta (14–30 Hz) bands) were calculated (Borghini et al., 2014; Otmani et al., 2005). Specifically, the log-transformed power was used to identify how drivers' EEG power changed across the whole process, and the power ratio was used to compare the proportion of different EEG components between different driver groups. A paired-samples *t*-test between Stage 1–1 and Stage 1–2 was performed to examine whether drivers' EEG activity remained stable during the normal driving stage. Repeated-measures ANOVA was applied to examine the LTP difference across the collision avoidance stages (Stage 1–2 to Stage 4), and Bonferroni's post-hoc pairwise comparisons were further performed for significant results. To identify the PR difference between driver groups (e.g., male vs. female, professional vs. non-professional), Hotelling's T-square test was first performed, and independent samples *t*-tests were further conducted to pinpoint the differences between every two groups. The two-layer examination helps avoid reporting results influenced by Type I error (Siegel, 1990). According to the collision avoidance outcome, drivers were divided into collision group and non-collision group. The PR difference between the two groups along with the collision avoidance stages was examined through Generalized Linear Mixed Models (GLMMs), which allow for independent variables from different distributions and account for both fixed and random effects. The statistical significance level used in the study was 0.05.

3. Results

3.1. Normal driving stage

The comparison of two normal driving stages was to identify whether drivers' EEG band power had significant changes without the pedestrian collision avoidance task. Table 1 lists the descriptive statistics (mean and S.D.) of the log-transformed power (LTP) of all four bands in the two normal driving stages and the paired-samples *t*-test between the two stages. The results showed that the LTP of all four bands had no significant change in the two normal driving stages, which means that drivers' EEG activity remained stable when drivers drove on a straight road without encountering the abrupt event (e.g., pedestrian crossing).

3.2. Collision avoidance stages

The descriptive statistics of four EEG bands LTP in four collision avoidance stages are listed in Table 2. The repeated-measures ANOVA results show that the LTP of all four bands changed significantly during the collision avoidance process. According to the post-hoc test results, the LTP of all four bands in the hazard perception stage (Stage 2) and post-hazard stage (Stage 4) was significantly larger than that in the normal driving stage (Stage 1) as shown in Fig. 4. The LTP of delta, theta, and alpha in the post-hazard stage (Stage 4) was also significantly larger than that in the evasive action stage (Stage 3). In addition, the LTP of beta increased significantly in the evasive action stage (Stage 3) compared to the normal driving stage (Stage 1).

3.3. Collision groups

According to the experimental results, 16 participants (42.1%) collided with the crossing pedestrian, and 22 participants (57.9%) avoided the collision successfully. The mean values and standard deviations of four EEG bands power ratio (PR) of two collision groups are listed in Table 3. Fig. 5 shows the change of power ratios between the two groups along with the collision avoidance stages.

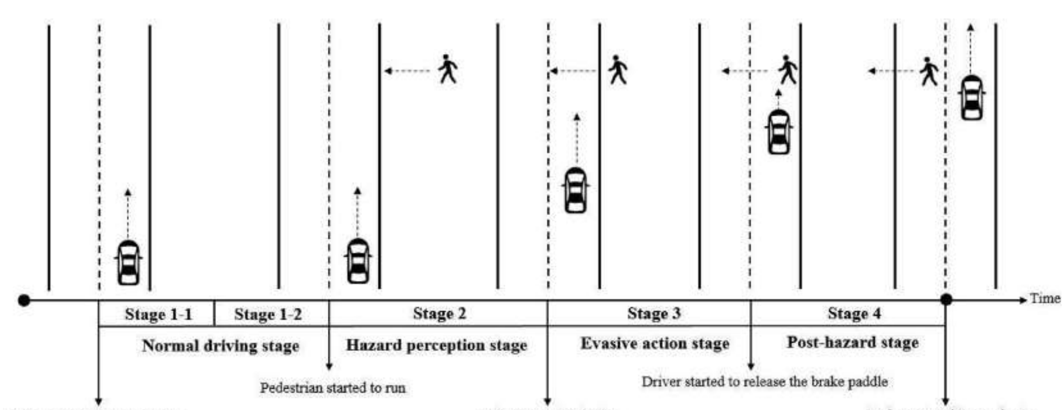


Fig. 3. The four-stage collision avoidance process.

Table 1
Descriptive statistics and paired-samples *t*-test of EEG band LTP in normal driving stage.

EEG band	Variables	Stage1-1	Stage1-2	*N*	df	t	*p*	95% CI
Delta	LTP	1.30 (0.47)	1.31 (0.43)	38	37	−0.096	0.924	[−0.08, 0.07]
Theta	LTP	0.66 (0.38)	0.65 (0.38)	38	37	0.375	0.710	[−0.04, 0.06]
Alpha	LTP	0.66 (0.34)	0.69 (0.34)	38	37	−1.382	0.175	[−0.09, 0.02]
Beta	LTP	0.84 (0.32)	0.87 (0.34)	38	37	−1.086	0.284	[−0.07, 0.02]

Table 2
Descriptive statistics and repeated-measures ANOVA results of EEG band LTP in four stages.

EEG band	Variables	Stage1	Stage2	Stage3	Stage4	*N*	df	*F*	*p*	η_p^2
Delta	LTP	1.31 (0.43)	1.52 (0.44)	1.50 (0.75)	1.74 (0.81)	38	3	7.880	<0.001	0.403
Theta	LTP	0.65 (0.38)	0.79 (0.35)	0.74 (0.49)	0.91 (0.55)	38	3	8.826	<0.001	0.431
Alpha	LTP	0.69 (0.34)	0.78 (0.30)	0.77 (0.41)	0.86 (0.43)	38	3	5.109	<0.01	0.305
Beta	LTP	0.87 (0.34)	0.95 (0.31)	1.01 (0.37)	1.07 (0.41)	38	3	7.281	<0.01	0.384

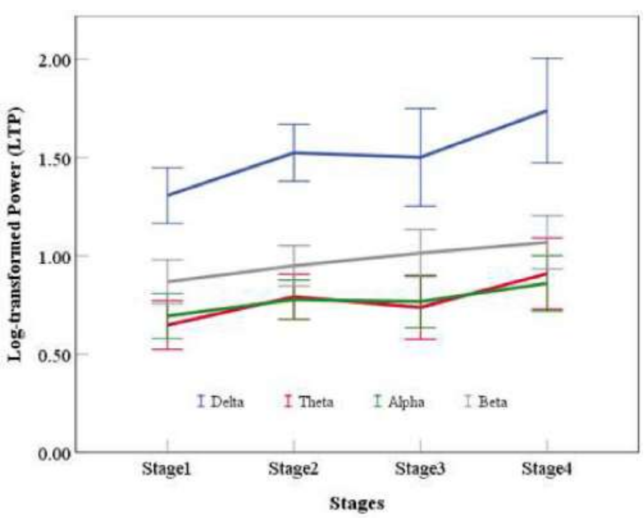


Fig. 4. The change of LTP of four EEG bands in four collision avoidance stages.

It can be observed that the collision group had a higher delta PR and lower beta PR during the collision avoidance process, except Stage 3 (evasive action stage). The theta PR started with a higher value in the non-collision group in Stage 1 and ended with a lower value in Stage 4 compared to the collision group, while the alpha PR showed an opposite pattern. The GLMMs results (see Table 4) showed that only the collision stage had significant main effects on the four EEG bands power ratio, while the main effect of the collision group and the interaction effect between collision group and collision stage were not significant. Regarding the effects of the collision stage, the power ratio of delta band in Stage 2 and Stage 4 was significantly higher than that in Stage 1. The theta and beta power ratios in Stage 4 were significantly lower than those in Stage 1, and the alpha power ratio in Stage 2 and Stage 4 significantly reduced compared to Stage 1.

3.4. Gender groups

Among the 19 female and 19 male drivers, nine females and seven males had collisions with the pedestrian. The Pearson Chi-Square test shows that driver gender was not associated with the collision occurrence (Pearson Chi-Square = 0.432, $p > 0.05$). Table 5 lists the descriptive statistics of the power ratio of four EEG bands and the Hotelling's T-square tests results of driver gender. The test results show that drivers' gender significantly influenced the power ratio of EEG bands in Stage 2 and Stage 3. Independent samples *t*-tests further show that the delta and alpha power ratios in Stage 2 and Stage 3 were significantly different between female and male drivers (see Table 6). Specifically, male drivers had a higher delta power ratio and a lower alpha power ratio than female drivers in these two stages (see Fig. 6).

3.5. Profession groups

Among the 20 professional and 18 non-professional drivers, 16 drivers in total with eight drivers in each group had collisions with the pedestrian. The Pearson Chi-Square test shows that driver profession was not associated with the collision occurrence (Pearson Chi-Square = 0.077, $p > 0.05$). The means and standard deviations of EEG bands power ratio of professional and non-professional drivers and the Hotelling's T-square test results of driver profession are shown in Table 7. The test results show that there was no significant effect of driver profession on the power ratio of four EEG bands in the collision avoidance process.

Table 3
Descriptive statistics of EEG band power ratio between collision and non-collision groups in different stages.

EEG band	Collision group	Stage 1	Stage 2	Stage 3	Stage 4
Delta	Collision	0.55 (0.14)	0.59 (0.13)	0.55 (0.19)	0.62 (0.18)
	Non-collision	0.51 (0.12)	0.57 (0.14)	0.55 (0.21)	0.60 (0.19)
Theta	Collision	0.12 (0.03)	0.11 (0.02)	0.11 (0.05)	0.11 (0.05)
	Non-collision	0.12 (0.03)	0.12 (0.05)	0.10 (0.05)	0.10 (0.03)
Alpha	Collision	0.14 (0.07)	0.13 (0.08)	0.11 (0.05)	0.09 (0.06)
	Non-collision	0.15 (0.06)	0.12 (0.05)	0.13 (0.07)	0.11 (0.06)
Beta	Collision	0.20 (0.08)	0.17 (0.07)	0.23 (0.13)	0.17 (0.09)
	Non-collision	0.22 (0.08)	0.19 (0.11)	0.23 (0.14)	0.18 (0.12)

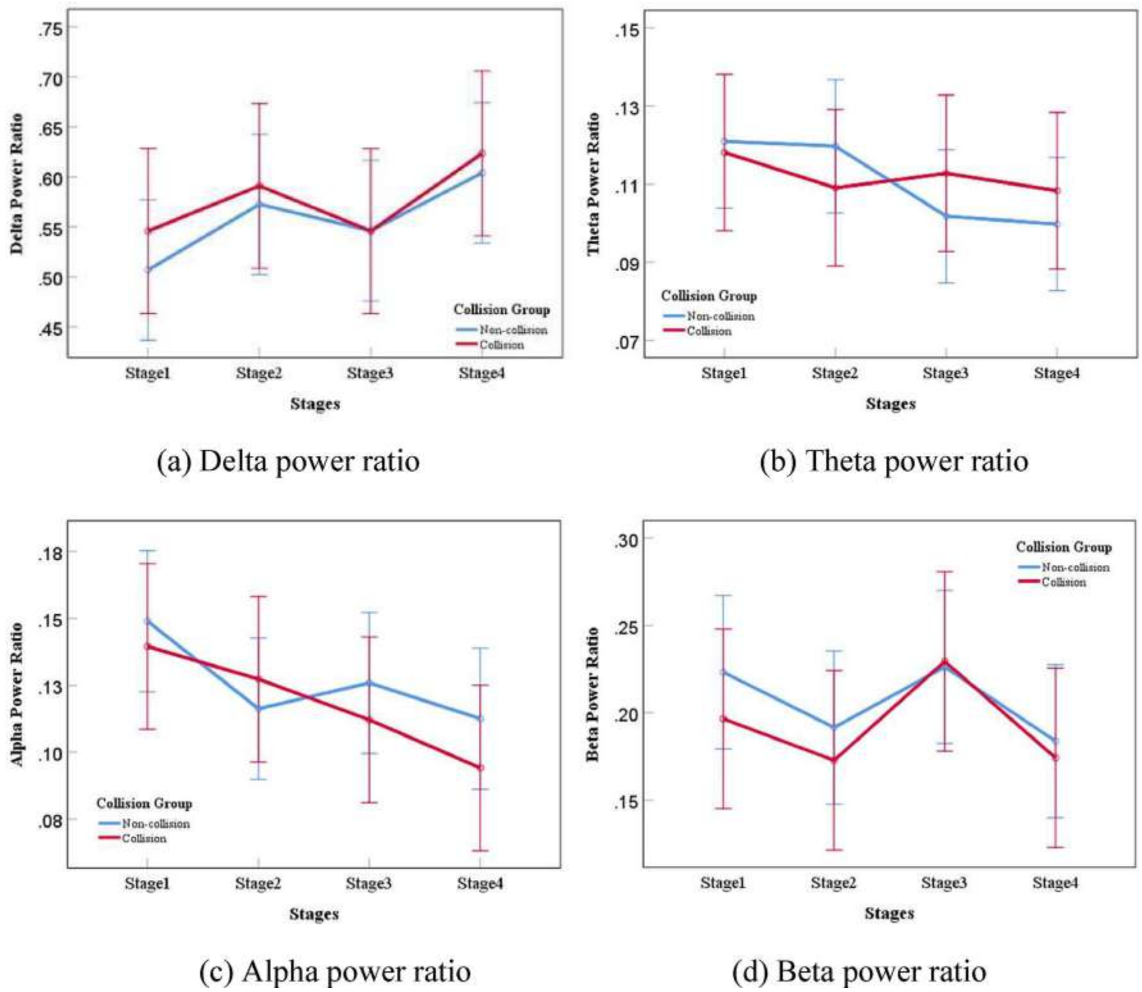


Fig. 5. EEG band power ratio of collision and non-collision group in different collision avoidance stages.

3.6. Subjective evaluation of simulator validity

A post-drive survey was used to collect the drivers' subjective ratings on the simulator validity. The survey contains seven aspects: overall vehicle control performance, throttle control, brake pedal control, steering wheel control, lane-change control, roadside buildings and surrounding vehicles. Participants rated on a 5-point Likert scale regarding their perceived sense of reality for each of the aspect, where 1 = very low, 2 = low, 3 = medium, 4 = high, 5 = very high. The mean score and S.D. of each aspect are listed in Table 8. The results indicate that participants were overall satisfied with the sense of reality in the simulator, with all aspects evaluated higher than a medium level. Specifically, the throttle control, roadside buildings and surrounding vehicles received ratings between high to very high, while the brake pedal control, steering wheel control and lane-change control received ratings between medium to high.

4. Discussion

The pedestrian collision evasive task in the study was designed as an emergent and safety–critical event with high collision risk. To successfully avoid the collision, drivers needed to perceive the pedestrian timely and take appropriate evasive actions. The task demands increased substantially during the collision avoidance process in comparison to a normal driving status without potential hazards. Previous research has reported that higher task demands led to an increase in the drivers' emotional and physiological arousal (Mehler et al., 2009; Schmidt-Daffy, 2013). Therefore, it is expected that during the process from hazard emergence to hazard avoidance, drivers' mental and physical arousal increased to a higher level. The fluctuation in arousal status was reflected in the changes in EEG activities. The power of all four EEG bands (i.e., delta, theta, alpha, beta) increased as the event developed. The result is partially in line with previous studies that reported an increased theta and beta power associated with increased task demands (Dussault et al., 2005; Fairclough et al., 2005), high time pressure (Fairclough et al., 2005), or high vigilance (Okogbaa et al., 1994). However, the consistent patterns of different EEG bands across stages observed in the present study were seldom reported in previous research. According to prior studies, especially those about fatigued driving, an increased activity of certain EEG bands (e.g., alpha, beta) usually appeared together with the suppressed activity of other bands (e.g., delta, theta; Jap et al., 2011; Zhao et al., 2012). Findings of the present study raised potential argument regarding the reverse relationship among different EEG bands reported in prior studies, such as the opposite changing pattern of alpha and beta in attention-demanding tasks (Jap et al., 2011; Schwartz et al., 1989) or the opposite pattern of delta and beta on fatigued participants (Jap et al., 2009). Although the present study did not focus on a specific driver status such as fatigue

Table 4
The GLMM results of four EEG band power ratio for collision group and collision stage.

EEG band	Model term	β	S.E.	t-value	p	95% CI
Delta	Intercept	0.507	0.034	14.723	<0.001	[0.439, 0.575]
	Collision = 1	0.039	0.053	0.734	0.464	[−0.066, 0.144]
	Stage = 2	0.066	0.031	2.133	<0.05	[0.005, 0.126]
	Stage = 3	0.039	0.043	0.907	0.366	[−0.046, 0.125]
	Stage = 4	0.097	0.038	2.585	<0.05	[0.023, 0.171]
	Collision = 1*Stage = 2	−0.021	0.047	−0.434	0.665	[−0.114, 0.073]
	Collision = 1*Stage = 3	−0.039	0.067	−0.590	0.556	[−0.171, 0.092]
	Collision = 1*Stage = 4	−0.020	0.058	−0.338	0.736	[−0.134, 0.095]
Theta	Intercept	0.121	0.007	16.805	<0.001	[0.107, 0.135]
	Collision = 1	−0.003	0.011	−0.257	0.798	[−0.025, 0.019]
	Stage = 2	−0.001	0.010	−0.127	0.899	[−0.020, 0.018]
	Stage = 3	−0.019	0.011	−1.701	0.091	[−0.041, 0.003]
	Stage = 4	−0.021	0.010	−2.076	<0.05	[−0.041, −0.001]
	Collision = 1*Stage = 2	−0.008	0.015	−0.522	0.603	[−0.037, 0.022]
	Collision = 1*Stage = 3	0.014	0.017	0.797	0.427	[−0.021, 0.048]
	Collision = 1*Stage = 4	0.011	0.016	0.725	0.470	[−0.020, 0.042]
Alpha	Intercept	0.149	0.015	9.877	<0.001	[0.119, 0.179]
	Collision = 1	−0.009	0.023	−0.406	0.685	[−0.055, 0.036]
	Stage = 2	−0.033	0.016	−2.067	<0.05	[−0.064, −0.001]
	Stage = 3	−0.023	0.015	−1.504	0.135	[−0.053, 0.007]
	Stage = 4	−0.036	0.015	−2.428	<0.05	[−0.066, −0.007]
	Collision = 1*Stage = 2	0.021	0.024	0.841	0.402	[−0.028, 0.069]
	Collision = 1*Stage = 3	−0.004	0.024	−0.185	0.854	[−0.051, 0.042]
	Collision = 1*Stage = 4	−0.009	0.023	−0.387	0.699	[−0.055, 0.037]
Beta	Intercept	0.233	0.023	9.801	<0.001	[0.178, 0.268]
	Collision = 1	−0.027	0.035	−0.760	0.449	[−0.096, 0.043]
	Stage = 2	−0.032	0.018	−1.806	0.073	[−0.066, 0.003]
	Stage = 3	0.003	0.024	0.125	0.901	[−0.045, 0.051]
	Stage = 4	−0.039	0.020	−1.989	<0.05	[−0.079, 0.000]
	Collision = 1*Stage = 2	0.008	0.027	0.290	0.772	[−0.046, 0.061]
	Collision = 1*Stage = 3	0.030	0.038	0.795	0.428	[−0.044, 0.104]
	Collision = 1*Stage = 4	0.017	0.031	0.559	0.557	[−0.043,0.078]

Note: Collision = 1 represents the collision group. The rest items including "Collision = 0", "Stage = 1" and the other interaction items are set as referential contrast with coefficient being 0 and are not listed in the table.

Table 5
Descriptive statistics and Hotelling's T-square test results of EEG band power ratio between gender groups.

Stage	EEG band	Male (n = 19)	Female (n = 19)	Hotelling's T^2	df	p	η_p^2
Stage1	Delta	0.55 (0.11)	0.50 (0.14)	1.656	34	0.671	0.044
	Theta	0.12 (0.03)	0.12 (0.03)				
	Alpha	0.14 (0.06)	0.15 (0.07)				
	Beta	0.20 (0.07)	0.23 (0.08)				
Stage2	Delta	0.63 (0.11)	0.53 (0.14)	9.432	34	<0.05	0.208
	Theta	0.11 (0.02)	0.12 (0.05)				
	Alpha	0.09 (0.03)	0.15 (0.07)				
	Beta	0.16 (0.09)	0.20 (0.10)				
Stage3	Delta	0.62 (0.22)	0.48 (0.15)	20.196	34	<0.01	0.360
	Theta	0.11 (0.06)	0.11 (0.04)				
	Alpha	0.09 (0.05)	0.15 (0.05)				
	Beta	0.19 (0.13)	0.27 (0.13)				
Stage4	Delta	0.69 (0.18)	0.53 (0.16)	8.604	34	0.06	0.193
	Theta	0.09 (0.04)	0.12 (0.04)				
	Alpha	0.08 (0.05)	0.13 (0.06)				
	Beta	0.14 (0.10)	0.22 (0.10)				

Table 6
Independent samples t-test results of EEG band power ratio between gender groups in Stage 2 and Stage 3.

Stage	EEG band	t-value	df	p	95% CI
Stage2	Delta	−2.579	36	<0.05	[−0.19, −0.02]
	Theta	1.014	36	0.318	[−0.01, 0.04]
	Alpha	2.970	36	<0.01	[0.02, 0.09]
	Beta	1.359	36	0.183	[−0.02, 0.10]
Stage3	Delta	−2.276	36	<0.05	[−0.26, 0.02]
	Theta	0.067	36	0.947	[−0.03, 0.03]
	Alpha	3.686	36	<0.01	[0.03, 0.10]
	Beta	1.841	36	0.074	[−0.01, 0.16]

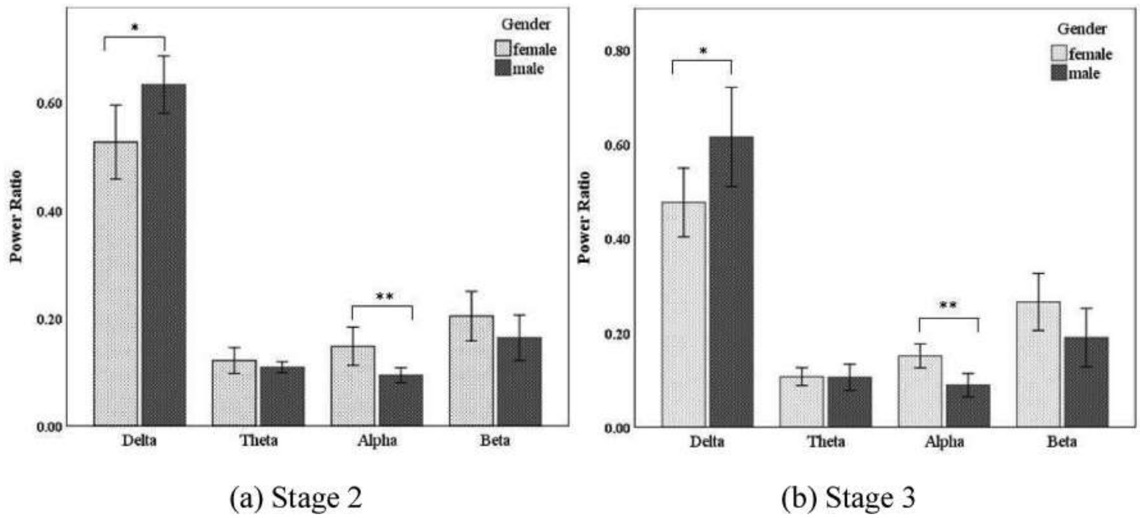


Fig. 6. EEG band power ratio between driver gender groups in Stage 2 and Stage 3 (* represents $p < 0.5$; ** represents $p < 0.01$).

Table 7
Descriptive statistics and Hotelling's T-square test results of EEG band power ratio between driver profession groups.

Stage	EEG band	Professional ($n = 20$)	Non-professional ($n = 18$)	Hotelling's T^2	df	p	η_p^2
Stage1	Delta	0.53 (0.14)	0.52 (0.12)	6.084	34	0.145	0.145
	Theta	0.11 (0.03)	0.13 (0.03)				
	Alpha	0.13 (0.06)	0.16 (0.07)				
	Beta	0.23 (0.09)	0.20 (0.05)				
Stage2	Delta	0.57 (0.13)	0.59 (0.15)	1.116	34	0.785	0.030
	Theta	0.11 (0.04)	0.12 (0.03)				
	Alpha	0.12 (0.05)	0.12 (0.07)				
	Beta	0.20 (0.11)	0.17 (0.07)				
Stage3	Delta	0.57 (0.21)	0.52 (0.19)	2.376	34	0.529	0.062
	Theta	0.10 (0.03)	0.12 (0.06)				
	Alpha	0.11 (0.05)	0.13 (0.07)				
	Beta	0.23 (0.15)	0.23 (0.11)				
Stage4	Delta	0.62 (0.19)	0.61 (0.18)	0.432	34	0.939	0.012
	Theta	0.10 (0.04)	0.10 (0.05)				
	Alpha	0.10 (0.06)	0.11 (0.07)				
	Beta	0.18 (0.12)	0.18 (0.09)				

Table 8
Subjective ratings on simulator validity.

Items	N	Mean	S.D.
Overall control performance	38	3.58	0.95
Throttle control	38	4.42	0.76
Brake pedal control	38	3.16	1.20
Steering wheel control	38	3.82	1.04
Lane-change control	38	3.95	1.11
Roadside buildings	38	4.03	1.08
Surrounding vehicles	38	4.03	1.15

or distraction, it is possible that enhanced activities of different EEG bands could be concurrent at some emergent and safety–critical situations.

The activities of four EEG bands between the collision and non-collision groups at different stages of collision avoidance were examined. However, no significant difference was observed between the two groups at all stages. It has been previously reported that the brain shows more fast-wave (e.g., beta) and less slow-wave (e.g., delta) activities when someone is alert (Kiymik et al., 2004). Moreover, beta activity in the frontal lobes was found to be associated with cognitive processes such as judgment, working memory, and decision making (Lin et al., 2011), while delta activity was associated more with driver fatigue and sleepiness (Lal & Craig, 2001, 2002;). A decrease in beta activity has been found to be related to worsening performance (Subasi, 2005). These previous findings might help explain the opposite patterns found in delta and beta power ratios across stages and the higher ratio of delta power and lower ratio of beta power observed on collision drivers in the early stages of the collision avoidance process.

Although driver gender was not associated with the collision outcomes, the study reported significant gender differences in EEG response during the collision avoidance process. Male drivers had a higher delta power ratio and lower alpha power ratio than female drivers in both hazard perception stage (Stage 2) and evasive action stage (Stage 3). According to prior studies, the emergence of slow waves such as delta was largely reported together with mental fatigue and drowsiness (Borghini et al., 2014; Lal & Craig, 2002). Regarding the alpha band, many studies reported that changes in alpha activities are related to tasks that require attentional process (Schier, 2000; Sonnleitner et al., 2014). However, the previous findings of the tendency of alpha change were not consistent. Some early studies referred to alpha activity as 'idling' activity, and attention-demand task could result in a reduction in alpha activity (i.e., 'alpha blocking;' Schwartz et al., 1989). Conversely, other studies reported an increment in specific alpha sub-band during judgment tasks, and the decrease in alpha activity indicated poor cognitive and memory performance (Klimesch, 1999). A more recent study investigated the alpha activity of distracted drivers in a simulator experiment and found that alpha spindle rate increased as drivers' cognitive load increased when driving with a secondary task (Sonnleitner et al., 2014). In terms of the gender difference observed in this study, it is highly suspected that the hazard evasive task resulted in different levels of mental and physical workload for female and male drivers, with females perceiving the task riskier and more effort-demanding than males. Thus, female drivers concentrated more attention on the task with more alpha activity and less delta activity observed in comparison to male drivers. No EEG difference from driver profession was observed in this study. This indicates that in a specific emergent event, the acquired experience, such as driving skills, played a less critical role in influencing EEG response compared to the innate driver characteristics such as gender.

The study has several limitations that should be noted. The study mainly focused on drivers' EEG response during a collision avoidance process. Behavioral performances such as speed control or steering wheel control in the process were not considered in the study. It might be interesting to investigate the relationship between EEG response and behaviors as well as the mutual-effect mechanism between them in future studies. Furthermore, the study only designed a single collision event to avoid participants' speculation and learning effect. It is suggested that future studies could compare other emergent and safety–critical events to validate the EEG patterns observed in this study. Finally, the sample size of the study was relatively small, considering the number of variables analyzed. A larger sample size is recommended for future studies to improve the generalizability of the results.

5. Conclusion

The study conducted a driving simulator experiment to investigate drivers' EEG response in a safety–critical situation that involved an impending collision. The log-transformed power (LTP) and power ratio (PR) of four EEG bands (i.e., delta, theta, alpha, and beta) were analyzed during the collision avoidance process that included normal driving stage (Stage 1), hazard perception stage (Stage 2), evasive action stage (Stage 3), and post-hazard stage (Stage 4). The EEG variables were compared between different collision groups (collision vs. non-collision) and driver characteristics groups (driver gender and profession). The study found consistent patterns of different EEG bands along with the development of the collision event. The log-transformed power of all EEG bands increased significantly from Stage 1 to Stage 4. Collision drivers and non-collision drivers did not show a significant difference in the four EEG band power ratios across different stages. Driver gender played a significant role in EEG response. Male drivers had a higher delta power ratio and lower alpha power ratio than female drivers in hazard perception and evasive action stages, which indicates that female drivers were more mentally aroused than male drivers. No difference from drivers' profession was found. The study provided more understanding of drivers' neurophysiological activities in a safety–critical event and demonstrated the role that different individuals played during the hazard avoidance process.

Conflict of interest

None.

Acknowledgement

This research was supported by the National Natural Science Foundation of China (71901036, 72001163).

References

Ako, M., Kawara, T., Uchida, S., Miyazaki, S., Nishihara, K., Mukai, J., ... Okubo, Y. (2003). Correlation between electroencephalography and heart rate variability during sleep. *Psychiatry and Clinical Neurosciences, 57*(1), 59–65.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews, 44*, 58–75.

Chen, T., Sze, N. N., Newnam, S., & Bai, L. (2021). Effectiveness of the compensatory strategy adopted by older drivers: difference between professional and non-professional drivers. *Transportation Research Part F: Traffic Psychology and Behaviour, 77*, 168–180.

Crundall, D., Chapman, P., Trawley, S., Collins, L., Van Loon, E., Andrews, B., & Underwood, G. (2012). Some hazards are more attractive than others: Drivers of varying experience respond differently to different types of hazard. *Accident Analysis & Prevention, 45*, 600–609.

Delorme, A., Fernsler, T., Serby, H., & Makeig, S. (2006). *EEGLAB tutorial*. University of San Diego California.

Dussault, C., Jouanin, J. C., Philippe, M., & Guezennec, C. Y. (2005). EEG and ECG changes during simulator operation reflect mental workload and vigilance. *Aviation, Space, and Environmental Medicine, 76*(4), 344–351.

Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology, 56*(2), 171–184.

Galán, F. C., & Beal, C. R. (2012). EEG estimates of engagement and cognitive workload predict math problem solving outcomes. In: Proceedings of the International Conference on User Modeling, Adaptation, and Personalisation, 51-62.

Guo, Z., Pan, Y., Zhao, G., Cao, S., & Zhang, J. (2018). Detection of driver vigilance level using EEG signals and driving contexts. *IEEE Transactions on Reliability, 67*(1), 370–380.

Harb, R., Yan, X., Radwan, E., & Su, X. (2009). Exploring precrash maneuvers using classification trees and random forests. *Accident Analysis & Prevention, 41*(1), 98–107.

Harris, C. R., & Jenkins, M. (2006). Gender differences in risk assessment: Why do women take fewer risks than men? *Judgement and Decision Making, 1*(1), 48–63.

Hernández, L. G., Mozos, O. M., Ferrández, J. M., & Antelis, J. M. (2018). EEG-based detection of braking intention under different car driving conditions. *Frontiers in Neuroinformatics, 12*, 29.

Jap, B. T., Lal, S., & Fischer, P. (2011). Comparing combinations of EEG activity in train drivers during monotonous driving. *Expert Systems with Applications, 38*(1), 996–1003.

Jones, M. P., Chapman, P., & Bailey, K. (2014). The influence of image valence on visual attention and perception of risk in drivers. *Accident Analysis & Prevention, 73*, 296–304.

Johansson, R. (2009). Vision zero-implementing a policy for traffic safety. *Safety Science, 47*(6), 826–831.

Kar, S., Bhagat, M., & Routray, A. (2010). EEG signal analysis for the assessment and quantification of driver's fatigue. *Transportation Research Part F: Traffic Psychology and Behaviour, 13*, 297–306.

Kiymik, M. K., Akin, M., & Subasi, A. (2004). Automatic recognition of alertness level by using wavelet transform and artificial neural network. *Journal of Neuroscience Methods, 139*(2), 231–240.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews, 29*(2–3), 169–195.

Lal, S. K., & Craig, A. (2002). Driver fatigue: Electroencephalography and psychological assessment. *Psychophysiology, 39*(3), 313–321.

Lal, S. K., & Craig, A. (2001). Electroencephalography activity associated with driver fatigue: Implications for a fatigue countermeasure device. *Journal of Psychophysiology, 15*(3), 183.

Li, X., Rakotonirainy, A., & Yan, X. (2019). How do drivers avoid collisions? A driving simulator-based study. *Journal of Safety Research, 70*, 89–96.

Li, X., Rakotonirainy, A., Yan, X., & Zhang, Y. (2018). Driver's visual performance in rear-end collision avoidance process under the influence of cell phone use. *Transportation Research Record, 2672*(37), 55–63.

Li, X., Yan, X., Wu, J., Radwan, E., & Zhang, Y. (2016). A rear-end collision risk assessment model based on drivers' collision avoidance process under influences of cell phone use and gender - a driving simulator based study. *Accident Analysis & Prevention, 97*, 1–18.

Lin, C., Chen, S., Ko, L., & Wang, Y., 2011. EEG-based brain dynamics of driving distraction. In: Proceedings of the 2011 International Joint Conference on Neural Networks, 1497–1500.

Markkula, G., Benderius, O., Wolff, K., & Wahde, M. (2012). A review of near-collision driver behavior models. *Human Factors, 54*(6), 1117–1143.

Massie, D. L., Green, P. E., & Campbell, K. L. (1997). Crash involvement rates by driver gender and the role of average annual mileage. *Accident Analysis & Prevention, 29*(5), 675–685.

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record, 2138*(1), 6–12.

Mognon, A., Jovicich, J., & Bruzzone, L. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology, 48*, 229–240.

Okogbaa, O. G., Shell, R. L., & Filipusic, D. (1994). On the investigation of the neurophysiological correlates of knowledge worker mental fatigue using the EEG signal. *Applied Ergonomics, 25*(6), 355–365.
Otmani, S., Pebayle, T., Roge, J., & Muzet, A. (2005). Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers. *Physiology & Behavior, 84*(5), 715–724.
Perrier, J., Jongen, S., Vuurman, E., Bocca, M. L., Ramaekers, J. G., & Vermeeren, A. (2016). Driving performance and EEG fluctuations during on-the-road driving following sleep deprivation. *Biological Psychology, 121*, 1–11.
Petridou, E., & Moustaki, M. (2000). Human factors in the causation of road traffic crashes. *European Journal of Epidemiology, 16*(9), 819–826.
Ray, W. J., & Cole, H. W. (1985). EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science, 228*(4700), 750–752.
Regev, S., Rolison, J. J., & Moutari, S. (2018). Crash risk by driver age, gender, and time of day using a new exposure methodology. *Journal of Safety Research, 66*, 131–140.
Schier, M. A. (2000). Changes in EEG alpha power during simulated driving: A demonstration. *International Journal of Psychophysiology, 37*(2), 155–162.
Schmidt-Daffy, M. (2013). Fear and anxiety while driving: Differential impact of task demands, speed and motivation. *Transportation Research Part F: Traffic Psychology and Behaviour, 16*, 14–28.
Schomer, D. L., & Da Silva, F. L. (2012). *Niedermeyer's electroencephalography: Basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.
Schwartz, B. J., Salustri, C., Kaufman, L., & Williamson, S. J. (1989). Alpha suppression related to a cognitive task. In *Advances in Biomagnetism* (pp. 237–240). Boston, MA: Springer.
Siegel, A. F. (1990). Research issues: Multiple tests: Some practical considerations. *TESOL Quarterly, 24*(4), 773–775.
Sonnleitner, A., Treder, M. S., Simon, M., Willmann, S., Ewald, A., Buchner, A., & Schrauf, M. (2014). EEG alpha spindles and prolonged brake reaction times during auditory distraction in an on-road driving study. *Accident Analysis & Prevention, 62*, 110–118.
Stenberg, G. (1992). Personality and the EEG: Arousal and emotional arousability. *Personality and Individual Differences, 13*(10), 1097–1113.
Subasi, A. (2005). Automatic recognition of alertness level from eeg by using neural network and wavelet coefficients. *Expert Systems with Applications, 28*(4), 701–711.
WHO (2018). Global status report on road safety 2018. Retrieved from https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/.
Xiaoli, X., Jiangbi, H., Xiaoming, L., Pingsheng, L., & Shuyun, W. (2009). The EEG changes during night-time driver fatigue. In *Proceedings of the 2009 IEEE intelligent vehicles symposium* (pp. 935–939).
Yang, L., Guan, W., Ma, R., & Li, X. (2019). Comparison among driving state prediction models for car-following condition based on EEG and driving features. *Accident Analysis & Prevention, 133*, 105296.
Yang, L., Ma, R., Zhang, H. M., Guan, W., & Jiang, S. (2018). Driving behavior recognition using EEG data from a simulated car-following experiment. *Accident Analysis & Prevention, 116*, 30–40.
Zhao, C., Zhao, M., Liu, J., & Zheng, C. (2012). Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accident Analysis & Prevention, 45*, 83–90.

Xiaomeng Li is a research fellow from the Centre for Accident Research and Road Safety – Queensland (CARRS-Q), Queensland University of Technology. She was awarded a PhD degree in January, 2017 from Beijing Jiaotong University. Her research mainly focuses on road traffic safety, driving performance, human factors, ITS and traffic engineering.

Yang Liu is a lecturer from School of Transportation, Wuhan University of Technology. Her research interests include driver behavior and Electroencephalography (EEG), driving simulator, traffic safety and autonomous vehicle.

Xuedong Yan is a professor from the MOT Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University. His research expertise includes advanced driving simulator research, driving behavior and human factor, GIS application on traffic engineering, traffic safety and security, statistical analysis, traffic simulation, highway geometric design, highway-rail grade crossing, ITS, and urban design and planning.

Journal of Safety Research 82 (2022) 93–101

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Are out-of-state drivers more seatbelt compliant than in-state drivers in the United States?

Kwaku F. Boakye

Transportation Engineer II, Arcadis U.S., Inc., 2839 Paces Ferry Rd SE, Suite 900, Atlanta, GA 30339, United States

ARTICLE INFO

Article history:
Received 16 August 2021
Received in revised form 3 December 2021
Accepted 3 May 2022
Available online 13 May 2022

Keywords:
Seatbelt use
Seatbelt laws
Fatal crash
Driver
Z-test
Logistic regression

ABSTRACT

Introduction: This study explored the seatbelt use among in-state and out-of-state drivers in relation to their personal (age, gender, license status, etc.) and crash characteristics (time, location, roadway factors, etc.) using crash data over a 10-year period (2010–2019) from the Fatality Analysis Reporting System (FARS). *Method:* Comparison of seatbelt use between the two groups (in-state vs. out-of-state drivers) were conducted using Z-test statistics. Logistic regression models were developed to examine the probability of seatbelt use among each group. *Results:* New findings in this study showed that out-of-state drivers were 5% more likely than in-state drivers to use seatbelts. Regardless of the driver's age, gender, license status, vehicle type, and injury severity, seatbelt use was significantly higher among out-of-state drivers. Moreover, irrespective of the location (rural or urban), the season (time, day, or month), road type (arterial, local streets, etc.), and jurisdictional seatbelt law (primary or secondary), out-of-state drivers were more seatbelt compliant than in-state drivers. Finally, out-of-state drivers traveling from states with secondary/no seatbelt laws exhibited higher seatbelt compliance rate in primary seatbelt law states than in states with less strict laws (i.e., secondary/no law). *Practical Applications:* The findings in this study are critical to addressing a myriad of policy questions related to seatbelt laws and seatbelt use. Future research should focus on the disparity in seatbelt use between the two groups and determine intervention strategies that are effective at promoting seatbelt use across the United States. Additionally, given the significant differences in driver seatbelt use behavior based on the type of seatbelt law, if states with less strict laws upgrade to primary seatbelt laws, there likely will be increases in seatbelt compliance in those states.

1. Introduction

The public health implications of motor-vehicle crashes tend to be severe. Worldwide, traffic crashes result in millions of deaths, serious injuries, property damages, and high societal costs. The United States has historically achieved low fatality and injury rates in motor-vehicle crashes. In 1975, traffic death rate in the United States was at 3.35 deaths per 100 million vehicle miles traveled (VMT) (National Center for statistics and Analysis, 2018). By the year 2000, the rate had decline to 1.53 deaths per 100 million VMT. Most current estimate (2019) shows new low level of 1.10 deaths per 100 million VMT (National Center for Statistics and Analysis, 2019). Despite the historic decline, a significant number of travelers are involved in fatal crashes each year. In 2019, an estimated 36,096 people were killed in motor-vehicle crashes in the United States (Kahane, 1960). Among the fatally injured with known restraint use, nearly half (47%) were unrestrained (National Center for Statistics and Analysis, 2017).

Wearing a seatbelt is crucial to surviving a crash. The seatbelt is the single most effective device that can keep occupants safe and prevent them from being totally ejected from their vehicles in the event of a crash. Research has shown that wearing a seatbelt in the front seat of a passenger car can reduce the risk of fatal injury by 45% and in a light truck by 60% (Kahane, 1960). In 2017 alone, seatbelts saved an estimated 14,955 lives and an additional 2,549 lives would have been saved if all vehicle occupants were restrained prior to the crash (National Center for Statistics and Analysis, 2017). Given the life-saving benefits of seatbelts, some passengers still travel without wearing seatbelts. For example, key populations including males, young adults, rear seat passengers, and rural residents are often associated with seatbelt non-use (Beck et al., 2019; Lipovac et al., 2015; Strine et al., 2010).

Seatbelt use rates have steadily increased over time in the United States. When the first comprehensive national seatbelt survey was conducted in 1994, use rate was at 58%. By 2002, belt use had

E-mail address: Kwaku.Boakye@arcadis.com

https://doi.org/10.1016/j.jsr.2022.05.001
0022-4375/

reached 75%, the highest rate since the beginning of the comprehensive national survey (Glassbrenner, 2002). Nationwide seatbelt use rate was 90.7% in 2019, but varied from as low as 70.7% in New Hampshire to as high as 97.1% in Hawaii (National Center for Statistics and Analysis, 2019). Several factors are known to account for the variation in seatbelt use among the states, however, the type of seatbelt law enacted in a particular state makes a significant difference. Research studies have demonstrated that states with primary seatbelt laws have higher seatbelt compliance rates than states with secondary or no seatbelt laws (Boakye & Nambisan, 2020; Shults et al., 2004). In a primary seatbelt law state, a driver can be pulled over and issued a violation ticket solely for not using a seatbelt, whereas, in a secondary seatbelt law jurisdiction, a driver can be given a ticket for not wearing a seatbelt only when there is another citable traffic infraction. Although mandatory seatbelt laws have positive influence on passengers' seatbelt use, as of July 2021, only 35 states, the District of Columbia, Guam, the Northern Mariana Islands, Puerto Rico, and the Virgin Islands have primary seatbelt laws for front-seat occupants (Governors Highway Safety Association, 2021).

In the United States, nationwide seatbelt use estimates are provided by the National Highway Traffic Safety Administration (NHTSA) through the National Occupant Protection Use Survey (NOPUS) – the only survey that provides nationwide probability-based observed data on seatbelt use. Although the surveys have yielded stable estimates over time and provided significant insights about the extent to which Americans are buckling up, several questions remain unanswered. For instance, how likely would out-of-state drivers who are residents from states with weaker laws (i.e., secondary/no laws) use seatbelts when traveling in states with more stringent laws (primary laws)? Moreover, does the seasonal factors (i.e., time of day, day of week, month of year) impact seatbelt use since these surveys (NOPUS) typically are conducted in the daytime (7 a.m. – 6 p.m.) around the month of June? The lack of adequate information or the impracticability of conducting observational surveys all year round limits the findings in the national surveys. In the absence of detailed data from roadside observations, the crash data can be used as a surrogate measure to fill in the gaps. An advantage of using crash data is that it provides more comprehensive information on vehicle occupant characteristics (e.g. age, licensing status, state of residence) that would be more difficult to collect from roadside observations.

In 2019, 47% of passenger vehicle occupants killed in crashes in the United States were unrestrained drivers and passengers (National Center for Statistics and Analysis, 2019). If nationwide seatbelt use rate was at 90% in 2019, then this means that the remaining 10% of the population accounts for almost half of the vehicle occupant deaths in the United States. Considering the high prevalence of unrestrained fatalities in motor-vehicle crashes, renewed attention is needed to promote seatbelt use across the United States. To help inform these efforts, the primary objective of this research study was to explore seatbelt use among in-state and out-of-state drivers and investigate the differences in seatbelt use between the two groups in terms of the personal, temporal, and spatial characteristics, utilizing crash data from the Fatality Analysis Reporting System (FARS).

2. Methodology

2.1. Data source

Research on seatbelt usage in the United States is provided by NHTSA through its nationwide seatbelt use surveys (NOPUS). Although the surveys have provided stable measurements over time, the results have some limitations due to the observational nature of the data collection. The lack of reliable data on vehicle occupants' personal characteristics, as well as the difficulty in tracking belt use at night, are some of the challenges in the NOPUS studies. Due to these limitations, researchers sometimes rely on crash data or self-reported surveys to fill in the research gaps (Beck et al., 2019; Harper & Strumpf, 2017). In this study, crash data from FARS are utilized as a surrogate measure to estimate adult seatbelt use. The FARS data contain information on vehicle crashes that resulted in at least one fatality. The crash data provides more comprehensive data on vehicle occupant characteristics (e.g., age, licensed status), as well as the characteristics of the vehicle (vehicle type, state of registration), the time, and location of the crash.

2.2. Data and parameters

The analysis in this study is based on a 10-year period (2010–2019) crash data. The data included a total of 483,431 cases (drivers involved in fatal crashes) from the 50 states and the District of Columbia (D.C.). After filtering out some cases to align the data with the research objective, a representative sample of 373,958 cases was used for the analysis. The variables of interest extracted from the crash data included personal, temporal, and spatial attributes, briefly described below:

1. *Driver seatbelt use:* This data element indicates whether the driver used shoulder belt, lap belt, both lap and shoulder belt, or none during the crash. This variable is used as a binary dependent variable with "Yes" for drivers wearing shoulder/lap belts or both lap and shoulder belts, and "No" for drivers not belted or used seatbelts improperly.
2. *Driver gender:* This variable shows the sex of the driver either as male or female.
3. *Driver age:* This variable indicates how old the driver was at the time of the crash. In the FARS database, it is recorded as a continuous variable. Driver age is categorized into four age groups – below 18, 18–35, 36–64, and 65-and-older age groups.
4. *Driver license:* This data element identifies the status of the driver's license at the time of the crash. The variable includes three categories – no license, invalid license (suspended, revoked, or expired), and valid license.
5. *In-State/Out-of-State Driver:* This variable indicates the state of resident of the driver at the time of the crash. This variable is derived using the zip code of the driver's address and the state of issuance of the license held by the driver. If the crash occurred in a state that has the same zip code as the driver's zip code address, the driver is considered an in-state driver or else out-of-state driver. Drivers who were not U.S. residents or its Territories and without zip code addresses were excluded from the analysis.
6. *Vehicle type*: This variable identifies a classification of vehicles based on their general body configuration, size, shape, doors, and so forth. In the FARS database, vehicles are classified under 99 different vehicle types. Based on vehicles with similar characteristics, five groups of vehicle types are created – (a) passenger cars, (b) sport utility vehicles and minivans (SUV-Van), (c) light/pickup trucks, (d) buses, and (e) large trucks. Special vehicles such as farm or construction equipment, golf cart, and three-wheel automobiles were excluded from the dataset.
7. *Injury severity:* This data element describes the severity of the injury to the driver in the crash. Vehicle occupant injury severity in FARS is reported by the *KABCO* scale: *K*-fatal injury, *A*-serious injury, *B*-minor injury, *C*-possible injury,

and *O*-no injury. To simplify the analysis, the crash outcomes are categorized into three groups: fatal injury (K), non-fatal injury (A, B, C) and no injury (O).

8. *Seatbelt law:* This variable identifies the type of seatbelt law enacted at the location where the crash occurred. The seatbelt laws and when they became effective in each of the 50 States and DC were retrieved from other sources (Governors Highway Safety Association, 2021; Institute, 2021) and paired with the crash data. Three categories of laws are identified, namely, no law, secondary law, and primary law.
9. *Land use:* This variable identifies the roadway segments on which the crashes occurred based on Federal Highway Administration (FHWA)-approved adjusted census boundaries of rural and urbanized areas. The variable is categorized into rural and urban areas.
10. *Road type:* This data element identifies the functional classification of the road segments on which the crashes occurred. Four categories of road types are considered, including, interstates, arterials, collectors, and local streets.
11. *Month:* This variable describes the month of the year (January through December) in which the crashes occurred.
12. *Day:* This data element records the days of the week (Sunday through Saturday) on which the crashes occurred.
13. *Time:* This variable records the hours at which the crashes occurred. The hourly times are separated into daytime (6 a.m. – 5:59 p.m.) and nighttime (6 p.m. – 5:59 a.m.) periods.

In processing the data for the analysis, cases with missing data or unknown attributes were excluded from the dataset, yielding a final sample size of 373,958 cases.

2.3. Statistical analysis

Descriptive statistics was used to assess the seatbelt usage among in-state and out-of-state drivers. Statistical tests were performed to compare the difference in seatbelt usage between the two groups in terms of the personal, temporal, and spatial variables. The statistical significance between the two groups was evaluated using Z-test statistic. Considering two proportions P_1 (out-of-state driver's seatbelt use) and P_2 (in-state driver's seatbelt use), the null and alternate hypotheses were stated as follows:

Null hypothesis: seatbelt use rate of out-of-state drivers is the same as that of in-state drivers. H_o: $P_1 = P_2$.
Alternate hypothesis: seatbelt use rate of out-of-state drivers differs from that of in-state drivers. H_a: $P_1 \neq P_2$.

The z-statistic test is shown in Eq. (1).

$$z_{cal} = \frac{(P_1 - P_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}} \quad (1)$$

where: z_{cal}= z test statistic compared to the standard normal deviate.
P_1 = seatbelt use rate of out-of-state drivers.
P_2 = seatbelt use rate of in-state drivers.
N_1 = the total number of out-of-state drivers.
N_2 = the total number of in-state drivers.
$\hat{p}$ = the estimated true proportion under the null hypothesis equal to $\left[\frac{P_1N_1+P_2N_2}{N_1+N_2}\right]$.

The calculated z test statistic (z_{cal}) is compared with z critical values on standard normal table. At 95% confidence level, the z critical value is 1.96 for a two-tail test. If z_{cal} is greater than z critical, then the null hypothesis is rejected, indicating that seatbelt use rate differs significantly between out-of-state and in-state drivers.

Logistic regression models were developed separately to examine the probability of seatbelt use among the two groups. The logistic regression does not assume equal distribution of the dependent variable for each level of independent variable, nor does it assume a normal distribution of the variables (Agresti, 2009). As such, it is one of the most widely used analytical tool in traffic safety research. A binary logistic model was fitted with driver seatbelt use (1 = seatbelt use, 0 = no seatbelt use) as the dependent variable and the personal, temporal and spatial parameters as the independent variables. The logistic regression model applies maximum likelihood estimation after transforming the categorical dependent variable into a logit variable. The logit model formulation is shown in Eq. (2), with Y indicating the dependent variable, X_i the independent variables and β_i the estimated parameter coefficients.

$$logit(Y) = \beta_0 + \beta_1X_1 + \beta_2X_2 + \cdots + \beta_nX_n \quad (2)$$

The logistic regression yields odds ratios for the individual variables. Odds ratio represent the odds of a driver belted relative to the odds of not belted, with a unit change in the independent variable. Odds ratio is often interpreted as risk ratio, but, only under certain conditions does the odds ratio approximate the risk ratio. When the outcome of interest in the study population is low (less than 10%), the odds ratio is close to the risk ratio. However, when the outcome is more frequent, the more the odds ratio will overestimate the risk ratio when it is more than 1 or underestimate the risk ratio when it is less than 1. To provide a measure that better represents the true relative risk, the adjusted odds ratios obtained from the logistic regression were corrected using the formula proposed by Zhang and Yu (Zhang & Yu, 1998) as shown in Eq. (3).

$$RR = \frac{OR}{(1-P_0)+(P_o \times OR)} \quad (3)$$

where RR represents risk ratio; OR, odds ratio; and P_o, outcome of interest (seatbelt use rate) in the reference group.

3. Results

3.1. Comparison of seatbelt use between out-of-state and in-state drivers based on personal characteristics

Table 1 shows the comparison of the average seatbelt use between out-of-state and in-state drivers in terms of their personal characteristics. Any significant difference between the two groups is indicated by the z-test statistics.

As shown in Table 1, the proportion of out-of-state male and female drivers belted were significantly higher than their corresponding in-state drivers. In both groups, female drivers showed higher seatbelt compliance rates compared to their male counterparts.

Regarding driver age, all age groups (i.e., 18–34 years, 35–64 years, and over 65 years) except teenagers (below 18 years) exhibited higher seatbelt use rate among out-of-state of drivers compared to their corresponding in-state drivers. Table 1 shows a strong correlation between seatbelt use and driver age as belt use increased with age. The proportions of belted older adults (35–54, >65 years) were greater than the younger age groups (<18, 18–34 years) for both in-state and out-of-state driver groups.

For driver license status, out-of-state drivers who possessed valid or invalid (suspended, revoked, expired, or canceled) licenses showed higher seatbelt compliance rates than their corresponding

Table 1
Seatbelt Use Based on Personal Characteristics.

Variable	Categories*	Out-of-State Drivers			In-State Drivers			Z-Value	P-value
		Sample (N_1)	Belted	%Belted (P_1)	Sample (N_2)	Belted	%Belted (P_2)		
Gender	Male	34,845	27,485	78.9%	228,164	163,528	71.7%	28.101	<0.001
	Female	10,214	8,367	81.9%	100,735	80,789	80.2%	4.163	<0.001
	All	45,059	35,852	79.6%	328,899	244,317	74.3%	24.264	<0.001
Age	Below 18 years	514	369	71.8%	9,247	6,791	73.4%	−0.824	0.410
	18–34 years	14,888	10,919	73.3%	123,974	85,544	69.0%	10.863	<0.001
	35–64 years	23,565	19,481	82.7%	146,673	112,810	76.9%	19.708	<0.001
	Over 65 years	6,092	5,083	83.4%	49,005	39,172	79.9%	6.485	<0.001
License Status	Unlicensed	1,188	709	59.7%	13,966	8,686	62.2%	−1.714	0.087
	Invalid	2,844	1,595	56.1%	23,687	12,362	52.2%	3.930	<0.001
	Valid	41,027	33,548	81.8%	291,246	223,269	76.7%	23.133	<0.001
Vehicle type	Car	15,076	11,496	76.3%	154,837	114,933	74.2%	5.440	<0.001
	SUV-Van	9,902	7,789	78.7%	79,144	59,767	75.5%	6.893	<0.001
	Pickup/light Truck	7,919	5,370	67.8%	67,718	45,372	67.0%	1.452	0.147
	Bus	203	186	91.6%	1,988	1,883	94.7%	−1.830	0.067
	Large Truck	11,959	11,011	92.1%	25,212	22,362	88.7%	10.042	<0.001
Severity	No injury	15,440	14,997	97.1%	97,797	94,057	96.2%	5.847	<0.001
	Non-fatal	13,588	11,849	87.2%	94,675	80,010	84.5%	8.183	<0.001
	Fatal	16,031	9,006	56.2%	136,427	70,250	51.5%	11.234	<0.001

*Chi-square test shows significant association between seatbelt use and all categorical variables at 95% confidence level.

in-state drivers. However, among unlicensed drivers, no significant difference was found between out-of-state and in-state drivers. In both groups, the proportions of belted drivers who possessed valid licenses were greater than the proportions of belted unlicensed drivers or those with invalid licenses.

Considering vehicle types, out-of-state drivers who traveled in passenger cars, SUVs and vans, and large trucks showed higher seatbelt use rates than their corresponding in-state drivers. Among occupants of pickup/light trucks and buses, no significant differences were found between the two groups.

From Table 1, there appears to be a strong relationship between driver seatbelt use and crash injury outcome. The proportions of drivers who were restrained at the time of the crash were higher (over 95% belted) among uninjured drivers than those who suffered fatal injuries (about 50% belted). For each injury category, out-of-state drivers recorded higher seatbelt use rates than their corresponding in-state drivers.

3.2. Comparison of seatbelt use between out-of-state and in-state drivers based on temporal and spatial characteristics

Table 2 shows the comparison of the average seatbelt use between out-of-state and in-state drivers based on several environmental factors.

As shown in Table 2, the seatbelt use rates for out-of-state drivers were significantly higher than in-state drivers regardless of the type of seatbelt law in the jurisdiction where the crash occurred. The point differences in seatbelt use between the two groups are noticeably higher (about 15% points difference) in states with secondary/no seatbelt laws. Further analysis to explore the seatbelt use of out-of-state drivers based on the seatbelt laws at their resident states is presented in Table 3. As shown in Table 3, 73.2% of out-of-state drivers from secondary/no seatbelt law states were belted in states with secondary/no laws, while 82.0% of out-of-state drivers from primary seatbelt law states were belted in states with secondary/no laws. These proportions (i.e., 73.2% and 82.0%) are significantly higher than the average seatbelt use of in-state drivers (64.7%). Likewise, considering crash locations where the seatbelt law was primary, 80.2% of out-of-state drivers from secondary/no seatbelt law states were belted while 79.4% of out-of-state drivers from primary seatbelt law states wore seatbelts. Again, these proportions are statistically significant compared to the average seatbelt use of in-state drivers in primary seatbelt states (76.7% belted). Out-of-state drivers from states with weaker laws (secondary/no laws) exhibited higher seatbelt compliance rates in states with stricter laws (primary laws).

Regarding land use, seatbelt use was lower in rural areas compared to urban areas. In both rural and urban areas, out-of-state drivers exhibited higher seatbelt compliance rate than their corresponding in-state drivers. Further analysis by road type showed significantly higher seatbelt use rates among out-of-state drivers than among in-state drivers on interstate and arterial roads. On collector streets, the difference was not significant. However, on local streets, seatbelt use was higher among in-state drivers compared to out-of-state drivers.

Analyses of seatbelt use by the seasonal factors showed that irrespective of the month of the year, day of the week or time of day, out-of-state drivers were more seatbelt compliant than their corresponding in-state drivers. The seatbelt use rate among out-of-state drivers ranged from as low as 78.7% in the month of November to as high as 80.4% in the month of April, whereas the rates among in-state drivers ranged from 73.5% in the month of January to 75.5% in the month of September. Both out-of-state and in-state drivers recorded lower seatbelt use rates during the weekends. Moreover, belt use was lower at nighttime than daytime for both out-of-state and in-state drivers.

3.3. Logistic regression model results

Table 4 shows the logistic regression models indicating the probability of seatbelt use among out-of-state and in-state drivers. Model-1 shows the results for the combined dataset. As shown in Model-1, out-of-state drivers were about 5% more likely than in-state drivers to wear seatbelts. Because of the differences observed between the two groups, separate logistic regression models were performed, one for out-of-state drivers (Model-2) and the other for in-state drivers (Model-3).

Results from the combined data (Model-1) show that the young age groups (<18, 18–34, and 35–64 years) compared to the older adults (>65 years) were less likely to wear seatbelts. Males compared to females were about 9% less likely to use seatbelts. Drivers who were not licensed, as well as those with invalid licenses (suspended, revoked, expired or canceled) were less likely to be restrained. Compared to drivers of passenger cars, drivers of pickup/light trucks, SUVs and vans showed lower likelihood of wearing seatbelts. The results for large trucks and buses were

Table 2
Seatbelt Use Based on Temporal and Spatial Factors.

Variable	Categories	Out-of-State Drivers			In-State Drivers			Z-Value	P-value
		Sample (N_1)	Belted	%Belted (P_1)	Sample (N_2)	Belted	%Belted (P_2)		
Seatbelt Law	No Law	245	154	62.9%	991	479	48.3%	4.072	**<0.001**
	Secondary	9,895	7,911	79.9%	65,954	42,853	65.0%	29.525	**<0.001**
	Primary	34,919	27,787	79.6%	261,954	200,985	76.7%	11.900	**<0.001**
Land Use	Rural	27,040	20,992	77.6%	161,143	110,244	68.4%	30.538	**<0.001**
	Urban	18,019	14,860	82.5%	167,756	134,073	79.9%	8.148	**<0.001**
Road Type	Interstate	31,001	25,202	81.3%	191,924	145,059	75.6%	21.971	**<0.001**
	Arterial	9,732	7,935	81.5%	83,650	65,482	78.3%	7.411	**<0.001**
	Collector	2,725	1,754	64.4%	27,549	17,339	62.9%	1.474	0.141
	Local	1,601	961	60.0%	25,776	16,437	63.8%	−3.020	**0.003**
Month	January	3,278	2,582	78.8%	25,904	19,047	73.5%	6.452	**<0.001**
	February	2,997	2,372	79.1%	23,013	16,938	73.6%	6.528	**<0.001**
	March	3,658	2,886	78.9%	25,950	19,190	73.9%	6.430	**<0.001**
	April	3,497	2,810	80.4%	25,665	18,911	73.7%	8.489	**<0.001**
	May	3,656	2,933	80.2%	27,495	20,339	74.0%	8.169	**<0.001**
	June	4,099	3,293	80.3%	27,253	20,380	74.8%	7.712	**<0.001**
	July	4,240	3,370	79.5%	27,835	20,714	74.4%	7.102	**<0.001**
	August	4,118	3,280	79.7%	28,438	21,162	74.4%	7.259	**<0.001**
	September	3,767	3,016	80.1%	28,482	21,499	75.5%	6.188	**<0.001**
	October	4,019	3,197	79.5%	30,348	22,698	74.8%	6.572	**<0.001**
	November	3,827	3,010	78.7%	29,448	21,909	74.4%	5.707	**<0.001**
	December	3,903	3,103	79.5%	29,068	21,530	74.1%	7.335	**<0.001**
Day	Monday	6,026	4,801	79.7%	44,362	33,606	75.8%	6.702	**<0.001**
	Tuesday	5,914	4,791	81.0%	43,193	32,995	76.4%	7.914	**<0.001**
	Wednesday	6,062	4,883	80.6%	43,690	33,319	76.3%	7.411	**<0.001**
	Thursday	6,397	5,168	80.8%	45,842	34,834	76.0%	8.493	**<0.001**
	Friday	7,299	5,881	80.6%	52,726	39,740	75.4%	9.753	**<0.001**
	Saturday	7,098	5,524	77.8%	53,331	37,816	70.9%	12.155	**<0.001**
	Sunday	6,263	4,804	76.7%	45,755	32,007	70.0%	11.017	**<0.001**
Time	Daytime	25,123	21,001	83.6%	179,902	140,981	78.4%	19.057	**<0.001**
	Nighttime	19,936	14,851	74.5%	148,997	103,336	69.4%	14.865	**<0.001**

*Chi-square test shows significant association between seatbelt use and all categorical variables at 95% confidence level.

Table 3
Seatbelt Use Based on Seatbelt Laws.

Seatbelt law at Crash Location	In-state Drivers		Out-of-State Drivers			
	No/Secondary/Primary law		Secondary/No law		Primary law	
	Total	%Belted*	Total	%Belted*	Total	%Belted*
Secondary/No law	66,945	[a]64.7%	2,870	[a]73.2%	7,270	[a]82.0%
Primary law	261,954	[b]76.7%	6,851	[b]80.2%	28,068	[b]79.4%

*Row proportions compared between in-state and out-of-state drivers are significant at 95% confidence level.
[a,b]Proportions in the same column are significantly different at 95% confidence level.

not significant. In reference to drivers who were not injured in the crashes, fatally injured drivers were about 48% less likely to wear seatbelt while non-fatally injured drivers were about 12% less likely to do so.

In terms of temporal and spatial characteristics, drivers who were involved in crashes in states with secondary seatbelt laws were about 17% less likely to wear seatbelts whereas those in the state with no law (New Hampshire) were about 42% less likely to do so compared to drivers involved in crashes in primary seatbelt law states. Drivers traveling in rural areas were about 4% less likely to buckle up compared to drivers traveling in urban areas. With regard to seatbelt use by road type, drivers were less likely to wear seatbelts on local, collector streets compared to interstate roads. From Table 4 (Model-1), the coefficient estimates for the months of January, February and December are negative and significant compared to the month of June. This implies that, in the winter months, drivers were less likely to wear seatbelts compared to the summer month in June. Seatbelt use by day of the week showed that drivers were about 3% less likely to use seatbelts on weekends (Saturday and Sunday) compared to Wednesday. Model-1 also shows that drivers were about 10% less likely to use seatbelts at night compared to daytime hours.

As shown in Table 4, the signs of the coefficient estimates and their significances are the same across the three models with few differences observed among out-of-state drivers (Model-2) and in-state drivers (Model 3). Considering Model-2, there was no strong evidence to support that out-of-state drivers' seatbelt use differed across locations (rural vs urban), on arterial roads, and across the days of the week or months of the year. Seatbelt use in secondary states was marginally significant in Model-2, however, among occupants of large trucks in contrast with passenger cars, the result was significant. Similarly, the model results for in-state drivers (Model-3) showed no statistically significant difference in seatbelt use in the month of December. Additionally, seatbelt use among in-state drivers of heavy trucks was significantly lower compared to passenger cars.

4. Discussions and limitations

4.1. Discussions

This study explored the seatbelt use of in-state and out-of-state drivers using crash data from FARS. New findings in this study show that out-of-state drivers are far more likely than in-state dri-

Table 4
Results of the Logistic Regression Models.

Parameter	Category	Model-1 (Combined Data)			Model-2 (Out-of-State Drivers)			Model-3 (In-State Drivers)		
		Coeff. (B)	Odds Ratio (95% CI)	†Risk Ratio (95% CI)	Coeff. (B)	Odds Ratio (95% CI)	†Risk Ratio (95% CI)	Coeff. (B)	Odds Ratio (95% CI)	†Risk Ratio (95% CI)
Intercept		5.013*	–	–	4.897*	–	–	5.056*	–	–
Driver	Out-of-state	0.216*	1.241 (1.206–1.277)	1.053 (1.046–1.059)						
	In-state	–	–	–						
Age	<18 years	−0.656*	0.519 (0.490–0.550)	0.846 (0.830–0.861)	−0.804*	0.448 (0.353–0.567)	0.831 (0.767–0.888)	−0.646*	0.524 (0.494–0.556)	0.846 (0.830–0.862)
	18–34 years	−0.902*	0.406 (0.395–0.417)	0.776 (0.768–0.784)	−0.861*	0.423 (0.387–0.462)	0.816 (0.792–0.838)	−0.907*	0.404 (0.392–0.416)	0.772 (0.763–0.780)
	35–64 years	−0.481*	0.618 (0.602–0.635)	0.892 (0.885–0.898)	−0.449*	0.639 (0.586–0.695)	0.914 (0.895–0.932)	−0.486*	0.615 (0.598–0.633)	0.888 (0.881–0.896)
	>65 years	–	–	–	–	–	–	–	–	–
Gender	Male	−0.400*	0.670 (0.657–0.684)	0.912 (0.907–0.917)	−0.348*	0.706 (0.66–0.755)	0.93 (0.915–0.945)	−0.407*	0.666 (0.652–0.680)	0.91 (0.904–0.915)
	Female	–	–	–	–	–	–	–	–	–
License	Unlicensed	−0.414*	0.661 (0.635–0.688)	0.896 (0.885–0.907)	−0.536*	0.585 (0.510–0.672)	0.885 (0.851–0.918)	−0.409*	0.664 (0.637–0.693)	0.894 (0.883–0.906)
	Invalid	−0.764*	0.466 (0.452–0.48)	0.794 (0.784–0.803)	−0.725*	0.484 (0.442–0.531)	0.837 (0.813–0.861)	−0.769*	0.464 (0.449–0.479)	0.788 (0.777–0.798)
	Valid	–	–	–	–	–	–	–	–	–
Vehicle	Large Truck	−0.034	0.967 (0.926–1.009)	0.991 (0.980–1.002)	0.152*	1.164 (1.060–1.278)	1.035 (1.014–1.054)	−0.126*	0.882 (0.840–0.926)	0.967 (0.953–0.980)
	Bus	0.014	1.014 (0.835–1.231)	1.004 (0.952–1.050)	−0.154	0.857 (0.501–1.466)	0.962 (0.809–1.082)	0.033	1.034 (0.840–1.273)	1.009 (0.953–1.059)
	Pickup/Light Truck	−0.626*	0.535 (0.523–0.547)	0.818 (0.811–0.825)	−0.683*	0.505 (0.470–0.543)	0.811 (0.789–0.833)	−0.617*	0.539 (0.526–0.553)	0.819 (0.812–0.828)
	SUV-Van	−0.260*	0.771 (0.754–0.788)	0.929 (0.923–0.936)	−0.198*	0.820 (0.765–0.880)	0.950 (0.932–0.969)	−0.269*	0.764 (0.746–0.782)	0.926 (0.919–0.933)
	Passenger car	–	–	–	–	–	–	–	–	–
Injury	Severity	Fatal	−3.270*	0.038 (0.037–0.039)	0.517 (0.510–0.523)	−3.245*	0.039 (0.035–0.043)	0.586 (0.558–0.610)	−3.276*	0.038 (0.036–0.039)
	0.508 (0.494–0.515)									
	Non-fatal	−1.520*	0.219 (0.211–0.227)	0.884 (0.879–0.888)	−1.522*	0.218 (0.196–0.244)	0.907 (0.895–0.918)	−1.520*	0.219 (0.211–0.227)	0.880 (0.875–0.885)
No injury	–	–	–	–	–	–	–	–	–	
Seatbelt Law	No law	−1.419*	0.242 (0.211–0.278)	0.582 (0.538–0.627)	−1.028*	0.358 (0.262–0.490)	0.732 (0.635–0.825)	−1.488*	0.226 (0.194–0.263)	0.556 (0.508–0.605)
	Secondary	−0.633*	0.531 (0.520–0.542)	0.832 (0.825–0.838)	−0.061	0.941 (0.882–1.005)	0.987 (0.973–1.001)	−0.707*	0.493 (0.482–0.504)	0.807 (0.800–0.814)
	Primary	–	–	–	–	–	–	–	–	–
Land Use	Rural	−0.171*	0.842 (0.827–0.858)	0.964 (0.960–0.968)	0.015	1.015 (0.959–1.075)	1.003 (0.993–1.012)	−0.199*	0.82 (0.804–0.836)	0.958 (0.953–0.962)
	Urban	–	–	–	–	–	–	–	–	–
Road Type	Local	−0.580*	0.560 (0.542–0.579)	0.843 (0.834–0.853)	−0.909*	0.403 (0.355–0.458)	0.783 (0.746–0.819)	−0.555*	0.574 (0.555–0.594)	0.847 (0.836–0.857)
	Collector	−0.358*	0.699 (0.678–0.720)	0.908 (0.899–0.916)	−0.590*	0.555 (0.502–0.613)	0.870 (0.843–0.894)	−0.323*	0.724 (0.701–0.747)	0.915 (0.906–0.924)
	Arterial	0.067	1.069 (1.046–1.092)	1.015 (1.010–1.020)	−0.049	0.952 (0.890–1.018)	0.991 (0.977–1.003)	0.082*	1.085 (1.061–1.110)	1.020 (1.014–1.025)
	Interstate	–	–	–	–	–	–	–	–	–
Month	January	−0.078*	0.925 (0.886–0.966)	0.981 (0.969–0.991)	−0.074	0.929 (0.813–1.061)	0.985 (0.957–1.011)	−0.077*	0.926 (0.884–0.969)	0.980 (0.968–0.992)
	February	−0.056*	0.946 (0.905–0.989)	0.986 (0.975–0.997)	−0.053	0.948 (0.828–1.087)	0.989 (0.961–1.016)	−0.055*	0.947 (0.903–0.992)	0.986 (0.974–0.998)
	March	−0.039	0.962 (0.922–1.005)	0.990 (0.980–1.001)	−0.061	0.941 (0.826–1.070)	0.988 (0.960–1.013)	−0.034	0.967 (0.924–1.012)	0.991 (0.980–1.003)
	April	−0.031	0.969 (0.928–1.012)	0.992 (0.981–1.003)	0.058	1.060 (0.929–1.210)	1.011 (0.985–1.035)	−0.041	0.959 (0.916–1.004)	0.989 (0.977–1.001)
	May	−0.024	0.976 (0.935–1.019)	0.994 (0.983–1.005)	0.019	1.019 (0.895–1.161)	1.004 (0.977–1.028)	−0.028	0.973 (0.930–1.017)	0.993 (0.981–1.004)
	July	−0.004	0.957 (0.918–0.998)	0.989 (0.979–1.000)	0.011	0.940 (0.827–1.067)	0.988 (0.960–1.013)	−0.006	0.961 (0.919–1.005)	0.990 (0.978–1.001)
	August	−0.018	0.996 (0.954–1.039)	0.999 (0.988–1.009)	−0.05	1.011 (0.893–1.146)	1.002 (0.977–1.026)	−0.015	0.994 (0.950–1.039)	0.998 (0.987–1.010)
	September	0.021	0.982 (0.942–1.025)	0.996 (0.985–1.006)	0.021	0.951 (0.838–1.079)	0.990 (0.963–1.015)	0.023	0.986 (0.942–1.031)	0.996 (0.985–1.008)
	October	−0.024	1.022 (0.979–1.066)	1.005 (0.995–1.015)	−0.041	1.022 (0.898–1.163)	1.004 (0.978–1.028)	−0.021	1.023 (0.978–1.070)	1.006 (0.994–1.017)
	November	−0.036	0.976 (0.936–1.018)	0.994 (0.984–1.004)	−0.099	0.960 (0.846–1.090)	0.992 (0.965–1.017)	−0.028	0.979 (0.937–1.024)	0.995 (0.983–1.006)

Table 4 (*continued*)

Parameter	Category	Model-1 (Combined Data)			Model-2 (Out-of-State Drivers)			Model-3 (In-State Drivers)		
		Coeff. (B)	Odds Ratio (95% CI)	†Risk Ratio (95% CI)	Coeff. (B)	Odds Ratio (95% CI)	†Risk Ratio (95% CI)	Coeff. (B)	Odds Ratio (95% CI)	†Risk Ratio (95% CI)
	December	−0.044*	0.964 (0.925–1.006)	0.991 (0.981–1.001)	−0.062	0.906 (0.797–1.029)	0.980 (0.952–1.006)	−0.04	0.972 (0.930–1.016)	0.993 (0.981–1.004)
	June	–	–	–	–	–	–	–	–	–
Day	Sunday	−0.132*	0.876 (0.847–0.906)	0.968 (0.960–0.976)	−0.046	0.955 (0.863–1.058)	0.991 (0.970–1.011)	−0.145*	0.865 (0.835–0.896)	0.964 (0.955–0.973)
	Monday	0.000	1.000 (0.966–1.034)	1.000 (0.992–1.008)	0.013	1.013 (0.914–1.124)	1.003 (0.982–1.022)	−0.002	0.998 (0.963–1.035)	1.000 (0.991–1.008)
	Tuesday	0.016	1.016 (0.982–1.052)	1.004 (0.996–1.012)	0.032	1.032 (0.929–1.147)	1.006 (0.985–1.026)	0.015	1.016 (0.979–1.053)	1.004 (0.995–1.012)
	Thursday	−0.005	0.995 (0.962–1.030)	0.999 (0.991–1.007)	0.038	1.038 (0.937–1.151)	1.007 (0.987–1.026)	−0.011	0.989 (0.954–1.025)	0.997 (0.989–1.006)
	Friday	−0.011	0.989 (0.957–1.022)	0.997 (0.990–1.005)	0.047	1.048 (0.949–1.159)	1.009 (0.990–1.027)	−0.017	0.983 (0.950–1.018)	0.996 (0.988–1.004)
	Saturday	−0.132*	0.876 (0.848–0.905)	0.968 (0.960–0.976)	−0.023	0.978 (0.885–1.080)	0.996 (0.975–1.015)	−0.146*	0.864 (0.835–0.894)	0.964 (0.955–0.973)
	Wednesday	–	–	–	–	–	–	–	–	–
Time	Night	−0.424*	0.655 (0.643–0.667)	0.900 (0.896–0.905)	−0.486*	0.615 (0.582–0.650)	0.907 (0.895–0.919)	−0.418*	0.658 (0.646–0.671)	0.899 (0.894–0.904)
	Day	–	–	–	–	–	–	–	–	–

*Parameter coefficient estimate significant at 0.05 alpha level.
†Risk ratio corrected by Eq. (3) using odds ratio from logistic regression results.
Goodness of fit (Model-1): Deviance value = 135,392.1; df = 144,901; value/df = 0.93; Log likelihood = −93,598.6.
Goodness of fit (Model-2): Deviance value = 24,957.6; df = 29,652; value/df = 0.84; Log likelihood = −14,357.3.
Goodness of fit (Model-3): Deviance value = 109,836.5; df = 115,213; value/df = 0.95; Log likelihood = −78,942.2.

vers to use seatbelts. A possible reason for the higher compliance rate among out-of-state drivers may be their exposure to traffic and travel distance. Gkritza and Mannering (2008), and Lipovac et al. (2015) found that drivers traveling longer distances especially on high-speed roads have higher likelihood of wearing seatbelt than those traveling shorter distances on low-speed facilities. Out-of-state drivers are mostly long-distance travelers on interstates and arterial roads (high-speed facilities) and have higher risk perceptions which probably make them feel the need to wear safety belts while traveling across states.

Strong associations between seatbelt use and various personal factors such as driver gender, age, vehicle type, license status and injury outcome were found among out-of-state and in-state drivers. Particularly, males compared to females were less likely to use seatbelts. This finding is consistent with numerous studies in the literature (Beck et al., 2019; Boakye et al., 2019a, 2019b). The gender difference in seatbelt use may be attributed to societal behaviors. As reported by Lipovac et al. (2015), male drivers tend to be riskier in order to show their self-confidence of being the "stronger" sex and, as a result, their inclination not to act in accordance with legal provisions when compared to female drivers.

Among in-state and out-of-state drivers, driver age was found to be positively associated with seatbelt use. The teen (<18 years), young (18–34 years) and middle-aged (35–64) drivers were less likely to use seatbelts compared to older adults. This finding supports previous studies (Beck et al., 2007; Kim & Kim, 2003; Lee & Schofer, 2003) where young drivers were found to exhibit risky driving behaviors including non-use of seatbelt and drunk-driving.

Results of the analysis also showed that unlicensed drivers and drivers with invalid licenses (i.e., suspended, revoked, or expired) were far less likely to use seatbelts compared to those with valid licenses. This was true among both in-state and out-of-state drivers. In a previous study by Kim and Kim (2003), unlicensed drivers were 1.4 times more likely than licensed drivers to be unbelted. These findings warrant the need to pay more attention to the relationship between unlicensed drivers and seatbelt non-use.

The logistic regression models for both out-of-state and in-state drivers showed that drivers of SUVs, vans, pickups, and light trucks were less likely to wear seatbelts compared to occupants of passenger cars. The findings are consistent with past studies (Boakye et al., 2019; Glassbrenner et al., 2004; Kim & Kim, 2003). As expected, occupants of pickup or light trucks have historically demonstrated lower seatbelt use rates. The false sense of increased safety in pickup trucks compared to passenger cars may be a contributory factor for the lower seatbelt use. Special educational and enforcement campaigns are needed to motivate seatbelt use among such driving populace (Boakye et al., 2019). For example, Nichols et al. demonstrated that combining media campaigns and high visibility programs such as "Buckle up in your truck" is one of the effective means of getting occupants of pickup trucks to buckle up (Nichols et al., 2009).

Seatbelt use among drivers who sustained no injuries in crashes was substantially higher compared to fatally or non-fatally injured drivers. The non-use of seatbelts among the fatally injured drivers may possibly have contributed to their deaths in the crash incidents. This finding provides strong evidence to support that the safety belt is one of the most effective devices that can reduce the severity of injuries in crashes. According to NHTSA, seatbelts, when properly used, can reduce fatal injuries to front seat car passengers by 45% and moderate-to-critical injuries by 50% (National Center for Statistics and Analysis, 2017). Other studies from different countries also show that the use of seatbelt decreases the probability of being killed in a traffic crash by about 40–50% for drivers and front-seat passengers and by about 25% for passengers traveling in the rear seat (Elvik, 2004).

The results in this study also support that the driving environment affects seatbelt use. States with primary seatbelt laws recorded higher seatbelt use rates than states with no or secondary seatbelt laws. Several studies found similar results whether the sampled data used for the analyses were observational surveys, self-reported surveys or crash data (Boakye & Nambisan, 2020; National Center for Statistics and Analysis, 2019; Shakya et al., 2020). Novel findings in this study showed that seatbelt use rates among out-of-state drivers were significantly higher than among in-state drivers regardless of the type of seatbelt law enacted at the residences of out-of-state drivers. For example, the seatbelt use of out-of-state drivers involved in crashes in secondary/no seatbelt law states was 73.2% for those traveling from secondary seatbelt law jurisdictions and 80.2% for those traveling from primary seatbelt law states. These averages were higher than the rates for in-state drivers (secondary/no law: 64.7%; primary law: 76.7%). The higher seatbelt use rate in states with primary laws confirms how important primary enforcement is in influencing driver behavior even among vehicle occupants from jurisdictions with less strict laws. Therefore, enacting stricter seatbelt laws, accompanied by effective enforcement programs and heightened public awareness campaigns could help increase adult seatbelt compliance, which in turn, would minimize the prevalence of unrestrained fatalities in motor vehicle crashes (Boakye, 2017; Shults et al., 2004; Vasudevan et al., 2009). Given the significant difference in driver seatbelt use behavior based on the type of seatbelt law, there is strong evidence to support that if states with weaker laws upgrade to primary laws, there likely will be increases in seatbelt use in those states.

Seatbelt use among in-state and out-of-state drivers were significantly higher in urban areas than in rural areas. The seatbelt use rate among out-of-state drivers was about 10 points different from in-state drivers in rural areas. However, in the logistic regression model for out-of-state drivers, the parameter land use (rural vs urban) was not a significant predicator unlike the model for in-state drivers. These findings are likely picking up differences in driver behavior in the effect of public awareness campaigns and enforcement programs (or lack of) across different geographic locations in the U.S. as expressed in past studies (Boakye et al., 2019; Gkritza & Mannering, 2008; Goetzke & Islam, 2015).

The parameter indicator variable for road type facilities showed that both in-state and out-of-state drivers were far less likely to use seatbelts traveling on local and collector streets compared to interstate roads. Previous studies have shown that drivers are more likely to be restrained traveling on high-speed road facilities and when traveling longer distances (Boakye et al., 2019; Gkritza & Mannering, 2008). The lower risk perception at lower-speed road facilities and when traveling shorter distances or drivers' consciousness of the dangers of risky driving on high-speed, high-volume expressways may be reasons for the varying differences in seatbelt use by road type.

The results of the effects of the seasonal and temporal factors (i.e., month of the year, day of the week, and time of day) on seatbelt use among in-state and out-of-state drivers were mixed. The parameter estimates for month and day were not statistically significant in the logistic regression model for out-of-state drivers. Thus, regardless of the month or day in which out-of-state drivers made their trips, their attitudes or behaviors towards seatbelt use remained consistent unlike in-state drivers who were less likely to wear seatbelts during the months of January and February, and on weekends. The model results for in-state drivers confirm similar findings in the literature where seatbelt use estimates were lower during the winter months (Boakye et al., 2019) and on weekends (Goetzke & Islam, 2015; Lipovac et al., 2015). In the logistic regression models, both in-state and out-of-state drivers were less likely to wear seatbelts at night supporting previous findings in the literature (Chaudhary & Preusser, 2006; Tison et al., 2010). The seatbelt non-use during night hours may be due to several confounding factors such as the perception of not being caught in the dark by enforcement officers or forgetfulness because of driving impairment. Considering this seasonality effect may be an important factor when developing intervention strategies and safety-belt use campaigns.

4.2. Limitations

The findings in this study are subject to several limitations. First, the study did not use data for the population of all vehicle occupants in the U.S. to measure the overall risk of unrestrained drivers in motor vehicle crashes. The FARS database contains information on all crashes that resulted in at least one fatality. Fatal crashes could be disproportionately related to factors such as impaired driving (e.g. alcohol or drug use), speeding, or other confounding factors which likely are bound to introduce some biasness in the estimation of seatbelt use. Since unbelted drivers are more likely to be killed in fatal crashes, seatbelt use derived from FARS data may underestimate actual use. However, since the focus of this study was to understand the differences in seatbelt use among in-state and out-of-state drivers, it is expected that the magnitude of such limitation will not detract from the objective of the analysis. The results show that out-of-state drivers do buckle up more frequently than in-state drivers, although it may or may not be the exact estimate using fatal crash data. This finding likely applies to the population even if the magnitude of the difference does not.

Second, people not killed in fatal crashes may inaccurately self-report their belt use, resulting in higher use rate than those fatally injured. As noted in this study, seatbelt use among out-of-state and in-state drivers who sustained no injuries were 97% and 96%, respectively, while corresponding rates among fatally injured drivers were 56% and 51%. Self-reporting of certain behaviors can be subject to social desirability bias. For example, in states with more stringent seatbelt laws, drivers are more likely to falsely report their seatbelt use status for the fear of being ticketed by police officers responding to and preparing the crash reports. Thus, seatbelt laws may also affect self-reported seatbelt use in the FARS data. Researchers should be cognizant of this fact when using crash data for seatbelt use estimates.

Lastly, seatbelt use or non-use may be associated with several factors that are not captured in this study. Although the enactment of seatbelt laws across the U.S. has played a vital role in promoting seatbelt use, the increases are likely influenced by enforcement and public awareness campaigns, and driver attitudinal and behavioral changes over the years.

5. Conclusion

Promoting seatbelt use to reduce traffic fatalities and serious injuries continues to be a national public health priority. Each year, almost half of vehicle occupants involved in fatal crashes in the U.S. are unrestrained. In light of the high prevalence of unbelted fatalities, enhanced strategies to increase seatbelt use are needed. This study investigated the differences in seatbelt use among out-of-state and in-state drivers using crash data from FARS.

New findings in this study showed that out-of-state drivers are about 5% more likely than in-state drivers to use seatbelts accounting for their personal (age, gender, license status, etc.) and crash characteristics (time, location, roadway factors, etc.). Moreover, out-of-state drivers traveling from states with secondary or no seatbelt laws showed higher seatbelt compliance rate in primary seatbelt law states than in states with less strict laws (secondary/no laws). The results in this study are critical to addressing a myriad of policy questions regarding seatbelt use and seatbelt laws. Future research should focus on the disparity in seatbelt use between the two groups (out-of-state and in-state drivers) and examine intervention strategies that are effective at promoting seatbelt use across the United States.

Declaration of interest

The views and opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of the institution where the individual is employed. The author thanks his current employer for supporting his research and dissemination efforts.

References

Elvik, R., & Vaa, T. (2004). *The Handbook of Road Safety Measures*.
Agresti, A. (2009). *An Introduction to Categorical Data Analysis. Second Edition*. John Wiley & Sond, Inc.
Beck, L. F., Kresnow, M., & Bergen, G. (2019). Belief about seat belt use and seat belt wearing behavior among front and rear seat passengers in the United States. *Journal of Safety Research, 68*, 81–88. https://doi.org/10.1016/j.jsr.2018.12.007.
Beck, L. F., Shults, R. A., Mack, K. A., & Ryan, G. W. (2007). Associations between sociodemographics and safety belt use in states with and without primary enforcement laws. *American Journal of Public Health, 97*(9), 1619–1624. https://doi.org/10.2105/AJPH.2006.092890.
Boakye, K. F. (2017). *Evaluation of Increased Targeted Enforcement and Community-Based Outreach and Education Programs to Increase Nighttime Seatbelt Use in East Tennessee*. University of Tennessee.
Boakye, K. F., Khattak, A., Everett, J., & Nambisan, S. (2019). Correlates of front-seat passengers' non-use of seatbelts at night. *Accident Analysis and Prevention, 130*, 30–37. https://doi.org/10.1016/j.aap.2018.04.006.
Boakye, K. F., & Nambisan, S. S. (2020). Seatbelt laws and seatbelt use among front- and rear-seat vehicle occupants in fatal crashes in the united states. *Case Studies on Transport Policy*. https://doi.org/10.1016/j.cstp.2020.04.003.
Boakye, K. F., Shults, R. A., & Everett, J. D. (2019). Nighttime seat belt use among front seat passengers: Does the driver's belt use matter? *Journal of Safety Research, 70*, 13–17. https://doi.org/10.1016/j.jsr.2019.04.004.
Chaudhary, N. K., & Preusser, D. F. (2006). Connecticut nighttime safety belt use. *Journal of Safety Research, 37*(4), 353–358. https://doi.org/10.1016/j.jsr.2006.05.005.
Gkritza, K., & Mannering, F. L. (2008). Mixed logit analysis of safety-belt use in single- and multi-occupant vehicles. *Accident Analysis & Prevention, 40*(2), 443–451. https://doi.org/10.1016/j.aap.2007.07.013.
Glassbrenner, D., Carra, J. S., & Nichols, J. (2004). Recent estimates of safety belt use. *Journal of Safety Research, 35*(2), 237–244. https://doi.org/10.1016/j.jsr.2004.03.010.
Glassbrenner, D. (2002). Safety Belt and Helmet Use in 2002 – Overall Results. Report No. DOT HS 809 500. National Center for Statistic s and Analysis, Advanced Research and Development, National Highway Traffic Safety Administration, Washington, D.C., September, 2002.
Goetzke, F., & Islam, S. (2015). Determinants of seat belt Use: A regression analysis with FARS data corrected for self-selection. *Journal of Safety Research, 55*, 7–12. https://doi.org/10.1016/j.jsr.2015.07.004.
Governors Highway Safety Association. Seat Belts. https://www.ghsa.org/state-laws/issues/Seat-Belts. Accessed May 31, 2021.
Harper, S., & Strumpf, E. C. (2017). Primary enforcement of mandatory seat belt laws and motor vehicle crash deaths. *American Journal of Preventive Medicine, 53*(2), 176–183. https://doi.org/10.1016/j.amepre.2017.02.003.
Insurance Institute for Highway Safety, and Highway Loss Data Institute. Seat Belt and Child Seat Laws by State. https://www.iihs.org/topics/seat-belts/seat-belt-law-table. Accessed May 31, 2021.
Kahane, C. J. (2015). *Lives Saved by Vehicle Safety Technologies and Associated Federal Motor Vehicle Safety Standards, 1960 to 2012 – Passenger Cars and LTVs – With Reviews of 26 FMVSS and the Effectiveness of Their Associated Safety Technologies in Reducing Fatalities, Injuries, and Crashes*. Report No. DOT HS 812 069. Washington, D.C.
Kim, S., & Kim, K. (2003). Personal, temporal and spatial characteristics of seriously injured crash-involved seat belt non-users in Hawaii. *Accident Analysis and Prevention, 35*(1), 121–130. https://doi.org/10.1016/S0001-4575(01)00097-5.
Lee, B.-H.-Y., & Schofer, J. L. (2003). Restraint use and age and sex characteristics of persons involved in fatal motor vehicle crashes. *Transportation Research Record: Journal of the Transportation Research Board, 1830*(1), 10–17. https://doi.org/10.3141/1830-02.
Lipovac, K., Tešić, M., Marić, B., & Đerić, M. (2015). Self-reported and observed seat belt use – A Case Study: Bosnia and Herzegovina. *Accident Analysis & Prevention, 84*, 74–82. https://doi.org/10.1016/j.aap.2015.08.010.
National Center for Statistics and Analysis (2019, March). *Lives Saved in 2017 by Restraint Use and Minimum-Drinking-Age Laws. Traffic Safety Facts Crash.Stats, Report No. DOT HS 812 683. National Highway Traffic Safety Administration, Washington, D.C., March*.
National Center for Statistics and Analysis (2019, April). *Occupant Protection in Passenger Vehicles: 2017 Data. Traffic Safety Facts. Report No. DOT HS 812 691. National Highway Traffic Safety Administration. Washington, D.C.*.
National Center for statistics and Analysis (2019, October). *2018 Fatal Motor Vehicle Crashes: Overview. Traffic Safety Facts Research Note. Report No. DOT HS 812 826. National Highway Traffic Safety Administration. Washington, D.C., October*.
National Center for Statistics and Analysis (2020, October). *Preview of Motor Vehicle Traffic Fatalities in 2019. Research Note. Report No. DOT HS 813 021.* National Highway Traffic Safety Administration.
National Center for Statistics and Analysis (2019, December). *Seat Belt Use in 2019 – Overall Results. Traffic Safety Facts. Research Note. Report No. DOT HS 812 875. National Highway Traffic Safety Administration. Washington, D.C.*
Nichols, J. L., Tison, J., Solomon, M. G., Ledingham, K. A., Preusser, D. F., & Siegler, J. N. (2009, June). *Evaluation of the Buckle Up in Your Truck Program. Final Report. Report No. DOT HS 811 131. National Highway Traffic Safety Administration*
Shakya, I., Shults, R. A., Stevens, M. R., Beck, L. F., & Sleet, D. A. (2020). State-level seat belt use in the United States, 2011–2016: Comparison of self-reported with observed use and use by fatally injured occupants. *Journal of Safety Research, 73*, 103–109. https://doi.org/10.1016/j.jsr.2020.02.015.
Shults, R. A., Elder, R. W., Sleet, D. A., Thompson, R. S., & Nichols, J. L. (2004). Primary enforcement seat belt laws are effective even in the face of rising belt use rates. *Accident Analysis & Prevention, 36*(3), 491–493. https://doi.org/10.1016/S0001-4575(03)00038-1.
Strine, T. W., Beck, L. F., Bolen, J., Okoro, C., Dhingra, S., & Balluz, L. (2010). Geographic and sociodemographic variation in self-reported seat belt use in the United States. *Accident Analysis and Prevention, 42*(4), 1066–1071. https://doi.org/10.1016/j.aap.2009.12.014.
Tison, J., Williams, A. F., & Chaudhary, N. K. (2010). *Daytime and Nighttime Seat Belt Use by Fatally Injured Passenger Vehicle Occupants*. Preusser Research Group, Inc.
Vasudevan, V., Nambisan, S. S., Singh, A. K., & Pearl, T. (2009). Effectiveness of media and enforcement campaigns in increasing seat belt usage rates in a state with a secondary seat belt law. *Traffic Injury Prevention, 10*(4), 330–339. https://doi.org/10.1080/15389580902995190.
Zhang, J., & Yu, K. F. (1998). What's the relative risk?: A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA, 280*(19), 1690–1691. https://doi.org/10.1001/JAMA.280.19.1690.

Kwaku Boakye is a certified Professional Engineer, Professional Traffic Operations Engineer and Road Safety Professional currently with Arcadis U.S., Inc – a global design, engineering, and management consulting company. He holds a doctorate degree in Civil Engineering (Transportation) and a master's degree in Statistics from the University of Tennessee, Knoxville. He obtained his bachelor's and master's degrees in Civil Engineering from Kwame Nkrumah University of Science and Technology, and the University of Kansas, respectively. His research interests include seatbelt use studies, crash analysis, intelligent transportation systems, and safety evaluation of countermeasures to support decisions and policies.

Journal of Safety Research 82 (2022) 261–274

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

A review of research on driving distraction based on bibliometrics and co-occurrence: Focus on driving distraction recognition methods

Ge Huimin *, Bo Yunyu, Sun Hui, Zheng Mingqiang, Lu Ying

School of Automotive and Traffic Engineering. Jiangsu University, Zhenjiang, Jiangsu, China

A R T I C L E I N F O

Article history:
Received 27 November 2020
Received in revised form 31 August 2021
Accepted 2 June 2022
Available online 14 June 2022

Keywords:
Driving behavior
Driving distraction recognition
Bibliometric analysis
Co-occurrence analysis

A B S T R A C T

Introduction: The existing selection of driving distraction recognition methods is based on a specific research perspective and does not provide comprehensive information on the entire field of view. *Method:* We conducted a systematic review of previous studies, aiming to come up with appropriate research methods to identify the driver's distraction state. First, this article selects four sets of search keywords related to driving distraction discrimination from five databases (Web of Science, ScienceDirect, Springer Link, IEEE, and TRID) and identifies 1,620 peer-reviewed documents from 2000 to 2020; these 1,620 documents underwent bibliographic analysis and co-occurrence network analysis. The co-occurrence coupling relationship is analyzed from the aspects of time, country, publication, author and keywords. Second, 37 papers published were screened, and the driving distraction recognition methods proposed by these 37 papers were summarized and analyzed. *Results:* The results show that this field has been prevalent since 2013; countries such as the United States, Britain, Germany, Australia, China, and Canada are in the forefront of research in this field, and the cooperation between related countries is relatively close. The cooperation between authors is characterized by aggregation, and the mobile phone as the main keyword is almost connected to other keyword nodes; the recognition model of deep learning algorithm based on video surveillance data sources has become the mainstream hot spot distraction recognition method. The recognition model of machine learning algorithm based on vehicle dynamics data, driver physiology, and eye movement data sources has specific advantages and disadvantages. *Practical Applications:* The results can help people to understand the current situation of driving distraction comprehensively and systematically, provide better theoretical support for researchers to choose the subsequent driving distraction recognition model, and provide research direction for driving distraction recognition in the future.

1. Introduction

The road traffic system is composed of people, motor vehicles, and road environment; motor-vehicle drivers are a weak link in the system, and human factors play a crucial role in about 90% of traffic crashes (Valero Mora, Ballestar Tarin, Tontsch, Pareja Montoro, & Sanchez Garcia, 2012). The driving process is complicated, including multiple aspects such as situational awareness, decision-making, and execution. It is inevitable that sudden abnormal situations such as distraction errors and recognition errors will directly lead to driving risks (Parker, Reason, Manstead, & Stradling, 1995). There are many reasons for driver errors, and distraction is one of the most important. The American Automobile Association Traffic Safety Foundation defines driving distraction status as the driver's attention not being focused on the driving task due to some events, activities, objects, or people inside or outside the car, resulting in a decrease of the driver's reaction ability and failing to deal with dangerous situations or improper behaviors in time (Dewar & Olson, 2002). A survey based on 1,367 drivers found that traffic crashes caused by distracted driving account for 14 to 33 percent of major accidents. (Mcevoy, Stevenson, & Woodward, 2007). Therefore, the recognition of driver's distraction state is of great significance to ensure driver, passenger, pedestrian/cyclist and property safety.

Driving distraction is a temporary shift of driver's attention (Yang, McDonald, & Zheng, 2012). The essence is that the driver's attention shifts from the driver's main task (driving) to thinking about something else. To identify the driver's distracted state in the driving process more accurately and timely, scholars in this field have done much research on identifying data sources, models and so on. Waard et al. collected vehicle dynamics data in

* Corresponding author.
E-mail address: hmge@ujs.edu.cn (H. Ge).

https://doi.org/10.1016/j.jsr.2022.06.002
0022-4375/

distracted state, and built a real-time recognition system based on statistics, artificial neural network, and fuzzy logic with a recognition accuracy rate of 89% (Waard, Brookhuis, & Hernandez-Gress, 2001). Masala et al. built a Sanger neural network recognition model by analyzing the video recording monitoring data of the driver's posture and eyes when driving was distracted, and the effective recognition accuracy rate was 92.7% (Masala & Grosso, 2014). Wang et al. conducted a driving distraction test by simulator, collected the electrical signals of the driver's brain, and constructed a support vector machine recognition model based on radial basis function, with a recognition accuracy of 86.2% (Wang, Jung, & Lin, 2015). Le et al. built a multi-scale fast rcnn(MS-FRCNN) recognition model based on deep learning using natural driving data, and the recognition accuracy rate reached 92.4% (Le, Zheng, & Zhu, 2016). Wollmer et al. built a driver's distraction detection model based on LSTM regression neural network by collecting vehicle dynamics and driver's head movement data in the state of driving distraction, and the recognition accuracy rate was 96.6% (Wöllmer, Blaschke, & Schindl, 2011). Dehzangi et al. collected the driver's skin electrical signals through driving distraction test and constructed a convolution neural network distraction recognition model, with an accuracy rate of 93.3% (Dehzangi & Taherisadr, 2018). Therefore, there is a lot of research on driving distraction status recognition by different methods, but there are still many problems and challenges.

Existing driver monitoring mainly focuses on low attention owing to fatigue, fatigue score, short distraction occurrence time, many affected factors and strong delay characteristics, which brings challenges to driver attention monitoring. In current vehicles, the source of distraction faced by drivers continues to exist, and it becomes more difficult to concentrate on driving. Various types of vehicle infotainment systems have developed rapidly. During driving, more and more people use various electronic devices in the car. Drivers are becoming ever more eager to keep in touch with others while driving; this information is the embodiment of modern people's lifestyle changes in driving. The use of in-car entertainment systems, real-time in-vehicle information systems, and smart phones to advance synchronously for multi-task operations has significantly increased (Wilson & Stimpson, 1999). Although some literature has analyzed emotional driving behaviors, there is little literature about specific driving distraction behaviors and discussing specific driving distraction recognition methods. Scott-Parker conducts research on young drivers and analyzes literature reviews on emotions and driving behaviors (Scott-Parker, 2017). Yusoff discussed some common measures of disturbing driving and five common methods for measuring driving distraction (Yusoff et al., 2017). In view of all the content mentioned, it can be assumed that reviews on the selection of driving distraction recognition methods have been extensively studied. However, these reviews do not include all research on driving distractions and do not provide a comprehensive understanding of the field. The purpose of driving distraction research is to better detect the state of driving attention deficit. At present, the recognition model constructed by various machine learning algorithms has become the mainstream method of driving distraction recognition. To better reveal the advantages and disadvantages of various recognition methods and their internal relations, it is necessary to conduct an in-depth discussion in this field. The combination of bibliography and co-occurrence network analysis can show more complex cooperation relations in this research field.

However, to the best of our knowledge, there is no literature that uses the co-occurrence network analysis method and literature measurement analysis method to systematically analyze the relationship between driving distraction studies (Gao, Sun, Geng, Wu, & Chen, 2016; Wu et al., 2020). Co-occurrence network distraction is a quantitative analysis method of co-occurrence information in various information carriers, which can reveal the content association of information and the co-occurrence relationship implied by feature items. Bibliometric analysis is a method of quantitatively analyzing all knowledge carriers by using mathematics and statistics. This method combines mathematics, statistics and philology, and can help readers to form a comprehensive knowledge system focusing on quantification. Therefore, this study attempts to fill this gap and provide new ideas when choosing driving distraction recognition methods. Overall, this article aims to provide a reference for future research on driving distraction through bibliography and co-occurrence networks. This paper expects to achieve four goals: (1) evaluate the current system research trends of driving distraction; (2) point out the keywords, countries, academic cooperation between authors and different clusters; (3) identify and sort out the existing literature on driving distraction recognition methods, including the detection targets of the recognition model and recognition methods; and (4) summarize and analyze the current recognition method and provide a theoretical basis for subsequent research. Through the completion of these research objectives, we can achieve a more comprehensive understanding of the breadth and depth of the whole driving distraction research field and have a more in-depth understanding of the most important problem in this field -- the selection of driving distraction recognition methods and future research trends -- and contribute to the development of high-level assisted driving systems and automatic driving systems.

2. Research methods

2.1. Literature retrieval strategy

Reporting criteria is based on systematic reviews and meta-analysis (PRIS-MA) (Page et al., 2020). In April 2020, the author retrieved academic papers related to driving distraction in the five databases of Web of Science, ScienceDirect, Springer Link, IEEE, and TRID. The first three are comprehensive databases, and the journals included in other databases are reputed. The search involves different disciplines such as electric, computer engineering, science and transportation. The search keywords are composed of 4 parts. At least 1 keyword related to each section. The 4 part keywords are: (1) driving, driver, driving age, novice driver, young driver, and driving psychology; (2) vehicles, passenger cars, trucks, electric vehicles and new energy vehicles; (3) road environment, urban roads, rural roads, mountain roads and highways; (4) distraction, factors, psychology, perception, subjectivity and attitude. According to the above keyword combination, the search language is adjusted according to different databases. Some important literature references are also included in the scope of reviewing search.

2.2. Screening criteria

The searched documents were further filtered to make the final documents meet the following conditions: (1) it must be published in a peer-reviewed English journal; (2) the research subject is mainly drivers; (3) there must be empirical data on the driving distraction test. In addition, this article contains only the literature that has quantitative analysis of the factors affecting the recognition of driving distraction, which will help the comprehensive comparative analysis of these documents.

2.3. Information extractions

This article uses the matrix method to extract standardized information tables from the literature being reviewed. The information extracted from each document mainly includes the following: (1) the characteristics of the literature (such as the location, year and author of the study); (2) the attributes of the respondent (such as the sample size, the investigator driving age distribution); (3) environmental factors affecting driving distraction; (4) driving distraction recognition methods and theoretical framework; and (5) main research conclusions of the literature. To ensure the reliability of information extraction, the two authors of this article randomly selected 50 articles from the literature for information extraction and found that the consistency of the extracted information reached 80%, and different information was unified through negotiation. This shows that this article has a high mutual reviewer reliability for the extraction of information in the literature.

2.4. Study collection

According to the above search strategy: 3,726 articles were retrieved from Web of Science, 8,946 articles were retrieved from ScienceDirect, 39,947 articles were retrieved from Springer Link, 1,276 articles were retrieved from IEEE, and 5,981 articles were retrieved from TRID. After removing the duplicates, a total of 59,876 articles were retrieved from these five databases. After manual selection of the title and abstract, 3,607 articles remained, of which 78 were unrelated to the driver and 1,460 were unrelated to driving distraction. Forty-five full texts were not available, and finally 2,024 full text articles were obtained. 246 of the 2,024 full text articles were review articles. Relevant literature in the articles was obtained, 158 non-empirical studies were removed, and 1,620 articles were finally obtained. Eventually, 1,620 available studies were left and satisfied the full criteria in this review, as shown in Fig. 1.

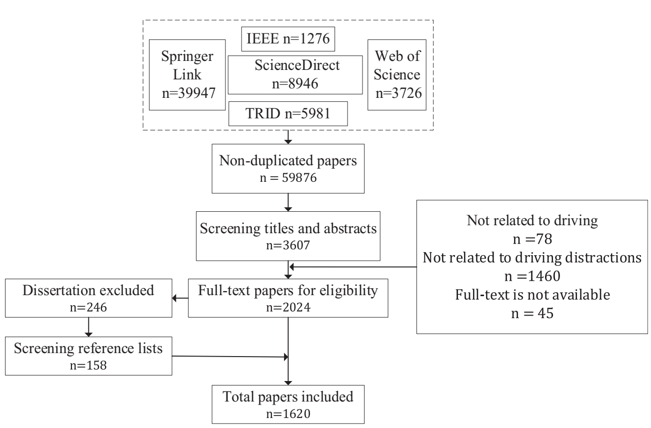


Fig. 1. Flowchart of the systematic review process.

3. Results

3.1. Overview of articles' development trends

Fig. 2 describes the increase in the number of articles (NO) included in this review from 2000 to 2020 after literature search and screening. NO indicates that the number of published articles is increasing with the progress of time. From the perspective of stages, the number of published articles in the first three years is less, which first decreases and then increases from 2004 to 2013, reaches the lowest value in 2009, and increases from 2014 to 2019; however, the number of published articles decreases in the last two years (the literature ends in June 2020). Fig. 2 also describes the changes of annual total cited quantity (TC) and annual average cited frequency per article (ACPP). The figure shows that TC fluctuated up and down before 2013, peaked in 2005, fluctuated greatly before 2009, fluctuated slightly between 2013, and declined year by year after 2014. ACPP showed two peaks in 2002 and 2006, respectively, and then declined slowly from 2009 to 2020.

3.2. Academic cooperation of different countries and areas

Promoting academic cooperation in different countries and regions is of great significance, which can promote the discovery of innovative points and the application of solutions.

Fig. 3 counts the number of articles published by various countries from 2000 to 2020. Fig. 3 shows that the United States is the country with the largest number of articles published in this field, which shows the attention and discourse power of the United States in the field of driving distraction. The articles published by U.S. scholars account for 36.5% of the total articles, while those published by British scholars account for 10% of the total number of articles, the articles published by other countries include Australia 8%, Germany 7%, China 7%, Canada 6%, and so forth. The articles published by scholars from the above six countries account for 74.5% of the total articles, while there are 233 countries and regions in the world, which shows that the research in this field is extremely unbalanced in the world at present.

Fig. 4 shows the cooperative relationship between the author's country and region. The thickness of lines and the size of nodes have a positive correlation with the degree of international academic cooperation; the larger the node, the more are the academic achievements, and the thicker the line, the closer is the academic cooperation. Fig. 4 shows that the United States is at the center of the network with the largest node, indicating that the United States is the first echelon of activity in this field internationally. From the line width point of view, the line width between the United States and China, Canada, and Britain is wider, indicating that the academic cooperation between the United States and these three countries is relatively close. From the overall cooperation network layout, the number of nodes is small, and the network is sparse, which means that there are few countries that pay attention to and deeply study this field and academic cooperation and exchange are not close. Therefore, to promote better and faster development in this field, close and in-depth cooperation between countries is indispensable.

3.3. Keyword co-occurrence analysis

Keywords are the concentrated expression of research content in a field. Through the analysis of keywords we can quickly understand the development and current hot spots of research in a field and deepen our understanding of this field.

Fig. 5 shows the keyword co-occurrence network in the driving distraction literature. Node size and line thickness are positively related to the keyword connection; the larger the node, the higher is the frequency of the keyword, and the thicker the line, the closer is the connection between the two topics. Keyword categories are expressed in different colors. Fig. 5 shows that keywords are divided into three categories: the first category is represented by distraction and driving; the second category is represented by driver distraction, driver simulator, and road safety; and the third category is represented by mobile phone and distracted driving. In addition, the secondary task keyword shows that it is mainly studied using the sub-task test at present; keywords such as eye movements and natural driving indicate that research is mainly conducted through eye movement data and natural driving data; mobile phone and cell phone show that there is much research on the influence of distraction in mobile phone driving, but there are still many sources of distraction in the driving process, which should be studied in all directions, in depth, and in a wide field;

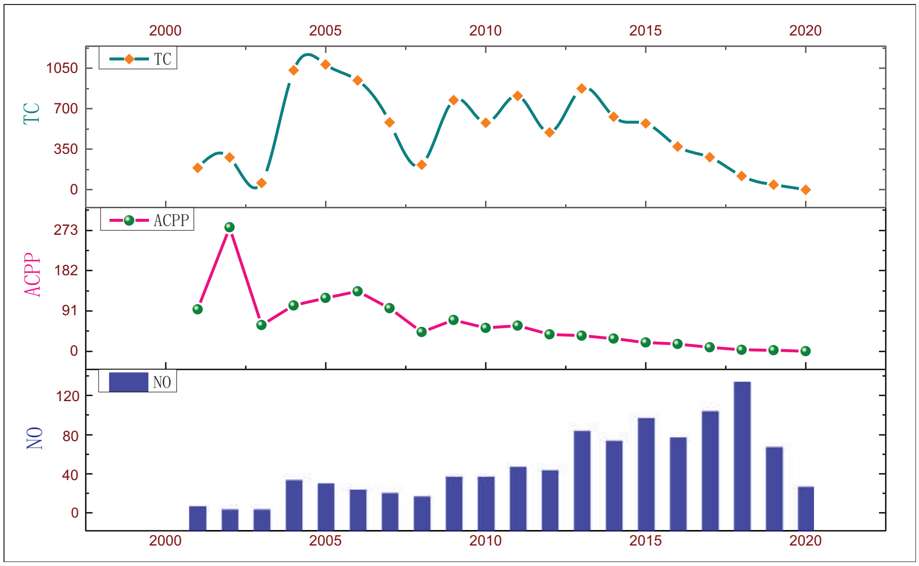


Fig. 2. Literature development trend.

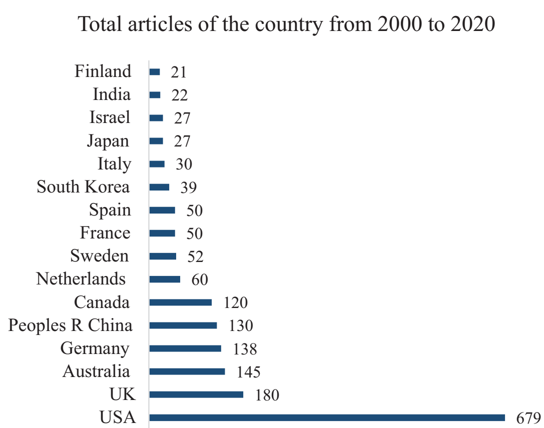


Fig. 3. Total articles of the country from 2000 to 2020.

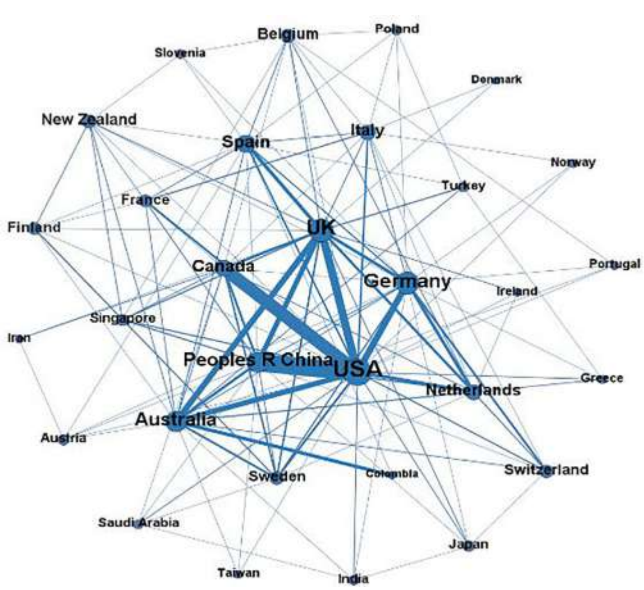


Fig. 4. Academic cooperation between different countries.

keywords such as older drivers and young drivers show the research objects of driving distraction, which are mainly analyzed from the age level, lacking the analysis of factors such as gender and personality. It also includes the research on driving distraction of autonomous vehicles represented by the keyword automated driving, but lacks the research on driving distraction of networked vehicles. To clearly show the meaning of Fig. 5, Table 1 shows the intensity between different keywords.

3.4. Author co-occurrence analysis

Co-occurrence network analysis is not only applied between countries and keywords, but also between authors. Zhao and Strotmann (2011) and Kim, Jeong, and Song (2016) applied the co-occurrence network to authors to analyze influential authors and their associations. Using Bibexcel to draw the relationship diagram and realize the visualization through Gephi, the author analyzes the relevant amount of articles issued by the author and the relationship between the co-authors. Finally, the author co-occurrence analysis showing more than 50 nodes is shown in Fig. 6. The size of the node corresponds to the amount of the post, and the thickness of the line indicates how closely the author cooperates. Color blocks of the same color indicate similarities between them.

To show the collaboration between authors more clearly and classify the research fields of coupling analysis of different color blocks, Fig. 7 expands these co-occurrence analysis graphs of the authors into six clusters. Although the research of each author is diversified, it can be seen from the cooperation between the authors that the authors collaborate in a cluster relationship. The dominant element of cluster1 is methods, the dominant element of cluster2 is driving mind, the dominant element of cluster3 is investment, the dominant element of cluster4 is cognitive distraction assessment, the dominant element of cluster5 is driving visual perception, and the dominant element of cluster6 is driving determinants. Fig. 7 shows that cluster1 with more research is dominated by Lee and Kolodge (2019), Donmez, Boyle, and Lee (2007), Reimer, D'Ambrosio, Coughlin, Kafrissen, and Biederman (2006), Mehler, Reimer, and Coughlin (2012) and so on. To more clearly show the cooperation relationship of the authors in the first category, Table 2 shows the cooperation intensity of the main authors in cluster1. Cluster1 focuses on the choice of methodology, includ-

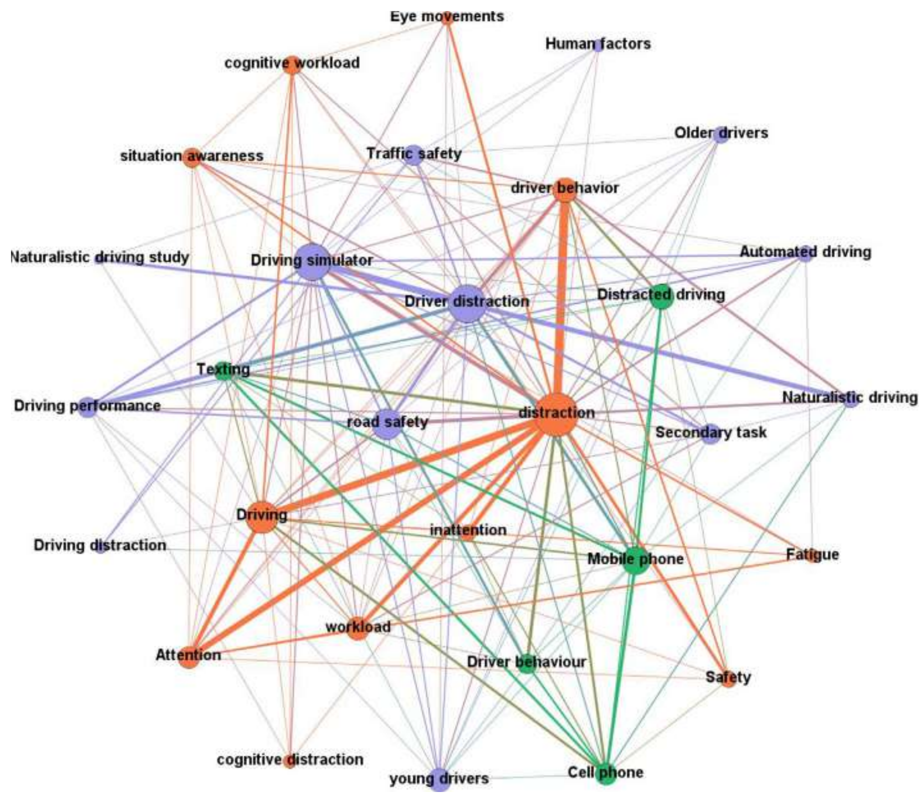


Fig. 5. Keyword co-occurrence network.

Table 1
Co-occurrence strength of the keyword.

Keyword1	Keyword2	Co-occurrence intensity	Keyword1	Keyword2	Co-occurrence intensity
distraction	driver behavior	14	distraction	Driving simulator	7
Driving	distraction	12	Driver distraction	Driving performance	7
Driver distraction	Driving simulator	11	distraction	workload	7
Attention	distraction	10	Driving	Attention	7
Driver distraction	Naturalistic driving	8	distraction	inattention	7

ing the use of eye movement data classification methods to study visual distraction, and the use of tactile detection methods to evaluate the distraction caused by in-car audio and video.

To focus on the study of driving distraction determination methods, 37 research studies were selected; each of these documents is related to the method of driving distraction recognition and is published in peer-reviewed journals. Table 3 provides information selected from 37 research studies. These studies are from 17 peer-reviewed journals in different disciplines such as transportation, biology, and accident analysis. Among them, 14 studies are from transportation-related journals. There are nine studies from journals in the comprehensive field, and three studies from journals in the field of biological health. All articles are published after 2015, and the number increased in the past two years. Table 3 shows that in recent years, the data sources of driving distraction research mainly include vehicle dynamics data, video image data, eye movement data, head and face data, physiological and psychological data, and multidimensional fusion of the above data. Because data acquisition is easy, vehicle dynamics data are most often used, but its recognition accuracy is not high. The recognition method mainly uses traditional machine learning algorithm and deep learning algorithm to build driving distraction recognition model, the recognition accuracy of deep learning model is high, but the real-time performance of the models is poor. Traditional machine learning recognition models represented by support vector machines have better real-time performance although the recognition accuracy is lacking.

3.5. Analysis of factors affecting driving distraction

The analysis of factors that affect driving distraction is the basis of research on driving distraction recognition methods. Different test environments and data types correspond to different recognition methods, and different recognition methods correspond to different recognition accuracy. Therefore, it is necessary to analyze the influencing factors of driving distraction through the study of different detection data types. Table 3 summarizes the influencing factors, recognition methods, and accuracy of driving distraction. Table 3 shows that the research conducted on the recognition of driving distraction is generally carried out from a real car test or a driving simulator test. The specific data detection types can be divided into video images, ECG, SC, EEG, vehicle power and Eye movement data.

With the increasing popularity and wide demand of vehicles, traffic crashes have become one of the most serious problems worldwide, among which driving accidents caused by driver distraction account for a large proportion (Mcevoy et al., 2007). Researchers and scholars are committed to reduce and alleviate

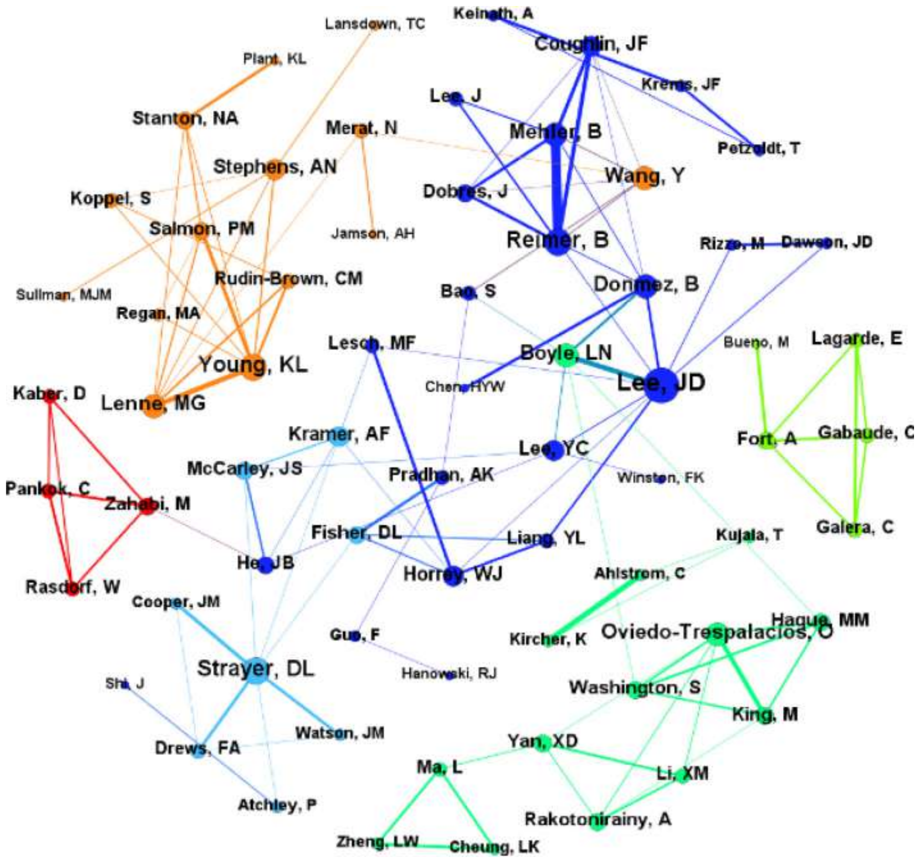


Fig. 6. Author co-occurrence network.

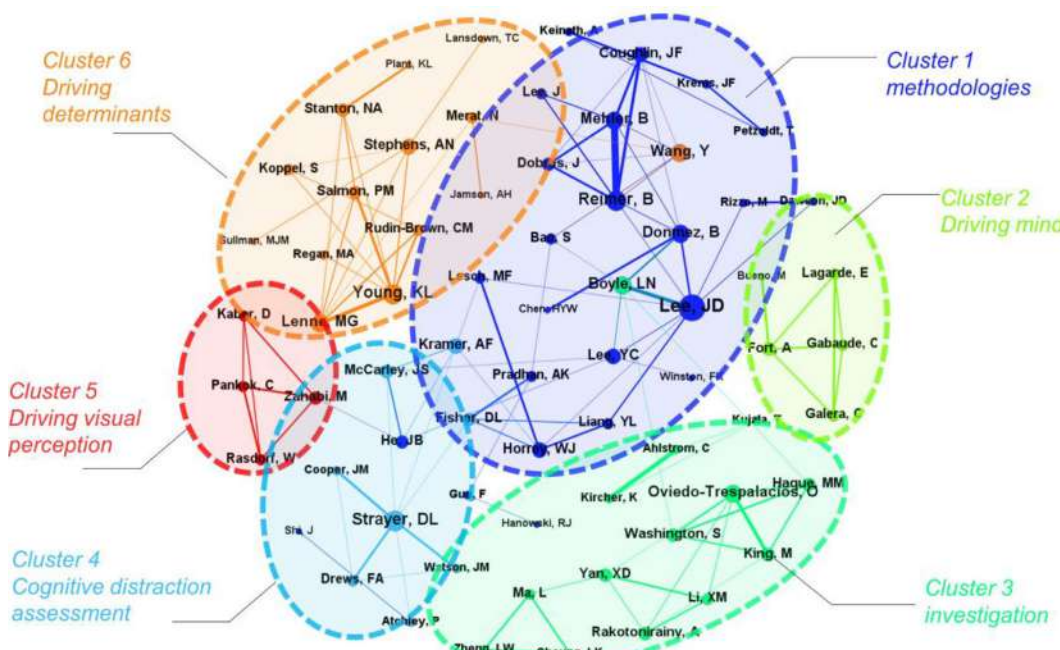


Fig. 7. Author Coupling Analysis Network2.5 Literature overview.

traffic crashes caused by driver distraction. Oviedo-Trespalacios, Afghari, and Haque (2020) developed a comprehensive multivariate ordered model in the Bayesian framework to study the risk compensation behavior of distracted drivers and designed a questionnaire to understand the risk compensation behavior of distracted drivers of mobile phones in Australia to empirically test the multi-ordered model. The results demonstrate that, for the targets corresponding to fixed parameters/fixed thresholds, the

Table 2
Co-occurrence strength of the author.

Author1	Author2	Co-occurrence intensity	Author 1	Author 2	Co-occurrence intensity
Reimer, B	Mehler, B	19	Boyle, LN	Lee, JD	10
Reimer, B	Coughlin, JF	8	Mehler, B	Coughlin, JF	7
Pradhan, AK	Fisher, DL	6	Horrey, WJ	Lesch, MF	6
Donmez, B	Lee, JD	6	Keinath, A	Krems, JF	6
Chen, HYW	Donmez, B	8	Mehler, B	Dobres, J	7
Reimer, B	Dobres,J	6	Dawson, JD	Rizzo, M	6

grouped random parameter/random threshold ordered model has a significantly improved fit, this proves that drivers making different types of risk compensation behaviors are interrelated. Similarly, Zhao, Xu, Ma, Li, and Chen (2019) also used a questionnaire survey method to analyze the characteristics of drivers with a focus on distracted driving behavior and developed a new method to reduce driver errors. Unlike Oviedo-trespalacios O, Zhao X uses a structural equation model (SEM) to grasp the complex relationship between related variables. The model can simultaneously solve the intricate relationship between endogenous variables and exogenous variables. Forty-four participants were recruited to complete the driving distraction test. In the experiment, driving behavior data were collected by driving simulator and driving posture data were collected by questionnaire survey. The conclusion shows that the main factor affecting illegal driving behavior is driving attitude, and the main factor affecting the performance of distracted drivers is the basic characteristics of drivers. In addition, the driver's training level is the most important factor that negatively affects the driver's basic characteristics (factor load = −0.91), and anger at slow driving is the main factor that affects the driver's attitude (factor load = 0.90).

Yang, Guan, Ma, and Li (2019), Atiquzzaman, Qi, and Fries (2018), and Stavrinos (2013) also used the driving simulator to study driving distraction. Yang et al. (2019) installed the electroencephalogram (EEG) detection module on the simulator and compared the model based on EEG features and based on mixed features (combination of driving features and EEG features) to determine the current driving state. Atiquzzaman et al. (2018) are committed to developing algorithms that only use vehicle dynamics data to detect interference tasks involving vision, cognition, and the physical workload. Stavrinos (2013) developed a system to detect inattention by using Electrocardiogram (ECG) and Surface Electromyography (SEMG) signals. This system requires the testers to use telephone and SMS services during driving to detect cognitive and visual inattention during driving. It can be seen by the scholars choosing the type of data collected by the simulation test, except for Chen, Wu, Zhong, Lyu, and Huang (2015) using smart cars to collect vehicle dynamics data. When research is needed on the vehicle dynamics information, most scholars often use simulation tests to collect data. While there is a gap in the authenticity of simulation tests and real car tests, the simulation test is easier to obtain vehicle dynamics test data compared with the actual car test. To ensure the safety of the test, if you want to carry out the real vehicle test, you often need to be equipped with a safety officer (Chen et al., 2015), and the simulation test is safe.

In the real vehicle test, there are also scholars who conduct research on driving distraction based on electrocardiogram (ECG) (Taherisadr, Asnani, Galster, & Dehzangi, 2018) and skin electrical (GSR) (Dehzangi, Sahu, Rajendra, & Taherisadr, 2019). Unlike the instruments used by Yang et al. (2019) to measure electroencephalograms (EEG), which require sensors to be installed on the driver's head, ECG and skin electrical instruments are smaller and can be worn on fingers, so that they can be used in real vehicle tests. Taherisadr et al. (2018) introduced a cepstrum representation of Mel frequency based on ECG and Convolutional Neural Network (CNN) driver distraction detection system. Spectrum cepstrum representation was used as the input of convolutional neural network. Dehzangi et al. (2019) proposed a less invasive wearable physiological sensor (available on smart watches) to quantify the skin conductance (SC) called Galvanic Skin Response (GSR) to characterize and identify natural interference during driving. Use of physiological signals compensates for early and safer distraction detection and has the advantages of low price and low invasiveness.

Depending on the driver's eye movement data, it is a very intuitive indicator to detect driving distraction. When the driver has a driving distraction, the driver's eye movement data often drifts. Topolšek, Areh, and Cvahte (2016) used eye tracking technology to develop a study of the effect of road signs or advertisements on the driver's attention. Xu et al. (2018) collected and analyzed eye movement data of drivers in simulated conflicts at different speeds, selected the peak point of the pupil diameter after wavelet processing, the first point to the left of the peak point and the first point after the peak point, and the construction conflict recognition method (CCFRM) is proposed based on the key point. Eye movement data-based research can more intuitively reflect the connection between driving distractions and roadside visual disturbances than other detection data.

Driving distraction detection based on video images is a more conventional detection method in recent years. Video image detection highlights the authenticity of videos and is often used in real vehicle test. Masood, Rai, Aggarwal, Doja, and Ahmad (2018) detected 22,424 sets of images, and proposed a machine learning model using convolutional neural networks. This model can not only detect distracted drivers but also can analyze the reasons for driving distraction through the images obtained by the camera module installed in the car. Subsequently, Shahverdy, Fathy, Berangi, and Sabokrou (2020) proposed an image deep neural network learning method to analyze five parameters including acceleration, gravity, throttle, speed and revolutions per minute (RPM), and used recursive graph technology to construct a two-dimensional convolutional neural network image based on driving signals. Xing, Lv, and Cao (2020) used a consumer-wide camera (Kinect) to monitor the driver and determine the driving tasks in the actual vehicle, seven common tasks performed by multiple drivers during driving were identified, use of feed-forward neural network (FFNN) was used to identify seven tasks, and finally, it was proven that the FFNN task detector is an effective model that can be used for real-time driver distraction and dangerous behavior recognition.

3.6. Recognition methods and accuracy

Distinct types of driving distraction detection data correspond to different recognition methods, and these different recognition methods correspond to different recognition accuracy. Therefore, it is necessary to analyze the different recognition methods, the

Table 3
Summary of research literature.

Author (year)	Country	Journal	Detection target	Method	Main research conclusion
Oviedo-Trespalacios et al. (2020)	Australia	Analytic Methods in Accident Research	Driving risk compensation behavior	Hierarchical Bayesian multivariate ordered model	Drivers'decisions to make different types of risk compensation are interrelated
Shahverdy et al. (2020)	Iran	Expert Systems with Applications	Acceleration, gravity, throttle, speed and revolutions per minute (RPM)	A two-dimensional convolutional neural network (CNN) on an image constructed from driving signals based on recursive graph technology	A novel and effective deep learning method is proposed for analyzing driver behavior
Xing et al. (2020)	Singapore	Actions for selected chapters	Seven common tasks performed during driving	Random forest and maximum information coefficient method	The average accuracy of the final test results of the seven driving tasks among the participants exceeded 80%
Li, Zhong, Hutmacher, Liang, Horrey, and Xu (2020)	China	Accident Analysis & Prevention	Driving video images	The first module predicts the bounding boxes of the driver's right hand and right ear from RGB images. The second module takes the bounding boxes as input and predicts the type of distraction	For overall distraction detection, it achieved F1-score of 0.74. The whole framework ran at 28 frames per second
Botta et al. (2019)	Italy	Knowledge and Information Systems	vehicle dynamics data and environmental data	single-layer feedforward neural network trained through pseudo-inversion methods	Propose an original approach which benefits from matrix sparsity, showing lower computational times with respect to standard implementations.
Le, Inagami, Hamada, Suzuki, and Aoki (2019)	Japan	Transportation Research Part F: Traffic Psychology and Behaviour	Driver eye tracking data	A model combining VOR and OKR	This model shows greater precision, reduces the effect of optic flow, and works well with changing gaze in the case of involuntary eye movement
Aksjonov, Nedoma, Vodovozov, Petlenkov, and Herrmann (2019)	Czech Republic	IEEE Transactions on Intelligent Transportation Systems	vehicle dynamics data	fuzzy logic and machine learning	The results presented in this research confirm its capability to detect and to precisely measure a level of abnormal driver performance
Chui, Alhalabi, and Liu (2019)	China	Data Technologies and Applications	Head motion coefficient	light computation power algorithm	Distraction detection using a light computation power algorithm is an appropriate direction and further work could be devoted on more scenarios as well as background light intensity and resolution of video frames
Riaz et al. (2019)	Pakistan	Ad Hoc & Sensor Wireless Networks	vehicle dynamics data	HUB-NET technology using an Exploratory Agent-Based Modeling level of a Cognitive Agent-based Computing (CABC) framework	The proposed driver distraction computing methodology and the Driver Distraction Detection Enabled-ADAS outperform simple ADAS
Eraqi, Abouelnaga, Saad, and Moustafa (2019)	Egypt	Journal of Advanced Transportation	face and hand localizations, and skin segmentation	an Ensemble of Convolutional Neural Networks	Propose a deep learning algorithm with 90% accuracy and a simplified version that can achieve 84.64% classification accuracy
Dehzangi et al. (2019)	USA	Smart Health	Galvanic Skin response (GSR) Skin Conductance (SC)	Embedded random forest feature selection and integrated bag classifier	Achieve minimal intrusion and achieve improved accuracy of 92.9% and 93.5%
Louie and Mouloua (2019)	USA	Applied Ergonomics	Working Memory Capacity (WMC)	Working memory diffuser	The weakening effect of distraction on braking response time is partially mediated by WMC
Yang et al. (2019)	China	Accident Analysis &	Electroencephalogram (EEG) characteristics and based on hybrid features (combination of driving	Two EEG analysis techniques (independent component analysis and brain source localization)	The EEG-based model has better performance than the driving data-based model, and the integrated model of

Table 3 (*continued*)

Author (year)	Country	Journal	Detection target	Method	Main research conclusion
		Prevention	features and EEG features)	and two signal processing methods (power spectrum analysis and wavelet analysis)	driving features and whole brain region features extracted through wavelet analysis is superior to other types of features. The highest accuracy is 86.27%.
Zhao et al. (2019)	China	Transportation Research Part F: Traffic Psychology and Behavior	Abnormal driving behavior	Structural Equation Model (SEM)	The main factor that affects illegal driving behavior is driving attitude, and the main factor that affects the performance of distracted drivers is basic driver characteristics
Masood et al. (2018)	India	Pattern Recognition Letters	Driving video images	CNN	The proposed model can greatly reduce training time and achieve 99% high detection accuracy on large data sets
Atiquzzaman et al. (2018)	USA	Transportation Research Part F: Traffic Psychology and Behavior	Vehicle dynamics data	Two linear (linear judgment analysis and logistic regression) and two nonlinear models (support vector machine and random forest)	The random forest algorithm has the best performance, it can detect texting and diet interference, and its accuracy is 85.38% and 81.26%, respectively
Aksjonov, Nedoma, Vodovozov, Petlenkov, and Herrmann (2018)	Czech Republic	IFAC-Papers Online	Changes in curvature and speed	Artificial neural network and adaptive neuro-fuzzy inference system	Although the prediction accuracy of the model depends on the algorithm specifications, compared with the adaptive neuro-fuzzy inference system, the artificial neural network is slightly more accurate in predicting driver performance
Taherisadr et al. (2018)	USA	Smart Health	Electrocardiogram (ECG)	Mel frequency cepstrum representation and convolutional neural network (CNN)	The proposed algorithm can achieve obvious classification accuracy between various topics
Ma, Hu, Chan, Qi, and Fan (2018)	China	Transportation Research Part D: Transport and Environment	Driving distance and vehicle speed	Forecasting model of driving task demand	Driving a vehicle is a multilevel task composed of various driving tasks and secondary tasks. The driver must assign attention to the task requirements in order to drive safely
Xu et al. (2018)	China	Accident Analysis & Prevention	Driver eye tracking data	construction conflict recognition method	The key points based on eye movements are proposed, the construction conflict period and the construction conflict are found, and a fast and effective identification method is proposed.
Duy, Ha, Sheng, Bai, & Chowdhary (2018)	USA	IET Intelligent Transport Systems	Driving video images	Four deep convolutional neural networks including VGG-16, AlexNet, GoogleNet, and residual network	GoogleNet is the best model out of the four for distraction detection in the driving simulator testbed
Son and Park (2018)	South Korea	International Journal of Automotive Technology	vehicle dynamics data	Radial Basis Probabilistic Neural Networks	The best performing model could detect distraction with an average accuracy of 78.0%
Ali and Hassan (2018)	Pakistan	KSII Transactions on Internet and Information Systems	Driver facial data	Active Shape Model (ASM) and Boosted Regression with Markov Networks (BoRMaN)	The approach that uses the novel ideas of motion vectors and interpolation outperforms other approaches in detection of driver's head rotation. We are able to achieve a percentage accuracy of 98.45 using Neural Network.
Li, Bao, Kolmanovsky and Yin (2018)	China	IEEE Transactions on Intelligent Transportation Systems	Vehicle dynamics data	Nonlinear autoregressive exogenous (NARX) driving model	Steering entropy and mean absolute speed prediction error from the NARX model are selected
Dehzangi, Rajendra, and	USA	Sensors	Skin electrical (GSR)	support vector machine recursive feature elimination	Demonstrated cross-validation accuracy of 94.81% using all the features and the accuracy of 93.01% using reduced

(*continued on next page*)

Table 3 (*continued*)

Author (year)	Country	Journal	Detection target	Method	Main research conclusion
Taherisadr (2018)					feature space
Choudhary and Velaga (2017)	India	Transportation Research Part F: Traffic Psychology and Behavior	Standard deviation of lane positioning, number of lane deviations, mean and standard deviation of lateral acceleration, mean and standard deviation of steering wheel angle and steering reversal rate	Repeated measurement ANOVA test	10° SRR can be provided in smart vehicle equipment to detect interference and alert the driver of its decentralized state
Savolainen (2016)	USA	Accident Analysis & Prevention	Vehicle speed and travel time	Random parameters and latent logit model	The goodness of fit between the random parameter model and the latent class model is very similar. While the random parameter model can use the standard to directly compare the statistical test with the merged logit model, you can choose between random parameters and latent parameters to a large extent depends on theoretical considerations
Topolšek et al. (2016)	Slovenia	Transportation Research Part F: Traffic Psychology and Behavior	Driver eye tracking data	Eye tracking technology	The age of the driver is independent of the number of roadside objects detected
Liao et al. (2016)	China	IEEE Transactions on Intelligent Transportation Systems	Vehicle dynamics and driver eye movement data	The support vector machine (SVM) recursive feature elimination algorithm	the classifier based on the fusion of driving performance and eye movement yields the best correct rate and F-measure
Liu, Yang, Huang, Yeo, and Lin (2016)	China	IEEE Transactions on Intelligent Transportation Systems	Dri'er's facial and eye movement data	Laplacian support vector machine and semi-supervised extreme learning machine	Semi-Supervised Machine Learning can enhance the efficiency of model development in terms of time and cost.
Chen et al. (2015)	China	Accident Analysis & Prevention	Standard deviation of acceleration rate (SDA) and standard deviation of yaw angular acceleration (SDY)	Wilcoxon rank sum test and double time window method	The optimized "parent window" and "child window" are 55 s and 6 s, respectively. The research results can be used to develop driver assistance systems
Sahayadhas et al. (2015)	India	Expert Systems with Applications	Electrocardiogram (ECG) and electromyography (EMG)	K-Nearest Neighbor (KNN)	The best combination of the features of ECG and EMG signals, the classification accuracy is 96%
Stavrinos (2013)	USA	Accident Analysis & Prevention	Speed fluctuation and lane change times	Repeated measurement multivariate variance analysis and generalized estimation equation Poisson model	Distracted driving will lead to traffic safety and traffic flow reduction, which will have a negative impact on traffic operation
Klauer (2014)	USA	New England journal of medicine	Vehicle dynamics data and video image data	Logistic regression analysis of mixed effects	With the execution of many secondary tasks, including texting and calling mobile phones, the risk of novice drivers crashing or approaching a crash will increase
Hickman and Hanowski (2012))	USA	Traffic injury prevention	Natural driving data	Advantage ratio analysis	Mobile phone use should not be regarded as a binary variable (yes/no), and the risks of different mobile phone subtasks are different
Foss and Goodwin (2014)	USA	Journal of Adolescent Health	Natural driving data (vehicle motion, video, audio)	Cluster analysis, variance analysis, univariate logistic regression estimation	The common assumption of adolescent drivers' distraction is only partially supported by in-car measurements, and the relationship between passengers and distraction seems to be more complicated than previously realized

Table 4
Influencing factors of driving distraction, recognition method and accuracy.

Author (year)	Participants	Test environment		Detection data type						Method	Accuracy
		Real car	Simulation	Video images	ECG	SC	EEG	Vehicle power	Eye movement data		
Shahverdy et al. (2020)	3	✓		✓						CNN	99.76%
Xing et al. (2020)	5	✓		✓						Random forest	>80%
Li et al. (2020)	20		✓	✓						CNN	92%
Aksjonov et al. (2019)	18		✓					✓		(ED) and (FZ)	> 90%
Eraqi et al. (2019)	44	✓		✓						CNN	98%
Dehzangi et al. (2019)	15	✓				✓				Random forest	93.50%
Yang et al. (2019)	52		✓				✓	✓		wavelet analysis	86.27%
Masood et al. (2018)	22,424	✓		✓						CNN	99%
Atiquzzaman et al. (2018)	35		✓					✓		Random forest	85.38%
Taherisadr et al. (2018)	10	✓			✓					CNN	95.51%
Duy et al. (2018)	2,000		✓	✓						CNN	92%
Son and Park, 2018)	15	✓						✓		RBPNN	78.0%
Ali and Hassan (2018)	4	✓		✓						Active Shape Model (ASM)	98.45%
Li et al. (2018))	10	✓						✓		SVM	95
Dehzangi et al. (2018)	10	✓				✓				SVM-RFE	93.01%
Liao et al. (2016)	27		✓					✓	✓	SVM-RFE	95.8%
Liu et al. (2016)	41	✓							✓	SVM	97.2%

resulting accuracy, and aggregate the distraction recognition method. Its accuracy is shown in Table 4.

Convolutional neural network (CNN) is a commonly used driving distraction recognition method in recent years, and the accuracy of using CNN to identify driving distraction can reach more than 95%, which is the highest accuracy recognition method in the summarized literature. Taherisadr et al. (2018) convened 10 testers to participate in the test, selected electrocardiogram (ECG) as the recognition factor, and adopted a cepstrum representation of Mel frequency based on ECG and Convolutional Neural Network (CNN) driver distraction detection system. The structure of the deep CNN can automatically learn the reliable recognition patterns in two-dimensional spatio-temporal spectral space as features, thereby replacing the traditional hand-made features when processing the recorded time series data sets, and the final recognition accuracy reaches 95.51%. Masood et al. (2018) selected 22,424 groups of video images for research, and proposed a machine learning model using convolutional neural networks. This model can not only detect distracted drivers, but also determine the reason of distraction by camera module installed in the car. The image obtained within the camera module determined the cause of its distraction. By learning spatial features from images, CNN can further examine them through a fully connected neural network, with a detection accuracy rate of 99%. Shahverdy et al. (2020) selected three drivers to participate in the actual car test, and used a two-dimensional convolutional neural network (CNN) on the image constructed from the driving signal based on the recursive graph technology to conduct research, the final recognition accuracy rate reached 99.76%. Through the research of these scholars, we can see that when CNN is applied to the video image detection driving distraction technology, its recognition accuracy is very high, and the data detection device has no contact load with the driver. When CNN is applied to the electroencephalogram (ECG) detection of driving distraction technology, the recognition accuracy is also very high. However, when CNN is used for video image detection to identify driving distraction, the problem of recognizing the distraction delay occurs, and especially when using video images with high accuracy, it is easy to lead to leakage of driving privacy, which provides obstacles for the widespread application of CNN.

Similar to CNN, Random Forest is also a commonly used driving distraction recognition method in recent year. Xing et al. (2020) used the Random Forest method to detect driving distraction in video images and selected five drivers to test seven common tasks during driving, and the average accuracy of the final detection results exceeded 80%. It also applies to video image detection for driving distraction. Random Forest method is lower in accuracy, but the Random Forest method overcomes the problem of CNN method recognition delay, and it has good real-time performance while retaining the advantages of video image detection without contacting with the driver. Also from the point of view of contactless drivers, to solve most real-life interference tasks, as well as the problem that eye, head or face tracking data is difficult to obtain in real time, Atiquzzaman et al. (2018) selected 35 drivers to conduct a real car test and used vehicle dynamics data to determine driving distraction. In this article, Atiquzzaman compares the performance of two linear (linear recognition analysis and logistic regression) and two nonlinear models (Support Vector Machine and Random Forest), and derives the Random Forest algorithm to measure the driving score in car dynamics. The conclusion of the best performance of the heart feature is that when Random Forest is used to detect the two driving distractions of texting and diet interference, the accuracy is 85.38% and 81.26%. Although the accuracy rate is lower than other methods, this detection method has no contact with the driver, and has low cost and strong practicality. It provides useful guidance for car manufacturers who integrate the distraction detection system into their vehicles. Dehzangi et al. (2019) also proposed a less invasive wearable physiological sensor (which can be used on smart watches) from the perspective of reducing body load. It quantifies Galvanic Skin Response (GSR) and Skin Conductance (SC) to characterize and identify interference during natural driving. The embedded Random Forest feature selection and integrated bag classifier only use 10-D and 15-D feature space to achieve 92.9% and 93.5% accuracy, respectively. Although the method proposed by Dehzangi has a load contact with the driver, this physiological sensor has a small load and can provide high-accuracy recognition.

Some scholars have adopted other recognition methods to detect driving distraction. Sahayadhas, Sundaraj, Murugappan, and Palaniappan (2015) selected 15 drivers for driving simulation and used KNN to develop a system that can detect inattention using ECG and SEMG signals. The results indicate that the overall maximum accuracy of the bispectrum feature on the ECG and EMG signals is 98.12% and 90.97%. Sahayadhas, Sundaraj, Murugappan, and Palaniappan (2015) used k-fold verification on the basis of their own research, and the overall maximum accuracy of bespectacled features on ECG and EMG signals was 96.75% and 92.31%. Sahayadhas also uses principal component analysis to fuse features of ECG and EMG signals, improving classification accuracy to 96%. Although this method has achieved good accuracy, the detection and acquisition of electrocardiogram (ECG) and surface electromyography (SEMG) signals have a heavy load on the driver. Yang et al. (2019) selected 52 drivers for simulated driving tests, based on EEG features and based on mixed features (a combination of driving features and EEG features) model, using two EEG analysis techniques (independent component analysis and brain source location), two signal processing methods (power spectrum analysis and wavelet analysis) to extract 12 kinds of EEG features and driving state prediction. The driving performance, EEG features and mixed features are evaluated and compared. The results show that the EEG-based model has better performance than the driving data-based model (accuracy rates are 83.84% and 71.59%, respectively). This method is based on mixed features for comparison, which has better recognition characteristics than single feature. However, the accuracy is lower than the other methods discussed, and there is load contact with the driver.

4. Discussion

4.1. Development of publications, countries and authors

Fig. 2 shows in detail the overall development trend of publications. The figure reveals that the growth of NO is substantial and has been very high since 2013, which is inseparable from the rapid development of using mobile phones and vehicle-mounted functions while driving. The keyword co-occurrence graph in Fig. 5 reflects this very well. In Fig. 5, the mobile phone is almost connected to other keyword nodes. In the statistics of the number of countries published in Fig. 3, the United States, Britain, Germany, Australia, China, and Canada are the top countries in terms of volume of documents. Interestingly, in the co-occurrence chart of countries in Fig. 4, these countries are also the countries with the closest cooperation. This shows that international academic cooperative research has a positive effect on article output, and academic research can promote the professional knowledge of different scholars, leading to more ideas and innovations. Fig. 6 shows the cooperation network between different authors. There is an aggregation feature in the cooperation between the authors, and there is a connection between the author's research field and the cooperation between the authors. Fig. 7 reveals the research directions that are more focused on when scholars collaborate, and focuses on the analysis of the methodology.

4.2. Summary of advantages and disadvantages of driving distraction recognition method

Research on driving distraction has different types of detection data, different recognition methods, and corresponding different recognition accuracy.

Detecting driving distraction based on surveillance video images has no contact burden on the driver. It has a high accuracy rate, but is prone to leakage of driving privacy, and the cost is relatively high. In the driving distraction detection method based on surveillance video images, CNN has higher accuracy than Random Forest, but Random Forest is better than CNN in real time.

The detection of distraction based on Electrocardiogram (ECG) and Galvanic Skin Response (GSR) has less burden on driver contact, and the overall accuracy is higher. However, these methods have the problem that the data are difficult to obtain in real time. The driving distraction detection method based on ECG and GSR is similar to that based on surveillance video images.

The detection of driving distraction based on automobile dynamics has no contact burden on the driver, the data are easily available and the cost is low, but the overall accuracy is low. The accuracy of power spectrum analysis, wavelet analysis, and random forest conversion used in the driving distraction detection method developed by automobile dynamics data is about 80%. Compared with other methods, this method still has room for improvement in accuracy.

Detection of distraction based on eye movement data has no contact burden on the driver, but the data is difficult to obtain in real time, and the cost is high. However, the collected data are the most intuitive.

4.3. The effect of driving fluctuation on driving distraction research

During the literature review, it was found that there were studies related to driving volatility in interconnected and natural driving systems. Driving volatility can be used as a measure of driving distraction behavior, which is of great significance to the study of driving distraction. Kim, Song, Rouphail, Aghdashi, Amaro, and Gonçalves (2016) collected the number of natural driving behaviors for 3 months based on on-board sensors, and studied the correlation between collision tendency and microscopic driving behaviors. Wali, Khattak, and Karnowski (2020) studied the relationship between the severity of collision injury and driving volatility by analyzing the collision event data set in the natural driving database. The results show that the greater driving fluctuation (longitudinal and transverse) increases the possibility of serious collision events. What is important is that the impact of longitudinal deceleration fluctuation is significantly greater than that of longitudinal acceleration fluctuation on collision results. Khattak and Wali (2017) also analyzed vehicle driving data, and captured the state degree of speed change through driving fluctuation, so as to distinguish normal driving from abnormal driving. Shangguan, Fu, Wang, and Luo (2021) put forward a method based on natural driving data, which integrates driving risk status recognition, feature extraction based on rolling time window, real-time driving risk status prediction, and driving risk influencing factor analysis, so as to accurately evaluate and predict real-time driving risk status. Feng, Bao, Sayer, Flannagan, Manser, and Wunderlich (2017) used vehicle sensor data in natural driving to study the variation characteristics of vehicle longitudinal acceleration, so as to identify aggressive drivers and driving behaviors. From the above research, it can be known that by studying driving volatility, abnormal driving behaviors such as normal driving and distraction can be distinguished, and they can also be linked with unsafe results in the real world (such as crash/near-crash). Therefore, we need to strengthen the research on the influence and function of driving volatility on driving distraction.

4.4. Limitations of the study

This article sorts out the literature of driving distraction recognition methods, which is limited by resources and ability, and has the following deficiencies:

(1) When searching for documents, only documents published in English were searched; hence, related documents published in other languages were excluded.
(2) When analyzing the author sharing network, due to the limitation of the database, only the literature data on the Web of Science are collected and analyzed.

(3) The selected driving distraction recognition methods are all based on traditional vehicles, and there also exists the absence of research on autonomous vehicles.
(4) The selected documents in this article focus on the recognition method of driving distraction, and other methods of recognition on abnormal driving behaviors such as road rage and driving fatigue are limited due to space limitations.

4.5. Current research deficiencies and future research trends

From the analysis of the current research situation of driving distraction, it can be seen that the acquisition method of driving distraction detection data is simple, the repeatability of test scenes is high, the research topics are concentrated, the detection data indicators are polarized, and the rationality analysis of detection data selection is less. The main problems existing in the method of driving distraction state recognition include: the types of driving distraction recognition are not comprehensive, the recognition model only recognizes the state, and the comparison between the advantages and disadvantages of the model method is not equal. With the development of artificial intelligence, image recognition technology becomes more mature, and image data analysis and processing will become a hot spot. At the same time, with the further development of technology, video data analysis will become simple and easy. At the same time, we should deeply research using big data forms such as cloud data and car networking data. Because driving distraction is a dangerous driving behavior, and natural driving data set is observed under normal driving behavior, simulation test is still the main data source for a long time, and future scene construction can be combined with holographic projection and VR technology to enhance the realism of tests. At present, the research on driving distraction mainly focuses on mobile phones, which should be decentralized to make the research topics more balanced. The current driving distraction research scenes are all manual driving scenes. With the development of intelligent transportation, there will be a man–machine driving scene for a long time before entering fully automatic unmanned driving. Follow-up research should be done on the influence of driving distraction in semi-automatic driving scenarios, such as the influence of vehicle–road cooperation on driving behavior, the study of task taking-over efficiency in man–machine co-driving, and the human factors analysis of different levels of self-driving vehicles with different mixing ratios in road network.

Combined with the characteristics of the selected indexes, a real-time recognition algorithm with better performance is constructed, and a perfect algorithm performance evaluation index is established. At present, the recognition effect of the same recognition model is uneven, which is mainly caused by the difference in the selection of recognition indicators. At present, the evaluation indicators of algorithms often use single indicators such as accuracy. In addition, the real-time requirements for state recognition are increasing day by day. Therefore, it is necessary to establish a complete comprehensive evaluation index of the recognition model, and compare and evaluate the performance of the model. Because the data used in the recognition model are different, and the quality of data plays an important role in the quality of the model, it is not absolutely fair to judge the quality of the recognition model by comparing the accuracy rate, and this problem can be solved by establishing a special model verification database. With the development of artificial intelligence, the construction of recognition model by deep learning has become a research hotspot in recent years, and will keep this trend in the future.

5. Conclusion

This research combines quantitative metrology and co-occurrence network analysis for the first time to carry out quantitative analysis of driving distraction literature in terms of time, country, publication, author, and keywords. It also analyzes the publications that study driving distraction recognition methods. This research can help researchers to understand the publication co-occurrence coupling relationship, understand the advantages and disadvantages of recognition methods and indicators, and develop more innovative ideas.

Funding

This work was supported in part by the National Natural Science Foundation of China under Grant 51905224.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aksjonov, A., Nedoma, P., Vodovozov, V., Petlenkov, E., & Herrmann, M. (2018). A novel driver performance model based on machine learning. *IFAC-Papers OnLine, 51*, 267–272.
Aksjonov, A., Nedoma, P., Vodovozov, V., Petlenkov, E., & Herrmann, M. (2019). Detection and evaluation of driver distraction using machine learning and fuzzy logic. *IEEE Transactions on Intelligent Transportation, 20*, 2048–2059.
Ali, S. F., & Hassan, M. T. (2018). Feature based techniques for a driver's distraction detection using supervised learning algorithms based on fixed monocular video camera. *KSII Transactions on Internet and Information Systems, 12*, 3820–3841.
Atiquzzaman, M., Qi, Y., & Fries, R. (2018). Real-time detection of drivers' texting and eating behavior based on vehicle dynamics. *Transportation Research Part F: Traffic Psychology and Behaviour, 58*, 594–604.
Botta, M., Cancelliere, R., Ghignone, L., Tango, F., Gallinari, P., & Luison, C. (2019). Real-time detection of driver distraction: Random projections for pseudo-inversion-based neural training. *Knowledge and Information Systems, 60*, 1549–1564.
Chen, Z., Wu, C., Zhong, M., Lyu, N., & Huang, Z. (2015). Identification of common features of vehicle motion under drowsy/distracted driving: A case study in Wuhan, China. *Accident Analysis & Prevention, 81*, 251–259.
Choudhary, P., & Velaga, N. R. (2017). Analysis of vehicle-based lateral performance measures during distracted driving due to phone use. *Transportation Research Part F: Traffic Psychology and Behaviour, 44*, 120–133.
Chui, K. T., Alhalabi, W., & Liu, R. W. (2019). Head motion coefficient-based algorithm for distracted driving detection. *Data Technologies and Applications, 53*, 171–188.
Dehzangi, O., Rajendra, V., & Taherisadr, M. (2018). Wearable driver distraction identification on-the-road via continuous decomposition of galvanic skin responses. *Sensors-Basel, 18*.
Dehzangi, O., Sahu, V., Rajendra, V., & Taherisadr, M. (2019). GSR-based distracted driving identification using discrete & continuous decomposition and wavelet packet transform. *Smart Health, 14*, 100085.
Dehzangi, O., & Taherisadr, M. (2018). Driver distraction detection using MEL cepstrum representation of galvanic skin responses and convolutional neural networks. In *IEEE. 24th international conference on pattern recognition* (pp. 1481–1486). New York: IEEE.
Dewar, R. E., & Olson, P. L. (2002). Human factors in traffic safety. *Publication of Lawyers & Judges Publishing Company Incorporated*.
Donmez, B., Boyle, L. N., & Lee, J. D. (2007). Safety implications of providing real-time feedback to distracted drivers. *Accident Analysis and Prevention, 39*, 581–590.
Duy, T., Ha, M. D., Sheng, W., Bai, H., & Chowdhary, G. (2018). Real-time detection of distracted driving based on deep learning. *IET Intelligent Transport Systems, 12*, 1210–1219.
Eraqi, H. M., Abouelnaga, Y., Saad, M. H., & Moustafa, M. N. (2019). Driver distraction identification with an ensemble of convolutional neural networks. *Journal of Advanced Transportation*.
Feng, F., Bao, S., Sayer, J. R., Flannagan, C., Manser, M., & Wunderlich, R. (2017). Can vehicle longitudinal jerk be used to identify aggressive drivers? *An examination using naturalistic driving data, Accident Analysis and Prevention, 104*, 125–136.

Foss, R. D., & Goodwin, A. H. (2014). Distracted driver behaviors and distracting conditions among adolescent drivers: Findings from a naturalistic driving study. *Journal of Adolescent Health, 54*(5), S50–S60.
Gao, C., Sun, M., Geng, Y., Wu, R., & Chen, W. (2016). A bibliometric analysis based review on wind power price. *Applied Energy, 182*, 602–612.
Hickman, J. S., & Hanowski, R. J. (2012). An assessment of commercial motor vehicle driver distraction using naturalistic driving data. *Traffic Injury Prevention, 13*(6), 612–619.
Khattak, A. J., & Wali, B. (2017). Analysis of volatility in driving regimes extracted from basic safety messages transmitted between connected vehicles. *Transportation Research Part C: Emerging Technologies, 84*, 48–73.
Kim, S., Song, T. J., Rouphail, N. M., Aghdashi, S., Amaro, A., & Gonçalves, G. (2016). Exploring the association of rear-end crash propensity and micro-scale driver behavior. *Safety Science, 89*, 45–54.
Kim, H. J., Jeong, Y. K., & Song, M. (2016). Content- and proximity-based author co-citation analysis using citation sentences. *Journal of Informetrics, 10*, 954–966.
Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A., Distracted driving and risk of road crashes among novice and experienced drivers, New England Journal of Medicine, 370(1), 54-59.
Le, T. H. N., Zheng, Y. T., & Zhu, C. C., et al. Multiple scale Faster-RCNN approach to driver's cell-phone usage and hands on steering wheel detection. In IEEE. IEEE Conference on Computer Vision and Pattern Recognition Workshops. New York: IEEE, 2016: 46-53.
Le, A. S., Inagami, M., Hamada, H., Suzuki, T., & Aoki, H. (2019). Towards online detection of driver distraction: Eye-movement simulation based on a combination of vestibulo-ocular reflex and optokinetic reflex models. *Transport Research F-Traffic, 65*, 716–729.
Lee, J. D., & Kolodge, K. (2019). Exploring trust in self-driving vehicles through text analysis. *Human Factors the Journal of the Human Factors & Ergonomics Society*. 001872081987267.
Li, Z., Bao, S., Kolmanovsky, I. V., & Yin, X. (2018). Visual-manual distraction detection using driving performance indicators with naturalistic driving data. *IEEE Transactions on Intelligent Transportation, 19*, 2528–2535.
Li, L., Zhong, Jr. B., Hutmacher, C., Liang, Y., Horrey, W.J., Xu, X. (2020). Detection of driver manual distraction via image-based hand and ear recognition, Accident Analysis Prevention, 137.
Liao, Y., Li, S. E., Wang, W., Wang, Y., Li, G., & Cheng, B. (2016). Detection of driver cognitive distraction: A comparison study of stop-controlled intersection and speed-limited highway. *IEEE Transactions on Intelligent Transportation, 17*, 1628–1637.
Liu, T., Yang, Y., Huang, G., Yeo, Y. K., & Lin, Z. (2016). Driver distraction detection using semi-supervised machine learning. *IEEE Transactions on Intelligent Transportation, 17*, 1108–1120.
Louie, J. F., & Mouloua, M. (2019). Predicting distracted driving: The role of individual differences in working memory. *Applied Ergonomics, 74*, 154–161.
Ma, Y., Hu, B., Chan, C., Qi, S., & Fan, L. (2018). Distractions intervention strategies for in-vehicle secondary tasks: An on-road test assessment of driving task demand based on real-time traffic environment. *Transportation Research Part D: Transport and Environment, 63*, 747–754.
Masala, G. L., & Grosso, E. (2014). Real time detection of driver attention: Emerging solutions based on robust iconic classifiers and dictionary of poses. *Transportation Research Part C-emerging Technologies, 49*, 32–42.
Masood, S., Rai, A., Aggarwal, A., Doja, M.N., & Ahmad, M. (2018). Detecting distraction of drivers using Convolutional Neural Network, Pattern Recognition Letter.
Mcevoy, S. P., Stevenson, M. R., & Woodward, M. (2007). The prevalence of, and factors associated with, serious crashes involving a distracting activity. *Accident Analysis & Prevention, 39*, 475–482.
Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: AN on-road study across three age groups. *Human Factors, 54*, 396–412.
Oviedo-Trespalacios, O., Afghari, A. P., & Haque, M. M. (2020). A hierarchical Bayesian multivariate ordered model of distracted drivers' decision to initiate risk-compensating behaviour. *Analytic Methods in Accident Research, 26*, 100121.
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T., Mulrow, C. D., et al. (2020). Mapping of reporting guidance for systematic reviews and meta-analyses generated a comprehensive item bank for future reporting guidelines. *Journal of Clinical Epidemiology, 118*, 60–68.
Parker, D., Reason, J., Manstead, A. R., & Stradling, S. (1995). Driving errors, driving violations and accident involvement. *Ergonomics, 38*, 1036–1048.
Reimer, B., D'Ambrosio, L. A., Coughlin, J. F., Kafrissen, M. E., & Biederman, J. (2006). Using self-reported data to assess the validity of driving simulation data. *Behav Res Methods, 38*, 314–324.
Riaz, F., Rathore, M. M., Sharif, T., Sajid, M., Ratyal, N., & Jararweh, Y. (2019). Internet of things based driver distraction detection and assistance system: A novel approach. *Ad Hoc & Sensor Wireless Networks, 44*, 305–337.
Sahayadhas, A., Sundaraj, K., Murugappan, M., & Palaniappan, R. (2015). A physiological measures-based method for detecting inattention in drivers using machine learning approach. *Biocybernetics and Biomedical Engineering, 35*, 198–205.
Sahayadhas, A., Sundaraj, K., Murugappan, M., & Palaniappan, R. (2015). Physiological signal based detection of driver hypovigilance using higher order spectra. *Expert Systems with Applications, 42*, 8669–8677.
Savolainen, P. T. (2016). Examining driver behavior at the onset of yellow in a traffic simulator environment: Comparisons between random parameters and latent class logit models. *Accident Analysis & Prevention, 96*, 300–307.
Scott-Parker, B. (2017). Emotions, behaviour, and the adolescent driver: A literature review. *Transportation Research Part F: Traffic Psychology & Behaviour, 50f*, 1–37.
Shahverdy, M., Fathy, M., Berangi, R., & Sabokrou, M. (2020). Driver behavior detection and classification using deep convolutional neural networks. *Expert Systems with Applications, 149*, 113240.
Shangguan, Q., Fu, T., Wang, J., & Luo, T. (2021). An integrated methodology for real-time driving risk status prediction using naturalistic driving data. *Accident Analysis and Prevention, 156* 106122.
Son, J., & Park, M. (2018). Detection of cognitive and visual distraction using radial basis probabilistic neural networks. *International Journal of Automotive Technology-Korea, 19*, 935–940.
Stavrinos, D. et al. (2013). Impact of distracted driving on safety and traffic flow. *Accident Analysis & Prevention, 61*, 63–70.
Taherisadr, M., Asnani, P., Galster, S., & Dehzangi, O. (2018). ECG-based driver inattention identification during naturalistic driving using Mel-frequency cepstrum 2-D transform and convolutional neural networks. *Smart Health, 9–10*, 50–61.
Topolšek, D., Areh, I., & Cvahte, T. (2016). Examination of driver detection of roadside traffic signs and advertisements using eye tracking. *Transportation Research Part F: Traffic Psychology and Behaviour, 43*, 212–224.
Valero Mora, P. M., Ballestar Tarin, M. L., Tontsch, A., Pareja Montoro, I., & Sanchez Garcia, M. (2012). Effects of manually entering navigator destinations while driving in a simulator. *IET Intelligent Transport Systems, 6*, 397–403.
Waard, D. D., Brookhuis, K. A., & Hernandez-Gress, N. (2001). The feasibility of detecting phone-use related driver distraction. *Vehicle Design, 26*(1), 85–95.
Wali, B., Khattak, A. J., & Karnowski, T. (2020). The relationship between driving volatility in time to collision and crash-injury severity in a naturalistic driving environment. *Analytic Methods in Accident Research, 28* 100136.
Wang, Y. K., Jung, T. P., & Lin, C. T. (2015). EEG-based attention tracking during distracted driving. *IEEE Transactions on Neural Systems & Rehabilitation Engineering, 23*(6), 1085–1094.
Wilson, F. A., & Stimpson, J. P. (1999). Trends in fatalities from distracted driving in the United States to 2008. *American Journal of Public Health, 100*(2010), 2213–2219.
Wöllmer, M., Blaschke, C., Schindl, T., et al. (2011). Online driver distraction detection using long short-term memory[J]. *IEEE Transactions on Intelligent Transportation Systems, 12*(2), 574–582.
Wu, L., Wang, W., Jing, P., Chen, Y., Zhan, F., Shi, Y., et al. (2020). Travel mode choice and their impacts on environment—a literature review based on bibliometric and content analysis, 2000–2018. *Journal of Cleaner Production, 249*, 119391.
Xing, Y., Lv, C., & Cao, D. (2020). *Chapter 5 - Driver behavior recognition in driver intention inference systems, advanced driver intention inference*. Elsevier.
Xu, Y., Li, S., Gao, S., Tan, D., Guo, D., & Wang, Y. (2018). Recognition method of construction conflict based on driver's eye movement. *Accident Analysis & Prevention, 113*, 193–201.
Yang, L., Guan, W., Ma, R., & Li, X. (2019). Comparison among driving state prediction models for car-following condition based on EEG and driving features. *Accident Analysis & Prevention, 133*, 105296.
Yang, Y., McDonald, M., & Zheng, P. (2012). Can drivers' eye movements be used to monitor their performance? A case study. *IET Intelligent Transport Systems, 6*, 444–452.
Yusoff, N. M., Ahmad, R. F., Guillet, C., Malik, A. S., Saad, N. M., & Merienne, F. (2017). Selection of measurement method for detection of driver visual cognitive distraction: A review. *IEEE Access, 5*, 22844–22854.
Zhao, D., & Strotmann, A. (2011). Counting first, last, or all authors in citation analysis: A comprehensive comparison in the highly collaborative stem cell research field. *Journal of the Association for Information Science & Technology, 62*, 654–676.
Zhao, X., Xu, W., Ma, J., Li, H., & Chen, Y. (2019). An analysis of the relationship between driver characteristics and driving safety using structural equation models. *Transportation Research Part F: Traffic Psychology and Behaviour, 62*, 529–545.

Huimin Ge received the Ph.D. degree from the School of Automobile and Traffic Engineering, Jiangsu University, Zhenjiang, China, in 2017, where she is currently an Associate Professor. Her research interests include transportation planning and safety, traffic conflicts and driving behavior research.

Journal of Safety Research 82 (2022) 409–416

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Assessing the Australian occupational driver behavior questionnaire in U.S. taxi drivers: Different country, different occupation and different worker population

Cammie Chaumont Menéndez [a,*], Richard Munoz [d], Timothy J. Walker [c], Benjamin C. Amick III [b]

[a] Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Division of Safety Research, 1095 Willowdale Road, Morgantown, WV 26505, United States
[b] Fay W. Boozman College of Public Health, University of Arkansas for Medical Sciences, 4301 West Markham #820, Little Rock, AK 72205, United States
[c] Department of Health Promotion and Behavioral Sciences, University of Texas Health Sciences Center at Houston School of Public Health, 1200 Pressler Street, Houston, TX 77067, United States
[d] Robert Stempel College of Public Health & Social Work, Florida International University, AHC5, 11200 SW 8th St #500, Miami, FL 33174, United States

ARTICLE INFO

Article history:
Received 20 November 2020
Received in revised form 17 January 2022
Accepted 13 July 2022
Available online 22 July 2022

Keywords:
ODBQ
Road safety
Questionnaire
Occupational
Motor vehicle crashes

ABSTRACT

Background: Promoting safe driver behaviors is an important aspect of road safety. To better understand road safety behaviors, there is a role for practical instruments that can validly measure typical road safety behaviors among occupational drivers. The Occupational Driver Behavior Questionnaire (ODBQ) was developed to assess road safety behaviors among home health nurses in Australia. *Methods:* We administered a cross-sectional survey to a sample of taxi drivers in two U.S. metropolitan areas. The survey included Newnam's ODBQ-12 and a study-specific 15-item version (ODBQ-15) assessing 4 different road safety behaviors with 3 more items added and motor-vehicle crashes in the past year. Logistic regression analyses examined the association of the road safety behaviors with motor vehicle crashes. A series of confirmatory factor analysis (CFA) models assessed the construct validity of the ODBQ-12 and ODBQ-15. *Results:* We pooled survey data from 497 Houston drivers and 500 Los Angeles drivers to assess study aims. CFA models examining the 12-item and the 15-item ODBQ versions had good model fit (Comparative Fit Index > 0.95, Tucker Lewis Index ≥ 0.95, root mean square error of approximation < 0.06, standardized root mean square residual ≤ 0.05). The ODBQ's road safety behaviors were significantly associated ($p < 0.001$) with crashes while working (ORs 0.51–0.75) and not working (ORs 0.57–0.84). *Conclusions:* The ODBQ-12 and ODBQ-15 were both significantly associated with motor vehicle crashes among taxicab drivers in two large U.S. metropolitan areas. Researchers studying occupational drivers who transport passengers may want to consider using the ODBQ-15. The 3 additional items are meaningful to this workforce and are priority areas for international road safety efforts.

1. Introduction

The United Nations passed a resolution in 2020 establishing a second *Decade (2021–2030) of Action for Road Safety* (United Nations, 2020), where road injury remains a leading cause of death for low and middle-income countries (World Health Organization, 2020) and the United States (CDC, 2021). Road injuries are estimated to cost the world economy almost $2 trillion dollars from 2015 to 2030 (Chen, Kuhn, Prettner, & Bloom, 2019). In the United States, transportation-related injuries are the leading cause of work-related death (Bureau of Labor Statistics, 2021a), with roadway incidents (e.g., collisions, running off the road) accounting for the vast majority of these fatalities (Bureau of Labor Statistics, 2021b). The annual economic burden to U.S. employers for work-related motor-vehicle crashes in 2015 was estimated at $25 billion (Network of Employers for Traffic Safety, 2016). Road safety continues to be a priority area among U.S. federal entities (e.g., Administration, 2021; Administration, 2021; National Institute for Occupational Safety and Health, 2021; Safety & Administration, 2012), the Academies (2016), non-governmental organizations (e.g., Council, 2021), academic institutions (e.g., National Institute for Occupational Safety and Health, 2021), and employers (Network of Employers for Traffic Safety, 2021).

* Corresponding author at: National Institute for Occupational Safety and Health, 1095 Willowdale Dr, MS 1811, Morgantown, WV 26505, United States.
E-mail address: cmenendez@cdc.gov (C. Chaumont Menéndez).

https://doi.org/10.1016/j.jsr.2022.07.008
0022-4375/

A key component of road safety research is improving driving behaviors. The roles of speeding (Bowie & Walz, 1994; Elvik, 2005; Speed, 2018; Joksch, 1993), distracted driving (Atchley, Tran, & Salehinejad, 2017; Caird, Johnston, Willness, Asbridge, & Steel, 2014), and fatigue (Robb, Sultana, Ameratunga, & Jackson, 2008) in motor-vehicle crash risks are well established. There is a wide range of measurement methods and tools with varying levels of sophistication and cost to assess these risk factors. Epidemiologic research in high-income countries focused on preventing crashes among occupational drivers routinely employs recent technological advances to directly measure risk factors for unsafe driving behaviors or crash outcomes (Bell, Taylor, Chen, Kirk, & Leatherman, 2017; Campbell, 2012; Chen, Fang, Guo, & Hanowski, 2016), with direct measurement considered the gold standard (Spielholz, Silverstein, Morgan, Checkoway, & Kaufman, 2001). However, research examining road safety behaviors among taxi and for-hire drivers continues to rely on self-report measures of drivers' perspectives of their driving behavior. Studies conducted across several continents assessing road safety behaviors span lower, middle, and high-income countries: Australia (Dalziel & Soames Job, 1997), Cameroon (Oyono et al., 2021), China (Meng et al., 2016; Routley, Ozanne-Smith, Qin, & Wu, 2009; Wang, Li, & et al., 2019, Wang, Zhang, & et al., 2019; Wu & Loo, 2016), Ethiopia (Asefa, Ingale, Shumey, & Yang, 2015; Hassen, Godesso, Abebe, & Girma, 2011), Iran (Dadipoor, Ranaei, Ghaffari, Rakhshanderou, & Safari-Moradabadi, 2020; Habibi, Haghi, & Maracy, 2014; Omidi, Mousavi, Moradi, & Taheri, 2021; Razmara, Aghamolaei, Madani, Hosseini, & Zare, 2018a; Razmara, Aghamolaei, Madani, Hosseini, & Zare, 2018b; Vehadi et al., 2018), Singapore (Lim & Chia, 2015), Tanzania (Nguyen et al., 2018), Uganda (Muni et al., 2019, 2020), United States (Hill, Baird, Torres, Obrochta, & Jain, 2021), and Vietnam (Hill et al., 2013). All but two of these studies were conducted in the past decade, while other, more sophisticated technologies to measure road safety behaviors have been available. There is a clear demand for cost-effective and practical road safety assessment measures in occupational safety and health research focused on taxi drivers, an occupation comparatively under-represented in road safety research.

Valid, reliable, and pragmatic measures are key for improving road safety research efforts when sophisticated technological tools are not feasible, not possible, or too far removed from understanding the drivers' behaviors. The ground transportation industry subsector that includes taxi and other for-hire drivers is a workforce with fewer road safety regulations, most frequently nontraditional employment arrangements (e.g., gig workers), and comprised of small fleets or operator-owned and managed (e.g., independent taxi drivers) with a very limited budget for road safety technology and generally no, if any, personnel dedicated to fleet safety. Road safety research involving interviewing occupational drivers about their driving behaviors continues to be conducted internationally in countries crossing all income levels because of their low cost and relative ease to administer among taxi (and other gig) drivers. There is a need for validated measurement instruments that can be used by companies or operators with very limited resources to support the veracity of research that is used to inform policy.

The Occupational Driver Behavior Questionnaire (ODBQ) is distinguished from the Driver Behavior Questionnaire as it recognizes the role of the occupational setting in road safety behaviors (Newnam & VonSchuckmann, 2012). The ODBQ is comprised of 12 questions asking about the frequency of drivers' road safety behaviors spanning traffic laws, speeding, fatigue, and distracted driving. It was intended to provide drivers and their management with proactive opportunities for improvements in road safety by using a survey tool designed for the occupational driving context. The ODBQ was developed and validated in an Australian community-based nursing organization and is a practical survey instrument for use in the field (Newnam, Greenslade, Newton, & Watson, 2011). The purpose of this analysis was to examine the psychometrics of the Occupational Driver Behavior Questionnaire among a population of U.S. taxi drivers in Houston and Los Angeles. This workforce differs socio-demographically, drives for longer hours, and works different shifts than the original population for which it was designed.

The occupational driving context is unique because of the added job demands and stress to the basic function of driving. Driving is leading the functioning of a vehicle while optimizing available cognitive resources to safely arrive at a location while navigating traffic, road signs/lights, and road and weather conditions during a dynamic process from departure to destination. For taxi drivers, job demands vary by both the number and duration of trips, providing customer service during the trip, and collecting payment at the end of the trip. Stress in the occupational context would include meeting these demands under too many or too little fares, the potential for passenger violence, fares late at night or early in the morning and/or after a long shift, and while tired. Adding items relevant to passenger safety captures unique elements of this domain of occupational driving jobs. Furthermore, items specific to drivers who exclusively drive passengers for a living are meaningful to company management and industry regulators who are faced with making decisions to strengthen existing policies or implement new ones in the midst of seemingly unrestricted Transportation Network Companies, which provide app-based ride sourcing services. In response to the current regulatory environment, three new items related to passenger safety while driving were added to the original scale. A secondary purpose of this analysis was to compare the performance of two versions of the ODBQ scale – the validated 12-item scale (ODBQ-12) and a 15-item version with additional items conceptually important for professional drivers who transport passengers (ODBQ-15).

2. Methods

2.1. Participants and procedures

Drivers licensed to drive a taxi for their city for 12 months or more were invited to participate in a cross-sectional study. Flyers describing the study were posted in break areas at airports and in regulatory offices. Additionally, ground transportation regulators in each city sent an email describing the study to licensed taxi drivers. Trained surveyors in sampling and administering the survey instrument conducted interviews of eligible taxi drivers at both international airports in Houston and a downtown location in Houston and the international airport in Los Angeles. Taxis were systematically approached within randomly selected parking lot lanes. Taxi drivers were informed the survey should take approximately 30 min and they would be provided with a $40 gift card. The drivers were told the study purpose was to ask about their work environment, experiences driving with passengers in the past year, and time spent driving. After agreeing to participate, taxi drivers meeting the inclusion criteria (licensed by the city for at least 12 months) and providing verbal assent after receiving and discussing the consent form were administered a 30-min survey. Drivers were remunerated the $40 gift card for their time. The National Institute for Occupational Safety and Health Institutional Review Board reviewed and approved the study protocol, data collection instruments, and consent process. The Office of Management and Budget reviewed and approved the project in accordance with the Paperwork Reduction Act.

Table 1
Occupational Driver Behavior Questionnaire subscales, items and corresponding descriptions.

Subscales	Items	Descriptions
Speeding[+]	SP1	How often do you exceed the speed limit on a residential road?
	SP2	How often do you exceed the speed limit on a highway or freeway?
	SP3	How often do you exceed the speed limit when traveling to do pickups[§]?
Rule Violation	RV1	How often do you not signal to change lanes when no other traffic is around?[¶]
	RV2	How often do you perform a U-turn in a non-designated zone?
	RV3	How often do you not come to a complete standstill at a stop sign?[¶]
Inattention	IN1	How often do you drive while thinking about how to get to your destination?
	IN2	How often do you drive while thinking about your next pickup[Ŧ] or work task?
	IN3	How often do you drive while thinking about your work-related problems/issues?
Driving While Fatigued	DF1	How often do you drive while tired?
	DF2	How often do you have difficulty driving because of tiredness or fatigue?
	DF3	How often do you find yourself nodding off while driving?
Additional Items[**]	SP4	How often do you exceed the speed limit when travelling with a passenger?
	RV4	How often do you wear your seat belt while driving?
	IN4	How often do you use a handheld cell phone while driving?

Note. *Items are Likert-based with 5 options, ranged from 1 to 5, 1 being "Rarely/Never" through 5 being "Very often/all the time." All items except for RV4 were reverse-coded. (Newnam et al., 2011).

[+] Every item on Speeding began with 'How often do you deliberately' in the original 12-item ODBQ. For the current study we dropped 'deliberately' from each question for both versions of the ODBQ.

[§] Modified from 'travelling to clients or the office'.

[¶] Removed 'fail' to not using a signal to change lanes and coming to a complete standstill for RV1 and RV3.

[Ŧ] Replaced 'next patient' with 'next pickup'.

[**] Items SP4, RV4, and IN4 were not a part of the 12-item ODBQ (Newnam et al., 2011).

2.2. Instrument

The ODBQ was administered as part of a larger survey developed to evaluate workplace violence and motor- vehicle crashes among taxi drivers, whose causes of work-related death are due almost exclusively to violence and crashes (BLS, 2018b). The 30-min overall survey included questions about the following topic areas: business-related aspects to driving a taxi, psychosocial work environment, passenger violence, motor-vehicle crashes, road safety behaviors (ODBQ), safety measures, and socio-demographics. Socio-demographic variables included: sex, age, race/ethnicity, nativity, educational attainment, and marital status. Participants provided exact age in years and designated sex as "Male" or "Female." When collecting data on race, surveyors showed respondents a card listing options for "White," "Black or African American," "Asian," "American Indian or Alaska Native," "Native Hawaiian or Pacific Islander," "Refused," and "Other"; drivers were asked to identify one or more as applicable. Participants responding "Other" were asked to specify. On the other side of the card were options to describe ethnicity with definitions for determining "Hispanic, Latino, or Spanish origin." Drivers responding "Yes" or "No" to the question"Were you born in the US" determined nativity. Educational attainment outlined the highest level of formal education completed as "Grade school," "Secondary school," "Some high school," "High school diploma," "Technical/-trade school," "Associate's degree," "Undergraduate degree," "Graduate degree, Master level," and "Graduate degree, Doctoral level." The responses "Married," "Not married, but in a long-term relationship," "Separated," "Divorced," "Widowed," and"Single" designated the driver's marital status.

Occupational Driver Behavior Questionnaire. The ODBQ-12 assesses four subscales: speeding, rule violations, inattention, and driving while tired (Newnam & VonSchuckmann, 2012; Newnam et al., 2011). Table 1 provides a list of the specific items according to their subscales. Three items were added to create the ODBQ-15 because they were conceptually meaningful, as described previously, to taxi driver safe driving behavior. The additional items were generated by subject matter experts knowledgeable of the road safety concerns (see Table 1). One item was added to the speeding subscale (speeding with a passenger), one to the rule violations subscale (wearing a seatbelt), and one to the inattention subscale (using a handheld mobile device while driving). All items were prefaced with"How often do you" and anchored by a 5-point Likert scale where 1 signifies "Rarely or never" and 5 signifies "Very often or all the time." Modifications to original item wording were minimal and done for clarity and relevance to the taxi driver population.

Two variables served as outcomes: (1) the frequency of motor-vehicle crashes in the past 12 months not related to driving a taxicab, and (2) the frequency of motor-vehicle crashes occurring while driving a taxi in the past 12 months; each was asked as an open-ended question and later coded as 1 (crash) and 0 (no crash).

2.3. Statistical analyses

Data collected from both cities, Houston and Los Angeles, were pooled together after it was determined there were no meaningful differences in demographic averages or inter-item correlation statistics between cities likely to affect the ODBQ psychometric properties. Descriptive statistics were calculated using SPSS, version 24 (IBM Corp, 2016).

We developed a strategic approach to selecting the ODBQ versions and scale forms for analyses. We calculated Cronbach's alpha to assess the ODBQ scale reliability (Cronbach, 1951). Cronbach's alpha values greater than 0.7 were considered acceptable for a minimum reliability threshold (Frost et al., 2007). We then performed Confirmatory factor analysis (CFA) using maximum likelihood estimation with robust standard errors using M-PLUS version 8 (Muthén & Muthén, 2017) to assess construct validity. We chose a confirmatory, instead of exploratory, approach because of existing empirical and theoretical knowledge about the ODBQ (Newnam & VonSchuckmann, 2012; Newnam et al., 2011). Notably, the ODBQ was developed to assess road safety behaviors across four subscales. In addition, existing research supports a 4-factor solution among drivers who drove at least once per week for occupational purposes (Newnam et al., 2011).

Given the existing information on home health care workers whose work-related driving was secondary to the primary job tasks of healthcare support, and the opportunity to test the ODBQ scale in a sample of taxi drivers whose primary job task is to drive passengers, we chose to examine the factor structure through a systematic process. We started with assessing a model fit for a 4-factor solution to compare findings with the original Australian sample. Testing the first-order model implies there is no meaningful conceptual difference between the subscales, which we do not believe, but we wanted to have data available to companies or road safety researchers and practitioners who embraced and moved forward with applying the ODBQ as a first-order model in practice. Finally, we assessed a second-order 4-factor model to provide a combination of practicality for users to obtain a single score but retain the ability to assess differences in the subscales after behavior-specific policies or targeted trainings are implemented.

We compared fit between models using Satorra Bentler's scaled chi-square difference test (Satorra & Bentler, 2001). The collective performance of the following indicators were used to assess model fit: overall Chi-square (non-significant value = good fit), comparative fit index (CFI, >0.90 = adequate fit and >0.95 = good fit), Tucker-Lewis index (TLI, >0.90 = adequate fit and >0.95 = good fit), root mean square error of approximation (RMSEA, 0.05–0.08 = adequate fit, <0.05 = good fit), and standardized root mean square residual (SRMR, 0.05–0.08 = adequate fit and <0.05 = good fit) (Bryne, 2012). CFAs were performed separately for the 12-item and 15-item ODBQ versions. We also examined the magnitudes of factor loadings and modification indices. Model adjustments based on modification indices were considered only if they indicated points of strain and were substantively meaningful.

Univariable and multivariable logistic regression models assessed the association of both versions of the ODBQ to each outcome. Odds ratios and 95% confidence intervals were calculated. The Wald chi-square (X^2) test statistic assessed model fit (Buse, 1982). All logistic regression models were conducted using SAS v9.4 (SAS Institute, Cary, NC).

3. Findings and results

3.1. Descriptive statistics

The average driver age was 43.9 years and the drivers were predominantly males (92%) (Table 2). More drivers were either Black (35%) or White (32%), with 12% identifying as Hispanic and 12% Asian. Just over half (53%) reported being born outside of the United States. Half of the study sample completed some college, and at least one-third attained a high school diploma. The majority (57%) of participating drivers were married or in a long-term relationship.

Item means ranged from 3.41-4.82 with no major floor or ceiling effects except for RV4, DF2, and DF3, which had ceilings of 86.7%, 68.2% and 77.3%, respectfully (Table 3). In addition, all items had complete data. There were two ODBQ items correlated at 0.85 or greater (Table 3). For speeding, 'how often do you exceed the speed limit on a highway or freeway' (SP2) and 'how often do you exceed the speed limit when traveling to do pickups' (SP3) were correlated at 0.87. This was a deciding point for dropping one of the items or keeping both. These two items were re-evaluated and considered equally valuable components of the speeding with a passenger subscale.

Table 2
Descriptive statistics for Houston and Los Angeles Taxi Cab Drivers for Years.

Individual factors	Mean ± SD or %
# of crashes for the past 12 months (n = 971)	
At least 1 crash (n = 148)	15%
Age (n = 971)	43.9 ± 10.1
Sex (n = 961)	
Male (n = 897)	92%
Race and Ethnicity (n = 835)	
Black or African American, non-Hispanic (n = 290)	35%
White, non-Hispanic (n = 266)	32%
Hispanic (n = 101)	12%
Asian, non-Hispanic (n = 100)	12%
American Indian/Alaska Native/Native Hawaiian/Pacific Islander, non-Hispanic (n = 30)	4%
Refused to answer question (n = 29)	4%
Nativity (n = 973)	
Born outside the U.S. (n = 514)	53%
Educational attainment (n = 970)	
Below High School (n = 128)	13%
High School (n = 333)	34%
Some College (n = 491)	50%
Graduate degree (n = 18)	2%
Relationship status (n = 971)	
Married or Long-term relationship (n = 559)	57%
Single (n = 273)	28%
Separated, Divorced, or Widowed (n = 129)	13%
Refused to answer question (n = 10)	1%

3.2. Examining ODBQ scale performance: Confirmatory factor analyses

The CFA results indicated the 4-factor models appeared to best fit the data (CFI ≥ 0.91, TLI ≥ 0.89, RMSEA ≤ 0.08, SRMR ≤ 0.08. (Table 4). The single factor models for both the 12- and 15-item ODBQ scales had poor fit (Table 4). The second-order single-factor models demonstrated satisfactory fit. More specifically, the CFI and TLI values were in the adequate range and the RMSEA and SRMR values were just outside the acceptable range (Table 4).

When examining local fit, the factor loadings were >0.50 and statistically significant for the first-order 4-factor ODBQ-12 model. The underlying factor ODBQ-12 model had factor loadings >0.50 with the exception of item 1 from Inattention (IN1 factor loading = 0.41) and items 2 and 3 from the fatigued driving construct (DF2, 0.43; DF3, 0.35). The second-order single-factor ODBQ-12 had factor loadings >0.50 with the exception of the fatigued driving construct (factor loading = 0.47).

When examining local fit for the ODBQ-15 models, the first-order 4-factor model had factor loadings >0.50 with the exception of the added item for Rule Violation: 'how often do you wear your seatbelt while driving' (factor loading = 0.02). There were five items with factor loadings <0.50 for the model examining one underlying factor: RV4 (0.03), IN1 (0.38), DF1 (0.47), DF2 (0.41), and DF3 (0.33). The second-order ODBQ-15 model had factor loadings >050 with the exception of RV4 (0.02) and the construct for fatigued driving (0.43).

Modification indices for the two 4-factor models indicated correlated residuals for items SP3 ('how often do you exceed the speed limit when traveling to do pickups') and SP2 ('how often do you exceed the speed limit on a highway or freeway'), and RV1 ('how often do you not signal to change lanes when no other traffic is around') and RV2 ('how often do you perform a U-turn in a non-designated zone'). Examining new models that included correlated residuals between items SP2&SP3 and RV1&RV2 indicated models with good fit indices for both the ODBQ-15 and the ODBQ-12 (CFI > 0.95, TLI ≥ 0.95, RMSEA ≤ 0.06, SRMR ≤ 0.05).

3.3. Examining ODBQ association with motor vehicle crashes: Logistic regression analyses

Scoring higher on the ODBQ-15 and the ODBQ-12 was significantly associated with not experiencing a motor- vehicle crash in

Table 3
Correlation matrix and descriptive statistics for ODBQ items.*

	Speeding				Rule violation					Inattention			Driving while fatigued		
	SP1	SP2	SP3	SP4	RV1	RV2	RV3	RV4	RV5	IN1	IN2	IN3	DF1	DF2	DF3
SP1	1														
SP2	**0.75**	1													
SP3	**0.74**	**0.87**	1												
SP4	**0.77**	**0.69**	**0.72**	1											
RV1	**0.68**	**0.62**	**0.63**	**0.68**	1										
RV2	**0.68**	**0.58**	**0.57**	**0.63**	**0.73**	1									
RV3	**0.67**	**0.57**	**0.57**	**0.61**	**0.67**	**0.77**	1								
RV4	0.03	0.06	**0.07**	0.06	0.04	0.04	0.00	1							
IN1	**0.20**	**0.29**	**0.30**	**0.18**	**0.21**	**0.18**	**0.19**	−0.03	1						
IN2	**0.33**	**0.44**	**0.46**	**0.33**	**0.34**	**0.31**	**0.29**	0.05	**0.72**	1					
IN3	**0.40**	**0.42**	**0.44**	**0.37**	**0.38**	**0.35**	**0.36**	0.05	**0.52**	**0.68**	1				
IN4	**0.45**	**0.46**	**0.46**	**0.44**	**0.42**	**0.45**	**0.47**	−0.01	**0.38**	**0.48**	**0.48**	1			
DF1	**0.32**	**0.41**	**0.43**	**0.31**	**0.37**	**0.38**	**0.33**	0.04	**0.38**	**0.46**	**0.52**	**0.30**	1		
DF2	**0.34**	**0.37**	**0.38**	**0.32**	**0.36**	**0.38**	**0.34**	−0.01	**0.32**	**0.40**	**0.45**	**0.34**	**0.68**	1	
DF3	**0.31**	**0.37**	**0.38**	**0.30**	**0.33**	**0.38**	**0.31**	0.05	**0.28**	**0.35**	**0.42**	**0.33**	**0.65**	**0.77**	1
Mean	3.69	3.46	3.52	3.89	3.90	4.05	3.99	4.82	3.41	3.59	3.86	3.89	4.08	4.48	4.63
SD	1.27	1.24	1.21	1.12	1.16	1.12	1.22	0.54	1.16	1.12	1.06	1.25	1.00	0.87	0.75
%Floor	9.4	9.6	8.0	4.7	5.6	4.7	6.9	0.6	7.8	6.1	3.9	6.6	2.7	1.2	0.4
%Ceiling	33.8	24.1	25.7	36.9	39.7	45.6	47.1	86.7	20.2	24.6	33.8	45.0	42.3	68.2	77.3
%Missing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Correlations are bivariate Pearson correlations. All correlations in bold were at least statistically significant at the < 0.05 observed significance level.
* Los Angeles N = 500, Houston N = 496. 19 cases removed in Houston dataset due to missingness; both datasets were combined. Final analytic sample n = 977.

Table 4
Model fit results, Confirmatory Factor Analysis (n = 977).

12-item model						
Models/Measures	χ^2	*df*	CFI	TLI	RMSEA	SRMR
1. First-order 4-factor	335.33	48	0.94	0.92	0.08	0.06
2. First-order 4-factor (MIA)	192.53	46	0.97	0.96	0.06	0.04
3. One underlying factor	1864.14	54	0.62	0.54	0.19	0.13
4. Second-order single-factor	395.16	50	0.93	0.90	0.08	0.08
15-item model						
Models/Measures	χ^2	*df*	CFI	TLI	RMSEA	SRMR
1. First-order 4-factor	631.14	84	0.91	0.89	0.08	0.07
2. First-order 4-factor (MIA)	351.99	82	0.96	0.95	0.06	0.05
3. One underlying factor	2256.61	90	0.66	0.60	0.16	0.12
4. Second-order single-factor	707.78	86	0.90	0.88	0.09	0.09

Note. χ^2, Chi-square; df, degrees of freedom; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; CFI, comparative fit index, TLI, Tucker-Lewis index.
12-item model. Models 1 and 4 were compared using the Satorra-Bentler (1994) scaled chi-square difference test, yielding statistically significant results [$\Delta\ \chi^2 = 56.88$, Δ df = 2, p < 0.001], suggesting that Model 1 fits the data better than 4.
15-item model. Models 1 and 4 were compared using the Satorra-Bentler (1994) scaled chi-square difference test, yielding statistically significant results [$\Delta\ \chi^2 = 80.85$, Δ df = 2, p < 0.001], suggesting that Model 1 fits the data better than 4.
MIA = modification indices applied. Both models with MIA had the highest two correlated residuals added, which were the same for both models, SP3 with SP2 and RV2 with RV1.
Every χ^2 p-value is < 0.001.

the past 12 months *while* driving a taxi (Table 5). Each subscale was inversely associated with the outcome and odds ratios ranged from 0.55 to 0.75(p < 0.001). Both full ODBQ scales were inversely associated with the outcome (ODBQ-12 OR = 0.52, 95% CI 0.41–0.66; ODBQ-15 OR = 0.51, 95% CI 0.40–0.66) (p < 0.001). When including all subscales in a single model to reflect a second-order 4-factor model, only Inattention (aORs = 0.73–0.75; p = 0.02) and Driving Fatigued (aORs-0.67–0.68; p < 0.01) remained associated with the outcome for both the ODBQ-12 and ODBQ-15 (Table 5).

For the second outcome, experiencing a motor-vehicle crash in the past 12 months *outside* of driving a taxi, both the ODBQ-15 and the ODBQ-12 were significantly associated with the outcome. Each subscale was inversely associated with the outcome. The odds ratios for each version of the scale were very similar for three of the subscales: Speeding, ORs = 0.67–0.69 (p < 0.001), Rule Violation, ORs = 0.58–0.66 (p < 0.001), and Driving Fatigued, ORs = 0.82, p = 0.05. The OR for Inattention was 0.84 (p = 0.05) for the ODBQ-12 and 0.78 (p = 0.01) for the ODBQ-15. Both full ODBQ scales were inversely associated with the outcome and similar in magnitude: ORs = 0.57–0.60 (p < 0.001). When including all subscales in a single model, only Rule Violation (aORs = 0.72–0.74; p < 0.05) remained associated with the outcome for either ODBQ scale version. The point estimates for Speeding were <1; ORs = 0.79–0.83. The point estimates for Inattention approximated 1; ORs = 0.99–1.01. The point estimates for Driving Fatigued were >1; ORs = 1.09.

4. Discussion

The focus of our analysis was to examine the validity of the original ODBQ in a new population of linguistically, racially, and

Table 5
Logistic Regression Models for Convergent Validity Testing.

Experiencing a Motor Vehicle Crash While Driving a Taxi < 12 months						
	ODBQ-15			ODBQ-12		
Model	O.R.	95% CI	Sig.	O.R.	95% CI	Sig.
Speeding	0.75	0.63, 0.88	<0.001	0.75	0.64, 0.89	<0.001
Inattention	0.59	0.48, 0.73	<0.001	0.61	0.50, 0.74	<0.001
Rule Violation	0.68	0.54, 0.85	<0.001	0.74	0.62, 0.87	<0.001
Driving Fatigued	0.55	0.45, 0.68	<0.001	0.55	0.45, 0.68	<0.001
Full Scale^	0.51	0.40, 0.66	<0.001	0.52	0.41, 0.66	<0.001
Speeding*	1.02	0.77, 1.36	0.89	1.03	0.79, 1.34	0.83
Inattention*	0.73	0.57, 0.95	0.02	0.75	0.59, 0.95	0.02
Rule Violation*	0.93	0.64, 1.34	0.68	0.90	0.69, 1.18	0.45
Driving Fatigued*	0.67	0.52, 0.87	<0.001	0.68	0.52, 0.88	<0.01
Experiencing a Motor Vehicle Crash Outside of Driving a Taxi < 12 months						
	ODBQ-15			ODBQ-12		
Model	O.R.	95% CI	Sig.	O.R.	95% CI	Sig.
Speeding	0.67	0.57, 0.78	<0.001	0.69	0.59, 0.80	<0.001
Inattention	0.78	0.65, 0.94	0.01	0.84	0.70, 1.00	0.05
Rule Violation	0.58	0.48, 0.71	<0.001	0.66	0.56, 0.77	<0.001
Driving Fatigued	0.82	0.66, 1.00	0.05	0.82	0.66, 1.00	0.05
Full Scale^	0.57	0.45, 0.72	<0.001	0.60	0.48, 0.74	<0.001
Speeding*	0.79	0.61, 1.02	0.07	0.83	0.66, 1.05	0.12
Inattention*	0.99	0.78, 1.25	0.92	1.01	0.81, 1.25	0.96
Rule Violation*	0.72	0.52, 0.99	0.04	0.74	0.58, 0.94	0.01
Driving Fatigued*	1.09	0.83, 1.42	0.53	1.09	0.83, 1.42	0.55

^ Overall survey score in one regression model.
* Subscales were combined into one regression model.

ethnically diverse occupational drivers and test the validity of an expanded version (ODBQ-15) with three added items meaningful to occupational drivers transporting passengers. Our findings indicated both versions of the Occupational Driver Behavior Questionnaire demonstrated good validity in a driving population that exclusively drives for a living. More specifically, confirmatory factor analysis testing revealed good construct validity for first-order 4-factor models, suggesting the ODBQ captures four distinct subscales: speeding, inattention, rule violation, and driving while fatigued. In addition, logistic regression models indicated good convergent validity for both versions of the ODBQ subscales. To our knowledge this is the first validation of the ODBQ outside of the original scale development work (Newnam et al., 2011; Newnam & Von Schuckmann, 2012), the first time in the United States among occupational drivers, and the first time in a field of research where cross-sectional study designs using surveys are the predominant road safety epidemiologic research tools and approaches.

Previous research examining the validity and reliability found best model fit for the 4-factor model of the ODBQ-12 with CFA testing revealing loading on the following constructs – speeding, rule violation, inattention, and fatigued driving -- and good model fit (Newnam et al., 2011). The original scale's development was conducted in Australia among occupational drivers who drove regularly over a week as part of their community nursing tasks. Our findings are consistent with that of Newnam et al. as our study indicated the ODBQ is made up of four subscales rather than one general scale.

Our study expands on this original work by examining the ODBQ in a new population, U.S. taxi drivers. The scale was originally conceptualized with the unique demands, timing, and work environment that add context to occupational drivers. Furthermore, taxi drivers ferry passengers around as their only work task on demand and can work long hours at times not supported by their body's circadian rhythm. To manage the stress and job demands, driving performance may be inadvertently protected at the expense of road safety behaviors known to be particularly crucial when transporting passengers (such as speeding, seatbelt use, and distracted driving). To this end, an item representing each of these road safety behaviors meaningful to the taxi industry was added to the ODBQ. Therefore, an additional objective for our study was to test a 15-item version of the ODBQ that included three questions about speeding (speeding with a passenger), rule violation (seatbelt use), and inattention (using a handheld device while driving). Our results indicated no major differences in fit between versions. Even though the 12-item version may be preferred because it has fewer items, we feel both versions are acceptable. Importantly, the 15-item version includes additional items that are conceptually important and indicate specific responses. For instance, every city requires some form of restricted cell phone use while driving a taxi. Every city requires its taxi drivers to fasten their seatbelt at all times. Taxi drivers' response to frequency of using a handheld device while driving or neglect to use their seatbelt should be 'never.' Any level of frequency of performing these behaviors is an opportunity to reinforce road safety behaviors by reminding taxi drivers of the importance of wearing their seatbelt and minimizing behaviors that lead to distracted driving and trying to understand and address the barriers to safe behaviors.

The convergent validity of both the 15-item and the 12-item versions of the scale was promising. Specifically, the overall scales were significantly associated with motor-vehicle crashes within and outside the work environment. The higher the score for road safety behavior subscales, the lower the odds ratio for experiencing a crash. This is a novel finding as previous evaluations of the ODBQ's validity did not examine convergent validity. Observing an association between both versions of the scale with relevant injury outcomes is a valuable contribution to the road safety literature, especially regarding the use of screening tools. Interestingly, when all subscales were included in multivariable models, only Inattention and Driving While Fatigued were associated with motor-vehicle crashes occurring while driving a taxi, whereas only Rule Violation was significantly associated with motor-vehicle crashes occurring while not driving a taxi. These findings suggest all of the subscales are important; however, Inattention and Driving While Fatigued may play more of a role in occupational crash outcomes and Rule Violation may play more of a role in non-occupational crash outcomes. It is worth mentioning Rule Violation and Speeding are correlated (Pearson's r ranging 0.74–0.76) and Speeding may play more of a role in non-occupational crash outcomes that is not observable. Overall, more work is necessary to examine the convergent validity of subscales with the same outcomes using objective measures or a study design that incorporates longitudinal data to gain a better understanding of the predictive validity.

The study has several strengths and limitations. A limitation is the cross-sectional study design provided concurrent validity, rather than the predictive validity obtained from a longitudinal study design, which limits the inferences that can be made about ODBQ scale performance in predicting motor-vehicle crashes and injuries. However, the robust population size and the rigor of the validity testing provided much needed insight into how well a valid scale was performing in a different worker population that spends more time driving. An important consideration for future applications of this tool is the training of management in its use for educational purposes rather than in a punitive context. For this study, all participants were informed their responses were confidential and none of the companies represented would be provided with any of the data. Drivers provided verbal assent and were given a copy of the consent form. In an environment where companies administer the ODBQ in house, if management has a history of punitive responses to unsafe road behaviors drivers would be less likely to respond to the questionnaire. Additional study strengths include the socio-demographic diversity of the population that increases the generalizability of the findings and the industry examined as novel contributions in the use and advancement of the ODBQ as a valid measurement tool for road safety among those who exclusively drive for a living.

Our findings contribute to the occupational road safety literature by identifying and further validating a practical and inexpensive measurement tool for road safety. We adapted the tool for use in a population of full-time drivers who transport passengers as their main job task. The constructs encompassed in the full scale are relevant to every aspect of driving a taxi – fatigue, distracted driving (inattention), obeying traffic laws (rule violation), and speeding. This versatile tool could be used by management to schedule driver refresher training or in research when expensive direct road safety measures are not feasible. The current findings support the use of either version of the ODBQ as a promising contribution to occupational road safety.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention.

References

Atchley, P., Tran, A. V., Salehinejad, M. (2017). Constructing a publically available distracted driving database and research tool. In *Accident analysis & prevention (Vol. 99, no. Part A, pp. 306–311)*.

Asefa, N. G., Ingale, L., Shumey, A., & Yang, H. (2015). Prevalence and factors associated with road traffic crash among taxi drivers in Mekelle Town, Northern Ethiopia, 2014: A cross sectional study. *PLoS ONE, 10*(3), e0118675.

Bell, J. L., Taylor, M. A., Chen, G.-X., Kirk, R. D., & Leatherman, E. R. (2017). Evaluation of an in-vehicle monitoring system (IVMS) to reduce risky driving behaviors in commercial drivers: Comparison of in-cab warning lights and supervisory coaching with videos of driving behavior. *Journal of Safety Research, 60*, 125–136.

BLS (2021a). Injuries, illnesses, and fatalities. Latest numbers: Fatal work-related injuries. www.bls.gov/iif.

BLS (2021b). All worker profile, 2003–2018. Fatal occupational injuries by selected characteristics, 2003–2018. https://stats.bls.gov/iif/oshcfoi1.htm.

Bowie, N., & Walz, M. (1994). Data analysis of the speed-related crash issue. *Auto & Traffic Safety, Vol. 1, no. 2*. Washington, DC: National Highway Traffic Safety Administration.

Bryne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York: Routledge.

Buse, A. (1982). The likelihood ratio, Wald and Lagrange multiplier tests: An expository note. *American Statistician, 36*, 153–157.

Caird, J. K., Johnston, K. A., Willness, C. R., Asbridge, M., & Steel, P. (2014). A meta-analysis of the effects of texting on driving. *Accident Analysis & Prevention, 71*, 311–318. https://doi.org/10.1016/j.aap.2014.06.005.

Campbell, K. L. (2012). The SHRP 2 naturalistic driving study: Addressing driver performance and behavior in traffic safety. *TR News, 282* https://insight.shrp2nds.us/documents/shrp2_background.pdf.

Chen, G. X., Fang, Y., Guo, F., & Hanowski, R. J. (2016). The influence of daily sleep patterns of commercial truck drivers on driving performance. *Accident Analysis & Prevention, 91*, 55–63.

Chen, S., Kuhn, M., Prettner, K., & Bloom, D. E. (2019). The global macroeconomic burden of road injuries: Estimates and projections for 166 countries. *The Lancet Planetary Health, 3*(9), e390–e398.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

Dadipoor, S., Ranaei, V., Ghaffari, M., Rakhshanderou, S., & Safari-Moradabadi, A. (2020). Safe driving behaviors among taxi drivers: A predictive cross-sectional study based on the health belief model. *Archives of Public Health, 78*, 82.

Dalziel, J. R., & Soames Job, J. R. (1997). Motor vehicle accidents, fatigue and optimism bias in taxi drivers. *Accident Analysis & Prevention, 29*(4), 489–494.

Elvik, R. (2005). Speed and road safety: Synthesis of evidence from evaluation studies. In *Transportation research record (Vol. 1908, pp. 59–69). Washington, DC: Transportation Research Board*.

Federal Motor Carriers Safety Administration (FMCSA) (2021). ETA: The motor carrier safety planner (accessed online September 16, 2021) https://csa.fmcsa.dot.gov/SafetyPlanner/Default.aspx.

Frost, M. H., Reeve, B. B., Liepa, A. M., Stauffer, J. W., Hays, R. D., & Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group (2007). What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value in Health, 10*(s2), S94–S105. https://doi.org/10.1111/j.1524-4733.2007.00272.x.

Habibi, E., Haghi, A., & Maracy, M. R. (2014). Investigating the predictive of risk-taking attitudes and behaviors among Iranian drivers. *Journal of Education and Health Promotion, 3*, 19.

Hassen, A., Godesso, A., Abebe, L., & Girma, E. (2011). Risky driving behaviors for road traffic accident among drivers in Mekele city, Northern Ethiopia. *BMC Research Notes, 45*, 35.

Hill, L., Baird, S., Torres, K., Obrochta, C., & Jain, P. (2021). A survey of distracted driving and electronic device use among app-based and taxi drivers. *Traffic Injury Prevention, 22*(1), S27–S31.

IBM Corp. Released 2016 IBM SPSS statistics for windows, Version 24.0 2016 IBM Corp Armonk, NY.

The International Transportation Forum, 2018. Speed and crash risk. https://www.itf-oecd.org/sites/default/files/docs/speed-crash-risk.pdf.

Joksch, H. C. (1993). Velocity change and fatality risk in a crash: A rule of thumb. *Accident Analysis & Prevention, 25*, 103–104.

Lim, S. M., & Chia, S. E. (2015). The prevalence of fatigue and associated health and safety risk factors. *Singapore Medical Journal, 56*(2), 92–97.

Meng, F., Li, S., Cao, L., Peng, Q., Li, M., Wang, C., et al. (2016). Designing fatigue warning systems: The perspective of professional drivers. *Applied Ergonomics, 53*, 122–130.

Muni, K., Kobusingye, O., Mock, C., Hughes, J. P., Hurvitz, P. M., & Guthrie, B. (2019). Motorcycle taxi programme is associated with reduced risk of road traffic crash among motorcycle taxi drivers in Kampala, Uganda. *International Journal of Injury Control and Safety Promotion, 26*(3), 294–301.

Muni, K., Kobusingye, O., Mock, C., Hughes, J. P., Hurvitz, P. M., & Guthrie, B. (2020). Motorcycle taxi programme increases safe riding behaviours among its drivers in Kampala, Uganda. *Injury Prevention, 26*, 5–10.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

National Academies of Sciences, Engineering, and Medicine (2016). Commercial motor vehicle driver fatigue, long-term health, and highway safety: Research needs. Washington, DC: The National Academies Press. https://doi.org/10.17226/21921.

National Highway Transportation Safety Administration (NHTSA). Road safety (accessed online September 16, 2021) https://www.nhtsa.gov/road-safety.

National Institute for Occupational Safety and Health (NIOSH). NIOSH strategic plan: FYs 2019–2023 (accessed online September 16, 2021) https://www.cdc.gov/niosh/about/strategicplan/.

National Safety Council. Defensive driving training for your employees (accessed online September 16, 2021) https://www.nsc.org/safety-training/defensive-driving/train-your-employees.

Network of Employers for Traffic Safety (NETS) (2016). Cost of motor vehicle crashes to employers, 2015. https://www.trafficsafety.org.

Network of Employers for Traffic Safety (NETS). Accessed online September 16, 2021 at https://www.trafficsafety.org.

Newnam, S., & VonSchuckmann, C. (2012). Identifying an appropriate driving behavior scale for the occupational driving context: The DBQ vs. the ODBQ. *Safety Science, 50*, 1268–1274.
Newnam, S., Greenslade, J., Newton, C., & Watson, B. (2011). Safety in occupational driving: Development of a driver behavior scale for the workplace context. *Applied Psychology, 60*(4), 576–599.
Nguyen, T., Vissoci, J. R., Joelson, T., Pesambili, M., Haglund, M., Gerardo, C. J., et al. (2018). Injury prevalence and safety habits of boda boda drivers in Moshi, Tanzania: A mixed methods study. *PLoS ONE, 13*(11), e0207570.
Occupational Safety and Health Administration (2012). Distracted driving. OSHA 3416-09R 2012 (accessed online September 16, 2021) https://www.osha.gov/Publications/3416distracted-driving-flyer.pdf.
Omidi, L., Mousavi, S., Moradi, G., & Taheri, F. (2021). Traffic climate, driver behavior and dangerous driving among taxi drivers. *International Journal of Occupational Safety Ergonomics*, 1–8.
Oyono, Y., Gjerde, H., Acha Asongalem, E., Kouomogne Nteungue, B. A., Bamuh, E., Deuwa Ngako, A., et al. (2021). Roadside surveys of drinking and driving in Cameroon. *Traffic Injury Prevention, 22*(5), 349–354.
Razmara, A., Aghamolaei, T., Madani, A., Hosseini, Z., & Zare, S. (2018a). Prediction of taxi drivers' safe-driving behaviors based on the theory of planned behavior: The role of habit. *Journal of Education and Health Promotion, 7*, 139.
Razmara, A., Aghamolaei, T., Madani, A., Hosseini, Z., & Zare, S. (2018b). Risky behaviors of taxi drivers in Bandar Abbas, Iran. *Electronic Physician, 10*(4), 6588–6595.
Robb, G., Sultana, S., Ameratunga, S., & Jackson, R. (2008). A systematic review of epidemiological studies investigating risk factors for work-related road traffic crashes and injuries. *Injury Prevention, 14*, 51–58.
Routley, V., Ozanne-Smith, J., Qin, Y., & Wu, M. (2009). Taxi driver seat belt wearing in Nanjing, China. *Journal of Safety Research, 40*, 449–454.
Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507–514.
Spielholz, P., Silverstein, B., Morgan, M., Checkoway, H., & Kaufman, J. (2001). Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. *Ergonomics, 44*, 588–613.
United Nations (2020). Resolution A/RES/74/299 "Improving global road safety" (accessed online September 24, 2021) https://undocs.org/en/A/RES/74/299.
Virginia Tech Transportation Institute (2018). Center for Truck and Bus Safety, Featured Projects. https://www. vtti.vt.edu/research/ctbs/index.php.
Wang, Y., Li, L., & Prato, C. G. (2019a). The relation between working conditions, aberrant driving behaviour and crash propensity among taxi drivers in China. *Accident Analysis and Prevention, 126*, 17–24.
Wang, Y., Zhang, Y., Li, L., & Liang, G. (2019b). Self-reports of workloads and aberrant driving behaviors as predictors of crash rate among taxi drivers: A cross-sectional study in China. *Traffic Injury Prevention, 20*(7), 738–743.
World Health Organization (2020). WHO global health estimates: Top 10 causes of death (accessed online September 24, 2021) https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.
Wu, C. Y. H., & Loo, B. P. Y. (2016). Motorcycle safety among motorcycle taxi drivers and nonoccupational motorcyclists in developing countries: A case study of Maoming, South China. *Traffic Injury Prevention, 17*(2), 170–175.

Cammie Chaumont Menendez, PhD, is an epidemiologist focused on conducting research examining the effectiveness of interventions and strategies designed to prevent or mitigate injuries in the workplace. Her interests are using Total Worker Health approaches to prevent injuries while promoting health and safety factors occurring both inside and outside the workplace, equity in implementing demonstrated occupational safety measures, and the health and safety of gig workers. She received her PhD from the NIOSH Education and Research Center in Injury Prevention at the University of Texas Health Sciences Center School of Public Health in Houston.

Benjamin C. Amick, III, PhD is Associate Dean for Research and Professor of Epidemiology at the University of Arkansas for Medical Sciences Fay W. Boozman College of Public Health. He has been involved in multiple health care, labor market and organizational program evaluations in the US, Canada and internationally. Currently he is collaborating with researchers at the University of Groningen Medical Center in the Netherlands on life course epidemiology research focused on cancer survivors and work role functioning and mental health and the school to work transition.

Timothy J. Walker, PhD is Assistant Professor of Health Promotion and Behavioral Sciences at the University of Texas Health Science Center at Houston, School of Public Health. His research interests include physical activity promotion, dissemination and implementation, and measure development and evaluation.

Richard Munoz is a Graduate Assistant and PhD candidate at Florida International University within the Robert Stempel College of Public Health & Social Work, pursuing his doctorate in Health Policy Management, with a specialization in Health Systems Research. He is currently involved with patient safety culture and risk management training within hospitals in Latin America, and volunteers with the Geriatric, Research, Education, and Clinical Center (GRECC) at the Department of Veteran Affairs Miami medical center in a multi-site, multi-VISN (Veterans Integrated Service Networks) survey project. His research interests are program evaluation, survey design, educational and psychological testing, and population-level geriatrics research.

Journal of Safety Research 82 (2022) 221–232

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Assessment of inequity in bicyclist crashes using bivariate Bayesian copulas

Bahar Dadashova [a,*], Eun Sug Park [a], Seyedeh Maryam Mousavi [a], Boya Dai [a], Rebecca Sanders [b]

[a] Texas A&M Transportation Institute, 1111 Rellis Pkwy, Bryan, TX 77807, United States
[b] Safe Streets Research & Consulting, 2641 SE Harrison Street, Portland, OR 97214, United States

ARTICLE INFO

Article history:
Received 24 April 2021
Received in revised form 27 November 2021
Accepted 25 May 2022
Available online 4 June 2022

Keywords:
Active transport
Bicyclist crashes
Bivariate Bayesian copulas
Equity

ABSTRACT

Introduction: Physical activity associated with active transport modes such as bicycling has major health benefits and can help to reduce health concerns related to sedentary lifestyles, such as cardiovascular disease, Type II diabetes, and obesity, as well as risks of colon and breast cancer, high blood pressure, lipid disorders, osteoporosis, depression, and anxiety. However, as a vulnerable user group, bicyclists experience negative health impacts of transportation policies and infrastructure, such as traffic crashes and exposure to air and noise pollution that is disproportionately distributed within low-income and underserved areas. *Method:* This study used aggregated (block-group) bicyclist crash data from Harris County, Texas, to analyze how various equity measures are associated with both fatal and injury (FI) and no injury (property damage only) bicyclist crashes that occurred from 2010 to 2017. We used Bayesian bivariate copula-based random effects regression analysis to evaluate these associations. In contrast to more traditional univariate analysis, this novel methodology can consider the effects of factors of interest across different severity levels or crash types to fully understand their effects and how they may differ across categories. *Results:* The analysis results indicate that the bicyclist exposure, vehicle exposure, population demographics, population density, the percentage of African-Americans, and households below the poverty level are associated with both FI and PDO bicyclist crashes. *Conclusions:* Although more location and context-specific analyses are required, this study's overall results once again conform with the findings and assumptions in bicycling safety literature that the low-income and racially diverse communities are prone to experience more bicyclist crashes. *Practical Applications:* The findings of this study may have implications for future transportation and planning policies. These findings can be used to guide the policies and strategies targeting the elimination of inequity in transportation-related health concerns.

1. Introduction

Bicycling, as an active mode of transport, is associated with many health and environmental benefits. Physical activity associated with bicycling can help to decrease health concerns emerging from sedentary lifestyles, such as cardiovascular disease, Type II diabetes, and obesity, as well as risks of colon and breast cancer, high blood pressure, lipid disorders, osteoporosis, depression, and anxiety (World Health Organization [WHO], 2002; Celis-Morales et al., 2017; Flint & Cummins, 2016). However, bicyclists are vulnerable road users at a higher risk of negative health impacts resulting from transportation policies and infrastructure, and this risk is exacerbated in low-income and ethnically-diverse communities (Barajas, 2020; Doran et al., 2021; Huang et al., 2010; Noland & Laham, 2018). Although rigorous research is required to disentangle the underlying causes of this disparity, the lack of access to active transport infrastructure has been documented as one of the major causes (Collins et al., 2012; Gordon-Larsen et al., 2006; Zhu et al., 2011). Low-income and minority communities have historically been co-located with high-capacity roadways (i.e., interstates and freeways) that carry high traffic volumes. By design, these roadway facilities are high-speed and allocate little to no space to bicyclists and pedestrians, hence impeding foot and other non-motorized traffic, exposing them to vehicular traffic, and isolating these users from the rest of the city or community (Avila, 2014). Further, low-quality roadway pedestrian and bicycle facilities increase perceived safety risks, thereby discouraging vulnerable populations such as older adults, children, women, and ethnic minorities from using active transport (Agyeman & Doran, 2021).

The disparity in active transport and its health impacts persists despite efforts to increase bicycling and improve bicyclist safety in

* Corresponding author.
E-mail address: B-Dadashova@tti.tamu.edu (B. Dadashova).

https://doi.org/10.1016/j.jsr.2022.05.013
0022-4375/

the US (Le et al., 2019). Fatal bicyclist crashes increased 36% from 2010 to 2019 (National Highway Traffic Safety Administration [NHTSA], 2020). High Injury/Crash Networks (HINs/HCNs) developed as part of Vision Zero plans often find that crashes and fatalities occur disproportionately near ''Communities of Concern'', which are typically lower-income and more diverse (Rebentisch et al., 2019). However, HIN analysis rarely goes deeper than roadway mapping, and it is not clear how these crashes are distributed among communities with diverse social, demographic and economic backgrounds. In turn, this gap in understanding may impact future data-driven transportation and planning policies and hinder the improvement of bicyclist safety on a system level.

One of the major challenges in implementing the data-driven methods for studying bicyclist crashes and exploring bicyclist crash-contributing factors is the scarcity of the data. Bicyclist crashes are frequently under-reported, and this is particularly true for single bicyclist crashes and crashes that are below a certain dollar value of property damage (Lopez et al., 2022). Small sample sizes of reported crashes can result in estimation bias by producing non-significant and biased estimates and affect the model performance and goodness of fit statistics, although statistical models based on Bayesian inference can help address these concerns (McNeish, 2016). Another limitation of broadly used crash prediction models such as Poisson and negative binomial models is that they usually do not account for the unobserved heterogeneity resulting from spatial and temporal instability in the crash data. Although this limitation has been addressed in the safety literature using the random parameter models (Mannering, 2018), these models work well with larger sample sizes, which is not the case for under-reported bicyclist crashes. Last but not least, the broadly used univariate crash prediction models do not account for the correlation between the response variables (e.g., property damage only and fatal and injury crashes).

To address these limitations in the bicyclist crash data, we used novel copula-based bivariate Bayesian regression model (Park et al., 2020) to evaluate the association of equity indicators with fatal and injury (FI) and no injury (also referred to as property damage only [PDO]) bicyclist crashes in Harris County, Texas, US. The copula-based multivariate models can handle both site-specific and outcome-specific unobserved heterogeneity in crash data and are able to address the correlation between the random response variables. Therefore, they have been frequently used in the safety literature as one of the effective approaches to model the multivariate outcomes (see Bhat & Sener, 2009; Zou et al., 2018). However, they have not been applied to scarce data such as bicyclist crashes previously. In this paper, we use the Bayesian inference to address the scarcity of the crash data and the copula-based bivariate modeling approach to address the unobserved heterogeneity and correlation between two random outcomes (i.e., FI and PDO bicyclist crashes). We measure equity by sociodemographic (e.g., age, ethnicity, and gender) and economic factors (e.g., household income and poverty), based on various sociodemographic and economic variables from the Centers for Disease Control and Prevention's (CDC's) social vulnerability index (SVI) (available at: https://www.atsdr.cdc.gov/placeandhealth/svi/index.html). The findings of this paper will shed light on crash fatality and injury disparities in bicycling to be able to more proactively address those disparities in the future.

2. Literature review

Our work builds on the findings of prior studies that have examined outcomes related to bicycling equity (or inequity), as described below.

2.1. Equity indicators

2.1.1. Demographic and socioeconomic factors

Demographic and socioeconomic factors such as race and income have been found to be associated with both frequency and severity of bicyclist crashes. Marshall et al. (2017) discovered disparities in road fatalities along racial and ethnic lines, particularly for pedestrians and bicyclists in predominantly black or Hispanic neighborhoods. Similarly, Lindsey et al. (2019) found that mean pedestrian and bicycle crash risk is higher in neighborhoods with lower incomes and ethnically diverse communities. Wu (2020) found a negative relationship between area median income and bicyclist and pedestrian crash rates.

Bicycle crash statistics show that bicyclists belonging to a certain age or gender were involved in more crashes than others. According to the National Center for Statistics and Analysis (2019), the average age of bicyclists involved in fatal crashes increased from 41 to 47 between 2008 and 2017; in 2017, the largest number of bicyclist fatalities was in the 50–54 age group. Additionally, the majority of bicyclists killed (89 percent) were males, and the population-based bicyclist fatality rate was eight times higher for males than for females. These findings are similar to those from a review of 20 bicycle safety papers published prior to 2015 (Vanparijs et al., 2015). In addition, studies suggested that age is positively related to increased crash severity, with senior cyclists having an elevated risk of serious or fatal injuries (Chen & Shen, 2016; Kröyer, 2015).

Though bicycling can benefit and improve the health of disabled people, disability is under-researched in bicycling studies. Many issues need to be resolved before bicycle infrastructure is accessible to (and usable by) disabled people (Clayton et al., 2017). Additional research is needed in this area.

2.1.2. Bicyclist exposure

Bicyclist exposure is a measure of the number of potential opportunities for a bicycle-involved crash to occur. Bicyclist volume and motor vehicle volume are the major indicators of bicyclist exposure. Exposure has been defined based on direct counts, population, hours of travel, miles of travel, and others. Bicyclist exposure can also be considered as a measure of inequity, in that people belonging to certain demographic groups may be more likely to bicycle. For example, Dadashova and Griffin (2020) found that household income and gender were important predictors of bicycle use in a block group. Income and gender should be considered as indicators of equity when assessing bicyclist safety.

The effect of bicyclist exposure on bicycle safety has been widely discussed in the literature, although the direction of the impact is unclear. Studies tend to agree that increasing the number of bicyclists on the road can help reduce the number and severity of crashes, a concept known as ''safety in numbers.'' A study by Marshall and Garrick (2011) suggests that a certain threshold of bicyclist volumes would compel drivers to drive slower, resulting in a safety for all road users. Other studies have also agreed that increasing the number of bicyclists may help to improve overall safety (Jacobsen, 2015; Kaplan & Giacomo Prato, 2015; Nordback & Marshall, 2011).

However, the role of bicycle infrastructure in contributing to ''safety in numbers'' is unclear. I.e., if changes in exposure are related to the installation of bicycle facilities, findings about ''safety in numbers'' may not hold in the absence of bicycling infrastructure. This question is particularly concerning given that the low-income and ethnically-diverse communities usually lack this type of infrastructure, as discussed below.

2.1.3. Bicycle infrastructure

The inequitable crash risk associated with different demographics and socioeconomic status (described in Section 1.1.1) may be largely attributable to the uneven distribution of bicycle infrastructure. Goodman et al. (2013) suggested that improved infrastructure might be used more often by residents with higher education or income level since the infrastructure is built in economically advantaged neighborhoods. Similarly, Smith et al. (2017) stated that there was some indication that infrastructure improvements might predominantly benefit socioeconomically-advantaged groups. These findings are supported by the work of Flanagan et al. (2016) and Hirsch et al. (2017), which revealed that in a combined total of five US cities, bike lanes tend to be built in socioeconomically advantaged areas.

In contrast, there has been an inadequate distribution of bicycle infrastructure in disadvantaged communities. A review of academic literature related to equity in bicycling revealed that the groups that have suffered disparities in access to satisfactory bicycle facilities are people with low incomes, immigrants and people of color, women, seniors, and children (Doran et al., 2021). Osmonson (2017) explored the relationship between an area's equity variables – the proportion of a Census Block Group population that was a racial minority, below the federal poverty line, under the age of 18, and 65 and older – and the area's bicycle infrastructure density. The results showed a lot of service gaps where many diverse and historically disadvantaged people live. Rebentisch et al. (2019) found that while higher-income and gentrified areas had better access to protected bicycle infrastructure, low-income and ethnically-diverse communities did not have access to such infrastructure; the latter communities were also overrepresented in fatal and severe injury crashes. Braun et al. (2019) examined cross-sectional associations between bike lanes and sociodemographic characteristics at the block group level for 22 large US cities. The findings further confirmed disparities in bike lane access.

2.2. Modeling techniques

In this paper we are conducting statistical analysis (using copula-based regression model) to assess the impact of various socio-economic factors on bicyclist crashes while accounting for correlation across FI and PDO crashes as well as unobserved heterogeneity across sites by random effects. Since the analysis performed in the paper are not spatial analysis but rather cross-sectional regression analysis (where an observation unit is a block group), the literature review does not discuss the spatial models.

2.2.1. Macro-modeling of bicyclist crashes

Previous studies have used various types of statistical models to explore macro-level bicyclist crash-contributing factors. These studies have used TAZ and other Census levels for studying bicyclist crashes. Amoh-Gyimah et al.(2016) conducted cross-comparison of various estimation methods such as non-spatial negative binomial, random parameter negative binomial, and Poisson-Gamma CAR (conditional autoregressive) for modeling bicyclist and pedestrian crashes that occurred at a Statistical Area level 2 (SA2), per the definition of Australian 2011 census classification. This study found that vehicle kilometers traveled (VKT), population, percentage of commuters bicycling or walking to work, percentage of households without motor vehicles and mixed land use have a significant and positive correlation with the number of pedestrian and bicyclist crashes.

Saha et al. (2018) also used CAR models within the Bayesian framework to evaluate bicyclist crashes at a block group level in Florida. This study found that population, daily vehicle miles traveled, age cohorts, household automobile ownership, the density of urban roads by functional class, bicycle trip miles, and bicycle trip numbers had increasing effects on both the total and fatal-and-severe crash models.

Guo et al. (2018) compared Poisson lognormal model (PLN), random intercepts PLN model (RIPLN), random parameters PLN model (RPPLN), and spatial PLN models to explore bicyclist crashes at the level of traffic analysis zones (TAZs). They found a positive association between bicyclist crashes and bike and vehicle exposure measures, households, commercial area density, and signal density. In contrast, bicyclist crashes were negatively associated with bike network indicators such as average edge (i.e., segment) length, average zonal slope, and off-street bike links.

2.2.2. Multivariate modeling of bicyclist crashes

Bicyclist crash data are usually collected by different severity levels or crash types, which are often correlated. In evaluating the socioeconomic factors associated with bicyclist crashes, it is important to consider the effects of those factors across different severity levels or crash types to fully understand their impacts, as they may vary with different severity levels or crash types.

In addition, jointly analyzing multiple severity levels or crash types, rather than conducting separate analyses for each severity level or crash type, leads to a more precise estimation of factor effects by borrowing information from other severity levels or crash types. Although multivariate models have been used to analyze vehicle crashes (see Park & Lord, 2007; Mannering et al., 2016), few studies have applied these models in bicyclist crash analysis, given the relative scarcity of bicycle crashes and the known limitations in bicyclist crash data, as discussed in the Introduction. Lee and Abdel-Aty (2018) implemented multivariate Bayesian Poisson lognormal CAR models to identify the bicyclist crash-contributing factors for (1) the number of bicyclist crashes in the crash location's ZIP code and (2) the number of crash-involved bicyclists in their residence's ZIP.

Recently, Park et al. (2020) proposed a copula-based Bayesian approach (copula-based random effects regression models) for modeling general correlation structures (regardless of correlations being positive or negative) in multivariate crash data. A copula is a flexible probabilistic tool for modeling the joint distribution of a random vector in two separate steps: selection of the marginal distribution and specification of a copula function which captures the correlation structure among the vector components. While most of the previous copula-based studies on crash count data predominantly used Archimedean copulas (which assume the same correlation structure for all pairs of dependent variables and allow only positive correlations), Park et al. (2020) introduced a more general multivariate approach using a Gaussian copula which offers great flexibility in modeling the dependence structure in crash counts. Overdispersion and general correlation structures, including both positive and negative correlations in multivariate crash counts, can easily be accounted for by this approach.

3. Data and methodology

3.1. Study area and data sources

In this study, we use bicyclist crash data from Harris County, Texas (Fig. 1), which is divided into 2,144 Census block groups with significant disparities in racial and ethnic groups.

Crash data were obtained from the Crash Record Information System (CRIS) database of Texas Department of Transportation (TxDOT) and roadway data were obtained from roadway inventory (RHiNO) of TxDOT (available at https://www.txdot.gov/inside-txdot/division/transportation-planning/roadway-inventory.html). To obtain a sufficient sample size, we aggregated the bicyclist crashes from 2010 to 2017. Bicyclist crash data are usually collected by different severity levels or crash types, which are often correlated. In evaluating the socioeconomic factors associated with

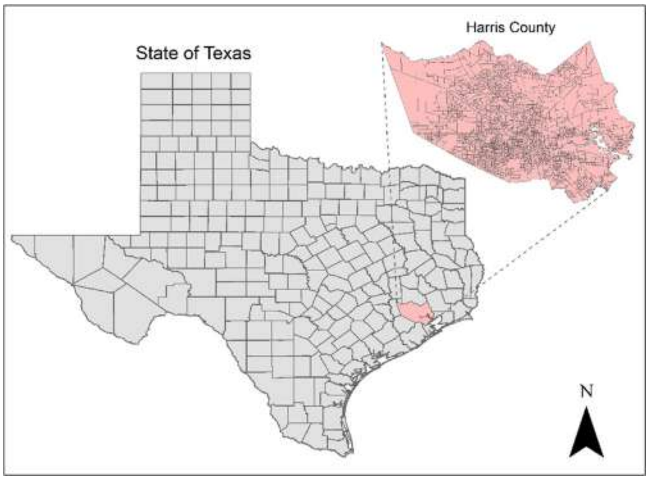


Fig. 1. Harris County, Texas.

bicyclist crashes, it is important to consider the effects of those factors across different severity levels or crash types to fully understand their effects, given that their effects may vary with different severity levels or crash types. While this study analyzed the impact of the equity indicators on bicyclist crashes by severity levels, we did not group bicyclist crashes into types or locations (e.g., intersection vs. segment) due to previously discussed limitations in the data. Because of the limitations of the data, we did not account for the bicycling infrastructure, such as the total length of bicycle lanes (such data is not readily available). Socioeconomic and demographic factors were collected from the American Community Survey (ACS, available at https://www.census.gov/programs-surveys/acs). Finally, bicyclist volumes were estimated using Strava data, based on the study by Dadashova et al. (2020). The ACS and Strava data used in the analysis were available from 2016 to 2017.

The three databases were integrated using ArcGIS tools. The geographic vector data used in this study are point (i.e., crashes), line (i.e., bicyclist exposure and traffic volume per segment and roadway design characteristics of the segment), and polygon (i.e., demographic and socioeconomic factors per ACS block group) data. We first joined point and line data to develop the final database by assigning every crash to the nearest roadway segment. We then joined the line and polygon data by assigning all the polygon attributes to the point that falls inside the polygon's boundaries. Finally, the line data were aggregated per polygons (e.g., total traffic and bicyclist volumes). The socioeconomic and demographic characteristics of these block groups remained unchanged, while we aggregated the number of crashes, bicyclist, and traffic volumes to estimate the vehicle miles traveled (VMT) and bicycle miles traveled (BMT) of the area as discussed below.

3.1.1. Bicycle miles traveled

We use the vehicle and bicycle miles traveled (BMT) to measure bicyclist exposure. In general, macro-level studies use traveled miles as the measure of exposure. Since bicyclist volumes are not readily available, we used direct-demand models developed in Dadashova et al. (2020). This study leverages the crowdsourced Strava data estimate of the average annual daily bicycles (AADB) by developing direct-demand models. The direct-demand models developed in this study account for the income level of the block group and roadway facility type. After estimating the AADB per roadway segment, we calculated the BMT using the following formulae:

$$BMT_l = \frac{AADB_l \times Length_l \times 365}{1,000,000} \tag{1}$$

where, BMT_l is the bicycle miles traveled on segment or link l, and $Length_l$ is the length of Strava segment. The BMT for the block group was calculated by summing up all the BMT_l of segments that are within the boundaries of a block group:

$$BMT_b = \sum_{l=1}^{L} BMT_l \tag{2}$$

where BMT_b is the BMT of a block group, and is the number of segments falling within the boundaries of block group.

3.1.2. Vehicle miles traveled

Similarly, vehicle volumes were measured using the vehicle miles traveled (VMT) which is calculated as:

$$VMT_l = \frac{AADT_l \times Length_l \times 365}{1,000,000} \tag{3}$$

where $AADT_l$ is the average annual daily traffic (AADT) of a segment l. AADT data from 2010 to 2018 were obtained from TxDOT's roadway inventory shapefile. VMT data were also summed up to calculate the VMT of a block group.

$$VMT_b = \sum_{l=1}^{L} VMT_l \tag{4}$$

3.2. Bayesian bivariate copula-based random effects regression

We employed the methodology developed by Park et al. (2020) to evaluate the association of equity indicators with bicyclist crash measures. The modeling framework for copula-based random-effects models proposed by Park et al. (2020) is briefly summarized below for bivariate crash counts. Interested readers may refer to Park et al. (2020) for a general case of multivariate crash counts as well as for more details.

Let $\boldsymbol{y_i} = (y_i^1, \ y_i^2)$ denote a bivariate crash count at site i ($i = 1, \ \cdots, \ I$). That is, y_i^j is the number of crashes of severity j occurring at site i. Let K be the number of covariates and $X_i = (1, X_{i1}, \ \cdots, \ X_{iK})$ be a $(K+1)$-dimensional row vector of covariates. Let $\boldsymbol{\beta^j} = \left(\beta_0^j, \ \beta_1^j, \ \cdots, \ \beta_K^j\right)'$ denote the $(K+1)$-dimensional column vector of the regression coefficients for the crash count of the jth severity level. Let $\boldsymbol{\upsilon_i} = (\upsilon_i^1, \ \upsilon_i^2)$ denote a vector of site and outcome-specific random effects corresponding to site i. Suppose that, conditional on υ_i^j and $\boldsymbol{\beta^j}$, the crash count of severity j at site i, y_i^j, follows a Poisson distribution with mean μ_i^j, i.e.,

$$y_i^j \middle| \upsilon_i^j, \boldsymbol{\beta^j} \sim Poisson\left(\mu_i^j\right) \tag{5}$$

where,

$$\mu_i^j = \upsilon_i^j \exp\left(X_i \boldsymbol{\beta^j}\right) \tag{6}$$

Here, υ_i^j handles (incorporates) both site- and outcome-specific heterogeneity. Let $\lambda_i^j = \exp\left(X_i \boldsymbol{\beta^j}\right)$.

For random effects $\boldsymbol{\upsilon_i}$, the Gaussian copula (which is easy to handle while being able to incorporate a general correlation structure) is assumed as follows:

$$\boldsymbol{\upsilon_i} = (\upsilon_i^1, \ \upsilon_i^2) \sim C_R^G(F_1, F_2) \tag{7}$$

where, C_R^G is a Gaussian copula and F_j is the marginal cdf of υ_i^j. The joint pdf of $\boldsymbol{\upsilon_i}$ is given as

$$h(\boldsymbol{\upsilon_i}) = \exp\left[-\frac{1}{2}z_i\left(R^{-1}-I\right)z_i'\right]\prod_{j=1}^{J}f_j\left(\upsilon_i^j\right) \tag{8}$$

where R is a correlation matrix, $z_i = (z_i^1, z_i^2)$, $z_i^j = \Phi^{-1}\left(F_j\left(\upsilon_i^j\right)\right)$, and f_j is the marginal pdf of υ_i^j. The marginal distribution of υ_i^j can be assumed to be a gamma distribution with mean 1 and variance $1/\eta^j$, i.e. $\upsilon_i^j \sim Gamma(\eta^j,\ 1/\eta^j)$, so that the marginal distribution of y_i^j is given as a negative binomial (NB) distribution with mean λ_i^j and variance $\lambda_i^j\left[1+\lambda_i^j/\eta^j\right]$.

For prior distributions of parameters $(\boldsymbol{\beta^1}, \boldsymbol{\beta^2}, \eta^1, \eta^2, R)$, the following distributions may be assumed:

$$\boldsymbol{\beta^j} \sim \mathrm{N}_{K+1}\left(b_0^j, B_0^{-1}\right),\ j = 1,2 \tag{9}$$

where,

$$\eta^j \sim Gamma\left(c_0^j, d_0^j\right),\ j = 1,2, \tag{10}$$

$$\pi(R) \propto I\{R_{jk} : R_{jk} = 1(j=k), |R_{jk}| < 1(j \neq k),\ R \text{ is positive definite}\}$$

$$j = 1,2,\quad k = 1,2 \tag{11}$$

where, $N_{K+1}\left(b_0^j, B_0^{-1}\right)$ is the (K + 1)-variate normal distribution with mean vector b_0^j and precision matrix B_0, $Gamma\left(c_0^j, d_0^j\right)$ is the Gamma distribution with a shape parameter c_0^j and a scale parameter d_0^j, having mean $c_0^j d_0^j$, and π denotes a prior distribution for R. Here, $\left(b_0^j, B_0^{-1}, c_0^j, d_0^j\right), j = 1,2$, are prespecified hyperparameters.

Estimation of the above model can be implemented by Markov Chain Monte Carlo (MCMC) methods (see, e.g., Gilks, Richardson, & Spiegelhalter, 1996; Gelman et al., 2013) using the algorithm given in Park et al. (2020), which simulates from the conditional distributions for $(\upsilon_i^1,\ \upsilon_i^2)$,$(\boldsymbol{\beta^1},\ \boldsymbol{\beta^2})$,$(\eta^1,\ \eta^2)$, and R iteratively.

4. Results and discussion

After combining the datasets from various sources, we conducted visual exploratory data analysis (EDA) and modeled the bicyclist crashes using the Bayesian bivariate copula-based random-effects regression approach. In this section, we provide a discussion of data analysis results.

4.1. Exploratory data analysis

4.1.1. Bicyclist crashes

In the CRIS database, bicyclist crashes are identified by selecting single-vehicle crashes and harmful event types. Bicycle crash data in Harris County were aggregated at the block group level based on crash severity. In this case, crash severity was classified as FI and PDO.

There were 4,007 FI and 3,785 PDO bicyclist crashes from 2010 to 2017 in the 2,144 block groups. Table 1 contains summary statistics of bivariate crash frequency by severity. Note that the unit of crash frequency in the table is the number of crashes per block group for ten years (from 2010 to 2017).

Fig. 2 shows the distribution of FI and PDO crashes per block group; blue indicates lower numbers, while red indicates higher numbers. As can be observed, neither FI nor PDO bicyclist crashes are equally distributed across the block groups, and certain block groups apparently have experienced higher numTexas bers of severe and/or PDO crashes.

4.1.2. Bicyclist exposure

Table 2 contains summary statistics of BMT and VMT. Note that on average, VMT is almost 165 times higher than the BMT, indicating that a very small percentage of roadway users are bicyclists.

Fig. 3 depicts the distribution of BMT and VMT per block group; the darker areas indicate higher numbers than the lighter areas. As observed, bicyclist exposure seems to be higher in suburban and downtown areas. VMT on the other hand, seems to be more uniformly distributed; very few block groups stand out. We can also observe a spatial difference between BMT and VMT, highlighting the importance of accounting for both exposure indicators in bicyclist crash analysis.

4.1.3. Socioeconomic and demographic factors

We used ACS 2016 data that summarizes 2010 to 201 estimates of various socioeconomic and demographic factors. We identified the list of variables based on the results of the literature and grouped them based on their context, such as (1) population demographics; (2) income; (3) household size; and (4) education. Because the ACS data is estimated using the population survey, in order to avoid uncertainty, we used the original ACS variables as much as possible. However, since some of the variables were correlated and/or their ACS definitions were obscure, we performed variables transformation for some variables to better interpret them and to improve the model performance. For example, we observed that using percentage instead of total numbers to explore the racial differences between the block groups helped improve the model performance. The percentage transformation of population demographics did not improve the model performance; hence they were left unchanged. We also did not define new variables, such as the age group of the total population, since the ACS does not provide this data. The log-transformation of continuous variables (i.e., BMT and VMT) was also considered. Table 3 contains summary statistics of original socioeconomic and demographic factors considered for the analysis.

Fig. 4 depicts some of the socioeconomic and demographic properties of block groups in Harris county; in all cases, darker areas indicate higher numbers than lighter ones. As observed, the population density is almost uniform in Harris county, with relatively higher density in central areas. Also, there is no significant difference between the adult male and female populations. However, there is a stark difference between the percentage of white and African American populations per block group. As observed, the white population is mainly concentrated in suburbs (which also had higher estimated BMT, see Fig. 3), while the African American population is mainly in central parts. We can also observe that median household income is higher in suburbs and south-west areas.

4.2. Bivariate copula-based random effects regression analysis

We conducted bivariate copula-based random effects regression analysis to explore the association of bicyclist and vehicle exposure, socioeconomic and demographic factors with FI and PDO bicyclist crashes. Recall that this analysis uses aggregated crash data per block group in order to produce statistically meaningful results.

We used copula-based random-effects regression models presented in Eqs. (5)–(11) to analyze the bicycle crash count data of two different severity levels (FI and PDO) in Harris county to account for a high correlation between FI and PDO crashes in estimation. We first selected the initial set of variables to be included in the regression model to assess the association of socioeconomic and demographic factors with bicyclist crashes in Harris County through extensive exploratory data analysis and correlation analysis. We then went through a screening process of significant vari-

Table 1
Summary Statistics of Harris County Bike Crash Data per Block Group, 2010–2017.

Dependent Variables	Total	Min	Max	Mean	Std. Dev.
Fatal and Injury (Y1)	4,007	0	23	1.87	5.23
Property Damage Only (Y2)	3,785	0	21	1.77	4.82

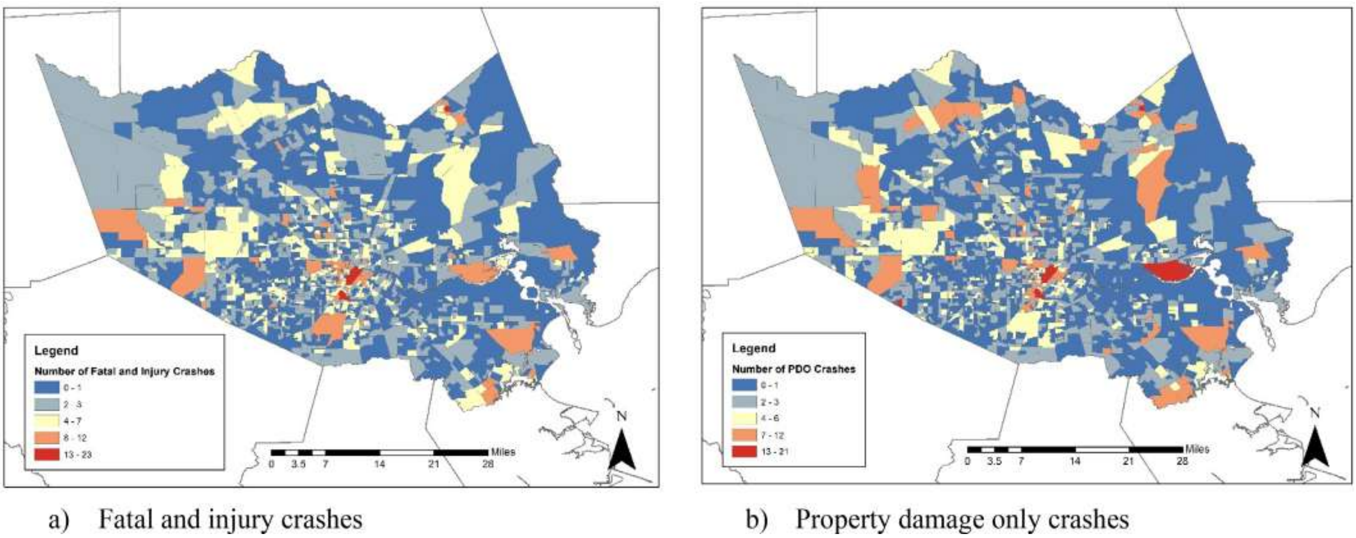


Fig. 2. Number of FI and PDO bicyclist crashes per Harris County block group.

Table 2
Summary Statistics of Bicyclist Exposure per Harris County Block Group (Estimated).

Variables	Min	Max	Mean	Std. Dev.
Bicycle Miles Traveled (BMT)	0.005	3.06	0.21	0.27
Vehicle Miles Traveled (VMT)	0.0003	1,016.14	34.46	79.74

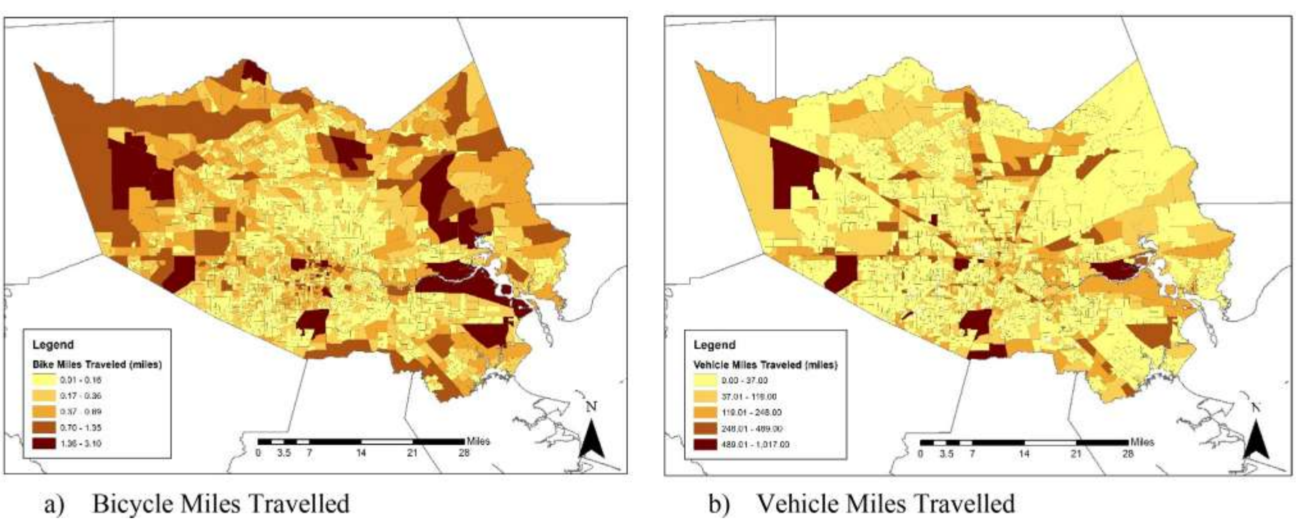


Fig. 3. Bicyclist exposure per block group.

ables that conventional multiple regression studies do in the preliminary analysis by implementing variable selection procedures based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) as well as considering correlation among predictors. We performed several Markov Chain Monte Carlo (MCMC) runs of copula models with different sets of variables and tuning parameter values and selected the model with good convergence results because non-convergence or slow mixing of MCMC often indicates inconsistency between the model and the data. The estimation of model parameters was implemented by the MCMC. The following hyperparameters were used for generating MCMC samples: $b_0^1 = [0.4 \quad \mathbf{0}_{1\times K}]'$, $b_0^2 = [0.3 \quad \mathbf{0}_{1\times K}]'$,

Table 3
Summary statistics of Harris county socio-economic and demographic factors, 2010–2015.

Variables	Min	Max	Mean	Std. Dev.
Population Density (per square mile)	4.16	66,674.40	6,568.64	6,332.08
Total Population	9	20,031	2068.22	1616.92
Number of Total Male Population	9	10,210	1,028.61	801.19
Number of Total Female Population	0	9,821	1,039.61	836.42
Number of Adult Males (Age 18–39)	0	3,985	348.46	305.50
Number of Middle-Aged Male (Age 40–65)	0	2,641	309.55	247.02
Number of Senior Male (Age > 65)	0	624	83.73	69.59
Number of Adult Female (Age 18–39)	0	3,379	338.34	301.68
Number of Middle-Aged Female (Age 40–65)	0	2,973	318.04	264.53
Number of Senior Female (Age > 65)	0	971	106.95	86.37
Percentage of White Population	0	1	0.64	0.24
Percentage of African American Population	0	1	0.19	0.23
Median Household Income	9,015	25,0001	63,371.34	42,050.07
Number of Households Below Poverty (Income is $10,000.00-$20,000.00)	0	796	116.17	101.68
Number of Low-Income Households (Income is $25,000.00-$40,000.00)	0	1,238	145.30	113.64
Number of Middle-Income Households (Income is $45,000.00-$125,000.00)	0	3,340	325.53	285.26
Number of High-Income Households (Income is >$150,000.00)	0	2,679	129.54	206.05
Number of Small Sized Household (Less than 2 family members)	0	1,551	171.84	138.48
Number of Medium Sized Household (Less than 3–5 family members)	0	3,096	277.46	265.43
Number of Large Sized Household (More than 6 family members)	0	581	38.51	50.32
Number of High-School Graduates	0	4,069	370.72	306.65
Number of College Graduates	0	7,851	673.39	703.02

$\boldsymbol{B}_0^{-1} = 3 \times \boldsymbol{I}_{(K+1)\times(K+1)}$, $[c_0^1, c_0^2] = [1, 1]$, and $\left[d_0^1, d_0^2\right] = [1.4, 1.6]$. The posterior samples are collected for 100,000 iterations, after the first 50,000 draws are discarded, subsampling every 50th value (resulting in 2,000 samples).

After exploring various models containing different combinations of variables chosen by variable selection procedures, as well as considering correlation among them in the preliminary analysis, a model with the following list of variables seems to be the best in terms of interpretability and the convergence of the MCMC simulations: BMT, VMT, population density, number of adult males, number of adult females, percentage of African American population, and number of households below poverty. Table 4 provides the estimates (point estimates, exponentiated transformations, and standard deviations) of the regression coefficients ($\boldsymbol{\beta}^1$, $\boldsymbol{\beta}^2$) for the variables of the final model.

We calculated the RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and MPB (Mean Prediction Bias), to check the accuracy of the estimated models:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (12)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (13)$$

$$MPB = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i) \quad (14)$$

where N is the number of observations, y_i is the observed crash count, and $\hat{y}_i$ is the predicted crash count. Lower values of all three statistics are associated with high performance.

As observed in Table 4, the errors are very low, indicating that the estimated models better explain the block-group level FI and PDO crashes in Harris county.

Table 5 contains the posterior mean (and posterior standard deviation in parenthesis) of the correlation matrix of the random effects. It can be seen that random effects of fatal and injury crashes and PDO crashes are very highly correlated for these block group level bike crash data.

We also conducted a prediction analysis to check the disaggregate level model accuracy further. Fig. 5 presents the plots of observed crash frequencies (denoted in blue dots) along with predicted crash frequencies (denoted in red lines) at block groups $i = 1, \ldots,$ 2,144. The plots show that the copula-based random effects regression model leads to reasonable predictions of bike crashes for the block groups in Harris County.

4.3. Discussion

The estimation results in Table 4 show that the overall magnitude of effects is quite small. This may be due to the fact that, even at the aggregated level, very few bicyclist crashes are reported to the police. It may also be explained by the lack of a strong bicycling culture in Harris County. For example, the City of Houston has an overall ridership score of 1.6 according to the People for Bikes city rating (available at: https://cityratings.peopleforbikes.org/compare/?c=1443). Although these coefficients provide insight into how various exposure and socioeconomic factors affect bicyclist crashes (i.e., the direction of the effect), the magnitude of the impacts should not be taken at face value (i.e., the coefficients may change in the face of previously unaccounted data).

Even with the limitations of crash data, however, BMT, VMT, number of adult males, and number of households below poverty are significantly positively associated with bicyclist crashes, while population density was significantly negatively (i.e., inversely related to crashes) associated with bicyclist crashes. Additionally, although not statistically significant, more bike crashes seem to be associated with block groups with a higher percentage of African Americans and the number of adult females in the population.

Among the exposure variables, the findings show that increasing BMT is associated with 1.9 and 2.2 percent increase in FI and PDO bicyclist crashes. This finding contradicts the "safety in numbers" principle and shows that increased exposure is associated with more bicyclist crashes, although the magnitude of the effect is higher for the PDO crashes. We urge caution with this interpretation, as our lack of accounting for bicycle facility presence means it is not clear whether these crashes occurred at bicycle facilities or outside. Moreover, our point-in-time exposure estimates may be biased and do not account for how bicycle facilities may have changed cycling safety over the eight-year crash period. VMT is also found to be associated with more bicyclist crashes. However, the magnitude of this impact is less than that of the BMT; 0.9 percent increase in FI and PDO bicyclist crashes as the VMT increases by one unit.

The next exposure variable found to be significantly associated with bicyclist crashes was population density. The estimation

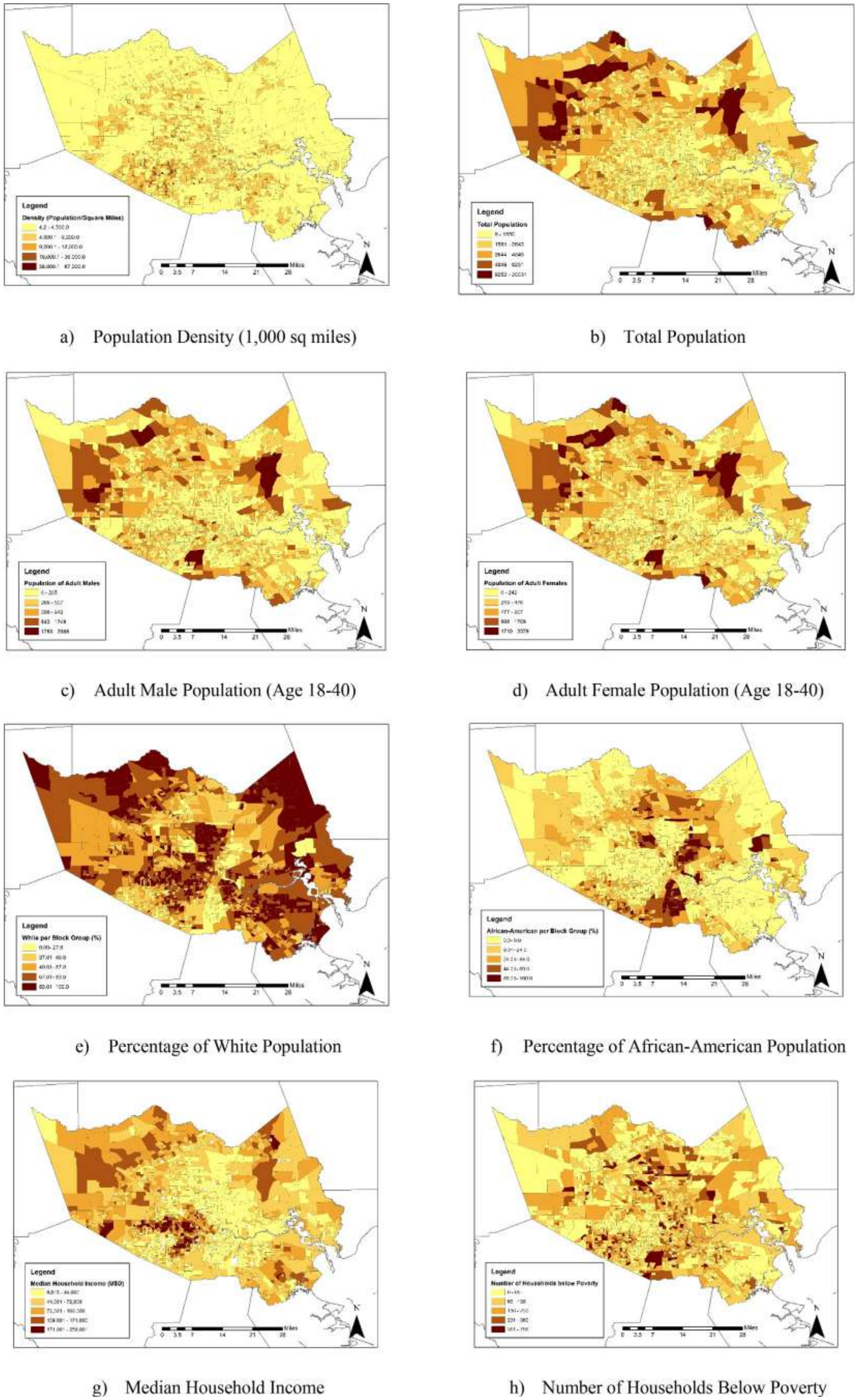


Fig. 4. Socioeconomic & demographic properties of Harris County block groups.

Table 4
Estimates of regression coefficients of copula-based random effects regression models.

Model Parameter	Fatal and Injury					Property Damage Only				
	Posterior Mean	Exp. (Posterior Mean)	Posterior Std. Dev.	95% Posterior Interval		Posterior Mean	Exp. (Posterior Mean)	Posterior Std. Dev.	95% Posterior Interval	
				Lower Limit	Upper Limit				Lower Limit	Upper Limit
Parameter Estimates										
Constant	**0.1855***	1.2%	0.0651	0.0527	0.3128	**0.1432**	1.2%	0.0643	0.0144	0.2682
Bicycle Miles Travelled	**0.6684**	1.9%	0.1307	0.4229	0.9460	**0.7701**	2.2%	0.1321	0.5213	1.0482
Vehicle Miles Travelled	**0.0016**	1.1%	0.0004	0.0009	0.0024	**0.0012**	1.1%	0.0004	0.0005	0.0020
Population Density	**−0.0172**	0.9%	0.0061	−0.0289	−0.0052	**−0.0216**	0.9%	0.0061	−0.0332	−0.0098
Number of Adult Male (Age 18–39)	**0.0005**	1.1%	0.0002	0.0002	0.0009	0.0005	1.5%	0.0002	0.0001	0.0009
Number of Adult Female (Age 18–39)	−0.0002	0.9%	0.0002	−0.0006	0.0002	−0.0001	0.9%	0.0002	−0.0004	0.0003
Percentage of African American Proportion	0.1427	1.2%	0.1241	−0.0967	0.3913	0.0346	1.1%	0.1259	−0.2129	0.2818
Number of Households Below Poverty	**0.0013**	1.1%	0.0003	0.0007	0.0019	**0.0010**	1.1%	0.0003	0.0004	0.0016
Goodness of Fit										
Root Mean Squared Error	0.6108					0.5976				
Mean Absolute Error	0.4491					0.4333				
Mean Prediction Bias	−0.0011					0.0007				

*Statistically significant effects at the 95% level (i.e., those for which the corresponding 95% posterior intervals do not contain 0) are shown in bold.

Table 5
Estimated correlation matrix of the random effects.

Crash Severity	FI (posterior st. dev.)	PDO (posterior st. dev.)
FI	1.0000	0.9874 (0.0042)
PDO	0.9874 (0.0042)	1.0000

results indicate that as the population density increases, the number of FI and PDO bicyclist crashes decreases by 0.9 percent. According to the exploratory data analysis, population density in Harris county is almost uniform, and few areas had significant differences. Therefore, this finding is an interesting one. The significant association of the increasing population density with the decreasing bicyclist crashes may be an indicator of the "safety in numbers", however, since we did not explore the relationship between bicyclist exposure and population density, we can only speculate about it.

The number of adult males and females in the block group was also significantly associated with bicyclist crashes. Note that this does not indicate that the bicyclist involved in a crash was necessarily a male or female; we are only comparing crashes with the demographics of the block group. The results show that as the block group's adult male population increases, the number of FI and PDO bicyclist crashes increases by 1.1 and 1.5 percent, while an increasing number of female adults is associated with a decreasing number of bicyclist crashes (0.9 percent for both FI and PDO crashes). This may be due to the behavior of adult males cyclists compared to that of female cyclists. Studies have found that, when accounting for other confounding factors, male cyclists were more likely to be involved in crashes than females (Prati et al., 2019).

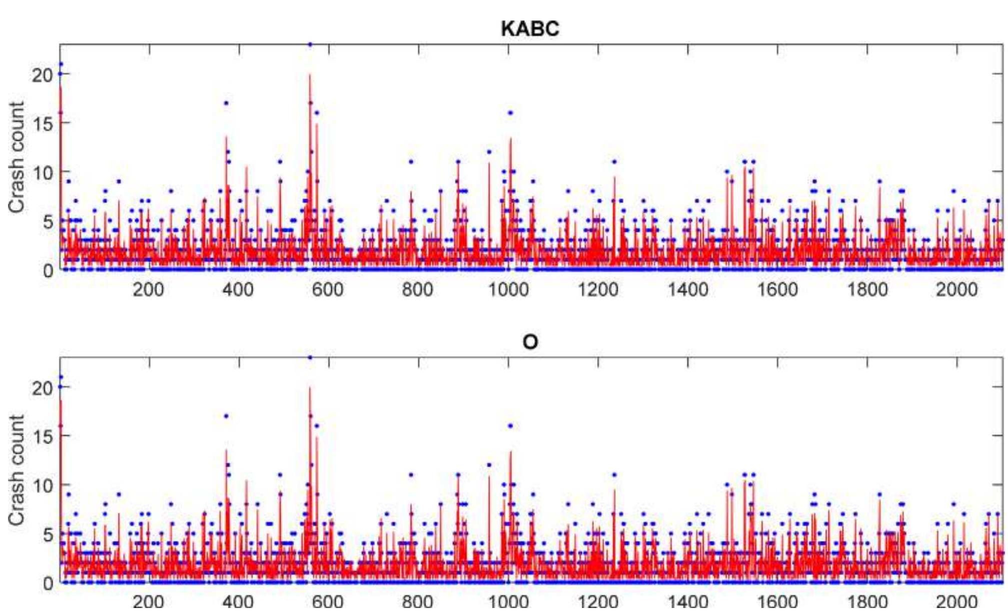


Fig. 5. Observed and predicted bike crash counts for 2,144 block groups in Harris County for each severity level.

As indicated earlier, Harris County is racially- and ethnically-diverse. Among the race and ethnicity variables, only the African-American population was found to have a meaningful association with the number of FI and PDO bicyclist crashes; the rest of the race and ethnicity variables were removed because they negatively affected the model convergence. Although the association of this factor with the bicyclist crashes is not significant, the results are intuitive and align with previous research. As the percentage of the African-American population in a block group increases, the number of FI and PDO bicyclist crashes appear to increase by 1.2 and 1.1 percent. Also, note that this is the only variable that has a higher association with FI crashes than PDO crashes, indicating that the block groups with a higher percentage of the African-American population may experience a higher number of fatal and injury crashes.

Among the fairly long list of income indicators, only the number of households below poverty was associated with a significant increase in bicyclist crashes. Results indicate that as the number of households below the poverty level increases, FI and PDO crashes increase by 1.1 percent. Although log-transformed median income was associated with bicyclist crashes during the early modeling process, this variable was not significant when accounting for the population density and had to be removed from the model to facilitate the model convergence.

Despite the study's findings, it had several limitations. First, bicycle crash data are limited in that agencies do not consistently record crash locations, and they differ with regard to the monetary (with regard to damage to a vehicle) or injury (with regard to a victim) threshold at which a bicycle crash is recorded, with some agencies require self-reporting of crashes. Furthermore, certain community members may not report bicycle crashes due to cultural barriers to interacting with the police (Lugo, 2018). Additionally, this study used estimated bicycle counts that may be biased towards a more active population (e.g., Strava users). Many studies focus on smaller spatial areas to address this limitation. Yet focusing on a smaller spatial area may not produce statistically meaningful results, so we sought to find a balance between estimated bicycle counts and sample size. Due to data limitations, we did not account for the bicyclist crashes per infrastructure type (e.g., presence and length/density of various bicycle facility types) due to data limitations. Moreover, we did not account for the demographic characteristics of bicyclists involved in the crashes, such as bicyclist's residence, zip code, age, gender, ethnicity, and race, since such information is rarely available in the police crash reports used in this study. Therefore, it is quite possible that the bicyclist involved in a crash at a certain block group did not reside in that area. Hence the results of this study should be used to assess the "place-based" rather than "person-based" inequity. For example, we can conclude that "bicyclists riding in block groups or communities with a higher percentage of African-American residents are more likely to be involved in fatal and injury crashes." We cannot conclude that "African-American bicyclists are more prone to being involved in fatal and injury crashes" since our findings do not support such claims. The paper can also be improved by accounting for other variables and their interaction effects whenever such variables are readily available. We urge caution not to use many estimated variables since it may affect the already high uncertainty in the model.

5. Summary and conclusions

Equity in traffic safety is still an emerging area with an ongoing effort to develop frameworks to assess transportation-related health inequity. There are also gaps concerning the methodological framework, influenced by the fact that most equity-related studies are policy-related and do not normally involve quantitative or another type of data-driven analysis. This is particularly the case for bicyclist crashes, which are over-represented in traffic crashes and may disproportionately – or at least substantially – include members of low-income and ethnically-diverse communities. However, there is limited research on equity in bicycling use and safety.

We aimed to help fill this research gap by assessing the association between equity indicators and bicyclist crashes in Harris County – one of the most economically and ethnically diverse counties in Texas, US. Our study has three main contributions. First, we developed a list of factors that can be used to explore equity and then identified the sources for collecting the data; this list can easily be replicated in other equity studies since the data are readily available and do not require additional effort to collect. Second, we used a novel copula-based bivariate regression model to assess equity indicators' impact on two measures of bicyclist safety for the first time. The copula-based bivariate model can help to address some of the major limitations found in the bicyclist crash data (e.g., scarcity, unobserved heterogeneity, and correlation between outcome variables). Finally, our findings contribute to the body of literature on bicyclist safety and transportation-related equity and provide fertile ground for future research.

The findings of this study may have implications for future transportation and planning policies. Although more location and context-specific analyses are required, this study's overall results once again conform with the findings and assumptions in bicycling safety literature that the low-income and racially diverse communities are prone to experience more bicyclist crashes. We also observe that communities with a higher percentage of African-Americans had experienced higher numbers of fatal and injury bicyclist crashes, in particular, instead of other demographic indicators that mostly seemed to affect non-injury (PDO) crashes. These findings can be used to guide the policies and strategies targeting the elimination of inequity in transportation-related health concerns.

Acknowledgements

The authors did not receive direct support for this study. However, parts of this study was built on the ongoing project that is supported by the Texas Department of Transportation Project 0-7043. The authors are responsible for the facts and the accuracy of the data presented herein. The authors would like to thank the feedback and comments received from two anonymous reviewers that helped to improve the quality of the paper.

References

Agyeman, J., & Doran, A. (2021). You want protected bike lanes, I want protected Black children. Let's link": Equity, justice, and the barriers to active transportation in North America. *Local Environment*, 1–18.

Amoh-Gyimah, R., Saberi, M., & Sarvi, M. (2016). Macroscopic modeling of pedestrian and bicyclist crashes: A cross-comparison of estimation methods. *Accident Analysis & Prevention, 93*, 147–159.

Avila, E. (2014). *The folklore of the freeway: Race and revolt in the modernist city*. U of Minnesota Press.

Barajas, J. M. (2020). Supplemental infrastructure: How community networks and immigrant identity influence cycling. *Transportation, 47*(3), 1251–1274.

Bhat, C. R., & Sener, I. N. (2009). A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. *Journal of Geographical Systems, 11*(3), 243–272.

Braun, L. M., Rodriguez, D. A., & Gordon-Larsen, P. (2019). Social (in) equity in access to cycling infrastructure: Cross-sectional associations between bike lanes and area-level sociodemographic characteristics in 22 large U.S. cities. *Journal of Transport Geography, 80*. https://doi.org/10.1016/j.jtrangeo.2019.102544 102544.

Celis-Morales, C. A., Lyall, D. M., Welsh, P., Anderson, J., Steell, L., Guo, Y., ... Gill, J. M. (2017). Association between active commuting and incident cardiovascular disease, cancer, and mortality: prospective cohort study. *BMJ, 357* j1456.

Chen, P., & Shen, Q. (2016). Built environment effects on cyclist injury severity in automobile-involved bicyclist crashes. *Accident Analysis & Prevention, 86*, 239–246. https://doi.org/10.1016/j.aap.2015.11.002.
Clayton, W., Parkin, J., & Billington, C. (2017). Cycling and disability: A call for further research. *Journal of Transport & Health, 6*, 452–462. https://doi.org/10.1016/j.jth.2017.01.013.
Collins, P., Al-Nakeeb, Y., Nevill, A., & Lyons, M. (2012). The impact of the built environment on young people's physical activity patterns: A suburban-rural comparison using GPS. *International Journal of Environmental Research and Public Health, 9*(9), 3030–3050.
Dadashova, B., Griffin, G. P., Das, S., Turner, S., & Sherman, B. (2020). Estimation of average annual daily bicycle counts using crowdsourced strava data. *Transportation Research Record., 2674*(11), 390–402. https://doi.org/10.1177/0361198120946016.
Dadashova, B., & Griffin, G. P. (2020). Random parameter models for estimating statewide daily bicycle counts using crowdsourced data. *Transportation research part D: transport and environment, 84* 102368.
Doran, A., El-Geneidy, A., & Manaugh, K. (2021). The pursuit of cycling equity: A review of Canadian transport plans. *Journal of Transport Geography, 90*. https://doi.org/10.1016/j.jtrangeo.2020.102927 102927.
Flanagan, E., Lachapelle, U., & El-Geneidy, A. (2016). Riding tandem: Does cycling infrastructure investment mirror gentrification and privilege in Portland, OR and Chicago, IL? *Research in Transportation Economics, 60*, 14–24. https://doi.org/10.1016/j.retrec.2016.07.027.
Flint, E., & Cummins, S. (2016). Active commuting and obesity in mid-life: cross-sectional, observational evidence from UK Biobank. *The lancet Diabetes & endocrinology, 4*(5), 420–435.
Gelman, A., Carlin, J.C., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2013), Bayesian Data Analysis, Chapman and Hall/CRC, 3rd edition.
Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. New York: Chapman & Hall.
Goodman, A., Sahlqvist, S., Ogilvie, D., & iConnect Consortium (2013). Who uses new walking and cycling infrastructure and how? Longitudinal results from the U.K. iConnect study. *Preventive Medicine, 57*(5), 518–524. https://doi.org/10.1016/j.ypmed.2013.07.007.
Gordon-Larsen, P., Nelson, M. C., Page, P., & Popkin, B. M. (2006). Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics, 117*(2), 417–424.
Guo, Y., Osama, A., & Sayed, T. (2018). A cross-comparison of different techniques for modeling macro-level cyclist crashes. *Accident Analysis & Prevention, 113*, 38–46.
Hirsch, J. A., Green, G. F., Peterson, M., Rodriguez, D. A., & Gordon-Larsen, P. (2017). Neighborhood sociodemographics and change in built infrastructure. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability, 10*(2), 181–197. https://doi.org/10.1080/17549175.2016.1212914.
Huang, H., Abdel-Aty, M. A., & Darwiche, A. L. (2010). County-level crash risk analysis in Florida: Bayesian spatial modeling. *Transportation Research Record, 2148*(1), 27–37.
Jacobsen, P. L. (2015). Safety in numbers: More walkers and bicyclists, safer walking and bicycling. *Injury Prevention, 21*(4), 271–275. https://doi.org/10.1136/ip.9.3.205rep.
Kaplan, S., & Giacomo Prato, C. (2015). A spatial analysis of land use and network effects on frequency and severity of cyclist–motorist crashes in the Copenhagen region. *Traffic Injury Prevention, 16*(7), 724–731. https://doi.org/10.1080/15389588.2014.1003818.
Kröyer, H. R. (2015). Accidents between pedestrians, bicyclists and motorized vehicles: Accident risk and injury severity. Faculty of Engineering, Lund University.
Lee, J., & Abdel-Aty, M. (2018). Macro-level analysis of bicycle safety: Focusing on the characteristics of both crash location and residence. *International Journal of Sustainable Transportation, 12*(8), 553–560.
Le, H. T., Buehler, R., & Hankey, S. (2019). Have walking and bicycling increased in the U.S.? A 13-year longitudinal analysis of traffic counts from 13 metropolitan areas. *Transportation Research Part D: Transport and Environment, 69*, 329–345.
Lindsey, G., Tao, T., Wang, J., & Cao, J. (2019). Pedestrian and Bicycle Crash Risk and Equity: Implications for Street Improvement Projects. Retrieved from - https://conservancy.umn.edu/handle/11299/203635.
Lopez, D., Malloy, L. C., & Arcoleo, K. (2022). Police narrative reports: Do they provide end-users with the data they need to help prevent bicycle crashes? *Accident Analysis & Prevention, 164*, 106475.
Lugo, A. E. (2018). *Bicycle/race: Transportation, culture, & resistance*. Microcosm Publishing.
Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research, 11*, 1–16.
Mannering, F. (2018). Temporal instability and the analysis of highway accident data. *Analytic Methods in Accident Research, 17*, 1–13.
Marshall, W. E., & Garrick, N. W. (2011). Evidence on why bike-friendly cities are safer for all road users. *Environmental Practice, 13*(1), 16–27. https://doi.org/10.1017/S1466046610000566.
Marshall, W. E., Piatkowski, D., & Johnson, A. (2017). Scofflaw bicycling: Illegal but rational. *Journal of transport and land use, 10*(1), 805–836.
McNeish, D. (2016). On using Bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(5), 750–773.
National Center for Statistics and Analysis. (2019). Bicyclists and other cyclists: 2017 data. (Traffic Safety Facts. Report No. DOT HS 812 765). Washington, DC – National Highway Traffic Safety Administration. Retrieved from – https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812765.
National Highway Traffic Safety Administration, 2020. Overview of Motor Vehicle Crashes in 2019. Available at https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813060.
Nordback, K., & Marshall, W. (2011). Effect of increased bicyclist volumes on individualized bicyclist. *Risk, No*, 11–4212.
Noland, R. B., & Laham, M. L. (2018). Are low income and minority households more likely to die from traffic-related crashes? *Accident Analysis & Prevention, 120*, 233–238.
Osmonson, B. L. (2017). An Equity Analysis of Bicycle Infrastructure Around Light Rail Stations in Seattle, WA (Doctoral dissertation).
Park, E. S., & Lord, D. (2007). Multivariate Poisson-Lognormal models for jointly modeling crash frequency by severity. *Transportation Research Record, 2019*, 1–6.
Park, E. S., Oh, R., Ahn, J. Y., & Oh, M. S. (2020). Bayesian analysis of multivariate crash counts using copulas. *Accident Analysis & Prevention*, 105431. https://doi.org/10.1016/j.aap.2019.105431.
Prati, G., Fraboni, F., De Angelis, M., & Pietrantoni, L. (2019). Gender differences in cyclists' crashes: An analysis of routinely recorded crash data. *International Journal of Injury Control and Safety Promotion, 26*(4), 391–398.
Rebentisch, H., Wasfi, R., Piatkowski, D. P., & Manaugh, K. (2019). Safe streets for all? Analyzing infrastructural response to pedestrian and cyclist crashes in New York City, 2009–2018. *Transportation Research Record, 2673*(2), 672–685. https://doi.org/10.1177/0361198118821672.
Saha, D., Alluri, P., Gan, A., & Wu, W. (2018). Spatial analysis of macro-level bicyclist crashes using the class of conditional autoregressive models. *Accident Analysis & Prevention, 118*, 166–177.
Smith, M., Hosking, J., Woodward, A., Witten, K., MacMillan, A., Field, A., & Mackie, H. (2017). Systematic literature review of built environment effects on physical activity and active transport–an update and new findings on health equity. *International Journal of Behavioral Nutrition and Physical Activity, 14*(1), 158. https://doi.org/10.1186/s12966-017-0613-9.
Vanparijs, J., Panis, L. I., Meeusen, R., & de Geus, B. (2015). Exposure measurement in bicycle safety analysis: A review of the literature. *Accident Analysis & Prevention, 84*, 9–19. https://doi.org/10.1016/j.aap.2015.08.007.
World Health Organization, 2002. Sedentary Lifestyle: A Global Public Health Problem. https://www.who.int/docstore/world-health-day/2002/fact_sheets4.en.pdf.
Wu, F. (2020). An Equity Analysis of Phoenix Bicyclist and Pedestrian Involved Crash Rates (Doctoral dissertation, Arizona State University).
Zhu, X., Lee, C., Kwok, O. M., & Varni, J. W. (2011). Context-specific correlates of walking behaviors to and from school: Do they vary across neighborhoods and populations? *Journal of Physical Activity and Health, 8*(s1), S59–S71.
Zou, Y., Ye, X., Henrickson, K., Tang, J., & Wang, Y. (2018). Jointly analyzing freeway traffic incident clearance and response time using a copula-based approach. *Transportation Research Part C: Emerging Technologies, 86*, 171–182.

Bahar Dadashova, Ph.D. - Dr. Dadashova is an interdisciplinary, associate research scientist at Texas A&M Transportation Institute. Her research focuses on linking the built environment and transportation infrastructure with highway safety, mobility, operation, public health outcomes and equity. She has significantly contributed to this research area by developing concepts and methodologies, research tools, guidelines, and supervising and mentoring graduate students.

Eun Sug Park, Ph.D. Dr. Park is a statistician and senior research scientist with the Texas A&M Transportation Institute. She co-authored a transportation statistics textbook 'Transportation Statistics and Microsimulation' as well as leading modeling and data analysis tasks in numerous transportation research studies for many state and federal agencies. She is a Fellow of the American Statistical Association (ASA), an Elected Member of the International Statistical Institute (ISI), a Member of the TRB Committee on Statistical Methods, and Editor for Statistics of the journal Chemometrics and Intelligent Laboratory Systems.

Seyedeh Maryam Mousavi, Ph.D. – Dr. Mousavi is an Assistant Research Scientist at Texas A&M Transportation Institute. She is an experienced researcher in evaluating and analyzing traffic safety, safety and operation of autonomous vehicles, safety and operation of various access management techniques, safety of vulnerable roadway users, and highway design. Moreover, she has been involved in developing new surrogate safety measures and safety performance functions as well as establishing design guidelines.

Boya Dai – Ms. Dai graduated from Texas A&M University in May of 2013, with an M.S. in Urban and Regional Planning. Her research is in the areas of traffic crash analysis and bicycle and pedestrian safety. While in school, she assisted TTI's Mobility Management Program on several safety-related projects for TxDOT, the Capital Area MPO, and the City of Austin. Her major work included the identification of crash patterns by analyzing the geographic, demographic, and physical features of traffic crash data in the Austin metropolitan region. She has also conducted hotspot analysis to identify prioritized locations for addressing bike and pedestrian safety, as well as literature reviews regarding traffic management and bike and pedestrian planning.

Rebecca Sanders, Ph.D., RSP_{2B} – Dr. Sanders is the Founder and Principal Investigator of Safe Streets Research & Consulting, LLC, a certified DBE specializing in rigorous crash and survey data analysis to provide insights into pedestrian, bicycle, e-scooter, and driver safety, behavior, and mobility. She has 15 years of experience spanning both the academic and private sectors, including research on national pedestrian safety trends, bicyclist and driver design preferences, and e-scooter user behavior. She has also contributed to national roadway design guidance and helped cities and counties across the country understand and address multi-modal safety through Vision Zero and systemic safety efforts. Dr. Sanders currently chairs TRB's Bicycle Transportation Committee and serves on the City of Portland's Pedestrian Advisory Committee.

Journal of Safety Research 82 (2022) 151–158

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Commentary

A tale of six climates: Reflections and learnings after the development of six industry-specific safety climate scales

Tristan W. Casey [a,*], Xiaowen Hu [b], Lisette Kanse [c], Angelica Varhammar [d]

[a] Safety Science Innovation Lab, Griffith University, Australia
[b] Business School, Queensland University of Technology, Australia
[c] School of Psychological Science, University of Western Australia, Australia
[d] Centre for WHS, NSW Government, Australia

ARTICLE INFO

Article history:
Received 28 April 2021
Received in revised form 9 December 2021
Accepted 11 May 2022
Available online 23 May 2022

Keywords:
Safety climate
Industry-specific safety climate
Leading indicators

ABSTRACT

Introduction: Researchers are finding merits in utilizing industry-specific safety climate scales that capture the nuances of context, and tend to show stronger associations with safety behavior and outcomes like incidents. Yet, to date, guidance around the practicalities of developing and validating such industry-specific scales is lacking in the safety science literature. *Method:* In this paper we outline our experiences developing six industry-specific safety climate scales and highlight strengths and limitations of our approach. We also briefly review the industry-specific safety climate literature and offer highlights for consideration when developing such scales. Our method to develop industry-specific safety climate scales followed an established best practice structure: literature review of existing published industry scales, collation and review of existing scale items, consultation interviews with industry experts, item drafting, exploratory and confirmatory statistical analyses, and finally, a real-world ecological validity test. *Results:* Our research highlighted the diversity of safety climate dimensions (both the conceptual and content domains of each dimension) when it is considered at an industry level. Also, the literature reviews revealed a dearth of industry-specific safety climate scales in the areas we engaged with, so our project filled a glaring gap in research and practice. Best practice safety climate scale development methods are provided to stimulate further research. *Conclusions:* We conclude with reflections on the nature of safety climate within and across industries, and offer suggestions for future lines of research across other contexts (e.g., national culture, geography, and regulatory settings). We suggest that industry-specific safety climate scales have a specific use case, such as identifying specific areas to improve and evaluating the impact of safety interventions. *Practical Applications:* This article provides applications for both applied researchers (to improve capabilities in safety climate scale development) and practitioners who wish to measure organisational safety climate and design effective interventions. Engaging with regulators to build safety climate scales is powerful because their personnel have rich experiences to share across multiple workplaces. Organisational researchers can engage with survey panels to build robust scales. Finally, industry-specific nuances can lead to richer insights into an organisation's safety climate.

1. Introduction

Safety climate is an established predictor of safety behavior (Clarke, 2010) and incidents at work (Bergman et al., 2014), and is defined as "shared perceptions of safety policies, procedures, and practices in an organization" (Zohar, 2011, p.143) that signal the importance and priority given to safety (Griffin & Curcuruto, 2016). Safety climate measurement also provides useful diagnostic information that businesses can use to improve safety management proficiency (Kim et al., 2019) and leadership capability (Zohar, 2002). Given that safety climate can be changed and is associated with enhanced safety performance (Lee et al., 2019), it represents a powerful lever for the reduction on injuries and illnesses in high-risk environments.

General survey scales measure safety climate in any workplace context where safety is an important organizational goal (e.g., Beus et al., 2019); therefore, it should apply in some form to all work-

* Corresponding author.
E-mail address: tristan.casey@griffith.edu.au (T.W. Casey).

https://doi.org/10.1016/j.jsr.2022.05.006
0022-4375/

places. However, nuanced scales that capture industry context tend to perform better, in terms of face validity and criterion validity, such as amplifying the relationship between safety climate and outcomes like safety behavior (Huang et al., 2013). Indeed, drawing on the emergence of various studies that adopted an industry-specific focus to assess safety climate (e.g., Glendon & Litherland, 2001; Niskanen, 1994; Wills et al., 2006), Zohar (2010) argued that industry-specific safety climate scales not only seem more face valid to the workers completing them, but also capture additional diagnostic and predictive information that can be used to improve safety performance and test hypotheses about safety climate formation. The rich dimensionality of contextualized safety climate scales helps practitioners identify and target specific interventions (Keiser & Payne, 2018; Lee et al., 2019) and measure change through more sensitive before/after measures.

Given the advantages of industry-specific safety climate measurement over generic scales, it is surprising that, to our knowledge, no published guidance yet exists. Furthermore, published studies tend to be highly inconsistent in their methods to develop contextualized scales (Keiser & Payne, 2018), and risk introducing safety climate dimensions that are outside the core construct definition (Guldenmund, 2007; Shea et al., 2021). Guidance on how to develop industry specific safety climate scales would be useful for several reasons: (1) a consistent approach would help to standardize the process, leading to more accurate operationalizations of safety climate; (2) it may encourage existing safety researchers to expand their programs into industry-specific safety climate, and (3) practitioners would benefit through building their capabilities to review, evaluate, and possibly adapt or refine existing safety climate scales. Also, although many industry-specific safety climate scales have been developed to date, the absence of any macro-level cross-industry reflections means that there has been limited opportunity to synthesize across different safety climate domains.

In this commentary, we aim to close these gaps by summarizing our experiences developing six industry-specific safety climate scales. Although there have been recent comprehensive reviews of safety climate scales (such as the one by Shea et al., 2021), these reviews have been largely silent on the issue of industry-specific safety climate. Our project, which we discuss in this paper, was part of an Australian regulator's efforts to provide evidence-based tools to assist businesses to measure and improve safety climate.

Our commentary is divided into three sections. First, we briefly summarize the current state of industry-specific safety climate research. Next, we discuss practicalities by describing the steps involved in our scale development process and provide advice for safety climate researchers. Finally, we reflect on contextual safety climate, and examine similarities and differences across the six industries we targeted, as well as offer recommendations for future research.

1.1. The current state of industry-specific safety climate research

The scholarly investigations of industry-specific versus general safety climate survey scales is ongoing. In this section, we briefly summarize key findings from the ongoing research program, which emphasize the inconsistency in how industry-specific safety climate scales are developed, along with the practical constraints reported by researchers. These issues are summarized into categories: incorrect operationalization, inconsistent degree of contextualization, limited evidence of incremental validity, diverse validation samples, and haphazard development of safety climate scales, including unclear articulation of scale purpose.

Incorrect operationalization of industry-specific safety climate. According to Zohar (2010) and Griffin and Curcuruto (2016), safety climate is formed from perceptions of safety policies, procedures, and practices that convey a sense of importance toward safety and are founded in items that reflect the discrepancies between espousals and enactments. Yet, many published industry specific safety climate surveys apply dimensions that are outside this scope. For instance, the survey developed by Parker, Tones, and Ritchie (2017) for the mining industry included constructs that have previously been deemed out-of-scope for safety climate (e.g., Zohar, 2010) such as risk perceptions and safety attitudes (Shea et al., 2021). Further, the dimensionality of safety climate across industries is likely to vary, with some consistent variables (e.g., management safety commitment) and some nuanced elements like 'schedule flexibility' and 'field orientation' (Huang et al., 2013b). Importantly, a recent review of safety climate scales highlighted the diversity of dimensions across measures and suggested that this issue creates a discrepancy between purported and actual scale content, compromising safety climate measurement (Shea et al., 2021).

Inconsistent contextualization. Keiser and Payne (2018) described the contextualization of psychological measures on a continuum, ranging from none (e.g., "Management is committed to safety") through to substantial (e.g., "Management does not expect support staff to put the needs of the client above their own personal safety"). Interestingly, safety climate measures can include a mixture of general and industry-specific items to achieve the 'best of both worlds' (e.g., Huang et al., 2013a). However, this is unconventional because contextualized scales are usually tailored to the industry setting in a wholesale fashion. The degree of contextualization in studies that develop industry specific safety climate scales varies considerably, with Huang and colleagues (2013a) developing new scales for lone-worker transportation from the ground up, whereas others studying construction safety climate, like Glendon and Litherland (2001) and Choudhry, Fang, and Lingard (2009) elected to modify existing general safety climate questionnaires. In the latter case, minor wording tweaks were made, and dimensions/questions deemed irrelevant to the context were dropped. This inconsistency in industry specific safety climate operationalization muddies the interpretation of findings given that in some industry settings, bespoke scales are used (with many industry-specific nuances) and in others, general questions are used. Future meta-analytic studies involving measures of safety climate should consider contextualization as a moderating factor, as it may show differential relationships with safety outcomes (Jiang, Lavaysse, & Probst, 2019).

Emerging incremental validity. Just two studies have so far been done to answer the question of predictive incremental validity regarding specific versus general safety climate survey scales. Keiser and Payne (2018) found that across five laboratory samples, contextualized safety climate added significant predictive power over a generalized scale (but only in three of the samples). In a comprehensive study involving 120 samples, Jiang and colleagues (2019) found meta-analytic evidence that industry-specific scales are better at predicting safety behavior and knowledge, whereas general safety climate was more predictive of adverse events like safety incidents. Inconsistencies in how industry safety climate scales are developed could be responsible for these effects, namely, by attenuating relationships with outcomes.

Diverse validation samples. Developing new safety climate scales is a challenging task that should not be undertaken lightly. Considerable investment in upfront data collection is required to ensure the scale demonstrates adequate psychometric performance. For instance, despite claiming to have developed contextualized laboratory safety climate scales, Kaiser and Payne (2018) used a generic safety climate scale and added 'laboratory specific information' from 'at least' one interview and conducted a subject matter expert review before deploying with the target population. Others, such as

Parker, Tones, and Ritchie (2017) developed a purportedly mining-specific safety climate scale that drew on existing general and cross-industry scales. No validation sample was recruited to test the instrument before administering it. This inconsistency in the validation of industry specific safety climate scales represents significant issues if they are to be used by other researchers, as the psychometric properties may not be adequately established. The current wide availability of good quality crowdsourced online samples means that safety climate researchers have no excuse to not use an appropriately rigorous process to validate new scales.

Mixed development of contextualized safety climate measures by industry. In a bibliometric review of 38 years of safety climate research, Bamel and colleagues (2020) discovered that construction, healthcare, and transportation have established programs of industry-specific research on safety climate. In their meta-analytic review covering research since 2000, Jiang and colleagues (2019) found 84 general safety climate samples and 36 industry-specific samples. From these studies, it is apparent that some industries are more represented in the domain-specific safety climate research program than others. For instance, construction, transportation, and healthcare seem particularly popular industries in which an industry-specific approach is taken. From a respondent engagement perspective, more face valid safety climate questions may be important in technical and high workload industries like healthcare because it reduces respondents' resistance to completing seemingly irrelevant surveys (Burns et al., 2008). Contextualized surveys often generate better quality data because the items prompt cognitive recall of information and reduce subjectivity (Tourangeau, Rips, & Rasinski, 2000). Also, and following the logic of Keiser and Payne (2018), some industries may benefit from contextualization more than others. Their study found that in certain laboratory contexts, more specific and nuanced wording of safety climate items added to the criterion validity, whereas in others it did not. The authors postulated that lower-risk industry settings may benefit more from contextualization because the additional item detail provides cognitive priming and more accurate responding given that the respondents are reminded about specific risks in their environment (of which they may have been unaware previously).

Unclear articulation of scale purpose. Safety climate scales have multiple purposes, and these purposes should inform the type of scale that is used as well as the dimensionality of the measure. For instance, a short generic safety climate scale such as the 6-item NIOSH scale developed by Hahn and Murphy (2008) might be suitable as a 'pulse check' included with other measures like employee engagement scales. Accordingly, the dimensionality of the scale focusses on a congeneric and global safety climate factor, as well as more conceptual item wording (as opposed to specific policies and practices). In contrast, a multi-dimensional and industry contextualized measure such as the one developed by Huang and colleagues (2013) is arguably more suited to organizational diagnostic work given the number of specific dimensions included. Drill-down analysis and specific recommendations are facilitated by a multidimensional measure. Another approach is the generic but detailed and multilevel set of safety climate scales developed by Zohar and Luria (2005). Such scales are sensitive to change, being based on perceptions of specific policies, procedures, and practices, so would be suited to intervention evaluation. Dimensionality in this case focusses on a set of two or more concepts such as safety prioritization and proactive safety practices. In general, safety climate research often fails to explicitly consider and describe the alignment between scale purpose and dimensionality.

Overall, the research on industry-specific safety climate is expanding but is considerably undermined by inconsistencies in methods. Furthermore, we are unaware of any existing commentary on cross-industry safety climate comparisons nor of any practical guidance to inform the development of industry-specific safety climate scales. We address these gaps in knowledge in the following sections with lessons learned from our industry project. Our objective is to provide practical guidance to safety climate researchers and to advance the state of industry-specific safety climate measurement and offer critical reflections on the conceptualization and operationalization of safety climate across the six industries we dealt with in this project. This discussion is aimed at a tactical level and a strategic level. Regarding the former, we offer practical ways to develop industry specific safety climate scales to eliminate the observed inconsistencies. Regarding the latter, we hope to simulate further commentary and thinking about safety climate across industries by examining macro trends.

1.2. Our approach

In our project, we were tasked with the development of six industry-specific safety climate survey scales. The industries were identified by the partner regulator as ones with elevated risk through a combination of inspection metrics and workers' compensation statistics (SafeWork NSW, 2018), so were prioritized for intervention using both compliance/enforcement and educational/guidance approaches. Through the project, we supported the development of self-help resources related to safety climate that managers within each industry could draw upon to measure their current state and implement improvements.

Our process to develop the industry specific scales followed the approach of established and leading scholars in safety climate, such as Huang and colleagues (2013) and Zohar (2010). Specifically, we conducted the following steps:

1. *Literature review of existing scales published within each industry.* This step was conducted to orient the research team to each industry context and provide any insights into the types of safety climate dimensions we could expect to generate.
2. *Collation of existing published industry-specific items (if any existed).* Most robust safety climate scale development studies begin with a collation of existing items (e.g., Huang et al., 2013). This strategy is widely considered best practice (Hinkin, 1998) to ensure any pre-validated items are included and improve efficiency. However, upon review of the industry specific literatures, only a handful of truly industry-specific safety climate items were located across the industry domains; mostly, past researchers made minor contextualizations to generic items.
3. *Consultation interviews with industry-specialist safety inspectors from the partner regulator.* These sessions were conducted online and were designed to identify broad industry-specific safety climate themes that could be probed and explored during the worker focus groups. Best practices for virtual interviews in qualitative data collection were followed, such as testing the technology ahead of time, develop technological backups (e.g., revert to mobile phone), plan for distractions, and effectively manage ethical informed consent procedures (Gray et al., 2020).
4. *Consultation focus groups with workers from each industry.* Themes from the regulator interviews were unpacked and explored, along with additional inductive work to identify new dimensions of safety climate that would not be visible to regulators (e.g., co-worker and supervisor safety practices). We followed virtual focus group best practice (Johnson & Odhner, 2021), such as ensuring a thorough briefing to reduce technical issues, utilizing commonly available technology (Microsoft Teams and Zoom), establishing ground rules and interaction protocols (e.g., use of raise hand function, chat box participation), and ensuring visual connection using web cameras.

5. *Item drafting and refinement by synthesizing literature review, interview, and focus group data.* Again, we followed best practice (Hinkin, 1998) such as developing roughly 2–3 times the number of items we were aiming for through an initial pool, leveraging independent subject matter experts to review and refine the scale items, and undertaking content validity sorting tasks to ensure all items were considered part of the safety climate construct.
6. *Exploratory analyses (EFAs) using online industry-specific samples from Prolific (*https://www.prolific.co) (one s*ample per industry of approximately 150 respondents each).* As outlined by Cortina et al. (2020), an exploratory phase should be conducted when developing new scales to identify the emerging factor structure. Where the EFA did not support the *a priori* factor structure identified through themes, we eliminated items and re-ran the EFAs until a clean solution was found, with all items loading 0.30 or greater onto their respective factors, no substantial cross-loadings, and at least three items per dimension (Tabachnick & Fidel, 2007). Only minimal scale modifications were required due to our robust item generation phase.
7. *Confirmatory factor analyses (CFAs) using additional online industry-specific samples from Prolific (one sample per industry of approximately 250 respondents each).* As recommended by Cortina and colleagues (2020), CFAs were conducted using a separate and independent sample to verify the factor structure. Model fit indices and factor loadings were examined to ensure every scale had acceptable construct validity. Further validity checks were done by including both convergent (general safety climate and safety leadership), divergent (personality and emotional regulation), and criterion (safety behavior) measures within the sample and examining pairwise Pearson correlation coefficients.
8. *'Real world' ecological validity test using a sample of 100–500 respondents each from companies operating within the industries (a total of eight companies across six industries).* This step helped to finalize the scales and develop useful supporting materials such as information flyers, templates, and analysis tools. Further, written testimonials were obtained from each organization to promote uptake once the scales were published by the partner regulator.

Our literature reviews revealed few or no existing contextualized safety climate scales for most of the industries we targeted (i.e., disability support, residential construction, meat processing, mixed sheep and cattle agriculture, long-distance transport, and law enforcement), except for some in the transportation and manufacturing settings. We found that commonly, researchers adapted existing general scales by using minor contextualization of language (e.g., replacing the word 'supervisor' with 'team leader' of 'foreman' to reflect common labels used in the industry). However, collating this existing literature was useful to inform targeted questioning and provide background knowledge for the industry consultation steps.

To our knowledge, no other safety climate researchers have consulted with safety regulators to develop their scales. Our experience was that this was a useful step given inspectors visit multiple workplaces and so have exposure to high level insights. Further, their industry-specific knowledge and prior working history (typically in the industries in which they inspect and regulate) are rich and useful to inform the creation of safety climate survey items. Partnering with a regulator in this way may help to decrease the gap between safety-as-legislated and safety-as-practiced by industry. Academics can facilitate two-way knowledge sharing between industry and government, building awareness and capability.

Regarding the specific method we employed during our industry stakeholder consultations, we derived the set of focus group questions from the safety climate theory outlined by Zohar (2010). Specifically, we considered the following: (1) the multilevel nature of safety climate, (2) the operationalization of safety climate as the perceived conflict or discrepancy between espousals and enactments, (3) competing goals and tensions that may affect safety, and (4) tangible practices and processes that signal the priority and importance of safety. The questions used in our investigations are shown below.

- How can you tell whether senior management is genuinely concerned about safety? What do they say or do?
- How can you tell whether a supervisor is genuinely concerned about safety? What do they say or do?
- (Inspectors only) When you visit a worksite, what signs do you look for to make an assessment of how safe the organization is?
- What are the signs of good safety in an organization from your industry?
- What are the signs of poor safety in an organization from your industry?
- Tell me about an occurrence you know of when an organization prioritized productivity or other work goals over safety? What happened? What were the outcomes?
- Tell me about an occurrence you know of when an organization prioritized safety over other work goals like productivity? What happened? What were the outcomes?
- In your industry, what goals or tensions might exist that affect the priority of employee safety? (For example: patient versus employee safety, food safety versus employee safety)

A strength of our focus group approach included the diversity of participants, as they typically included a range of government and industry participants across multiple organizations within each industry. Between five to seven inspectors were consulted for each industry (resulting in a total of 30 inspectors participating across all industries), and between 8 and 10 workers and supervisors (separated into different focus groups to avoid hierarchy or power/influence effects) were involved (resulting in a total of 60 industry representatives across all industries). Table 1 shows a summary of participants in our study. Focus groups were also an efficient means to consult with a sizeable number of people. All focus groups were conducted virtually (primarily due to COVID-19), which, despite our attempts to manage dynamics, may have impeded the participation of some group members. Nevertheless, virtual focus groups tend to result in better discussions due to an increased willingness to disagree and engage in productive conflict, increased self-disclosure due to perceived anonymity and reduced social presence, and significantly more new ideas shared within the group (Reid & Reid, 2005). Group dynamics may have resulted in a natural consensus emerging around core themes, an issue that would have been reduced had we engaged in a one-on-one interview format. Our experiences with using virtual interviews and focus groups were positive, perhaps due to the facilitation skills of the first author who engaged in strategies such as referring to all participants by name (which is facilitated online through a participant list), permitting and encouraging participants to use both the text-based chat function and verbal communication (enabling greater sharing of information), capturing key comments and themes live via screen-sharing functions, and significant pre-session preparation in the form of a briefing and detailed informed consent statement and study document.

Tripartite collaboration between government, industry, and academics was a hallmark of this project and our experiences offer insight into how to develop and maintain the important relation-

Table 1
Summary of project participants involved in industry-specific safety climate survey development.

Phase of Development	Disability support	Residential construction	Meat processing	Mixed sheep and cattle agriculture	Long-distance freight transport	Law enforcement
Initial development (qualitative phase)	5 specialist regulator inspectors; 1 focus group with 6 workers	5 specialist regulator inspectors; 1 focus group with 8 workers and supervisors	4 specialist regulator inspectors; 1 focus group with 10 workers and supervisors	6 specialist regulator inspectors; 5 interviews with workers, managers, and consultants	5 specialist regulator inspectors; 2 focus groups with 8 workers and supervisors	4 specialist regulator inspectors; 16 interviews with police officers from different jurisdictions and levels (ranging from constables to senior managers)
Exploratory phase (EFA)	*N* = 150 disability care workers from Prolific	*N* = 175 construction workers from Prolific	*N* = 150 manufacturing workers from Prolific	*N* = 150 agriculture workers from Prolific	*N* = 220 transport workers from Prolific	*N* = 206 police and military personnel from Prolific
Confirmatory phase (CFA)	*N* = 250 disability care workers from Prolific	*N* = 275 construction workers from Prolific	*N* = 250 manufacturing workers from Prolific	*N* = 250 agriculture workers from Prolific	*N* = 250 transport workers from Prolific	*N* = 300 police and military personnel from Prolific
Ecological validity phase	*N* = 140 workers from a disability care organization in Australia	*N* = 166 construction workers from a national company in Australia	*N* = 106 workers from four different sites in a meat processing company in Australia	*N* = 130 agriculture workers from a national company in Australia	*N* = 109 workers from two transport companies in Australia	*N* = 253 police officers from a state-level police organization in Australia

ships between parties (Casey et al., 2019). Industry, through a real-world pilot test of the developed scales, was engaged through providing access to research and consulting expertise at no cost and emphasizing the practical benefits of participation. To this end, flyers were developed for industry that briefly described the project, its value, and potential business impacts. The flyers also contained the safety climate survey and simple user instructions. Personal contacts and government networks were instrumental in generating interest and ensuring a wide recruitment reach. We used social media platforms to identify participating companies. Considerable effort was invested in providing the participants with practical summaries of the research and 1:1 senior management debriefing sessions to ensure value was created through the safety climate diagnostic process.

Finally, our applied approach to safety climate scale validation is a useful template that other researchers may benefit from. Specifically, we used an online panel platform (Prolific.co) to recruit exploratory and confirmatory samples, which expedited the scale development process and allowed us to pre-screen participants such that only those from the targeted industry could be involved. Although using online samples have been shown to be viable in terms of evaluating preliminary psychometric performance (Palan & Schitter, 2018), there are nuances within online samples that require stringent data cleaning and checking to ensure elimination of problematic cases that may accentuate correlations between items and distort factor structures. In our research, we included best-practice techniques such as attention check items and warning participants that data quality would be evaluated (Abbey & Meloy, 2017).

Following each validation with a real-world evaluation using actual companies in the targeted industries was logistically challenging, but enabled us to confirm factor structure, demonstrate incremental validity over a general safety climate scale, and evaluate the feedback and dissemination steps. A limitation was the smaller sample sizes (typically between 100–200 respondents) achievable within the industry samples and some restrictions on survey length. These constraints limited the use of stringent CFA techniques for replication of factor structure and limited the number of criterion validity variables we could include. Further, tests of aggregation and operationalization of safety climate strength (Casey, Griffin, Flatau Harrison, & Neal, 2017) were not possible due to a lack of team-level identifiers, and so will be a useful avenue for further research.

1.3. Collation of best practices to develop industry-specific safety climate scales

Leveraging what already exists. Scanning the industry-specific safety climate literature was initially disappointing given the lack of existing scales. Instead, most researchers simply adapted existing generic scales through minor changes and did not explain why/how these changes were made. Nevertheless, as more safety climate scales are developed specific to industries, researchers will benefit. Indeed, researchers should consider scanning the grey literatures as projects such as ours will be included. This initial scanning step will prevent overlap with other industry specific scales, in turn avoiding confusion and inefficiencies.

Qualitative explorations. Typically, industry safety climate researchers conduct many interviews, sometimes in excess of 50 (Huang et al., 2013a). Our experience in this project is that this many interviews may not be required. Qualitative researchers suggest that data saturation is typically achieved between 12 and 14 interviews (Silverman, 2013), which we agree with based on our project experiences. We found that the choice of interviewees was essential to form a broad appreciation of industry specific nuances. Regulators, unions, and associations are well positioned to comment on industry characteristics given their specializations and oversight across multiple organizations. Combined with a sample of workers from across multiple organizations in a focus group format, broad themes identified initially can be refined and focused quite readily.

Draw on survey panels for initial validation. Our use of online survey panels led to an efficient scale development process. As online survey panels like Prolific.co become more widespread and adopted by a broader range of Internet and industry specific users, data quality will further improve. Recent studies that compare different platforms have shown Prolific.co to generate the highest quality response data and also provide representative industry-level samples (Peer et al., 2021). We urge researchers to 'do their homework' on survey panels before undertaking safety climate development research as it is possible that some panels will be less appropriate than others.

Emphasize industry specific nuances. To fully reap the benefits of an industry specific safety climate scale, we encourage researchers to partner with qualitative experts and mine their interview and focus group data for specific themes. We recommend avoiding generically worded scale items for dimensions such as manage-

Table 2
Summary of safety climate dimensions for each industry, with each cell summarizing the industry-specific nuances of the concept.

Dimensions	Disability support	Residential construction	Meat processing	Mixed sheep and cattle agriculture	Long-distance freight transport	Law enforcement
Management safety commitment	Making financial investments for safety and leveraging inspection/audit data to make safety decisions.	Safety is factored into all project stages and not compromised due to client pressures.	Willingness of management to halt the production line for safety and openness to improving safety.	N/A (management are not generally recognized in this industry as supervisors are often also the 'boss')	Appropriate scheduling and flexibility due to sickness or traffic delays (absence of pressure to rush).	Genuineness of commanders' intentions towards health and safety (perceived authenticity).
Supervisor safety practices	Helping team members to manage their stress and wellbeing. Also, a second 'proactivity' dimension that relates to following up on raised safety concerns and setting clear safety expectations for staff.	Safety-related communication such as highlighting risks and conducting Toolbox Talks.	Building staff knowledge to work safely through communication and informal job coaching/training.	Ensuring only competent workers are allocated high-risk jobs. Also, a second dimension relating to care and concern through managing fatigue (ensuring breaks are taken, checking in on welfare) .	Genuine care and concern for drivers; supervisors maintaining close contact with drivers on the road through mobile phone.	Ensuring equipment is functional and safe to use for staff. Also, a second dimension relating to results pressure (i.e., pressure to achieve various targets and KPIs) .
Co-worker safety practices	N/A	Openness to change so safety can be improved and importance of showing 'respect' for working safely.	N/A (due to noise and the fast paced environment, workers can find it difficult to interact with each other)	Drawing attention to land and livestock hazards (risk-related communication).	N/A (truck drivers often work alone with limited opportunity to interact with others)	Ensuring less experienced staff are monitored and supported; using safety practices to embed learning among colleagues.
Safety procedures and processes (e.g., safety training)	N/A	Processes in place to learn from incidents and accepting feedback from external inspectors.	Setup of production lines to optimize efficiency and safety. Also, a second 'documentation' dimension that related to ease of use and accessibility of safety protocols and machine operating standards.	N/A (most safety procedures are informal as the industry is starting to mature and implement safety management systems only recently).	N/A	Regular refreshing and reviewing of operational safety procedures.
Resourcing practices	Staff are given adequate time to complete safety activities and there are enough staff allocate to shifts.	N/A	N/A	Appropriate and well-maintained tools are given to staff.	Safe and maintained vehicles are provided for use.	Quality and adequacy of safety equipment, such as first-aid kits and general investments in safety gear.
Physical environment	N/A	N/A	N/A	N/A	Design of the truck depot for safety and housekeeping quality.	N/A

ment safety commitment, as having more specific items that reflect the policies and practices of management that are unique to the industry will likely increase face validity, respondent engagement, statistical relationships with outcomes, and intervention targeting. For instance, in our project, the police safety climate management commitment dimension included a theme related to authenticity and genuineness, whereas for residential construction, the dimension was more focused on resourcing and pushing back on client pressure. Understanding how each dimension should be conceptualized, defined, and operationalized in specific industries will be an important outcome of initial qualitative work to build these scales.

1.4. Reflections on industry-level findings from safety climate research

Turning our attention to the second objective of this paper (i.e., the cross-industry comparison), when adopting a macro perspective across the six industries, several themes were apparent. First, we discuss the consistencies and differences in safety climate dimensionality across industries. Second, we highlight the differences in referent targets for safety climate items, which we believe emphasizes the relative importance of different stakeholder groups in informing safety climate perceptions. Next, we highlight the nuances within dimensions, with similarly labeled dimensions such as management safety commitment being operationalized in different ways. Fourth, we identified instances where different stakeholders who participated in the inductive phase (i.e., government regulators vs. industry stakeholders) revealed different dimensions of safety climate.

Consistencies and differences in dimensionality. Table 2 shows a summary of the dimensions we discovered for each industry. In line with seminal safety climate research (Flin et al., 2000), management safety commitment and supervisor safety practices were identified as relevant across the most industries. Co-worker safety practices was the next most-commonly identified dimension, which fits with the observation that most industries perform work interdependently in teams. Meat processing and distance-trucking each had distinct barriers to interacting with co-workers—noise levels (inhibiting communication in manufacturing settings) and limited opportunities for social connection (in the case of long-distance transportation). Our findings matched those from Huang et al. (2013a), who did not find evidence of a co-worker specific safety climate dimension but did find that supervisor safety practices were relevant. Some novel and industry-specific dimensions also appeared, namely, 'depot design for safety' in transportation, and the emphasis on staff mental health by supervisors in disability support. Notably, despite the consistency in labeling of each identified dimension, there were differences in how these dimensions should be conceptualized. These differences could be due to industry nuances around how various perceptions are informed. For instance, management safety commitment, although labeled consistently across multiple safety climate scales, appears to differ in terms of how it should be operationalized. According to novel research by Fruhen, Griffin, and Andrei (2019), management safety commitment perceptions are informed by specific practices and actions undertaken by organizational leaders. Although the researchers identified six categories of commitment practices, it is noted that the sample was drawn only from the mining and oil and gas sectors. Thus, the types of commitment-inducing practices may be specific to this setting. Consequently, there may be an opportunity to develop more nuanced operationalizations of management safety commitment across industries, which reflect the specific salient opportunities that leaders have to signal their support for safety.

Different referents. Notably, in agriculture the term 'management' was not well recognized. Supervisors can often also be farm owners and senior managers, outside of corporate entities. This has practical implications for the aggregation of safety climate measures as people may adopt dual roles in this industry. Being clear about the type of key referents, the correct labels used to refer to them (e.g., boss, management, commander, owner/operator) will ensure the integrity of the safety climate scale.

Conceptual nuances within dimensions. Also shown by Table 2 are the nuances in conceptualization that exist between different industry dimensions. These differences can be considered variants of safety climate in that the same dimensions will be expressed through slightly different signals from the various referents. This finding is in line with Zohar (2010) who indicated that industries have unique practices that convey the importance of health and safety. For instance, in the policing sector, interviewees highlighted that management's authenticity was a symbol of safety commitment, whereas, in residential construction, pushing back on client pressures was a sign of commitment. Representing more specific aspects of safety climate within industries is likely to lead to greater face validity, respondent engagement, and higher quality response data.

Stakeholder perspectives on safety climate. Importantly, our study revealed some evidence that different stakeholders provide insights into unique safety climate dimensions. Regulators (who visit many hundreds of workplaces over a year and have a high-level understanding of trends within industries) have a unique perspective. In our study, regulators identified a specific nuance to the organizational learning dimension within residential construction. Specifically, regulators emphasized that 'safe' construction companies are open to feedback from external parties and consider how safety can be improved based on inspection and audit results. Anecdotally, the construction industry is often averse to feedback from external parties as the heavily unionized environment pushes back against stakeholders such as employer associations and regulators. It is possible that safety climate within organizations can in fact be shaped and influenced by parties external to organizations, like accreditation certifiers and regulation inspectors.

2. Conclusions

In this paper, we briefly summarized existing research on industry-specific safety climate scales. Even though the utility of industry specific scales was supported by our work, we do not recommend a wholesale replacement of general scales but rather to make use of each type depending on the situation. If an organization wishes to 'pulse check' (Hahn & Murphy, 2008) or explore associations with major adverse events rather than safety behavior (Jiang et al., 2019), a general safety climate scale is probably most appropriate. A short and well-established general scale for pulse checking is the NIOSH short-form safety climate measure (Hahn & Murphy, 2008) and a longer 'cross-industry safety climate scale' with multiple sub-dimensions was developed by Beus and colleagues (2019) and is likely more suitable for diagnostic work. An industry-specific scale appears most useful when identifying factors for improvement (given the richer level of feedback possible to organizations), evaluating the impact of safety interventions (as context-specific measures are more sensitive to change), and comparing against industry-specific norms and benchmarks. Our experiences with developing six industry-specific safety climate scales confirm the importance of identifying nuances in how the construct should be operationalized, with particular attention paid to the nature of scale dimensions. For instance, management safety commitment in one industry may be measured through items that focus on financial investment in safety, whereas in others, practices like visibility participation in safety activities may be more appropriate. Finally, a key contribution of this commentary is a detailed set of principles and practices that safety climate researchers and practitioners alike can use to develop, evaluate, or refine industry specific safety climate scales. Our hope is that this paper stimulates further research and evidence-based practice in health and safety settings.

Acknowledgements

Center for Work, Health & Safety, NSW Government, funding reference: SWNSW/8775/2018 CWHS_RP_014 safety culture driving machine safety.

References

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management, 53*, 63–70.

Bamel, U. K., Pandey, R., & Gupta, A. (2020). Safety climate: Systematic literature network analysis of 38 years (1980–2018) of research. *Accident Analysis & Prevention, 135* 105387.

Bergman, M. E., Payne, S. C., Taylor, A. B., & Beus, J. M. (2014). The shelf life of a safety climate assessment: How long until the relationship with safety–critical incidents expires? *Journal of Business and Psychology, 29*(4), 519–540.

Beus, J. M., Payne, S. C., Arthur, W., Jr, & Muñoz, G. J. (2019). The development and validation of a cross-industry safety climate measure: Resolving conceptual and operational issues. *Journal of Management, 45*(5), 1987–2013.

Burns, K. E., Duffett, M., Kho, M. E., Meade, M. O., Adhikari, N. K., Sinuff, T., & Cook, D. J. (2008). A guide for the design and conduct of self-administered surveys of clinicians. *CMAJ, 179*(3), 245–252.

Casey, T., Griffin, M. A., Flatau Harrison, H., & Neal, A. (2017). Safety climate and culture: Integrating psychological and systems perspectives. *Journal of Occupational Health Psychology, 22*(3), 341–353. https://doi.org/10.1037/ocp0000072.

Casey, T. W., Neal, A., & Griffin, M. (2019). LEAD operational safety: Development and validation of a tool to measure safety control strategies. *Safety Science, 118*, 1–14.

Choudhry, R. M., Fang, D., & Lingard, H. (2009). Measuring safety climate of a construction company. *Journal of Construction Engineering and Management, 135* (9), 890–899.

Clarke, S. (2010). An integrative model of safety climate: Linking psychological climate and work attitudes to individual safety outcomes using meta-analysis. *Journal of Occupational and Organizational Psychology, 83*(3), 553–578.
Cortina, J. M., Sheng, Z., Keener, S. K., Keeler, K. R., Grubb, L. K., Schmitt, N., ... Banks, G. C. (2020). From alpha to omega and beyond! A look at the past, present, and (possible) future of psychometric soundness in the Journal of Applied Psychology. *Journal of Applied Psychology, 105*(12), 1351.
Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Measuring safety climate: Identifying the common features. *Safety Science, 34*(1–3), 177–192.
Fruhen, L. S., Griffin, M. A., & Andrei, D. M. (2019). What does safety commitment mean to leaders? A multi-method investigation. *Journal of safety research, 68*, 203–214.
Glendon, A. I., & Litherland, D. K. (2001). Safety climate factors, group differences and safety behavior in road construction. *Safety Science, 39*(3), 157–188.
Gray, L. M., Wong-Wylie, G., Rempel, G. R., & Cook, K. (2020). Expanding qualitative research interviewing strategies: Zoom video communications. *The Qualitative Report, 25*(5), 1292–1301.
Griffin, M. A., & Curcuruto, M. (2016). Safety climate in organizations. *Annual Review of Organizational Psychology and Organizational Behavior, 3*, 191–212.
Guldenmund, F. W. (2007). The use of questionnaires in safety culture research–an evaluation. *Safety Science, 45*(6), 723–743.
Hahn, S. E., & Murphy, L. R. (2008). A short scale for measuring safety climate. *Safety Science, 46*(7), 1047–1066.
Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104–121.
Huang, Y. H., Zohar, D., Robertson, M. M., Garabet, A., Lee, J., & Murphy, L. A. (2013). Development and validation of safety climate scales for lone workers using truck drivers as exemplar. *Transportation Research Part F: Traffic Psychology and Behaviour, 17*, 5–19.
Huang, Y. H., Zohar, D., Robertson, M. M., Garabet, A., Murphy, L. A., & Lee, J. (2013). Development and validation of safety climate scales for mobile remote workers using utility/electrical workers as exemplar. *Accident Analysis & Prevention, 59*, 76–86.
Jiang, L., Lavaysse, L. M., & Probst, T. M. (2019). Safety climate and safety outcomes: A meta-analytic comparison of universal vs. industry-specific safety climate predictive validity. *Work & Stress, 33*(1), 41–57.
Johnson, B., & Odhner, K. (2021). Focus groups from home: Conducting virtual focus groups during the COVID-19 pandemic and beyond. *College & Research Libraries News, 82*(6), 258.
Keiser, N. L., & Payne, S. C. (2018). Safety climate measurement: An empirical test of context-specific vs. general assessments. *Journal of Business and Psychology, 33*(4), 479–494.
Kim, N. K., Rahim, N. F. A., Iranmanesh, M., & Foroughi, B. (2019). The role of the safety climate in the successful implementation of safety management systems. *Safety Science, 118*, 48–56.
Lee, J., Huang, Y. H., Cheung, J. H., Chen, Z., & Shaw, W. S. (2019). A systematic review of the safety climate intervention literature: Past trends and future directions. *Journal of Occupational Health Psychology, 24*(1), 66.
Niskanen, T. (1994). Safety climate in the road administration. *Safety Science, 17*(4), 237–255.
Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27.
Parker, A. W., Tones, M. J., & Ritchie, G. E. (2017). Development of a multilevel health and safety climate survey tool within a mining setting. *Journal of Safety Research, 62*, 173–180.
Peer, E., Rothschild, D. M., Evernden, Z., Gordon, A., & Damer, E. (2021). MTurk, Prolific or panels? Choosing the right audience for online research. Choosing the right audience for online research. *Social Science Research Network*. https://doi.org/10.2139/ssrn.3765448.
Reid, D. J., & Reid, F. J. (2005). Online focus groups: An in-depth comparison of computer-mediated and conventional focus group discussions. *International Journal of Market Research, 47*(2), 131–162.
SafeWork NSW. (2018). Work health and safety roadmap 2022. Retrieved from https://www.safework.nsw.gov.au/roadmap.
Shea, T., De Cieri, H., Vu, T., & Pettit, T. (2021). How is safety climate measured? *A review and evaluation. Safety science, 143* 105413.
Silverman, D. (2013). *Doing qualitative research: A practical handbook*. Sage.
Tabachnick, B. G., & Fidel, L. S. (2007). *Using multivariate statistics*. USA: Person Education Inc.
Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.
Wills, A. R., Watson, B., & Biggs, H. C. (2006). Comparing safety climate factors as predictors of work-related driving behavior. *Journal of Safety Research, 37*(4), 375–383.
Zohar, D. (2002). The effects of leadership dimensions, safety climate, and assigned priorities on minor injuries in work groups. *Journal of Organizational Behavior, 23*(1), 75–92.
Zohar, D. (2011). Safety climate: Conceptual and measurement issues. In J. C. Quick & L. E. Tetrick (Eds.), *Handbook of occupational health psychology*. Washington DC: American Psychological Association.
Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. *Accident Analysis & Prevention, 42*(5), 1517–1522.
Zohar, D., & Luria, G. (2005). A multilevel model of safety climate: cross-level relationships between organization and group-level climates. *Journal of applied psychology, 90*(4), 616.

Tristan Casey Dr Tristan Casey (Work Science & Griffith University) is a scientist-practitioner who straddles academia, consulting/industry, and government. He is an expert in project management, relationship formation and maintenance, and intervention design and evaluation. He is internationally renowned for his work on dynamic and situational models of safety leadership, developing the LEAD model and implementing it with over 50 organisations globally. He is also a thought-leader in safety training transfer, safety climate, and organisational culture. Tristan is exceptionally well-connected internationally to OHS institutions and organisations, having worked with a number of professional associations, government bodies, and large multinational companies.

Xiaowen Hu Dr Xiaowen Hu from Queensland University Technology is an emerging thought leader in safety science who brings organisational theory and organisational psychology to address pressing safety challenges. She has published extensively in the area of safety climate and safety leadership. Her current research focuses on cultivating both/and thinking among leaders, safety professionals, and frontline employees to effectively navigate through to AI and digital technology transformation.

Lisette Kanse Lisette's expertise covers human and organisational factors and particularly work design characteristics that influence people's work practices and performance. She has worked in research, training and consultancy in a variety of safety-critical industry settings across Europe and Australia, including chemical industry, oil and gas, mining, rail transport, and hospital settings. Currently, as a senior lecturer at the University of Western Australia, Lisette teaches into several postgraduate courses across psychology, business and engineering, and coordinates the Master of Business Psychology and online Graduate Certificate in Business Psychology courses, and continues her research in work design and work health and safety.

Angelica Varhammar Angelica is a senior research officer with experience in academic and government research, data management and analysis, visualisation and reporting, across both natural and social sciences. This includes agriculture climate change resilience, sustainable development and environmental management, and more recently, behaviour change, regulation and Work Health and Safety.

Journal of Safety Research 82 (2022) 402–408

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Convergent validity of vision based technology (VBT) among professional bus drivers ☆

Rachel Shichrur [a,b,*], Navah Z. Ratzon [b]

[a] Ariel University, School of Health Sciences, the Occupational Therapy Department, Ariel, Israel
[b] Tel Aviv University, Sackler Faculty of Medicine, School of Health Professions, Department of Occupational Therapy, Tel Aviv, Israel

ARTICLE INFO

Article history:
Received 10 August 2021
Received in revised form 18 January 2022
Accepted 12 July 2022
Available online 22 July 2022

Keywords:
In vehicle data recorder
Naturalistic driving
Crash involvement
Professional drivers
Binary-logistic regression model

ABSTRACT

Introduction: Due to the relative rarity of crashes, researchers use traffic offenses, police records, public complaints, and In-Vehicle Data Recorder (IVDR) data as proxies for assessing crash risk. In this study, a unique IVDR system, called Vision-Based Technology [(VBT), (Mobileye Inc.)] was used to monitor perilous naturalistic driving events, such as insufficient distance from other vehicles and pedestrian or bicycle rider near-misses. The study aimed to test the convergent validity of VBT as an indicator of crash involvement risk. *Methods:* Data from 61 professional drivers working for a large bus company were analyzed (16 of 77 in the original data cohort were excluded for insufficient VBT data). Data included: recorded VBT data, objective data collected from official records (crash records provided by the bus company, and public complaints of reckless driving), self-report data regarding crash involvement, and police tickets. The correlation between VBT, objective and self-reported data was analyzed. Binary-logistic regression modeling (BLM) was used to calculate the odds ratio (OR) for participants involved in a car crash. *Results:* Correlations were found between the total VBT risk score and official crash records, public complaints, and self-reports of crash involvement. The BLM correctly classified 90% of those who were involved in a crash (sensitivity) and 60% of those who were "crash-free" (specificity). The VBT total risk score was the only significant contributing factor to crash risk, and for each point of increase, the odds of being involved in a crash increased by a factor of 1.55. *Conclusions:* It is the first study to provide empirical evidence validating the VBT as an indicator of crash involvement and driver safety among professional bus drivers. *Practical Applications:* VBT technology can provide researchers and clinicians a better understanding of bus drivers' risky driving behaviors- a valuable contribution to road safety interventions for this target group.

1. Introduction

Road crashes are frequently characterized as rare events (comparing to near-crashes), requiring several years to gather sufficient data for analysis (Barraclough, af Wahlberg, Freeman, Watson, & Watson, 2016; Wu, Aguero-Valverde, & Jovanis, 2014). Over the last 40 years, the prevalence of events with attributes similar to crashes (including near-crashes observed in intersections, traffic conflicts counts, or evasive actions of drivers; Wu & Jovanis, 2012) has been studied as an alternative indicator of crash risk (e.g., Perkins & Harris, 1968; Evans & Wasielewski, 1982; Reason, 1990; Tarko, Davis, Saunier, Sayed, & Washington, 2009). The study of Wu et al. (2014) reviewed the disadvantages of safety-related event studies, claiming poor-quality data, lack of available and useful exposure measures linked to the observations, subjective evaluations of conflicts, and difficulty assessing culpability. In light of these mentioned disadvantages, Saunier and Sayed (2008) recommended using an automated objective system to evaluate traffic conflicts rather than relying on subjective judgment or witness recollection.

With advances in technology, assessing objective on-road driving safety data is becoming increasingly available with In-Vehicle Data Recorders (IVDRs). IVDR technology examines naturalistic behavior (actual driving) in naturalistic contexts of real traffic, by analyzing the rich stimuli of genuine traffic situations on a regular basis (monitoring hours of continuous driving) (Schmuckler, 2001). As such, data collected by IVDR technology best meets the criteria

☆ This research was supported by the Israel National Road Safety Authority (RSA) [4500386443] and the Ministry of Senior Citizens [3-9673].

* Corresponding author at: Ariel University, School of Health Sciences, the Occupational Therapy Department, Ariel, Israel.

E-mail addresses: rachelsh@ariel.ac.il (R. Shichrur), navah@tauex.tau.ac.il (N.Z. Ratzon).

https://doi.org/10.1016/j.jsr.2022.07.007
0022-4375/© 2022 National Safety Council and Elsevier Ltd. All rights reserved.

of high ecological validity, meaning that the IVDR findings can be generalized to real-life settings in a useful way.

Indeed, data collected by IVDR has been found to be more reliable than other data sources, provides a richer picture of driver behavior in space and time, is less stressful than a driving test, and provides accurate and objective data (Dingus, Neale, Klauer, Petersen, & Carroll, 2006; Ellison, Greaves, & Bliemer, 2015; Shichrur, Sarid, & Ratzon, 2014). The convergent validity of IVDR technology in promoting safety has already been established by several studies. Significant positive correlations were reported between past crash involvement and estimated crash risk as determined by driving profiles derived from IVDR data based on GPS receivers and accelerometers (Musicant, Lotan, & Toledo, 2007; Toledo, Musicant, & Lotan, 2008). Toledo and Lotan (2007) concluded, therefore, that IVDR data are a reliable source for studying driving behavior.

Although these studies established the convergent validity of IVDR technology, they utilized systems that measure speeds and G-forces applied to the vehicle, and evaluated driving behavior only through various events related to these measurements. Ellison et al. (2015) reviewed literature on naturalistic driving studies, including those that employed GPS, accelerometers, video cameras, distance sensors, and on-board diagnostics (OBD). The authors concluded that one of the main drawbacks of naturalistic driving studies is their susceptibility to noise from exogenous factors that may not be measured by any of the sensors in the vehicle, such as the road environment (including the presence of other vehicles on the road).

To address the challenge of including the environmental component along with naturalistic driving data, the present study utilized a unique emerging IVDR technology based on "computer vision." The innovation of computer Vision Based Technology (VBT) lies in examining the relationships within the driver-vehicle-environment complex and building an accurate map of the environment in real-time. VBT is manufactured by Mobileye® and is becoming increasingly used in transportation as the basis for autonomous driving. VBT is designed to measure the distance to other vehicles, lane markings, speed limit signs, pedestrians, bicycle riders, and quickly identify potentially dangerous driving events and situations in the environment as the vehicle moves forward.

One of the most common safety measures used as a crash surrogate in traffic safety assessment is Time-to-Collision (TTC). The computation of the TTC is not trivial, and its definition has changed since the 1970 s. The classical "lane-based" definition of TTC was "*the time required for two vehicles to collide if they continue at their present speed and on the same path*" (Hayward, 1972), while new computational procedures, which take into account lane-changing maneuvers and trajectory conflicts, suggest a modified definition of "*the time it will take a subject vehicle to collide with another vehicle in its immediate vicinity if the present trajectories continue to be followed*" (Hou, List, & Guo, 2014).

Indeed, a forward collision warning component of the VBT detects whether a crash is imminent by computing the TTC, taking into account the host vehicle speed, relative speed, and relative acceleration. The latter two are measured based on the change of the image size of the target (scale change), creating a comprehensive human-level perception of the vehicle's environment and all actionable cues within it. One of the innovations of the VBT is that it functions in a similar way to the human eye, as the scale change of the VBT imitates optical variables of retinal image growth; a driver uses optical information contained in retinal image transformations that occur with movement through the environment to determine they are rapidly approaching a lead vehicle and that the time-to-collision (TTC) is short (Green, 2012; Weinberger, 1971).

Using TTC and the unique features of naturalistic driving enabled by VBT meets the demands of desirable criteria for a crash surrogate, as identified by Wu and Jovanis (2012), such as: having a short period of data collection, being more frequent than car crashes, and having contributing factors similar to a crash. They also defined surrogates as "markers" correlated to a crash, with a time scale underpinning and having a statistical and causal relationship to crashes. Although VBT has been in widespread use over the last 20 years, there is limited evidence in the literature to support the construct validity of this technology. Convergent validity is a subtype of construct validity and it refers to the degree to which two measures of constructs that theoretically should be related, are in fact related. In the present study we test the convergent validity of VBT and its ability to significantly correlate with bus crashes and be used as a safety surrogate measure for crash risk.

Collisions involving bus drivers can endanger dozens of passengers and cause severe direct and indirect damages (Broughton, Baughan, Pearce, Smith, & Buckle, 2003; Mallia, Lazuras, Violani, & Lucidi, 2015). Given the high risk to bus drivers, their passengers, and other road users, it is essential to identify risky driving characteristics and improve the ability to identify risk factors for bus crashes.

2. Materials and methods

2.1. Participants

This is a study of driving data from 77 professional male bus drivers working for a large bus company. The drivers were 27–69 years old ($M = 52.3$, $SD = 9.3$), possessed a driver's license for an average of 32.3 years (range = 12–50, $SD = 9.7$), and worked as a bus driver for an average of 20.9 years (range = 1–45, $SD = 12.7$). The drivers drove approximately-six days per week, working mainly in an urban area. Females did not respond to the invitation to participate in the study, probably due to the low prevalence of females in this profession (only about 1–2%). Inclusion criteria were having a valid bus driver's license and working for the specific bus company. VBT's were installed in 77 buses. Drivers with less than 50 hours of collected VBT data were excluded; 16 drivers were ultimately excluded from the statistical analysis due to insufficient VBT data, leaving a final cohort size of 61 bus drivers ($n = 61$). No statistical difference was found between the two groups with or without the VBT data.

2.2. Measures

2.2.1. IVDR: VBt

The VBT (https://www.mobileye.com/) is an Advanced Driver Assistance System (ADAS) equipped with motion detection algorithms and artificial vision technology, which functions as a "third eye" on the road (Dagan, Mano, Stein, & Shashua, 2004). VBT identifies undesirable events by analyzing raw measurements, and uses this information to indicate the overall trip safety. It can identify objects that may pose a threat to the vehicle, such as other vehicles, bicycles, motorcycles, and pedestrians, in both daytime and nighttime conditions (Gat, Benady, & Shashua, 2005). The technology includes two components: a high-resolution vision sensor and a visual display. The vision sensor, which is a black box about the size of a road toll "tag," is mounted on the inside of the vehicle's front windshield. The black box emits audio warnings for the driver in case of a predicted crash, enabling reaction. Although VBT can provide real-time warning alerts to drivers, the present study focused on the various events recorded by the IVDR operating in "no feedback" mode, in order to test its ability to be used as a proxy for the assessment of the level of risk to which the driver is exposed.

The VBT system continuously measures the distance and relative speeds of objects on the road, predicting their path and calculating the risk of the vehicle colliding with them (Table 1). The VBT total risk score in this study refers to the total mean of all types of VBT events. The VBT is able to function speedily in real time because it detects reflecting light from objects in the environment, as opposed to radar-based technology that is based on detecting radio waves reflected by objects.

2.3. Data collection

The primary source of data was the four event categories of VBT IVDR data: Forward Collision Monitoring (FCM), Urban Forward Collision Monitoring (UFCM), Unsafe Headway Monitoring (HM), and Unsignaled Lane Deviations (LD), as shown in Table 1. Events were recorded as the number of occurrences per hour.

The VBT data were evaluated in comparison to objective data collected from official records and to self-reported data.

2.3.1. Objective risk measures

1. Collision records (number of crashes officially recorded in the previous year prior to the beginning of the study). These data are based on the claims submitted by the bus company to their insurance company, including vehicle damage and injury to people as a result of bus collisions. The measure of "bus crashes" includes only involvement in crashes while driving buses on the job and not while driving a personal passenger vehicle.

2. Public complaints about reckless driving, traffic offenses, or inadequate service ("How am I driving?" sticker with a phone number for reporting attached to the back of the company buses) submitted to the bus company's office. These data refer to the number of complaints recorded with this bus company in the last year prior to the beginning of the study.

2.3.2. Subjective risk measures

Self-reported data obtained from a demographic questionnaire and a driving history profile questionnaire (e.g., driving habits and patterns) were the collected subjective risk data. Participants were also given a list of major violations (such as running a red light, ignoring a stop sign, etc.) and asked to report every police ticket they received since obtaining their driver's license (the variable "total police tickets"). Self-reported data regarding car crash involvement included the number of at-fault crashes with car damage and with injury to people.

2.4. Procedure

The study was approved by a local university Helsinki ethics committee (RMC-0103-10). A random convenience sample of 77 drivers was recruited from a large bus company. All volunteers were informed about the nature and purpose of the study, that participation will not present any risk in terms of their employment, and they gave their informed consent to participate. Following recruitment, bus drivers were asked to complete the demographic and driving history profile questionnaires. The participants were compensated monetarily for their time and traveling expenses related to study participation. VBT IVDRs were installed by the bus company technicians in the participants' vehicles, after which detailed information about unsafe events that occurred during driving was recorded. The buses were monitored during the working shifts while driving on the regular urban routes.

Although the VBT IVDR systems were installed in the buses of all 77 drivers participating in this study, some of the drivers were required, incidentally, by the company to switch to buses without monitoring devices, which resulted in missing data for these drivers. In addition, in several buses, technical problems were encountered as a result of improper installation. Accordingly, IVDR data for a large percentage of drivers in each of the IVDR event categories were lacking (21%–23% of drivers). Sixteen drivers without complete IVDR monitoring data were excluded from the relevant data analyses. To ensure that only the participant drove that vehicle and no other drivers, each driver entered their personal code before driving and all data recorded was synchronized with both vehicle-specific and driver-specific codes.

2.5. Data analysis

A univariate analysis was performed to analyze the correlation between IVDR data, crash records, and traffic safety-related events. An additional sub-analysis was performed to analyze the correlation between VBT total risk score, crash record, and traffic safety-related events, as a factor of driver age. Spearman correlations were used as most variables did not have a normal distribution. We used a binary-logistic regression model (BLM) to study the odds ratio (OR) that a participant would be involved in a car crash. The dependent variable, crash incidence, is presented dichotomously (0 = not involved in a work-related bus crash during the last year before participating in the study, and 1 = involved in one or more work-related bus crashes during the last year before participating in the study). In addition, the independent variable age was transformed to dichotomic variable age (0 below 60 and 1 above 60 years).

ROC curves were derived for significant contributing factors in order to assess their "predictive" power for past crash involvement. To prevent a problem of endogeneity, the collection of crash records was performed the year before entering the study, which was prior to the IVDR data collection.

3. Results

Descriptive statistics of objective risk measures and self-reported data of bus drivers are presented in Table 2. The table presents all drivers participating in the study as no statistical difference was found between the two groups with or without the VBT data. As Table 2 demonstrates, although the number of crashes and public complaints were relatively high, the means were relatively low, amounting to an average of two registered crashes

Table 1
List of potential events that the VBT system records.

Forward Collision Monitoring (FCM)	Urban Forward Collision Monitoring (UFCM)	Unsafe Headway Monitoring (HM)	Unsignaled Lane Deviations (LD)
Obstacle in front of vehicle is < 2.7 sec away or < 1.6 m	Obstacle in front of the vehicle is < 2.7 sec away or < 1.6 m	Obstacle in front of the vehicle is < 1 sec away	Unplanned deviation without signaling
Active at speeds of ≥ 30 kph	Active at speeds of ≤ 30 kph; better adapted for slow speeds / city traffic jams	Active at speeds of ≥ 30 kph	Active at speeds of ≥ 55 kph; better adapted for long-distance travel

Table 2
Descriptive statistics of objective risk measures and self-reported data of bus drivers.

	n	*M*	*SD*	Min	Max
Bus crash records	77	1.96	1.94	0	10
Public complaints	68	1.03	1.41	0	7
LD[1]	59	2.11	1.92	0	8
HM[2]	61	4.90	6.81	0	30.3
UFCM[3]	61	8.99	6.24	0	26.2
FCM[4]	61	4.41	4.88	0	24.3
Age	77	52.28	9.29	27	69
Years of private license	76	32.30	9.67	12	50
Years of public license	75	20.87	12.7	1	45
Days driven per week	77	6.19	0.62	5	7
Total Self-report crashes (with vehicle damage only)	77	8.51	9.52	0	50
Total Self-report at-fault crashes (with vehicle damage only)	76	3.24	5.31	0	40
Total Self-report crashes (with injury to people)	77	1.49	3.02	0	20
Total Self-report at-fault crashes (with injury to people)	77	0.51	1.36	0	8
Total Self-report police tickets	77	5.21	5.26	0	28

[1] LD = lane deviation (lane departure without turn signal).
[2] HM = unsafe headway distance monitoring between the vehicle and other bodies (vehicles, pedestrians etc.).
[3] UFCM = urban forward collision monitoring.
[4] FCM = forward collision monitoring.

and one complaint for each driver per year. Among the four IVDR event categories, the rate of Urban Forward Collision Monitoring (UFCM) was the highest, with a mean of about nine events registered per hour; and unsignaled Lane Deviations (LD) was lowest, with a mean of about two events per hour. In addition, the total number of all crashes (i.e., with vehicle damage and injuries to people) was three times the total number of at-fault crashes.

3.1. Relation between VBT data, objective, and subjective risk measures

As shown in Table 3, a low positive correlation was found between the VBT total risk score and official bus crash records ($r = 0.25$, $p = 0.05$), public complaints ($r = 0.25$, $p = 0.05$), and self-reported at-fault crashes (with vehicle damage only) ($r = 0.33$, $p = 0.009$).

The specific VBT events that were found to be significantly associated with crash involvement were: (a) UFCM (operating at up to 30 kph), which was correlated with crash records ($r = 0.30$, $p = 0.02$) and, (b) unsafe HM, which was correlated with car crash self-reports ($r = 0.39$, $p = 0.002$) and self-reported at-fault crashes (with car damage only) ($r = 0.26$, $p = 0.04$). LD events correlated with self-reported number of police tickets ($r = 0.29$, $p = 0.03$).

For the older drivers (>60 years old), significant positive correlations were found between the VBT total risk score and self-reported total police tickets ($r = 0.57$, $p = 0.01$), self-reported at-fault crashes (car damage only) ($r = 0.58$, $p = 0.009$), and self-reported at-fault crashes (injury to people) ($r = 0.50$, $p = 0.03$).

A significant positive correlation was found between the VBT total risk score and official crash records ($r = 0.32$, $p = 0.04$), and a negative correlation between the VBT total risk score and self-reported total police tickets ($r = -0.32$, $p = 0.04$) for the younger group of drivers (<60 years old).

Next, BLM analysis was used to estimate the odds ratio of participants involved in a crash. The variables included in the model were VBT total risk score, UFCM, age (above and below 60 years),

Table 3
Correlations between VBT data, objective and subjective risk measures ($n = 61$).

Spearman's rho		Objective data		Self-reported data		
		Crash records	Public complaints	Police tickets	At-fault crashes (with vehicle damage only)	At-fault crashes (with injury to people)
VBT[1]	Correlation coefficient	**0.25***	**0.25***	0.01-	**0.33****	0.18
	Sig. (2-tailed)	0.05	0.05	0.91	0.009	0.17
LD[2]	Correlation coefficient	0.02	0.21	**0.29***	0.24	0.23
	Sig. (2-tailed)	0.91	0.12	0.03	0.07	0.07
HM[3]	Correlation coefficient	0.15	0.10	0.15	**0.39****	**0.26***
	Sig. (2-tailed)	0.26	0.48	0.24	0.002	0.04
UFCM[4]	Correlation coefficient	**0.30***	0.18	−0.06	0.23	0.12
	Sig. (2-tailed)	0.02	0.18	0.66	0.07	0.38
FCM[5]	Correlation coefficient	0.22	0.13	−0.19	0.18	0.11
	Sig. (2-tailed)	0.08	0.33	0.15	0.16	0.41

* $p < 0.05$.
** $p < 0.001$.
[1] Vision Based Technology total risk score.
[2] LD = unsignaled lane deviation (lane departure without turn signal).
[3] HM = unsafe headway distance monitoring between the vehicle and other bodies (vehicles, pedestrians etc.).
[4] UFCM = urban forward collision monitoring.
[5] FCM = forward collision monitoring.

Table 4
Logistic regression predicting past bus crash involvement with VBT total risk score and safety-related events ($n = 61$).

Predictor	*B*	*SE*	Wald	*P*	Odds ratio	95% *CI* for EXP(*B*)	
						Lower	Upper
Complaints	0.23	0.27	0.74	0.39	1.26	0.74	2.14
AGE > 60	0.06	0.94	0.00	0.95	1.06	0.17	6.78
At-fault car damage	−8.14	5.31	2.35	0.13	0.00	0.00	9.60
Police tickets (total)	−1.84	2.92	0.40	0.53	0.16	0.00	48.61
VBT total risk score	0.44	0.17	6.93	**0.01**	**1.55**	1.12	2.16
UFCM	0.03	0.10	0.07	0.80	1.03	0.84	1.25
Constant	−0.76	0.96	0.63	0.43	0.47		

Nagelkerke $R^2 = 0.495$, $p < 0.001$.

and total numbers of public complaints, police tickets, and at-fault crashes (with car damage only) Table 4.

A test of the full model versus a model with an intercept only was statistically significant, $p < 0.001$. The model explained 50% (Nagelkerke R^2) of the variance in participants' crash involvement. Although UFCM is one of the components of the VBT total risk score, it was entered in the model after making sure that there was no suspicion of multicollinearity ($r = 0.48$).

Adopting a 0.05 criterion of statistical significance, the VBT total risk score had the only significant partial effects. For each point of increase on the VBT total risk score, the odds of being involved in a crash increased by a multiplicative factor of 1.55.

The model succeeded in classifying correctly 90% of those who were involved in a crash (sensitivity) and 60% of those who were "crash-free" (specificity), with an overall success rate of 81.8%.

In order to check the "predictive" power of the VBT total risk score for past crash involvement, a ROC curve was computed, as shown in Fig. 1, presenting a large area under the curve equal to 0.79, which identified good sensitivity at 94% ($p = 0.00$, $CI = 0.68$–0.91).

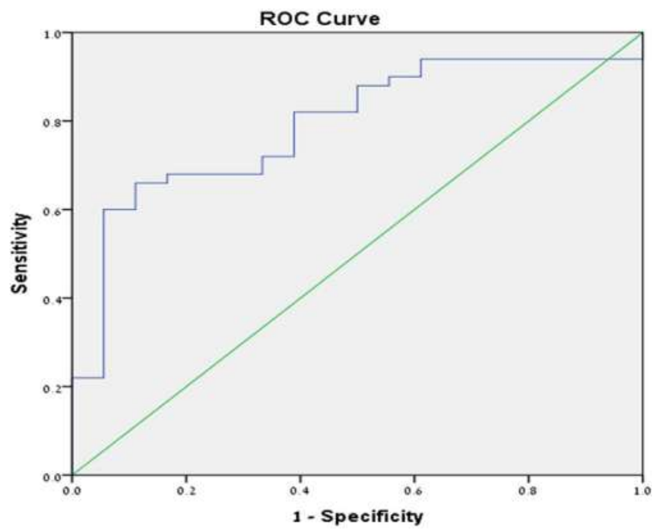


Fig. 1. ROC curve predicting safe and unsafe driver ratings (0 crashes or greater than 0 crashes) for the VBT total risk score.

4. Discussion

This study presents the overview of an IVDR system, VBT, which integrates environmental factors into its analysis. VBT technology incorporates real-time visual recognition and scene interpretation, which help identify objects in the path of the vehicle that may pose threats, especially those common in dense urban areas, in both day- and night-time conditions.

The study findings reflect the convergent validity of the VBT data, as evidenced by the fact that the VBT total risk score and urban forward collision monitoring were significantly correlated with past crash records. All correlations were significant but low.

We used logistic regression analysis to determine the odds ratio of participants being involved in a car crash. According to Jonasson and Rootzén (2014), odds ratios may be suited for external validation using variables that are available in both real crashes and in naturalistic driving studies. The VBT total risk score was the only significant contributing factor of past bus crash involvement, presenting a large area under the curve, equal to 0.79; this identified a large proportion of unsafe drivers, with a sensitivity of 94%.

Validation of the VBT system was carried out on a sample of professional bus drivers who spend many hours on the road. Mallia et al. (2015) proposed that improving public transportation and road safety should be accomplished through a combination of primary elements: (a) *vehicle*—attending to technical aspects, such as improving vehicle safety features; (b) *environment*—improving the bus lanes, traffic conditions, and reducing road congestion; and (c) *person*—driver characteristics and attributes. The VBT system adds a new dimension to the assessment of naturalistic driving by addressing not only the vehicle and person aspects, but also the environment. The importance of this is reinforced by the study of Ellison et al. (2015), whose results showed that even after controlling for the influence of the road environment, environmental factors remain the strongest predictors of driver behavior, suggesting that different spatiotemporal environments elicit a variety of psychological responses in drivers.

Distribution of the data by age (above and below 60 years old) contributed to a better understanding of the correlations between the VBT total risk score and other traffic safety-related events. For the younger group of bus drivers (below 60 years old), the VBT total risk score was positively associated with the number of past crashes, but negatively associated with the number of self-reported police tickets. The negative correlation may be related to a phenomenon reported in the literature of a tendency of male drivers to over-estimate their driving skills. In a study by McKenna, Stanier, and Lewis (1991), male drivers' estimation of their driving skills compared to those of "an average driver" were affected by positive self–bias.

Among older drivers, the VBT total risk score was positively associated with the self-reported total number of police tickets and number of at-fault crashes. Limitations of self-reported behavior data, particularly social desirability and recall biases, are well known (Lajunen & Summala, 2003; Blanchard, Myers, & Porter, 2010; Classen et al., 2010). Given these limitations, many researchers have called for studies to use more objective measures of exposure. The findings of the present study demonstrate that objective VBT data are correlated with other measures of driver safety behavior and crash risk, and can discriminate between self-reports of different age groups. The study results support the significance of the VBT system as a valid indicator of driving safety and crash involvement among professional bus drivers.

Therefore, VBT can upgrade the set of driving assessment tools currently available. At the practical level, using VBT technology can provide better understanding of bus drivers' risky driving behaviors- a valuable contribution to road safety interventions for this target group.

5. Limitations

The study had a few limitations. It was conducted among a small homogenous sample of bus drivers, and therefore should be conducted with other driving populations and a larger sample size in future research to generalize the results better. Additionally, it is not clear to what extent the results are transferable from buses to private vehicles and this should be checked in future studies. Future research is also needed to explore the potential impact of VBT's "immediate feedback" intervention (not utilized in the current study) on reduction of the number of unsafe driving events and to determine if it enables better overall performance of professional drivers in real time. We are aware that gathering official crash data and complaint records for several years prior to the beginning of the study may have provided a more accurate assessment of objective crash history and current driving behavior. However, limiting the data only to the previous year was made to exclude a significant percentage of relatively new drivers (10% of the bus drivers had less than 3 years of possession of a professional driver's license). Finally, although the correlation coefficients of the study variables were statistically significant, some of the values were quite "small" and their interpretation might need to be considered with caution.

6. Conclusions

The current study's findings indicate that VBT system data correlated with other measures of driver safety behavior and crash involvement risk. The BLM explained 50% of the variance in participants' crash involvement. For each point of increase on the VBT total risk score, the odds of being involved in a crash increased by a multiplicative factor of 1.55. The model succeeded in correctly classifying 90% of those who were involved in a crash (sensitivity) and only 60% of those who were "crash-free" (specificity), with an overall success rate of 81.8%. The area under the ROC curve identified good sensitivity at 94% ($p = 0.00$, $CI = 0.68–0.91$) for the VBT total risk score as a contributing factor for past crash involvement. The findings of the present study demonstrate that objective VBT data are correlated with other measures of driver safety behavior and crash risk, and can discriminate between self-reports of different age groups among professional bus drivers.

7. Practical applications

Naturalistic driving experiments are considered more appropriate and accurate for assessing driving behavior than questionnaire surveys and traditional research methods (Papadimitriou, Tselentis, & Yannis, 2018; Tselentis, Vlahogianni, & Yannis, 2018). The future trend appears to be more real-time recording and receiving increased development with the advent of automated vehicles (Ziakopoulos, Tselentis, Kontaxi, & Yannis, 2020). The current research is the first study to provide empirical evidence for validating the VBT IVDR as an indicator for crash involvement and bus driver safety. The VBT total risk score may be a valuable tool for researchers and clinicians to identify risky bus drivers' behavior and safety performance -- a helpful contribution to road safety interventions for this target group.

Acknowledgements

This research was conducted with the financial support of the Israel National Road Safety Authority (RSA) and the ministry of senior citizens. The authors wish to thank Michal Dahari, Shlomit Cohen, Efi Starick, Rannin Sackas-Dibbiny (Tel-Aviv University), Moshe Soroker, Tzuri Barda, and David Fitussi for their assistance and contribution to this research.

References

Barraclough, P., af Wahlberg, A., Freeman, J., Watson, B., & Watson, A. (rraclough et al. 2016). Predicting Crashes Using Traffic Offences. A Meta-Analysis that Examines Potential Bias between Self-Report and Archival Data Retrieved from. *PLoS ONE, 11*(4) https://link-gale-com.mgs.ariel.ac.il/apps/doc/A453417950/AONE?u=cojs&sid=AONE&xid=8e8dfe65.

Blanchard, R. A., Myers, A. M., & Porter, M. M. (2010). Correspondence between self-reported and objective measures of driving exposure and patterns in older drivers. *Accident Analysis and Prevention, 42*, 523–529.

Broughton, J., Baughan. C. J., Pearce, L., Smith, L., & Buckle, G. (2003) *Work-related Road accidents*, Prepared for Road Safety Division, Department for Transport, TRL Report TRL582, ISSN 0968-4107, TRL Limited 2003.

Classen, S., Winter, S. M., Velozo, C. A., Bedard, M., Lanford, D. N., Brumback, B., & Lutz, B. J. (2010). Item Development and Validity Testing for a Self- and Proxy Report: The Safe Driving Behavior Measure. *American Journal of Occupational Therapy, 64*(2), 296–305.

Dagan, E., Mano, O., Stein, G. P., & Shashua, A. (2004). Forward collision warning with a single camera. *Proceedings of the IEEE Intelligent Vehicles Symposium, 37*(42), 14–17.

Dingus, T. A., Neale, V. L., Klauer, S. G., Petersen, A. D., & Carroll, R. J. (2006). The development of a naturalistic data collection system to perform critical incident analysis: An investigation of safety and fatigue issues in long-haul trucking. *Accident Analysis and Prevention, 38*, 1127–1136.

Ellison, A. B., Greaves, S. P., & Bliemer, M. C. (2015). Driver behaviour profiles for road safety analysis. *Accident Analysis and Prevention, 76*, 118–132.

Evans, L., & Wasielewski, P. (1982). Do accident-involved drivers exhibit riskier every-day driving behavior. *Accident Analysis and Prevention, 14*, 57–64.

Gat, I., Benady, M., & Shashua, A. (2005). A Monocular Vision Advance Warning System for the Automotive Aftermarket. *SAE Transactions, 114*(7), 403–410. https://doi.org/10.4271/2005-01-1470.

Green, M. (2012). The Psychology of Rear End Collisions: Looming. Retrieved in 2016 from: http://www.visualexpert.com/Resources/rearendcollision.html.

Hayward, J. C. (1972). Near-miss determination through use of a scale of danger. *Highway Research Record, 384*, 24–34.

Hou, J., List, G. F., & Guo, X. (2014). New algorithms for computing the time-to-collision in freeway traffic simulation models. *Computational intelligence and neuroscience, 2014*, 1–8. https://doi.org/10.1155/2014/761047.

Jonasson, J. K., & Rootzén, H. (2014). Internal validation of near-crashes in naturalistic driving studies: A continuous and multivariate approach. *Accident Analysis and Prevention, 62*, 102–109.

Lajunen, T., & Summala, H. (2003). Can we trust self-reports of driving? Effects of impression management on driver behavior questionnaire responses. *Transportation Research Part F: Traffic Psychology and Behaviour, 6*(2), 97–107.

Mallia, L., Lazuras, L., Violani, C., & Lucidi, F. (2015). Crash risk and aberrant driving behaviors among bus drivers: The role of personality and attitudes towards traffic safety. *Accident Analysis and Prevention, 79*, 145–151.

McKenna, F. P., Stanier, R. A., & Lewis, C. (1991). Factors underlying illusory self-assessment of driving skill in males and females. *Accident Analysis and Prevention, 23*, 45–52.

Musicant, O., Lotan, T., & Toledo, T. (2007). Safety correlation and implication of an In-Vehicle Data Recorder on driver behavior. Preprints of the 86th *Transp Res Board Ann Meet*. Washington DC.

Papadimitriou, E., Tselentis, D. I., & Yannis, G. (2018). Analysis of driving behaviour characteristics based on smartphone data. In *Proceedings of 7th Transport Research Arena TRA* (pp. 16–19).

Perkins, S. R., & Harris, J. L. (1968). Traffic Conflict Characteristics: Accident Potential at Intersections. *Highway Research Board, 225*, 35–43.

Reason, J. T. (1990). *Human Error*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139062367.

Shichrur, R., Sarid, A., & Ratzon, N. Z. (2014). Determining the Sampling Time Frame for In-Vehicle Data Recorder Measurement in Assessing Drivers. *Transportation Research Part C: Emerging Technologies, 42*, 99–106.

Schmuckler, M. (2001). What is ecological validity? A dimensional analysis. *Infancy, 2*(4), 419–436.

Tarko, A. P., Davis, G., Saunier, N., Sayed, T., & Washington, S. (2009). *Surrogate Measures of Safety. White Paper. Subcommittee on Surrogate Measures of Safety and Committee on Safety Data Evaluation and Analysis*. Washington D.C: Transportation Research Board.

Toledo, T., & Lotan, T. (2006). An In-Vehicle Data Recorder for Evaluation of Driving Behavior and Safety. *Transportation Research Record: Journal of the Transportation Research Board, 1953*(1), 112–119.

Toledo, T., Musicant, O., & Lotan, T. (2008). In-vehicle data recorders for monitoring and feedback on drivers' behavior. *Transportation Research Part C: Emerging Technologies, 16*, 320–331.
Tselentis, D. I., Vlahogianni, E. I., & Yannis, G. (2018). Comparative evaluation of driving efficiency using smartphone data. *Methodology, 32*, 33.
Weinberger, H. (1971). Conjecture on the visual estimation of relative radial motion. *Nature, 229*. 562–562.
Wu, K. F., & Jovanis, P. P. (2012). The relationship between crashes and crash-surrogate events in naturalistic driving data. *Accident Analysis and Prevention, 45*, 507–516.
Wu, K. F., Aguero-Valverde, J., & Jovanis, P. P. (2014). Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level. *Accident Analysis and Prevention, 72*, 210–218.
Ziakopoulos, A., Tselentis, D., Kontaxi, A., & Yannis, G. (2020). A critical overview of driver recording tools. *Journal of safety research, 72*, 203–212.

Rachel Shichrur is a Lecturer and a researcher at the Occupational Therapy Department of Ariel University in Israel. Her researches are in the field of driving safety and advanced technologies such as In-Vehicle Data Recorder (IVDR) and in-vehicle cameras. She completed her postdoctoral studies in 2018 in the department of Industrial Engineering and Management at Ben Gurion University in Beer Sheva combined with the Department of Occupational Therapy at Tel Aviv University. Her areas of expertise include the statistical analysis of crash data sets and naturalistic data, experimental design, evaluation of driving-safety technologies, measurement of driver performance, and human-machine interface.

Navah, Z. Ratzon is a Full Professor in the Occupational Therapy Department at the Tel Aviv University, Israel, and since 2020 Head of the Stanley Steyer School of Health Professions Sackler Faculty of Medicine at Tel Aviv University. She has an Occupational Therapy diploma from the Hebrew University of Jerusalem (HUJ), Israel (1978), a MA (1988) from the New York University, New York, U.S.A., and a Ph.D. from the HUJ (1996). Her work implements advanced technology such as driving simulators and In-Vehicle Data Recorders (IVDRS), which enables off-road and on-road evaluation and monitoring of long periods of driving. Her studies refer to various populations such as professional drivers, youth with ADHD, people with mental illnesses, and people with motor disabilities. She is the author of more than 100 articles. Prof. Ratzon was a recipient of the Israeli Occupational Therapy Association Award for Excellence in 2012.

Journal of Safety Research 82 (2022) 293–300

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Derived patterns of musculoskeletal symptoms and their relationships with ergonomic factors among electronic assembly workers: A latent class analysis

Yidan Dong [a], Ping Jiang [a], Xu Jin [a], Nazhakaiti Maimaiti [a], Shijuan Wang [a], Liyun Yang [b,c], Mikael Forsman [b,c], Lihua He [a,*]

[a] Department of Occupational and Environmental Health, School of Public Health, Peking University, Beijing 100191, China
[b] Institute of Environmental Medicine, Karolinska Institutet, 17177 Stockholm, Sweden
[c] Division of Ergonomics, KTH Royal Institute of Technology, 14157 Huddinge, Sweden

A R T I C L E I N F O

Article history:
Received 23 June 2021
Received in revised form 8 January 2022
Accepted 7 June 2022
Available online 18 June 2022

Keywords:
Musculoskeletal symptoms
Latent class analysis
Electronic assembly workers
Pain patterns
Risk factors

A B S T R A C T

Introduction: Multi-site musculoskeletal symptoms (MSS) are considered to be more common and have more serious consequences than single-site MSS. This study aimed to determine whether derived patterns of MSS may be identified in electronic assembly workers and if extracted MSS classes are associated with personal and work-related factors. *Method:* A cross-sectional questionnaire study was performed with 700 participating electronic assembly workers. The questionnaire included individual factors, psychosocial and physical exposures, and MSS. The derived patterns of MSS and their relationships with ergonomic factors were analyzed using latent class analysis (LCA) and multinomial logistic regression models (MLRM). *Results:* The 1-year prevalence of MSS affecting only one body site or two or more body sites was 14.9% and 32.7%, respectively. The results of LCA showed three distinct classes of MSS patterns, which were labelled 'MSS in most sites' (5.0%), 'MSS in neck and shoulder' (27.0%), and 'MSS in one or no site' (68.0%). The results of MLRM showed that the 'MSS in neck and shoulder' was associated with job tenure (OR 5.579, 95% CI 2.488–12.511), excessive dynamic and static loads (OR 3.868, 95% CI 1.702–8.793 and OR 5.270, 95% CI 2.020–13.747, respectively); while the 'MSS in most sites' was associated with high job demands (OR 4.528, 95% CI 1.647–12.445) and excessive dynamic loads (OR 111.554, 95% CI 4.996–2490.793). *Conclusions:* The results showed unique patterns of MSS among electronic assembly workers that were associated with personal and work-related factors. *Practical applications:* The findings highlight that the high prevalence of multi-site MSS in this group should be a focus. It also provides further evidence that LCA considering the number and location of anatomical sites involving MSS can be used to determine distinct classes of MSS patterns, which is of great significance for the epidemiological study and management of MSS in the future.

1. Introduction

Work-related musculoskeletal disorders and symptoms (MSS) are common public health problems. They are impairments of the bodily structures, such as muscles, joints, tendons, ligaments, nerves, bones and the localized blood circulation system, which are caused or aggravated primarily by the performance of work and by the effects of the immediate environment in which work is carried out (European Agency for Safety and Health at Work, 2019). Around three out of every five workers in the European Union (EU)-28 reported musculoskeletal complaints in the year 2015 (European Agency for Safety and Health at Work, 2019). These diseases will not only lead to sickness absence and incapacity for work, but also impart a substantial economic burden on society. It is estimated that musculoskeletal disorders affect at least 100 million people in Europe, accounting for 50% of all European absences from work and for 60% of permanent work incapacity (Cammarota, 2007). The total costs of lost productivity attributable to musculoskeletal disorders was approximately 240 billion euros in the EU, 980.1 billion dollars in the United States, and 6.89 billion dollars in Korea (Bevan, 2015; Oh, Yoon, Seo, Kim, & Kim, 2011; Yelin, & Cisternas, 2014).

It is noteworthy that the majority of available studies concentrating on the occurrence of MSS have focused on a specific

* Corresponding author.
E-mail address: alihe2009@126.com (L. He).

https://doi.org/10.1016/j.jsr.2022.06.004
0022-4375/

anatomical site. However, MSS usually occur in several anatomical locations and MSS at one site are associated with an increased occurrence of MSS at another site (Haukka et al., 2006). Recent studies have highlighted the investigation of multi-site MSS in the general and occupational population, indicating the moderate prevalence of single-site MSS (estimated prevalence of 16.8–20.3% in different studies) and the obvious prevalence of multi-site MSS (estimated prevalence of 19.7–53.0% in the general population and 41.3–73.0% in the occupational population, respectively) (Hartvigsen, Davidsen, Hestbaek, Søgaard, & Roos, 2013; Haukka et al., 2006; Kamaleri, Natvig, Ihlebaek, & Bruusgaard, 2008; Larsen, Andersson, Tranberg, & Ramstrand, 2018; Neupane, Nygård, & Oakman, 2016). Furthermore, multi-site MSS are considered to have more serious consequences than single-site MSS. Some studies have shown that the association between MSS and poor work ability would be stronger when the number of pain sites increased (Neupane, Miranda, Virtanen, Siukola, & Nygård, 2011; Phongamwong & Deema, 2015). A prospective study among workers in the food industry demonstrated that sickness absences due to MSS increase with the increase in the number of pain sites (Neupane et al., 2015). Another prospective study found a strong "dose–response" relationship between the number of pain sites and disability with a 10-fold increase from 0 to 9–10 pain sites (Kamaleri, Natvig, Ihlebaek, & Bruusgaard, 2009).

It is generally believed that musculoskeletal disorders are of multi-factorial origin, which are associated with individual characteristics as well as biomechanical and psychosocial factors (da Costa & Vieira, 2010; Gatchel & Schultz, 2012; Yu et al., 2012). Although a large number of studies have provided reasonable evidence that work-related physical and psychosocial factors are associated with musculoskeletal disorders, the relationship between these factors and multi-site musculoskeletal disorders is rarely reported. Previous studies have used 2 by 2 combinations or the number of pain sites as the classification indicator of multi-site MSS (Freimann, Coggon, Merisalu, Animägi, & Pääsuke, 2013; Haukkal et al., 2011; Hoe, Kelsall, Urquhart, & Sim, 2012; S. Neupane et al., 2016). However, different studies have a different cut-off point for multi-site MSS ($\geq$2 or $\geq$3 or $\geq$4 pain sites), and the cut-off point for multi-site MSS lacks a unified standard (Freimann et al., 2013; Haukkal et al., 2011; Neupane et al., 2016; Oakman, de Wind, van den Heuvel, & van der Beek, 2017). In addition, some studies have suggested that there is a correlation between pain sites, and hence multi-site MSS should not be defined as a simple accumulation of each site (de Cássia Pereira Fernandes, da Silva Pataro, de Carvalho, & Burdorf, 2016).

Latent class analysis (LCA) is a statistical method used for identifying the latent structures (or classes) in data, which assumes that observed variables are indicators of an unobserved, latent variable and attempts to explain this relationship in terms of a small number of subgroups or classes (Hagenaars, & McCutcheon, 2002; Carragher, Adamson, Bunting, & McCann, 2009). In recent years, LCA has been widely used in psychology, sociology, and preventive medicine (Carragher et al., 2009; Lanza & Rhoades, 2013; Wang & Hanges, 2010). However, few studies have analyzed major patterns of MSS and classified the study population into several more homogenous subgroups using LCA (Hartvigsen et al., 2013; Molgaard Nielsen, Hestbaek, Vach, Kent, & Kongsted, 2017; Yazdi, Karimi Zeverdegani, & MollaAghaBabaee, 2019). Therefore, this study aims to assess whether specific 'classes' of MSS may be identified in electronic assembly workers. A secondary aim is to examine whether the extracted MSS classes are associated with personal and work-related factors.

2. Methods

2.1. Participants

This cross-sectional questionnaire was conducted among 928 participants in three electronic accessories processing enterprises in Beijing between June 2017 and July 2017. Further information about the cross-sectional study has been reported previously (Maimaiti et al., 2019).

In this study, participants over the age of 18 years who had worked in the industry for at least 1 year were recruited. Those who have been diagnosed with musculoskeletal injuries, rheumatoid arthritis, tumors, tuberculosis, infections, autoimmune diseases, and other diseases affecting the musculoskeletal system were excluded. A total of 928 participants were eligible to participate, and 752 of them gave informed consent and returned filled questionnaires. Of the 752 returned questionnaires, 700 questionnaires were valid, giving a 93.1% efficient rate. Approval for this study was obtained from the Medical Ethics Committee of Peking University (IRB0000105216015).

2.2. The questionnaire

The self-administered Chinese Musculoskeletal Questionnaire (CMQ) was used for evaluating MSS and ergonomic factors in the workplace, which has previously been tested for reliability and validity (Wang et al., 2017). The questionnaire includes personal factors, work-related factors, and MSS. The variables are reported as follows.

Personal factor variables: gender (male, female), age (years old), job tenure (years), body mass index (BMI) (kg/m^2), education (junior middle school or below, senior high school, junior college, bachelor degree or above), monthly income ($\leq$2,000 RMB, 2,001–4,000 RMB, 4,001–5,000 RMB, $\geq$5,001 RMB), physical exercise (never, 1–3 times/quarter, 2–3 times/month, 1–2 times/week, more than 3 times/week), smoking (yes, no), and drinking behaviors (yes, no).

Work-related factor variables: postural factors, psychosocial factors, and work environmental factors. Postural factors were composed of several items on each body region, which were modified from Rapid Upper Limb Assessment (Stanton, Hedge, Hendrick, Salas, & Brookhuis, 2004). Psychosocial factors mainly included job demands, social support, and job control, which were selected from the full version of the Karasek Job Content Questionnaire (Karasek et al., 1998). Job demands, social support, and job control were dichotomized at the 75th percentile into 'high' and 'low' exposures; all values at the 75th percentile and above were considered as high exposure (Sembajwe et al., 2013). All responses were made on a 5-point scale (ranging from 'never' to 'always'). Work environmental factors were collected by asking the participants about every aspect of the work environment with the 5-point scale (ranging from 'very dissatisfied' to 'very satisfied'). The questionnaire also asks whether the participants use hand-held vibration tools at work.

MSS variables: participants were asked if they experienced MSS such as pain, discomfort, numbness, or limitation of movement during the past 12 months in a body map with nine body sites (neck, shoulders, upper back, low back, elbows, wrists/hands, hips/thighs, knees and ankles/feet). Symptoms in the past 12 months were assessed by self-reported symptom frequency (no pain, 1–2 times/year, 1–2 times/quarter, 1–2 times/month, once a week, almost every day), symptom duration (no pain, less than an hour, less than a day, less than a week, less than a month, more than a month) and symptom intensity (a 0–10 visual analogue scale: 0 mark as be painless, 1 to 3 marks as mild pain, 4 to 6 marks as moderate pain, 7 to 9 severe pain, and 10 marks as maximum pain). The design of this domain was in accordance with the Standardized Nordic Musculoskeletal Questionnaire (Crawford, 2007). MSS were defined as positive if participants had MSS such as pain, discomfort, numbness, or limitation of movement during the past 12 months, which lasted for more than 24 h and had no relief after rest.

2.3. Data analyses

LCA was conducted with Mplus 7.0 (Muthén and Muthén, Los Angeles, CA, USA) to identify derived patterns of MSS. First, a series of models were run to evaluate the potential classes within the data. The optimal number of classes was determined by comparing the Bayes information criterion (BIC), the sample-size adjusted BIC (aBIC), the Akaike's information criterion (AIC), Pearson chi-square test (Pearson χ^2), likelihood ratio chi-square test (G^2), entropy and the Lo-Mendell-Rubin likelihood ratio test (LMR *P*-value) (Lanza, Collins, Lemmon, & Schafer, 2007). Recommendations suggest that the model with the smallest BIC, aBIC, AIC, Pearson χ^2 and G^2, significant LMR *P*-value comparing the k and $k-1$ class model, and Entropy with values closer to 1 (range, 0–1) should be selected (Collins & Lanza, 2010; Lanza et al., 2007).

For each class in the chosen LCA model, item conditional probabilities of pain sites give the probabilities that a participant in that class reported pain at specific site(s). These probabilities were examined to determine the class-specific characteristics of pain sites, with each class allocated informative names according to an arbitrary cut-off of probability of pain at each site of $\geq$0.5 (Lacey et al., 2015).

After selecting the optimal model and number of classes, participants were assigned to the class for which they have maximum posterior probability (Lanza et al., 2007). Class distinction was measured using class average posterior probabilities, where a value of above 0.7 indicates clear separation (Lacey et al., 2015). Class membership variables were retained for subsequent analyses.

Descriptive statistics and multinomial logistic regression were performed using SPSS 23.0 (IBM, New York, NY, USA). The demographic distribution of MSS classes was presented as frequencies and percentages. The prevalence of MSS classes in different groups was compared with Chi-square test. The association of MSS classes with personal and work-related factors was analyzed using multinomial logistic regression models (MLRM), presented as adjusted odd ratios (ORs) with 95% confidence intervals (CIs).

3. Results

3.1. Descriptive analysis

Of the study participants, 51.7% (362/689) were female. The median age of participants was 26.5 years with an interquartile range (IQR) of 23–30 years. About 89.0% of participants (544/611) worked in their current positions for less than 5 years. The number of participants who were overweight and obese accounted for 18.9% (131/693) and 5.5% (38/693), respectively. It was observed that 93.9% of the participants (636/677) were less-educated with an educational level below college, and 93.3% of participants (596/639) had monthly income of less than 5,000 RMB. Physical exercise was reported by 72.4% of participants (470/649). In this study, 20.5% of participants (143/697) had smoking behaviors and 18.7% of participants (129/691) had drinking behaviors.

The prevalence of MSS in only one site and multiple body sites is shown in Fig. 1. The 1-year prevalence of MSS affecting only one body site and two or more body sites was 14.9% (104/700) and 32.7% (229/700), respectively. The prevalence of MSS in two or more body sites was two times more common than prevalence of MSS in only one site.

3.2. Characteristics of the classes

The MSS major patterns and latent structure or unobserved heterogeneity of the study population were recognized using LCA. Table 1 provides fit statistics of LCA model for different numbers of classes. A three-class model was identified as the optimal model, with BIC being lowest for this solution (BIC = 4264.703).

Fig. 2 shows the class-specific probability for having symptoms at each site, given membership for each of the three classes. Class 1 (n = 35, 5.0%) had high probabilities of symptoms in all sites, was characterized by participants with MSS in most body sites (median 8 sites, IQR 7–9), and was labelled 'MSS in most sites.' Class 2 (n = 198, 28.3%) was represented by participants with MSS in a median of 3 (IQR 2–4) body sites, and a high probability of having neck and shoulder symptoms, so was labelled 'MSS in neck and shoulder.' And, class 3 (n = 467, 66.7%) accounted for more than half of the sample, which had low probabilities of symptoms in all sites, was marked by participants with MSS in one or no body site (median 0 sites, IQR 0-0), and was labelled 'MSS in one or no site.'

The actual classification results of participants are presented in Table 2. Take the first participant as an example, the posterior probability of class 3 was the highest, so the first participant was assigned to the 'MSS in one or no site.' Actual class counts and proportions based on the estimated posterior probabilities were as follows: there were 35 (5.0%) participants in class 1, 189 (27.0%) participants in class 2, and 476 (68.0%) participants in class 3. The average posterior probabilities for all three classes exceeded 0.7 (0.936 for class 1, 0.939 for class 2, and 0.962 for class 3), indicating accurate classification of the participants to the correct class.

3.3. Association of classes with personal and work-related factors

See Table 3 for the demographic distribution of MSS patterns. The results of chi-square test showed that there were statistically significant differences in gender, job tenure, and smoking among different classes (p < 0.05).

Taking all the personal and work-related factors into consideration, the results of MLRM showed that there were five variables with statistical significance in the final model, which are presented in Table 4. The risk of MSS in neck and shoulder were associated with: 6–10 years of tenure (OR = 5.579, 95%CI = 2.488–12.511), at work often twisting arms frequently (OR = 3.868, 95%CI = 1.702–8.793), always bending neck in a forward posture for long periods (OR = 5.270, 95%CI = 2.020–13.747), and often bending neck in a forward posture for long periods (OR = 4.150, 95%CI = 1.744–9.877). The risk of MSS in most sites was associated with high job demand (OR = 4.528, 95%CI = 1.647–12.445), always raising arms frequently (OR = 111.554, 95%CI = 4.996–2490.793), and seldom raising arms frequently (OR = 4.307, 95%CI = 1.072–17.293).

4. Discussion

In this study, we estimated the prevalence of multi-site MSS, and evaluated major patterns of MSS and their association with personal and work-related factors among electronic assembly workers. The total 12-month prevalence of MSS in multiple body sites (two or more) was more prevalent than that in only one body site, which was in agreement with other studies, revealing that it is necessary to pay attention to multi-site MSS in future studies (Bæk

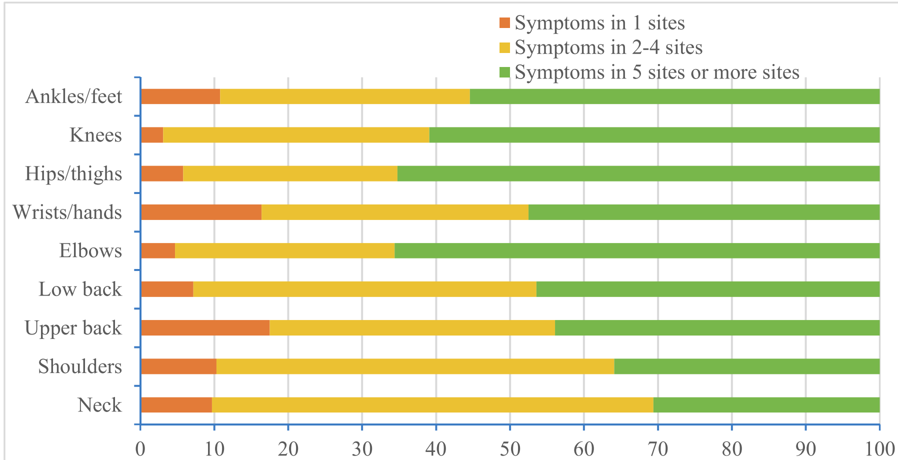


Fig. 1. Prevalence of MSS by the number of sites involving with symptoms: only one site, two to four sites, and five or more body sites.

Table 1
Fit indices of different latent class analyses.

Number of latent classes	Number of parameters estimated	AIC	BIC	aBIC	Pearson χ^2	G^2	Entropy	LMR *P*-value
1	9	5339.227	5380.187	5351.610	2440.972	900.412	–	–
2	19	4286.257	4372.720	4312.399	1715.727	509.589	0.877	<0.001
3	29	4132.722	4264.703	4172.622	658.292	325.699	0.887	<0.001
4	39	4105.066	4282.558	4158.726	474.813	277.966	0.901	0.075

Abbreviations: AIC: Akaike's information criterion, BIC: Bayes information criterion, aBIC: sample-size adjusted BIC, Pearson χ^2: Pearson chi-square test, G^2: likelihood ratio chi-square test, LMR: Lo-Mendell-Rubin likelihood ratio test.

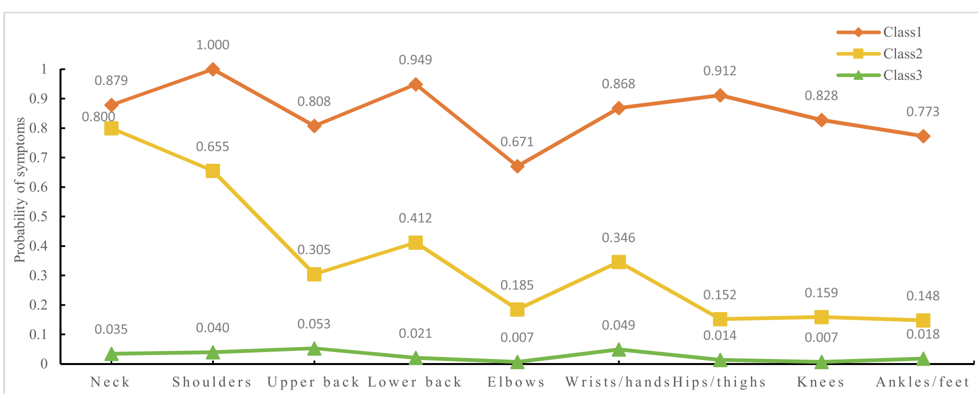


Fig. 2. Class-specific probability of symptoms (0–1) in different body sites according to class membership.

Larsen, Ramstrand, & Fransson, 2018; de Cássia Pereira Fernandes et al., 2016; Neupane et al., 2016). One explanation for this phenomenon may be that shared risk factors for musculoskeletal pain act at more than one body site, and another explanation may be an underlying generalized vulnerability to chronic pain (Croft, Dunn, & Von Korff, 2007).

A new finding in our study pointed out that electronic assembly workers can be divided into three distinct classes using LCA, based on spatial patterns and numbers of body sites involving with symptoms. Class 1 was characterized by participants with high probabilities of symptoms in all sites and was labelled 'MSS in most sites.' Class 2 was recognized by participants with high probabilities of symptoms in the neck and shoulder, was labelled 'MSS in neck and shoulder.' Class 3 was represented by participants with low probabilities of symptoms in all sites and was labelled 'MSS in one or no site.' The proportion of actual classification from class 1 to class 3 was 5.0%, 27.0%, and 68.0%, respectively, while the corresponding latent class probability was 5.0%, 28.3%, and 66.7%, respectively, indicating there may be a misclassification in class 2 and class 3. However, Lubke et al. pointed out that entropy values around 0.80 and above are related to at least 90% correct assignment (Lubke & Muthén, 2007). In addition, some studies showed that average posterior probabilities above 0.7 represent that the classes are clearly separated (Lacey et al., 2015). In this study, the entropy was greater than 0.80 and the average posterior probabilities of three classes were greater than 0.90. Therefore, although there were some misclassifications, it could be considered that the classification result was ideal. Compared with workers having MSS in most sites, the proportion of workers having MSS in neck and shoulder was higher, which may partly depend on their occupational characteristics (Maimaiti et al., 2019).

Table 2
The actual classification results of participants based on the estimated posterior probabilities.

Original data of 9 sites	Posterior probability			Classification results
	Class 1	Class 2	Class 3	
000,000,000	0.000	0.005	0.995	Class 3
000,000,000	0.000	0.005	0.995	Class 3
000,000,000	0.000	0.005	0.995	Class 3
000,000,000	0.000	0.005	0.995	Class 3
101,100,000	0.000	0.993	0.007	Class 2
000,000,000	0.000	0.005	0.995	Class 3
000,000,000	0.000	0.005	0.995	Class 3
000,000,000	0.000	0.005	0.995	Class 3
000,000,000	0.000	0.005	0.995	Class 3
010,000,000	0.000	0.182	0.818	Class 3
010,100,000	0.000	0.880	0.120	Class 2
000,000,000	0.000	0.005	0.995	Class 3
001,000,000	0.000	0.037	0.963	Class 3
000,000,000	0.000	0.005	0.995	Class 3
001,000,000	0.000	0.037	0.963	Class 3
000,000,000	0.000	0.005	0.995	Class 3
000,000,000	0.000	0.005	0.995	Class 3
000,000,000	0.000	0.005	0.995	Class 3
000,000,000	0.000	0.005	0.995	Class 3
110,111,111	0.996	0.004	0.000	Class 1
111,011,000	0.004	0.996	0.000	Class 2
111,100,010	0.021	0.979	0.000	Class 2
. . ..				
000,001,000	0.000	0.049	0.951	Class 3

Some studies have pointed out that multi-site MSS are a continuum of single-site MSS, maintained by exposure to several risk factors, rather than the result of a specific risk factor that initiates the multi-site MSS but not single-site MSS; this is similar to the findings of our study, showing that multi-site MSS are associated with biomechanical, psychosocial, and personal factors (de Cássia Pereira Fernandes et al., 2016; Herin et al., 2014; Neupane, Miranda, Virtanen, Siukola, & Nygård, 2013). As for biomechanical factors, there is evidence that exposure to awkward postures, repetitive motion, and forceful exertions were associated with musculoskeletal disorders at one or more anatomical site (da Costa & Vieira, 2010; Neupane et al, 2013; Punnett & Wegman, 2004). Our findings provide new support to the link between awkward postures and the occurrence of multi-site MSS. The results of MLRM in our study showed that the occurrence of MSS in neck and shoulder was associated with bending neck forward for long periods and twisting one's arms frequently; and the occurrence of MSS in most sites was associated with raising one's arms frequently, which is in accordance with earlier studies on sedentary workers (Dong, Zhang, Liu, & Shao, 2020; Keester & Sommerich, 2017; Pope-Ford & Jiang, 2015). A plausible hypothesis is that electronic assembly workers usually maintain seated positions, which may elevate upper trapezius exertion and contribute to the development of musculoskeletal disorders (Pope-Ford & Jiang, 2015). In addition, a person working in an awkward posture will need to use more force to finish the same amount of task, which in turn increases the muscle loading and compressive stress on the vertebral disc (Anderson, Chaffin, & Herrin, 1986).

With regard to psychosocial factors, it has been suggested that high job demands, low job control, and low social support may increase strain and subsequently increase muscle tension or other physiological reactions that put individuals at a greater risk for developing musculoskeletal disorders (Bongers, de Winter, Kompier, & Hildebrandt, 1993). In our study, although we did not observe that the occurrence of MSS in neck and shoulder was associated with job demands, there was a positive relationship between the occurrence of MSS in most sites and job demands.

Table 3
The demographic distribution of classes.

Variables	Class 1 MSS in most sites (n = 35)	Class 2 MSS in neck and shoulder (n = 189)	Class 3 MSS in one or no site (n = 476)	χ^2	P-value
Gender [a]				26.402	<0.001***
Male	17 (5.2)	59 (18.0)	251 (76.8)		
Female	18 (5.0)	128 (35.4)	216 (59.7)		
Age (years old) [a]				3.140	0.535
≤20	5 (5.4)	20 (21.7)	67 (72.8)		
21–30	23 (4.9)	123 (26.2)	323 (68.9)		
≥31	7 (5.2)	43 (31.9)	85 (63.0)		
Job tenure (years) [a]				12.032	0.012*
1–5	29 (5.3)	139 (25.6)	376 (69.1)		
6–10	1 (2.3)	21 (47.7)	22 (50.0)		
≥11	2(8.7)	9 (39.1)	12 (52.2)		
BMI (kg/m^2) [a]				5.534	0.477
<18.5	2 (2.4)	21 (25.0)	61 (72.6)		
18.5–23.9	27 (6.1)	119 (27.0)	294 (66.8)		
24–27.9	5 (3.8)	33 (25.2)	93 (71.0)		
≥28	1 (2.6)	14 (36.8)	23 (60.5)		
Education [a]				12.209	0.057
Junior middle school or below	8 (6.2)	32 (24.6)	90 (69.2)		
Senior high school	24 (5.4)	116 (25.9)	308 (68.8)		
Junior college	1 (1.7)	17 (29.3)	40 (69.0)		
Bachelor degree or above	1 (2.4)	20 (48.8)	20 (48.8)		
Monthly income (RMB)				7.640	0.227
≤2,000	0	1 (20.0)	4 (80.0)		
2,001–4,000	24 (4.8)	136 (27.4)	336 (67.7)		
4,001–5,000	7 (7.4)	22 (23.2)	66 (69.5)		
≥5,001	0	18 (41.9)	25 (58.1)		
Exercise [a]				8.150	0.419
Never	14 (7.8)	48 (26.8)	117 (65.4)		
1–3 times/ quarter	3 (3.3)	32 (35.6)	55 (61.1)		
2–3 times/month	6 (4.4)	36 (26.7)	93 (68.9)		
1–2 times/ week	8 (4.7)	42 (24.6)	121 (70.8)		
More than 3 times/week	2 (2.7)	19 (25.7)	53 (71.6)		
Smoking [a]				7.230	0.027*
No	26 (4.7)	162 (29.2)	366 (66.1)		
Yes	9 (6.3)	26 (18.2)	108 (75.5)		
Drinking [a]				3.456	0.178
No	24 (4.3)	157 (27.9)	381 (67.8)		
Yes	10 (7.8)	30 (23.3)	89 (69.0)		

Abbreviations: BMI: body mass index.
[a] Variables with missing values.
* $p < 0.05$.
*** $p < 0.001$.

Our results indicated that high job demands may affect multiple anatomical sites simultaneously, which is closely related to the occurrence of multi-site MSS, not just a single site. For the other

Table 4
The results of multinomial logistic regression describing the association between influencing factors and classes.

Variables	Categories	Class 2 MSS in neck and shoulder (n = 189) OR (95%CI) [a]	Class 1 MSS in most sites (n = 476) OR (95%CI) [a]
Job tenure (1–5 years)	≥11 years	2.357(0.789–7.042)	1.970 (0.347–11.172)
	6–10 years	5.579(2.488–12.511) ***	0.761 (0.088–6.564)
Job demands (low)	high	0.935 (0.536–1.631)	4.528 (1.647–12.445) **
Twisting one's arms frequently (never)	always	1.307 (0.347–4.923)	0.579 (0.046–7.274)
	often	3.868 (1.702–8.793) **	0.522 (0.096–2.834)
	sometimes	0.631 (0.246–1.623)	0.575 (0.121–2.737)
	seldom	0.951 (0.472–1.915)	0.276 (0.063–1.210)
Raising one's arms frequently (never)	always	10.559 (0.978–114.048)	111.554 (4.996–2490.793) **
	often	1.833 (0.681–5.203)	5.525 (0.769–39.714)
	sometimes	1.318 (0.562–3.094)	4.488 (0.963–20.923)
	seldom	1.548 (0.800–2.997)	4.307 (1.072–17.293) *
Bending neck in a forward posture for long periods (never)	always	5.270 (2.020–13.747) **	3.446 (0.321–36.982)
	often	4.150 (1.744–9.877) **	3.899 (0.420–36.225)
	sometimes	0.970 (0.345–2.730)	4.826 (0.485–47.997)
	seldom	1.090 (0.408–2.909)	3.162 (0.309–32.306)

[a] Compared with class 3 (MSS in one or no site, n = 35).
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

psychosocial factors assessed in the present study, low level of job control and social support were not associated with multi-site MSS, which is inconsistent with other studies (de Cássia Pereira Fernandes et al., 2016; Haukkal et al., 2011). This result needs to be confirmed in further population studies.

The occurrence of MSS in neck and shoulder was also influenced by job tenure, which has also been reported in previous studies (Warren, Dussetschleger, Punnett, & Cherniack, 2015). It is generally accepted that with the increase of employment length, the longer the workers are exposed to occupational risk factors, the higher the prevalence of MSS in neck and shoulder (Constantino Coledam, Júnior, Ribeiro, & de Oliveira, 2019). In our study, the Chi-square test showed the prevalence of MSS in neck and shoulder in females was higher than that in males, which may be because the differences between males and females in muscle strength, motor control, fatigue response mechanism, and reaction to stress and pain (Côté, 2012). However, considering the interaction between various factors, the result of MLRM showed that the occurrence of MSS in neck and shoulder was not associated with gender. This result needs to be explored in further studies.

Considering the diversity and complexity of the combination of anatomical sites involving with symptoms, this study is the first to use LCA to explain the relationship between MSS in nine specific body sites with a small number of classes. The number of body sites involving with symptoms was suggested as a classification indicator for multi-site MSS in previous studies (Haukkal et al., 2011). There were also studies using statistical approaches (e.g., clustering and structural equation mixture modeling [SEMM]), to classify individuals into homogenous groups according to MSS (Gold et al., 2010; Yazdi, Feizi, Hassanzadeh Keshteli, Afshar, & Adibi, 2018). In a Dutch study of patients with upper extremity MSS, cluster analysis was employed to classify the patients into eight classes based primarily on symptom quality and severity (Gold et al., 2010). In a study on the Iranian general population, the population was classified into two major subgroups using SEMM based on some neuro-skeletal indicators (Yazdi et al., 2018). The occurrence patterns of MSS were classified into three classes using LCA in the current study, which simultaneously achieved dimensionality reduction and clustering. In addition, LCA took both the location and the number of body sites involved with symptoms into account.

However, some limitations should be acknowledged when interpreting these results. First, the data were collected by self-reported questionnaires, which may have memory bias and lead to misclassification of participants to the correct class. Second, although the probability of misclassification of the identified classes determined by LCA is low, not all participants have 100% probability of belonging to their assigned class, and this uncertainty could not be accounted for in the subsequent analyses. Third, the cross-sectional design has limitations in inferring cause–effect association from the findings. Therefore, our results should be interpreted with caution and need to be further verified in cohort or experimental studies.

5. Conclusions and practical applications

This is the first investigation to use LCA to determine whether unique patterns of MSS may be identified in electronic assembly workers. The results of the study suggested three distinct classes of MSS patterns in electronic assembly workers. In addition, this study supports previous general epidemiological MSS studies, indicating that derived patterns of MSS were associated with personal, physical, and psychosocial factors.

The findings of this study highlight that the high prevalence of multi-site MSS in this group should be a focus. It also provides further evidence that LCA considering the number and location of anatomical sites involved with MSS can be used to determine distinct classes of MSS patterns, which is of great significance for the epidemiological study and management of MSS in the future.

Funding

This work was supported by the National Key Technologies Research and Development Program of China [grant number 2016YFC0801700].

7. Data availability statement

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Author contributions

Yidan Dong performed the statistical analysis for this study and drafted the manuscript. Lihua He designed the project. Nazhakaiti Maimaiti and Shijuan Wang completed the survey and data collection. Mikael Forsman and Liyun Yang contributed to the design of the study and critically reviewed the manuscript. Ping Jiang and Xu Jin reviewed and edited the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest in this work.

References

Anderson, C. K., Chaffin, D. B., & Herrin, G. D. (1986). A study of lumbosacral orientation under varied static loads. *Spine, 11*(5), 456–462. https://doi.org/10.1097/00007632-198606000-00012.

Bæk Larsen, L., Ramstrand, N., & Fransson, E. (2018). Psychosocial job demand and control: Multi-site musculoskeletal pain in Swedish police. *Scandinavian Journal of Public Health, 47*(3), 318–325. https://doi.org/10.1177/1403494818801507.

Bevan, S. (2015). Economic impact of musculoskeletal disorders (MSDs) on work in Europe. *Best Practice & Research Clinical Rheumatology, 29*(3), 356–373. https://doi.org/10.1016/j.berh.2015.08.002.

Bongers, P. M., de Winter, C. R., Kompier, M. A., & Hildebrandt, V. H. (1993). Psychosocial factors at work and musculoskeletal disease. *Scandinavian Journal of Work, Environment & Health, 19*(5), 297–312. https://doi.org/10.5271/sjweh.1470.

Cammarota, A. (2007). The European commission initiative on WRMSDs: recent developments, presentation to EUROFOUND conference on 'Musculoskeletal disorders'.

Carragher, N., Adamson, G., Bunting, B., & McCann, S. (2009). Subtypes of depression in a nationally representative sample. *Journal of Affective Disorders, 113*(1–2), 88–99. https://doi.org/10.1016/j.jad.2008.05.015.

Collins, L., & Lanza, S. (2010). Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences. New York, NY: John Wiley & Sons, Inc. https://doi.org/10.1002/9780470567333.ch1.

Constantino Coledam, D. H., Júnior, R. P., Ribeiro, E. A. G., & de Oliveira, A. R. (2019). Factors associated with musculoskeletal disorders and disability in elementary teachers: A cross-sectional study. *Journal of Bodywork and Movement Therapies, 23*(3), 658–665. https://doi.org/10.1016/j.jbmt.2018.05.009.

Côté, J. N. (2012). A critical review on physical factors and functional characteristics that may explain a sex/gender difference in work-related neck/shoulder disorders. *Ergonomics, 55*(2), 173–182. https://doi.org/10.1080/00140139.2011.586061.

Crawford, J. (2007). The Nordic Musculoskeletal Questionnaire. *Occupational Medicine, 57*(4), 300–301. https://doi.org/10.1093/occmed/kqm036.

Croft, P., Dunn, K. M., & Von Korff, M. (2007). Chronic pain syndromes: You can't have one without another. *Pain, 131*(3), 237–238. https://doi.org/10.1016/j.pain.2007.07.013.

da Costa, B. R., & Vieira, E. R. (2010). Risk factors for work-related musculoskeletal disorders: A systematic review of recent longitudinal studies. *American Journal of Industrial Medicine, 53*(3), 285–323. https://doi.org/10.1002/ajim.20750.

de Cássia Pereira Fernandes, R., da Silva Pataro, S. M., de Carvalho, R. B., & Burdorf, A. (2016). The concurrence of musculoskeletal pain and associated work-related factors: a cross sectional study. *BMC Public Health, 16*, 628-636. https://doi.org/10.1186/s12889-016-3306-4.

Dong, H. Y., Zhang, Q., Liu, G. Z., & Shao, T. G. (2020). Prevalence of neck/shoulder pain among public hospital workers in China and its associated factors: A cross-sectional study. *Scientific Reports, 10*(1), 12311. https://doi.org/10.1038/s41598-020-69382-4.

European Agency for Safety and Health at Work. (2019). Work-related MSDs: prevalence, costs and demographics in the EU. https://osha.europa.eu/en/publications/work-related-musculoskeletal-disorders-prevalence-costs-and-demographics-eu/view.

Freimann, T., Coggon, D., Merisalu, E., Animägi, L., & Pääsuke, M. (2013). Risk factors for musculoskeletal pain amongst nurses in Estonia: A cross-sectional study. *BMC Musculoskeletal Disorders, 14*, 334–340. https://doi.org/10.1186/1471-2474-14-334.

Gatchel, R. J., & Schultz, I. Z. (2012). Handbook of Occupational Health and Wellness. New York, NY: Springer Science. https://doi.org/10.1007/978-1-4614-4839-6.

Gold, J. E., Piligian, G., Glutting, J. J., Hanlon, A., Frings-Dresen, M. H., & Sluiter, J. K. (2010). Cluster analysis of symptoms among patients with upper extremity musculoskeletal disorders. *Journal of Occupational Rehabilitation, 20*(4), 526–536. https://doi.org/10.1007/s10926-010-9240-x.

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied latent class analysis*. Cambridge, UK: Cambridge University Press.

Hartvigsen, J., Davidsen, M., Hestbaek, L., Sogaard, K., & Roos, E. M. (2013). Patterns of musculoskeletal pain in the population: A latent class analysis using a nationally representative interviewer-based survey of 4817 Danes. *European Journal of Pain, 17*(3), 452–460. https://doi.org/10.1002/j.1532-2149.2012.00225.x.

Haukka, E., Leino-Arjas, P., Solovieva, S., Ranta, R., Viikari-Juntura, E., & Riihimäki, H. (2006). Co-occurrence of musculoskeletal pain among female kitchen workers. *International Archives of Occupational and Environmental Health, 80*(2), 141–148. https://doi.org/10.1007/s00420-006-0113-8.

Haukkal, E., Leino-Arjasl, P., Ojajärvil, A., Takalal, E. P., Viikari-Juntural, E., & Riihimäkil, H. (2011). Mental stress and psychosocial factors at work in relation to multiple-site musculoskeletal pain: A longitudinal study of kitchen workers. *European Journal of Pain, 15*(4), 432–438. https://doi.org/10.1016/j.ejpain.2010.09.005.

Herin, F., Vézina, M., Thaon, I., Soulat, J. M., Paris, C., ESTEV & group. (2014). Predictive risk factors for chronic regional and multi-site musculoskeletal pain: A 5-year prospective study in a working population. *Pain, 155*(5), 937–943. https://doi.org/10.1016/j.pain.2014.01.033.

Hoe, V. C., Kelsall, H. L., Urquhart, D. M., & Sim, M. R. (2012). Risk factors for musculoskeletal symptoms of the neck or shoulder alone or neck and shoulder among hospital nurses. *Occupational & Environmental Medicine, 69*(3), 198–204. https://doi.org/10.1136/oemed-2011-100302.

Kamaleri, Y., Natvig, B., Ihlebaek, C. M., & Bruusgaard, D. (2008). Localized or widespread musculoskeletal pain: Does it matter? *Pain, 138*(1), 41–46. https://doi.org/10.1016/j.pain.2007.11.002.

Kamaleri, Y., Natvig, B., Ihlebaek, C. M., & Bruusgaard, D. (2009). Does the number of musculoskeletal pain sites predict work disability? A 14-year prospective study. *European Journal of Pain, 13*(4), 426–430. https://doi.org/10.1016/j.ejpain.2008.05.009.

Karasek, R., Brisson, C., Kawakami, N., Houtman, I., Bongers, P., & Amick, B. (1998). The Job Content Questionnaire (JCQ): An instrument for internationally comparative assessments of psychosocial job characteristics. *Journal of Occupational Health Psychology, 3*(4), 322–355. https://doi.org/10.1037/1076-8998.3.4.322.

Keester, D. L., & Sommerich, C. M. (2017). Investigation of musculoskeletal discomfort, work postures, and muscle activation among practicing tattoo artists. *Applied Ergonomics, 58*, 137–143. https://doi.org/10.1016/j.apergo.2016.06.006.

Lacey, R. J., Strauss, V. Y., Rathod, T., Belcher, J., Croft, P. R., Natvig, B., ... McBeth, J. (2015). Clustering of pain and its associations with health in people aged 50 years and older: Cross-sectional results from the North Staffordshire Osteoarthritis Project. *BMJ Open, 5*(11), e008389.

Lanza, S. T., Collins, L. M., Lemmon, D. R., & Schafer, J. L. (2007). PROC LCA: A SAS procedure for latent class analysis. *Structural Equation Modeling, 14*(4), 671–694. https://doi.org/10.1080/10705510701575602.

Lanza, S. T., & Rhoades, B. L. (2013). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science, 14*(2), 157–168. https://doi.org/10.1007/s11121-011-0201-1.

Larsen, L. B., Andersson, E. E., Tranberg, R., & Ramstrand, N. (2018). Multi-site musculoskeletal pain in Swedish police: Associations with discomfort from wearing mandatory equipment and prolonged sitting. *International Archives of Occupational and Environmental Health, 91*(4), 425–433. https://doi.org/10.1007/s00420-018-1292-9.

Lubke, G., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling, 14*(1), 26–47. https://doi.org/10.1080/10705510709336735.

Maimaiti, N., Wang, J. J., Jin, X. N., Wang, S. J., Qin, D. L., He, L. H., ... Yang, H. J. (2019). Cervical musculoskeletal disorders and their relationships with personal and work-related factors among electronic assembly workers. *Journal of Safety Research, 71*, 79–85. https://doi.org/10.1016/j.jsr.2019.09.018.

Molgaard Nielsen, A., Hestbaek, L., Vach, W., Kent, P., & Kongsted, A. (2017). Latent class analysis derived subgroups of low back pain patients – do they have prognostic capacity? *BMC Musculoskeletal Disorders, 18*, 345–363. https://doi.org/10.1186/s12891-017-1708-9.

Neupane, S., Leino-Arjas, P., Nygård, C. H., Miranda, H., Siukola, A., & Virtanen, P. (2015). Does the association between musculoskeletal pain and sickness absence due to musculoskeletal diagnoses depend on biomechanical working conditions? *International Archives of Occupational and Environmental Health, 88*(3), 273–279. https://doi.org/10.1007/s00420-014-0957-2.

Neupane, S., Miranda, H., Virtanen, P., Siukola, A., & Nygård, C. H. (2011). Multi-site pain and work ability among an industrial population. *Occupational Medicine, 61*(8), 563–569. https://doi.org/10.1093/occmed/kqr130.

Neupane, S., Miranda, H., Virtanen, P., Siukola, A., & Nygård, C. H. (2013). Do physical or psychosocial factors at work predict multi-site musculoskeletal pain? A 4-year follow-up study in an industrial population. *International Archives of Occupational and Environmental Health, 86*(5), 581–589. https://doi.org/10.1007/s00420-012-0792-2.

Neupane, S., Nygård, C. H., & Oakman, J. (2016). Work-related determinants of multi-site musculoskeletal pain among employees in the health care sector. *Work, 54*(3), 689–697. https://doi.org/10.3233/wor-162320.

Oakman, J., de Wind, A., van den Heuvel, S. G., & van der Beek, A. J. (2017). Work characteristics predict the development of multi-site musculoskeletal pain. *International Archives of Occupational and Environmental Health, 90*(7), 653–661. https://doi.org/10.1007/s00420-017-1228-9.
Oh, I. H., Yoon, S. J., Seo, H. Y., Kim, E. J., & Kim, Y. A. (2011). The economic burden of musculoskeletal disease in Korea: A cross sectional study. *BMC Musculoskeletal Disorders, 12*, 157–165. https://doi.org/10.1186/1471-2474-12-157.
Phongamwong, C., & Deema, H. (2015). The impact of multi-site musculoskeletal pain on work ability among health care providers. *Journal of Occupational Medicine and Toxicology, 10*, 21–25. https://doi.org/10.1186/s12995-015-0063-8.
Pope-Ford, R., & Jiang, Z. L. (2015). Neck and shoulder muscle activation patterns among dentists during common dental procedures. *Work, 51*(3), 391–399. https://doi.org/10.3233/WOR-141883.
Punnett, L., & Wegman, D. H. (2004). Work-related musculoskeletal disorders: The epidemiologic evidence and the debate. *Journal of Electromyography and Kinesiol, 14*(1), 13–23. https://doi.org/10.1016/j.jelekin.2003.09.015.
Sembajwe, G., Tveito, T. H., Hopcia, K., Kenwood, C., O'Day, E. T., Stoddard, A. M., ... Sorensen, G. (2013). Psychosocial stress and multi-site musculoskeletal pain: A cross-sectional survey of patient care workers. *Workplace Health & Safety, 61*(3), 117–125. https://doi.org/10.1177/216507991306100304.
Stanton, N., Hedge, A., Hendrick, H. W., Salas, E., & Brookhuis, K. (2004). *Handbook of human factors and ergonomics methods*. London, UK: Taylor & Francis Inc..
Wang, M., & Hanges, P. J. (2010). Latent class procedures: Applications to organizational research. *Organizational Research Methods, 14*(1), 24–31. https://doi.org/10.1177/1094428110383988.
Warren, N., Dussetschleger, J., Punnett, L., & Cherniack, M. G. (2015). Musculoskeletal disorder symptoms in correction officers: Why do they increase rapidly with job tenure? *Human Factors, 57*(2), 262–275. https://doi.org/10.1177/0018720814547569.
Yazdi, M., Feizi, A., Hassanzadeh Keshteli, A., Afshar, H., & Adibi, P. (2018). A classification analysis of musculoskeletal complaints and its association with anxiety, depression and psychological distress: Results from a large-scale cross-sectional study of adult Iranian population. *Advances in Human Biology, 8*(3), 180–185. https://doi.org/10.4103/AIHB.AIHB_13_18.
Yazdi, M., Karimi Zeverdegani, S., & MollaAghaBabaee, A. H. (2019). Association of derived patterns of musculoskeletal disorders with psychological problems: A latent class analysis. *Environmental Health and Preventive Medicine, 24*, 34–42. https://doi.org/10.1186/s12199-019-0784-x.
Yelin, E. H., & Cisternas, M. Total economic impact on the US economy. https://www.boneandjointburden.org/fourth-edition/viiie1/total-economic-impact-us-economy. Accessed on 2014.
Yu, S., Lu, M. L., Gu, G., Zhou, W., He, L., & Wang, S. (2012). Musculoskeletal symptoms and associated risk factors in a large sample of Chinese workers in Henan province of China. *American Journal of Industrial Medicine, 55*(3), 281–293. https://doi.org/10.1002/ajim.21037.
Wang, J. J., Cui, Y., He, L. H., Xu, X. R., Yuan, Z. W., Jin, X. N., & Li, Z. M. (2017). Work-related musculoskeletal disorders and risk factors among Chinese medical staff of obstetrics and gynecology. *International Journal of Environmental Research and Public Health, 14*, 562–574. https://doi.org/10.3390/ijerph14060562.

Yidan Dong is a PhD student at Department of Occupational and Environmental Health, School of Public Health, Peking University. She has been engaged in the State 13th Five-year Scientific and Technological Support Projects. She has published 6 articles in domestic and foreign journals with 1 article covered in SCI.

Ping Jiang is a Master student at Department of Occupational and Environmental Health, School of Public Health, Peking University. She has been engaged in the State 13th Five-year Scientific and Technological Support Projects.

Xu Jin is a Master student at Department of Occupational and Environmental Health, School of Public Health, Peking University. He has been engaged in the State 13th Five-year Scientific and Technological Support Projects.

Nazhakaiti Maimaiti is a Master student at Department of Occupational and Environmental Health, School of Public Health, Peking University. She has been engaged in the State 13th Five-year Scientific and Technological Support Projects.

Shijuan Wang is Master student at Department of Occupational and Environmental Health, School of Public Health, Peking University. She has been engaged in the State 13th Five-year Scientific and Technological Support Projects.

Liyun Yang, PhD, is a Postdoctoral researcher at Karolinska Institutet. She received her PhD in a joint doctoral program in Medical Technology (KTH Royal Institute of Technology) and Medical Science (Karolinska Institutet). Her research focuses on ergonomics risk analysis, wearable technology as methods for ergonomics, work technique training, and usability studies. With her research activities, she aims to contribute to a better working environment and sustainable working lives.

Mikael Forsman, PhD, is Professor of Ergonomics at Karolinska Institutet and KTH Royal Institute of Technology, Stockholm, Sweden. His area of expertise includes quantification of biomechanical loads on muscles and joints, industrial ergonomics, participative ergonomics and signal processing. Prof. Forsman is the author of about 250 publications and member of the editorial board of Applied Ergonomics. In addition to his research activities, Forsman is contributing to the area of occupational medicine by supporting local work health actors.

Lihua He, PhD, is a professor and doctoral supervisor at Department of Occupational and Environmental Health, School of Public Health, Peking University. She majors in ergonomics and occupational hazards of physical factors and has been engaged in teaching and research work on occupational health over the years. She has made major contributions to several National Natural Science Foundation Projects as well as the State Five-year Scientific and Technological Support Projects. She has published more than 190 articles in domestic and foreign journals with 30 articles covered in SCI.

Journal of Safety Research 82 (2022) 367–370

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Differences in fall-related emergency departments visits with and without an Injury, 2018

Briana L. Moreland [a,b,*], Elizabeth R. Burns [b], Yara K. Haddad [b]

[a] Cherokee Nation Operational Solutions, Atlanta, GA, United States
[b] Division of Injury Prevention, National Center of Injury Prevention and Control, Centers for Disease Control and Prevention, Atlanta, GA, United States

ARTICLE INFO

Article history:
Received 7 September 2021
Received in revised form 29 October 2021
Accepted 7 July 2022
Available online 18 July 2022

Keywords:
Elderly
Fall injuries
Falls
ICD-10-CM
Older adults

ABSTRACT

Background: Falls, with or without an injury, often affect the health of older adults (65+). *Methods:* We used the 2018 Healthcare Cost and Utilization Project to describe older adults' fall-related ED visits. We defined fall-related ED visits as those with a fall external cause of morbidity code and fall-injury related ED visits as those with an injury diagnosis code and a fall external cause of morbidity code. Percentages of fall-related and fall-injury related ED visits were analyzed by select characteristics. *Results:* Over 86% of fall-related ED visits were fall-injury related. A higher percentage of females (87%) and rural (88%) older adults' fall-related ED visits were fall-injury related compared to males (85%) and urban older adults (86%). A higher percentage of fall-related ED visits without a coded injury (33%) were hospitalized compared to those with a coded injury (29%). *Conclusion:* The majority of fall-related ED visits included an injury diagnosis. *Practical applications:* Researchers can consider which method of measuring ED visits related to falls is most appropriate for their study. Limiting fall-related ED visits to only those where an injury diagnosis is also present may underestimate the number of fall-related ED visits but may be appropriate for researchers specifically interested in fall injuries.

1. Introduction

Falls among older adults (aged 65 and older) impose a significant burden on healthcare systems. Older adult falls result in over three million emergency department (ED) visits annually (Centers for Disease Control and Prevention [CDC] 2013). Injuries from falls include bruises and scrapes and more serious injuries such as traumatic brain injuries and hip fractures (Bentler et al., 2009; Haddad, Shakya, Moreland, Kakara, & Bergen, 2020; Peterson, Xu, Daugherty, & Breiding, 2019). Falls that do not result in physical injury can result in psychological effects, such as fear of falling, and may produce prolonged disability, loss of independence, and declining mental and physical health (Scheffer, Schuurmans, van Dijk, van der Hooft, & de Rooij, 2008; Schoene et al., 2019).

Surveillance of older adult falls rely on records from ED visits and hospitalizations. The International Classification of Disease 10th revision Clinical Modification (ICD-10-CM) is used in healthcare systems to document morbidity. Older adult falls are characterized by two types of ICD-10-CM codes: injury diagnosis codes and external cause of morbidity codes. Injury diagnosis codes provide information about the type of injury (e.g., hip fracture) and external cause of morbidity codes describe the mechanism of injury (e.g., fall, motor vehicle crash). External cause of morbidity codes mainly provide information about injuries, however they can be used with any diagnosis code. For example, overexertion leading to a heart attack may have a diagnosis code of heart attack and an external cause of morbidity code of overexertion, without a documented injury (Centers for Medicare and Medicaid Services [CMS] and National Center for Health Statistics [NCHS], 2021). The surveillance case definition for fall-related ED visits using ICD-10-CM includes any visit with an external cause of morbidity code for a fall, regardless if there is an injury diagnosis code documented (Hedegaard, Johnson, & Ballesteros, 2017). This differs from previous definitions that define fall-related ED visits as visits where both an injury diagnosis and a fall external cause of morbidity code are included (Moreland, Burns, & Haddad, 2021). The objective of this study is to describe the difference between fall-related ED visits (without recorded injury) and fall-injury related ED visits (with coded injury) among older adults.

2. Methods

We analyzed data from the 2018 Healthcare Cost and Utilization Project (HCUP), Nationwide Emergency Department Sample

* Corresponding author at: 4770 Buford Hwy, NE, Atlanta, GA 30341, United States.

E-mail address: bmoreland@cdc.gov (B.L. Moreland).

https://doi.org/10.1016/j.jsr.2022.07.002
0022-4375/

(NEDS), to describe fall-related ED visits among older adults (Agency for Healthcare Research and Quality [AHRQ], 2021). HCUP-NEDS is a nationally representative sample and includes data from over 35 million ED visits from 990 EDs in the United States (AHRQ, 2021). When weighted, this represents approximately 145 million ED visits. Data were obtained from the Agency for Healthcare Research and Quality (AHRQ). Institutional Review Board (IRB) approval is not required because HCUP databases are classified as limited data sets. HCUP data is publicly available for purchase for researchers who complete a data use agreement. Fall-related ED visits were analyzed by sex, age group, urban/rural residence, and disposition.

Fall-injury related ED visits were defined as ED visits with an injury diagnosis code (S00-S99; T07-T34; T36-T50 with a sixth character of 1,2,3,4 except T36.9, T37.9, T39.9, T41.4, T42.7, T43.9, T45.9, T47.9, and T49.9 with a fifth character of 1,2,3,4; T51-T76; T79; and M97; all codes had a 7th character of A,B,C or missing to reflect an initial encounter) in any position. Fall-injury related ED visits also required a fall external cause of morbidity code (V00.11-V00.89 with sixth character of1; W00-W17 where the W16 codes include a sixth character of 2 except W16.4 and W16.9 which must include a fifth character of 2; W18.1-W18.3; and W19; all codes had a 7th character of A or missing to reflect an initial encounter) in any position (Table 1) (Hedegaard et al., 2017; Moreland et al., 2021).

Fall-related ED visits were defined as a visit with a fall documented as an external cause of morbidity in any position, regardless if an injury diagnosis code was present (Table 1). We classified the patient's county of residence urban or rural based on federal classifications (Ingram & Franco, 2013). ED disposition was categorized as the following: treated and released, hospitalized (same hospital or transferred to a different hospital), home healthcare, other transfers (skilled nursing facility, intermediate care, other facilities), or other/unknown (unknown disposition, left against medical advice, transferred to law enforcement). Analysis was limited to adults aged $\geq$65 years. ED visits in which a patient died were excluded.

Weights from NEDS (AHRQ, 2018) were used to calculate weighted estimates and corresponding 95% confidence intervals using SAS 9.4 survey procedures and SUDAAN (version 11; Research Triangle Institute) to produce estimates representative of the U.S. population and to account for the complex survey design. Rates were calculated using vintage 2018 bridged race population estimates from NCHS (NCHS, 2019). Two-sample t-tests were used to compare rates and Wald chi-squares were used to compare percentages.

3. Results

In 2018, there were an estimated 3,531,165 fall-related ED visits among older adults. Of these, approximately 3,049,421 visits (86.4% of fall-related ED visits) had an injury diagnosis coded in the record (Table 2). Females had a higher rate of fall-injury ED visits (6,850 95% CI: 6,477, 7,223 per 100,000) and fall-related ED visits (7,868 95% CI: 7,442, 8,294 per 100,000) than males (Fall-injury ED visits: 4,523 95% CI: 4,278, 4,768 per 100,000; fall-related ED visits: 5,318 95 %CI: 5,032, 5,604 per 100,000). A larger percentage of older female fall-related ED visits (87.1%; 95% CI 86.6%, 87.5%) included an injury diagnosis code compared to older males (85.1%; 95% CI: 84.5%, 85.6%) (Table 2).

The rates of fall-injury ED visits (3,389 per 100,000 adults aged 65–74 years; 6,816 per 100,000 adults aged 75–84 year; 14,771 per 100,000 adults aged 85 years and over) and the rates of fall-related ED visits (3,934 per 100,000 adults aged 65–74 years; 7,890 per 100,000 adults aged 75–84 years; 17,069 per 100,000 adults aged 85 and over) increased with age (Table 2). The percentages of fall-related ED visits with an injury code did not significantly vary by age group.

Rural older adults had a higher rate of fall-injury ED visits (6,526 95% CI: 6,098, 6,954) compared to urban older adults (5,651 95% CI: 5,290, 6,012). A higher percentage of rural older adults fall-related ED visits included an injury diagnosis (87.5%; 95% CI: 86.8%, 88.1%) compared to urban older adults (86.1%; 95% CI: 85.6%, 86.6%) (Table 2).

A higher percentage of older adults with a fall-injury visit (65.3%; 95% CI: 64.6%, 66.0%) were treated in the emergency department and released compared to those with a fall-related ED visit without a coded injury (57.9%; 95% CI: 56.5%, 59.4%) (Fig. 1). A higher percentage of fall-related ED visits among older adults without an injury coded were hospitalized (32.7%; 95% CI:31.2%, 34.2%) or were transferred to another facility such as a skilled nursing facility (7.1%; 95% CI: 6.5%, 7.6%) compared to those with a fall-injury visit (hospitalized: 28.5%; 95% CI:27.9%, 29.1%; transferred to other facility: 5.0%; 95% CI:4.7%, 5.3%) (Fig. 1).

4. Discussion

Over 86% of fall-related ED visits were among older adults who had an injury diagnosis. Defining an ED visit as fall-related only when an injury diagnosis code is included may underestimate the burden falls have on EDs up to 14%. Current coding guidance allows for use of external cause of morbidity codes with any diag-

Table 1
International classification of diseases, 10th revisions, clinical modification (ICD-10-CM) coding for fall-related and fall-injury related ED visits.

	Diagnosis Codes [a]	External Cause of Morbidity Codes [b]
Fall-injury ED Visits[c]	• S00-S99 • T07-T34 • T36-T50*Must include a sixth character of 1,2,3,4 except T36.9, T37.9, T39.9, T41.4, T42.7, T43.9, T45.9, T47.9, and T49.9 which must include a fifth character of 1,2,3,4* • T51-T76 • T79 • M97	• V00.11-V00.89 *Must include sixth character of 1* • W00-W17 *W16 codes must include a sixth character of 2 except W16.4 and W16.9 which must include a fifth character of 2.* • W18.1-W18.3 • W19
Fall-related ED Visits[c]	Any Diagnosis Code	• V00.11-V00.89 *Must include sixth character of 1* • W00-W17 *W16 codes must include a sixth character of 2 except W16.4 and W16.9 which must include a fifth character of 2.* • W18.1-W18.3 • W19

[a] 7th character of A, B, C, or missing reflects an initial encounter.
[b] 7th character of A or missing reflects an initial encounter.
[c] Diagnosis and external cause of morbidity codes can occur in any position.

Table 2
Rates and percentages of nonfatal fall-related emergency department (ED) visits with and without an injury diagnosis among older adults, Healthcare Cost and Utilization Project-Nationwide Emergency Department Sample, 2018.

	Fall-injury ED visits			Fall-related ED visits without injuries			Total fall-related ED visits	
	Weighted N	Rate* (95% CI)	% (95% CI)	Weighted N	Rate* (95% CI)	% (95% CI)	Weighted N	Rate* (95% CI)
Total**	**3,049,421**	**5,816 (5,501, 6,131)**	**86.4 (85.9, 86.8)**	**481,744**	**919 (861, 977)**	**13.6 (13.2, 14.1)**	**3,531,165**	**6,735 (6,373, 7,097)**
Sex								
Male	1,054,262	4,523 (4,278, 4,768)	85.1 (84.5, 85.6)	185,295	795 (744, 846)	14.9 (14.4, 15.5)	1,239,557	5,318 (5,032, 5,604)
Female	1,995,050	6,850 (6,477, 7,223)	87.1 (86.6, 87.5)	296,432	1,018 (953, 1,083)	12.9 (12.5, 13.4)	2,291,482	7,868 (7,442, 8,294)
Age Group								
65–74	1,033,431	3,389 (3,216, 3,563)	86.2 (85.6, 86.7)	166,099	545 (508, 581)	13.8 (13.3, 14.4)	1,199,530	3,934 (3,733, 4,135)
75–84	1,049,307	6,816 (6,444, 7,188)	86.4 (85.9, 86.9)	165,262	1,074 (1,005, 1,142)	13.6 (13.1, 14.1)	1,214,569	7,890 (7,462, 8,317)
85+	966,683	14,771 (13,885, 15,657)	86.5 (86.1, 87.0)	150,383	2,298 (2,147, 2,449)	13.5 (13.1, 14.1)	1,117,066	17,069 (16,054, 18,083)
Urban/Rural								
Urban	2,456,575	5,651 (5,290, 6,012)	86.1 (85.6, 86.6)	396,596	912 (847, 978)	13.9 (13.4, 14.4)	2,853,171	6,564 (6,149, 6,978)
Rural	584,844	6,526 (6,098, 6,954)	87.5 (86.8, 88.1)	83,884	936 (853, 1,019)	12.5 (11.9, 13.2)	668,729	7,462 (6,970, 7,954)

* Crude rate per 100,000 older adults.
** Columns may not add to the total due to missing values for sex or urban/rural status.

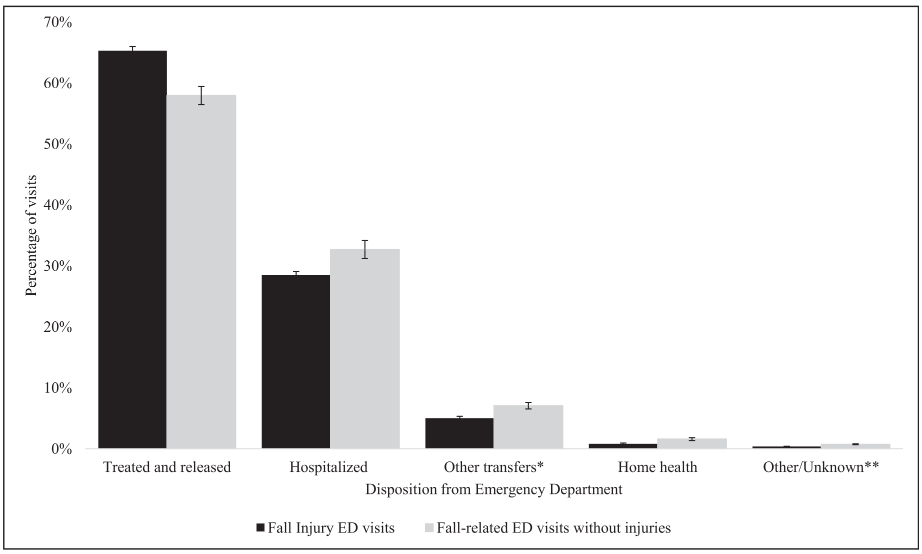


Fig. 1. Percentage of fall-related ED visits with and without an injury diagnosis among older adults by disposition, Healthcare Cost and Utilization Project-Nationwide Emergency Department Sample, 2018, *Transfer to skilled nursing facility, intermediate care, or other facility, **Unknown disposition, against medical advice, or discharged/transferred to court/law enforcement.

nosis code (CMS & NCHS, 2021). In a study of sport-related injuries, about 6% of sport-related ED visits were coded using external cause of morbidity codes without an injury diagnosis (Weiss & Elixhauser, 2016). Common non-injury diagnoses coded along with sports-related injuries included joint disorders, headache, and syncope. These diagnosis codes could be associated with fall external cause of morbidity codes in the absence of a coded injury. For example, if the fall resulted in joint pain but not a contusion or fracture it would only be included if fall-related ED visits were not limited to those with a diagnosed injury.

We found a higher percentage of fall-related ED visits without an injury diagnosis were hospitalized compared to fall-injury ED visits with an injury diagnosis included. In 2018, the leading causes of hospitalizations among older adults were septicemia and heart failure (AHRQ, 2018). It is possible that these conditions may cause the patient to collapse, which could be documented as a fall. This could contribute to the higher percentage of non-injurious falls requiring hospitalization that we observed in the data. These types of falls may not be prevented by traditional fall prevention interventions such as strength and balance exercises, or home modifications (Gillespie et al., 2012).

We found that a higher percent of female fall-related ED visits and a higher percent of rural older adults fall-related ED visits were associated with injury diagnosis compared to males and urban older adults, respectively. Previous studies reported that most fall-related ED visits in community-dwelling older adults were for females (Haddad et al., 2020), and that females have higher rates of reported fall injuries compared to males (Haddad et al., 2020; Moreland, Kakara, & Henry, 2020). Rural older adults also report higher rates of falls compared to urban adults (Moreland et al., 2020). Rural adults may lack access to primary care services and rely more on the ED compared to urban counterparts or may be more likely to delay or avoid care for non-urgent injuries (Venkatesh et al., 2020).

This study has several limitations. First, by design, our study excluded deaths and limited visits to only initial encounters, therefore our results should not be used to estimate the overall burden of fall-related ED visits. Additionally, external cause of morbidity codes are missing for about 10% of injury-related ED visits leading to further underestimations of fall-related ED visits, regardless if an injury diagnosis code was present. Second, because we did not limit disposition to treated and released, our findings may differ from other studies using the same dataset. Third, data from EDs, including a patient's chief complaint and notes from providers could better describe circumstances around fall-related ED visits without an injury diagnosis, however these data are not available from HCUP-NEDS.

5. Conclusion

Most fall-related ED visits included an injury diagnosis. A higher percentage of older female fall-related ED visits included an injury diagnosis compared to older male visits. Additionally, a higher percentage of rural older adult's fall-related ED visits included an injury compared to urban older adult's visits. A higher percentage of older adults with a visit without an injury coded were hospitalized compared to older adults with a fall injury visit. Future studies that investigate chief complaints, notes from providers, and additional diagnosis codes used in conjunction with fall cause of morbidity codes are needed to better describe fall-related ED visits in the absence of an injury diagnosis code.

Practical applications

Including only visits with a documented injury diagnosis may underestimate the number of fall-related ED visits among older adults but may be appropriate if researchers are measuring the impact of fall injuries specifically. Researchers using ICD-10-CM codes to analyze fall-related ED visits should consider which method is most appropriate for their study.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Previous presentations of the work

None.

Funding

No outside funding was used to support this work.

Acknowledgement

None.

References

Agency for Healthcare Research and Quality (AHRQ), 2021. HCUP User Support (HCUP-US). (Online). Available from: URL: https://www.hcup-us.ahrq.gov/ (accessed: May 2021).

Agency for Healthcare Research and Quality (AHRQ), 2018. Healthcare Cost and Utilization Project HCUPnet. Rockville, MD. (Online). Available from: URL: https://hcupnet.ahrq.gov/ (accessed June 2021).

Agency for Healthcare Research and Quality (AHRQ), 2018. Producing National HCUP Estimates. Healthcare Cost and Utilization Project (HCUP). (Online). Available from: URL: www.hcup-us.ahrq.gov/tech_assist/nationalestimates/508_course/508course_2018.jsp. Accessed October 2021.

Bentler, S. E., Liu, L., Obrizan, M., Cook, E. A., Wright, K. B., Geweke, J. F., & Wolinsky, F. D. (2009). The aftermath of hip fracture: discharge placement, functional status change, and mortality. *American Journal of Epidemiology, 170*(10), 1290–1299. https://doi.org/10.1093/aje/kwp266.

Centers for Disease Control and Prevention, 2003. Web-based Injury Statistics Query and Reporting System (WISQARS) [Online]. National Center for Injury Prevention and Control, Centers for Disease Control and Prevention (producer). Available from: URL: www.cdc.gov/ncipc/wisqars. (accessed June 2021).

Centers for Medicare and Medicaid Services (CMS) and the National Center for Health Statistics (NCHS), 2021. ICD-10-CM Official Guidelines for Coding and Reporting. [Online]. Available from: URL: https://www.cdc.gov/nchs/data/icd/10cmguidelines-FY2021.pdf. (accessed June 2021).

Gillespie, L. D., Robertson, M. C., Gillespie, W. J., Sherrington, C., Gates, S., Clemson, L. M., & Lamb, S. E. (2012). Interventions for preventing falls in older people living in the community. *Cochrane Database of Systematic Review, 2012*(9), Cd007146. https://doi.org/10.1002/14651858.CD007146.pub3.

Haddad, Y. K., Shakya, I., Moreland, B. L., Kakara, R., & Bergen, G. (2020). Injury diagnosis and affected body part for nonfatal fall-related injuries in community-dwelling older adults treated in emergency departments. *Journal of Aging and Health, 32*(10), 1433–1442. https://doi.org/10.1177/0898264320932045.

Hedegaard, H. B., Johnson, R. L., & Ballesteros, M. F. (2017). Proposed ICD-10-CM surveillance case definitions for injury hospitalizations and emergency department visits. *National Health Statistics Reports* (100), 1–8.

Ingram, D. D., & Franco, S. J. (2013). NCHS urban–rural classification scheme for counties. National Center for Health Statistics (Online) (accessed June 2021). *Vital Health Statistics, 2*(166), 2014 https://www.cdc.gov/nchs/data/series/sr_02/sr02_166.pdf.

Moreland, B., Kakara, R., & Henry, A. (2020). Trends in nonfatal falls and fall-related injuries among adults aged ≥65 years — United States, 2012–2018. *Morbidity and Mortality Weekly Report, 69*, 875–881. https://doi.org/10.15585/mmwr.mm6927a5.

Moreland, B. L., Burns, E. R., & Haddad, Y. K. (2021). National rates of non-fatal emergency department visits and hospitalisations due to fall-related injuries in older adults 2010–2014 and 2016: Transitioning from ICD-9-CM to ICD-10-CM. *Injury Prevention, 27*(Suppl 1), i75–i78. https://doi.org/10.1136/injuryprev-2019-043516.

National Center for Health Statistics (NCHS), 2019. Bridged Race Population Estimates, United States, July 1st resident population compiled from bridged-race vintage 2018 (2010–2018) postcensal population estimates. [online]. Available from: URL: http://wonder.cdc.gov/bridged-race-v2018.html (accessed December, 2020).

Peterson, A., Xu, L., Daugherty, J., & Breiding, M. (2019). Surveillance report of traumatic brain injury-related emergency department visits, hospitalizations, and deaths—United States, 2014. https://www.cdc.gov/traumaticbraininjury/pdf/TBISurveillance-Report-508.pdf.

Scheffer, A. C., Schuurmans, M. J., van Dijk, N., van der Hooft, T., & de Rooij, S. E. (2008). Fear of falling: Measurement strategy, prevalence, risk factors and consequences among older persons. *Age and Ageing, 37*(1), 19–24. https://doi.org/10.1093/ageing/afm169.

Schoene, D., Heller, C., Aung, Y. N., Sieber, C. C., Kemmler, W., & Freiberger, E. (2019). A systematic review on the influence of fear of falling on quality of life in older people: Is there a role for falls? *Clinical Interventions in Aging, 14*, 701–719. https://doi.org/10.2147/CIA.S197857.

Venkatesh, A. K., Greenwood-Ericksen, M. B., Mei, H., Rothenberg, C., Lin, Z., & Krumholz, H. M. (2020). Unscheduled care access in the United States – A tale of two emergency departments. *The American Journal of Emergency Medicine*. https://doi.org/10.1016/j.ajem.2020.08.095.

Weiss, A. J., Elixhauser, A. (2016). Sports-related emergency department visits and hospital inpatient stays, 2013. HCUP Statistical Brief #207. July 2016. Agency for Healthcare Research and Quality, Rockville, MD. http://www.hcup-us.ahrq.gov/reports/statbriefs/sb207-SportsHospital-Emergency-Department-2013.pdf.

Briana L. Moreland MPH, is a contractor at Cherokee Nation Operational Solutions for the Division of Injury Prevention at CDC's National Center for Injury Prevention and Control. Her research focuses on older adult fall prevention.

Elizabeth R. Burns MPH, is a health scientist for the Division of Injury Prevention at CDC's National Center for Injury Prevention and Control. Her research focuses on clinical fall prevention.

Yara K. Haddad PharmD, MPH, is a geriatric pharmacist and epidemiologist at the Centers for Disease Control and Prevention, Division of Injury Prevention. Her areas of concentration are geriatric care, older adult injury prevention, promoting older adult safe mobility and aging without injury, and exploring effects of medications and polypharmacy on older adult safety

Journal of Safety Research 82 (2022) 275–282

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Effects of large vehicles on pedestrian and pedalcyclist injury severity

Mickey Edwards *, Daniel Leonard

Visiting Research Specialist, University of Illinois at Springfield, United States

A R T I C L E I N F O

Article history:
Received 12 October 2021
Received in revised form 22 November 2021
Accepted 8 June 2022
Available online 18 June 2022

Keywords:
Pedestrian
Cycling
Crash
Large vehicles
Fatalities

A B S T R A C T

Introduction: Fatal pedestrian and pedalcyclist crashes have been on the rise in the United States since 2009. This rise in fatalities coincides with the rise of large vehicles on American roadways, continuing a trend that began years earlier. *Method:* Through rare access to both crash and hospital records, this report investigates the relationship between striking vehicle type and medical outcomes of pedestrian and pedalcyclist cases. *Results:* Results suggest that children are eight times more likely to die when struck by a SUV compared to those struck by a passenger car. Passenger cars were the striking vehicle in most fatal pedestrian and pedalcyclist crashes, though they were underrepresented relative to the proportion of all crashes in which they were involved. Though pickup trucks were the striking vehicle in just 5.6% of pedestrian and pedalcyclist crashes, they were involved in 12.6% of fatalities. SUVs were similarly overrepresented in fatalities relative to the proportion of their involvement in all crashes. SUVs struck 14.7% of the pedestrians and pedalcyclists investigated here, but were involved in 25.4% of the fatalities. Head and thorax injury severities are examined by vehicle type and age. Hospital charges of pedestrian and pedalcycle crash victims are also analyzed by striking vehicle type and victim age. *Practical applications:* Findings suggest larger vehicles are involved in pedestrian and pedalcyclist crashes with more severe injuries that result in higher hospital charges. By race, Blacks are also found to be overrepresented as pedestrian and pedalcyclist crash victims.

1. Introduction

Pedestrian and pedalcyclist fatalities in the United States have been rising in recent years even as fatalities for motorists has declined or remained relatively flat (Coleman & Mizenko, 2018). Fatalities among motorists, pedestrians, and pedalcyclists alike were on the decline in the United States starting around 1980. But in 2009 the data trend lines diverged as pedestrian and pedalcyclist fatalities began rising (Arias et al., 2021). In fact, from 2010 to 2019 pedestrian fatalities increased by 46% to 6,301 deaths in 2019 (GHSA, 2021). Further, in 2016, which corresponds to this study's timeframe, 4,074 children were killed in motor-vehicle crashes in the United States – making crashes the number one killer of American children (Cunningham et al., 2018).

The 1980's saw the production and sales volume of large motor vehicles begin to command an increasing share of the U.S. automobile market (Fig. 1). Large vehicles are commonly considered those classified as light trucks: pickup trucks, SUVs, and vans/minivans – a consideration also applied here. But is the growing American predilection for large vehicles associated with increased pedestrian and pedalcyclist fatalities? By one estimate, the external cost of one person choosing to drive one large motor vehicle rather than a passenger car on pedestrian death risk alone is $75, $98, and $114 per year for each SUV, pickup truck, and van/minivan, respectively (Tyndall, 2021). Still, a growing number of Americans are choosing large motor vehicles over the traditional passenger car. Fig. 1 depicts the diminishing share of car production over the past decades as it gave way to the growing "Truck SUV," "Car SUV," and "Pickup" categories (EPA, 2021). In fact, passenger car sales in the United States dropped at an annual rate of 2.4% from 2008 to 2018 alone, while pickup truck sales increased at an annual rate of 6.4% (Davis & Boundy, 2020). In 2008 light trucks were about 40% of light vehicles produced, in 2018 they were nearly half (Ibid).

Larger vehicles not only produce excess carbon emissions but may also pose a greater threat to pedestrian and pedalcyclist safety. Tyndall (2021) uses pedestrian fatality data from across the United States to estimate that a 100 kg increase in average motor-vehicle weight correlates with a 2.4% increase in pedestrian fatalities for a median fatality rate region. He further finds that converting 10% of a regional vehicle fleet from cars to light trucks correlates with a 3.6% increase in fatal pedestrian crashes (Ibid). Desapriya et al. (2010) estimate in their meta-analysis that pedestrians struck by a pickup truck were 50% more likely to be killed compared to those struck by a passenger car. Roudsari et al. (2004) find that those hit

* Corresponding author.
E-mail addresses: Hedwa4@uis.edu (M. Edwards), Dleon01s@uis.edu (D. Leonard).

https://doi.org/10.1016/j.jsr.2022.06.005
0022-4375/

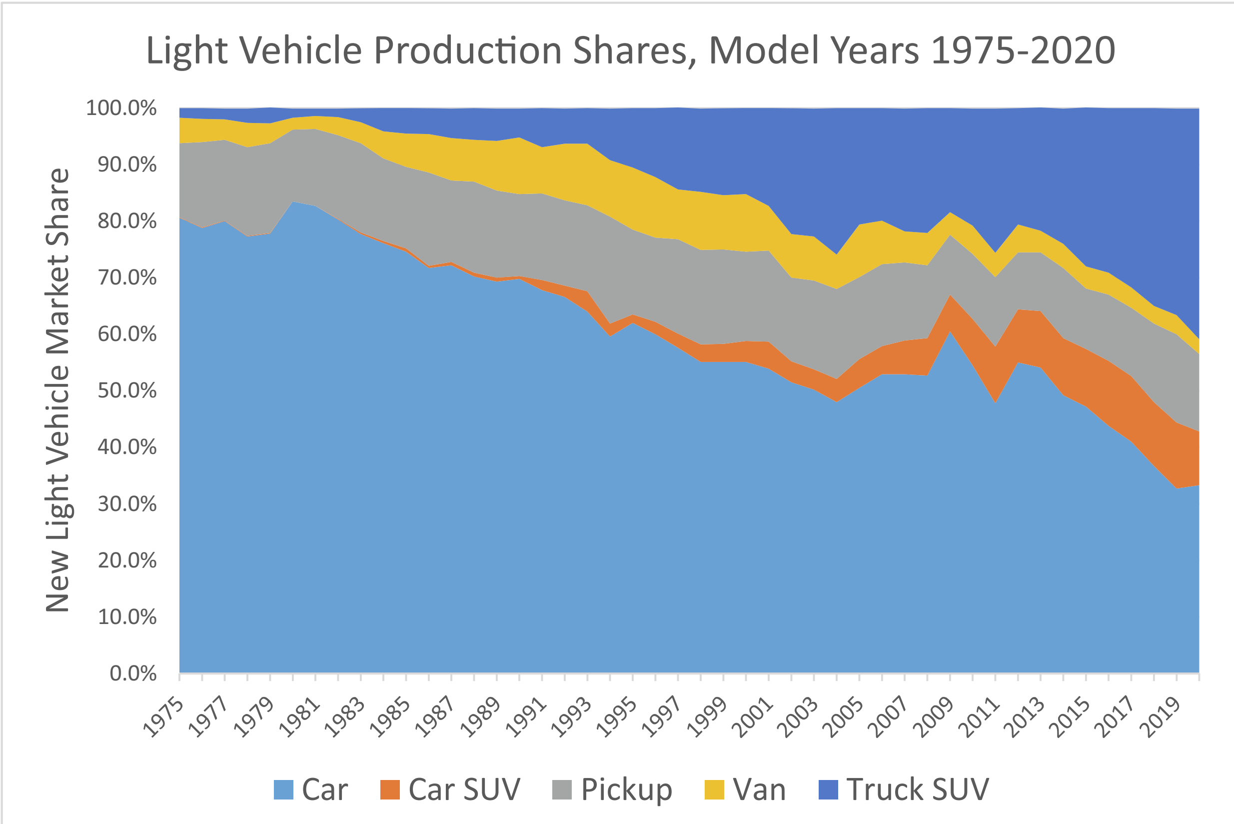


Fig. 1. Light vehicle production shares, model years 1975–2020*.

by light trucks (including SUVs, vans/minivans, and pickup trucks) had higher rates of severe brain injury (33%) relative to those hit by cars. In their review of the pedestrian safety literature, Doggett et al. (2018) find that unreported pedestrian and pedalcyclist crashes underestimate injuries by 21%. They find that crashes are more likely to go unreported if the pedestrian or pedalcyclist is: less likely to receive an insurance payout, Black, or male. Crashes that happened on a state road, Y-intersection, or divided highway were also less likely to be reported. These studies imply that the true scope of pedestrian and pedalcyclist crashes is likely worse than can be estimated by relying only on crash databases. What we *do* know by relying on crash databases is that pedestrian and pedalcyclist crashes disproportionately affect poor and minority communities, perhaps to a greater degree than previously believed (GHSA, 2021). Further, Braun et al. (2021) find that the health risks associated with cycling in Los Angeles (pollution, injury, fatality) are disproportionately high among communities of lower income, lower educational attainment, and greater proportions of racial/ethnic minorities.

There is still much to learn. This paper aims to contribute by utilizing uncommon government agency access to crash and hospital records to investigate those who are disproportionately affected by pedestrian and pedalcyclist crashes – especially those most vulnerable among us.

2. Sources and methods

2.1. Linking crash and hospital data files

Funded by a grant from the Centers for Disease Control and Prevention, the Illinois Department of Public Health in collaboration with the Illinois Department of Transportation and the University of Illinois at Springfield successfully linked Illinois crash and hospital records for the years 2016 through 2018. The linkage was accomplished using an advanced method developed in the National Highway Traffic Safety Administration's Crash Outcome Data Evaluation System program (McGlincy, 2021). Using LinkSolv software to complete the linkage, a combination of data fields were identified as those with the highest success rate: county, victim age, crash date, victim date of birth, and victim sex. Cook County, home to Chicago, is where some 40% of the Illinois population resides, effectively making county a relatively indiscriminate match field – which is a factor controlled for in the LinkSolv software. Two versions of data are applied here and communicated in the text of which version is used for each analysis. The report begins with an analysis of the crash file alone to investigate the effect of vehicle type on pedestrian and pedalcyclist fatalities, as the crash file is considered the authoritative source for fatalities. The second half of the report employs only linked crash and hospital data, as this permits a higher fidelity investigation of bodily injury location, severity, and ultimately hospital charges. These linked files are critical in our understanding of the effects of motor vehicles on the lives of the citizens of Illinois. Such an investigation as presented here would not be possible without the successful linkage of the disparate crash and hospital files.

2.2. Data independence and strength of association

Some 69.4% of linked crash and hospital pedestrian and pedalcyclist records occurred in Cook County, and an additional 11% of linked records are in counties bordering Cook County. This fact may result in some data bias, manifesting in the analysis results skewing toward the characteristics of Cook County. When appropriate throughout this paper cases are disaggregated between Chicago and the rest of Illinois for analysis. Given the nature of data linkage and innate inaccuracy in records, the data are likely incomplete and may contain mismatched records despite using advanced linking software and methods. Still, the Pearson's chi-squared alpha value and Cramer's-V are presented in each table to demonstrate both variable independence and the strength of association between variables.

2.3. Logistic regression model

A logistic regression was performed to estimate the effects of vehicle type, road conditions, and victim demographics on the likelihood of a fatal pedestrian or pedalcyclist crash event occurring. Full results are presented in the Results portion of this manuscript. The logistic regression model was statistically significant with a Chi-square value of 146, 18 degrees of freedom, and a *p* value of less than 0.000. The model correctly classified 98% of cases and explains about 14% (Nagelkerke R^2) of the variance in fatal pedestrian and pedalcyclist crash events.

3. Results

3.1. Summary statistics

The analysis begins within the unlinked crash file, which contains an aggregated 23,090 pedestrian and pedalcyclist cases across 2016, 2017, and 2018. Some 14,552 cases (63%) involved a pedestrian, and 8,538 cases (37%) involved a pedalcyclist. Pedestrians were overrepresented in the 477 fatalities reported by police with 85.5% of deaths; the remaining 14.5% were pedalcyclists. Of note, if not specified throughout this report, pedestrians and pedalcyclists are considered together. In 14,324 cases (62%) the striking vehicle was classified as a passenger vehicle, commonly referred to as a car. In 3,396 (14.7%) cases the striking vehicle was classified as a sport utility vehicle (SUV). Vans/minivans and pickup trucks were the striking vehicle classifications in 1,343 (5.8%) and 1,291 (5.6%) cases, respectively. Thirteen various other motor-vehicle classifications comprised the remaining 11.8% of cases.

3.2. Injury severity by striking vehicle type

Taller and heavier vehicle types (like pickup trucks, SUVs, and vans/minivans) combined to make up just 26.1% of pedestrian and pedalcyclist crashes, but were the striking vehicle in 44.1% of fatalities (Table 1). SUVs were especially overrepresented in fatalities. Though SUVs were the striking vehicle in 14.7% of cases, they were involved in greater than one-in-four (25.4%) fatalities. Pickup trucks were also overrepresented in fatal pedestrian and pedalcyclist crashes relative to the proportion of all cases. Of all pedestrian and pedalcycle fatalities, 12.6% involved a pickup truck – some two and a quarter times the proportion of all cases involving a pickup. Conversely, though passenger cars were the striking vehicle in 62% of cases, they were involved in just 38.4% of fatalities. Though males made up about 62% of pedestrian and pedalcyclist crash victims, they were overrepresented in fatalities at about 72% of cases.

3.3. Injury severity by age distribution

Fig. 2 demonstrates the bimodal age distribution of pedestrian and pedalcyclist crash victims. Victim age distribution roughly follows the shape of age distribution across Illinois, though the peaks in either mode are taller here. One explanation, among others, for the tall bimodal distribution may be that those at the top and bottom of the age distribution have a diminished capacity to drive, and thus walk and cycle more frequently while increasing their exposure to motor-vehicle traffic. Further, the most frequently occurring victim age was 15 followed by age 14, representing 2.8% and 2.7% of all pedestrian and pedalcycle crash victims, respectively. The following section investigates the distribution of injury severity across age groups in more detail.

3.4. Injury severity by age and striking vehicle type

A useful way to study the danger posed to pedestrians and pedalcyclists by vehicle type is to investigate whether certain types cause more severe injuries more frequently. One would expect injury severity levels to be roughly evenly distributed across vehicle types in proportion to the frequency with which each vehicle type is involved in a pedestrian or pedalcyclist crash. But injury severities were not roughly evenly distributed. Large and heavy vehicles caused much more damage to human bodies relative to passenger cars; though passenger cars also caused many severe injuries.

A child (under age 18) struck by a SUV was eight times more likely to be killed than a child struck by a passenger car (Table 2). An adult (aged 18–64) struck by a pickup truck was four times more likely to be killed than an adult struck by a passenger car. And a senior (aged 65 and over) struck by a pickup truck was nearly three times more likely to be killed compared to a senior struck by a passenger car.

In every age group, passenger cars represented the greatest proportion of fatalities, though they were underrepresented relative to the proportion of cases in which they were involved. For example, passenger cars were the striking vehicle in almost 62% of pedestrian and pedalcyclist crashes involving children, but just about 19% of childhood fatalities.

In contrast, the proportion of fatalities involving pickup trucks was more than double the overall proportion of pickup trucks involved in pedestrian and pedalcyclist crashes for all age groups. For example, pickup trucks were the striking vehicle in 6.1% of all cases involving seniors, but represent 13.5% of all senior pedestrian and pedalcyclist fatalities.

SUVs were particularly deadly for children. SUVs were the striking vehicle in greater than 40% of childhood fatalities, even though SUVs were involved in just 16.9% of childhood cases. Further, children represented 21% of all pedestrian and pedalcyclist crash victims but 26.1% of cases involving SUVs – implying SUVs were not only more deadly, but also disproportionately struck children. Vans/minivans are also overrepresented in cases involving childhood fatalities. Just under 6% of child pedestrians and pedalcyclists were struck by a van/minivan, but 12.5% of childhood fatalities involved a van/minivan. Together, SUVs, pickup trucks, and vans/minivans combined to cause two-thirds of fatalities involving child pedestrians and pedalcyclists (Table 2).

3.5. Logistic regression model results

The logistic regression model used the linked data file, so both crash and hospital records were used to estimate the effects of crash characteristics on a fatal event occurring. Six variables are predicted to add significantly to the model at the 5% level: pickup truck, SUV, van/minivan, intersection, child, and senior. Since the explanatory variables are binary, the odds ratio value is the true value of the estimated effect on the likelihood of a fatal crash. For example, the model estimates that a pedestrian or pedalcyclist struck by a pickup truck was 4.7 times more likely die as a result. Those struck by a SUV or van were 3.37 times and 4.58 times more likely to be killed, respectively.

Difficulty in recording accurate crash data may contribute to some of the variables not significantly contributing to the logistic

Table 1
Injury severity distribution of pedestrians and pedalcyclists by striking vehicle type cross tabulated and totaled by columns for Illinois 2016–18 (counts in parentheses)*.

	Injury Severity					
Vehicle Type	**O**	**C**	**B**	**A**	**K**	**% of all**
Passenger Car	64.7% (534)	64.7% (4,099)	62.5% (7,257)	58.6% (2,251)	38.4% (183)	62% (14,324)
Pickup Truck	4.5% (37)	5.1% (323)	5.2% (603)	7.0% (268)	12.6% (60)	5.6% (1,291)
SUV	14.9% (123)	14.2% (897)	14.4% (1,673)	15.1% (582)	25.4% (121)	14.7% (3,396)
Van/Minivan	4.4% (36)	5.6% (353)	6.0% (693)	6.0% (232)	6.1% (29)	5.8% (1,343)

*Other vehicle types not presented in table (columns do not total 100%); K: Fatality,
A: Incapacitating Injury, B: Non-incapacitating injury, C: Possible injury, O: No indication of injury; Pearson chi-square alpha value < 0.01, Cramer's-V = 0.15.

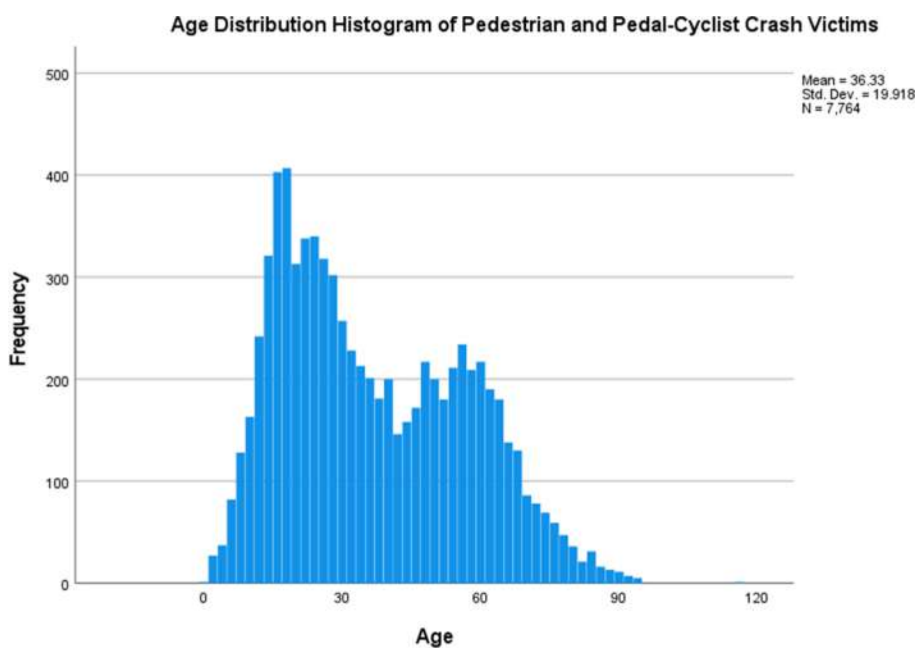


Fig. 2. Age distribution of linked pedestrian and pedalcyclist crash victims in Illinois 2016–18.

model. The day of week of the crash variable is a relatively easy data point to collect at the crash scene, yet does not add significantly to the model. Yet other crash factors that also do not add significantly to the model are relatively more difficult to accurately record in the crash report. Fields such as "distracted," "impaired," and "speed" are intuitively related to a higher propensity for a fatal crash event occurring, but the model does not reflect this. A driver could simply self-report to police at the crash scene that they were neither distracted nor speeding, and without sufficient evidence to the contrary the crash file would record such an assertion. Both the crash file and the hospital file record fields indicative of substance use impairment. Police at the crash scene record whether the evidence present warrants flagging the case as impaired. The hospital records reflect a medical substance test conducted at the hospital and the subsequent positive or negative diagnoses. Several variations of the impaired field variable from both the crash and hospital files were attempted in preparation for the logistic regression model, none were estimated to add significantly to the model. That impairment does not significantly add to the model is probably more indicative of administrative record-keeping shortcomings than of a diminished role of impairment on pedestrian and pedalcyclist crash fatalities. The model results presented in Table 3 uses the binary impaired field from the crash file.

The logistic regression model suggests that intersections are somewhat protective of pedestrians and pedalcyclists – estimating a diminished likelihood of a fatal crash occurring at an intersection. Succeeding sections of this paper demonstrate that relatively large vehicles were deadlier for all crash victims, including children. Yet, intuitively, the model implies that being a child is associated with a greater likelihood of surviving being struck by any vehicle type, while being a senior is associated with an increased likelihood (4.72 times) of death when struck.

Though the logistic regression model is useful in estimating statistically significant factors associated with an event occurring, fatal pedestrian and pedalcyclist crashes in this case, it is somewhat of a blunt instrument. A finer, more nuanced investigation is necessary to achieve a richer understanding of the factors related to these fatal crashes. The following sections are an attempt at deepening that understanding.

3.6. Head and thorax injury severity by vehicle type

Similar to the logistic regression model presented previously, this section uses only those pedestrian and pedalcyclist cases in which the crash and hospital files were successfully linked (7,764 cases; 36%). Pedalcyclists made up about 34% of linked cases, and

Table 2
Injury severity distribution by striking vehicle type and age group*.

Age	Vehicle type		Injury severity scale K	A	B	C	O	Total
Under 18	Passenger Car	Within Cars	0.2%	12.9%	55.7%	28.1%	3.1%	100%
		Within Severity	18.8%	57.8%	61.9%	64.8%	60.9%	61.8%
	Pickup Truck	Within Pickups	1.4%	19.0%	52.0%	24.5%	3.1%	100%
		Within Severity	12.5%	8.4%	5.7%	5.6%	6.0%	6.1%
	SUV	Within %Vs	1.6%	14.3%	57.4%	23.6%	3.1%	100%
		Within Severity	40.6%	17.5%	17.4%	14.9%	16.6%	16.9%
	Van/mini van	Within Vans	1.4%	15.9%	55.1%	25.1%	2.5%	100%
		With in Severity	12.5%	6.8%	5.8%	5.5%	4.6%	5.9%
	Counts		32	664	2,679	1,290	151	
18–14	Passenger Car	Within Cars	1.2%	16.1%	50.4%	29.2%	3.1%	100%
		Within Severity	36.9%	58.5%	62.7%	64.8%	65.5%	62.1%
	Pickup	Truck Within Pickups	4.8%	21.8%	45.3%	25.2%	2.8%	100%
		With in Severity	12.6%	7.0%	5.0%	4.9%	5.2%	5.5%
	SUV	Within SUVs	3.6%	18.0%	47.9%	27.9%	2.7%	100%
		Within Severity	23.7%	14.6%	13.3%	13.8%	12.7%	13.9%
	Van/Minivan	Within Vans	2.0%	16.8%	52.1%	27.1%	2.0%	100%
		Within Severity	5.5%	5.7%	6.0%	5.6%	3.9%	5.8%
	Counts		325	2,657	7,776	4,364	458	
65+	Passenger Car	Within Cars	4.6%	22.4%	47.7%	23.9%	1.5%	100%
		Within Severity	47.7%	60.8%	62.7%	63.7%	60.7%	61.6%
	Pickup Truck	Within Pickups	13.2%	16.7%	43.9%	25.4%	0.9%	100%
		Within Severity	13.5%	4.5%	5.7%	6.7%	3.6%	6.1%
	5LIV	Within SUVs	9.8%	21.7%	45.8%	21.0%	1.7%	100%
		Within Severity	26.1%	15.0%	15.3%	14.3%	17.9%	15.7%
	Van/mini van	Within Vans	5.5%	27.6%	40.9%	23.6%	2.4%	100%
		Within Severity	6.3%	8.2%	5.9%	6.9%	10.7%	6.8%
	Counts		111	426	881	435	28	

K: Fatality, A: Incapacitating Injury, B: Non-incapacitating injury, C: Possible injury, O: No indication of injury; All Pearson chi-square alpha values <0.01, Cramer's-*V*: Under 18: 0.191, 18–64: 0.17, 65+: 0.15.

* Other vehicle types not presented in table (columns do not total 100%).

were victims in about 37% of crashes reported by police. This proportionate representation of pedestrians and pedalcyclists implies the linked data were not biased toward either transport mode.

Since large motor vehicles are commonly taller than a passenger car, one would expect more severe injuries to a pedestrian's or pedalcyclist's thorax and head. This is because when a relatively short passenger car strikes a pedestrian it likely makes contact with the victim's legs while forcing the upper body onto the car hood. When a taller, heavier motor vehicle strikes a pedestrian it is more likely to strike the victim's body (thorax) and head while also knocking them to the street and potentially running them over (Roudsari et al., 2004).

Table 4 shows that each vehicle type studied is overrepresented in their proportion of non-minor head injuries relative to their overall involvement in striking pedestrians and pedalcyclists. (Other vehicle types not analyzed here, like ATV or Farm Equipment, are likely underrepresented in non-minor head injuries.) For example, SUVs were involved in 13.5% of linked pedestrian and pedalcyclist crashes, but 16.2% of cases with a non-minor head injury. Looking at the proportion of non-minor head injuries within vehicle types offers a bit more insight. For example, greater than 11% of pedestrians and pedalcyclists struck by a pickup truck had a non-minor head injury compared to 9.5% of those struck by a passenger car. For children, 12.7% and 11% of cases involving a van/minivan and a pickup truck resulted in a non-minor head injury, respectively.

Pedestrians and pedalcyclists struck by a large motor vehicle were more likely to suffer moderate or worse injuries to their thorax compared to those struck by a passenger car (Table 4). Though the proportion of pickup trucks involved in all cases examined here was 5.6%, that proportion nearly doubles to 11.1% of all non-minor thorax injuries. Further, nearly 10% of all occurrences of a pickup truck striking a pedestrian or pedalcyclists resulted in a non-minor thorax injury. For passenger cars, only 3.8% of occurrences resulted in such injuries. Finally, though passenger cars represent 54.1% of all cases here, they are underrepresented as causing non-minor thorax injuries at 42.1% of all such injuries.

Table 3
Logistic regression modeling the likelihood of a fatal pedestrian or pedalcyclist crash occurring*.

Variable	*Coefficient*	*Odds Ratio*	*Significance*
Passenger Car	0.146	1.16	0.651
Pickup Truck	**1.55**	**4.70**	**0.000**
SUV	**1.22**	**3.37**	**0.000**
Van/Minivan	**1.52**	**4.58**	**0.000**
Weekend	0.111	1.12	0.609
Rural	−0.309	0.734	0.271
Distracted	0.187	1.21	0.695
Impaired	0.029	1.03	0.952
Speed	0.350	1.42	0.128
Intersection	**−0.707**	**0.493**	**0.000**
Female	−0.366	0.694	0.077
Child	**−1.13**	**0.322**	**0.005**
Senior	**1.55**	**4.72**	**0.000**
Race: White	0.081	1.09	0.760
Race: Black	−0.687	0.503	0.063
Race: Asian	−0.277	0.758	0.630
Ethnicity: Hispanic	−1.12	0.326	0.155

* Variables that add significantly to the model at the 5% level appear in **bold**.

3.7. Hospital charges by vehicle type and age

Fig. 3 demonstrates that not only are pedestrians and pedalcyclists more likely to be more severely injured when struck by a large motor vehicle relative to a passenger car, but those more severe injuries, intuitively, result in higher hospital charges. Both

Table 4
Distribution of moderate (greater than or equal to 2 on the abbreviated injury scale) and worse head and thorax injuries by vehicle type*.

		Head Injury Severity		Thorax Injury Severity	
Vehicle Type	*Proportion Involved in Crashes*	*Proportion of all ≥ 2 Head Injuries*	*Proportion of ≥ 2 Within Vehicle Type*	*Proportion of all ≥ 2 Thorax Injuries*	*Proportion of ≥ 2 Within Vehicle Type*
Passenger Car	54.1%	58.6%	9.5%	42.1%	3.8%
Pickup Truck	5.6%	7%	11.1%	11.1%	9.7%
SUV	13.5%	16.2%	9.9%	17.4%	6.3%
Van/Minivan	5.7%	8.2%	10.6%	7.6%	6.6%

Abbreviated Injury Scale (AIS) 1: Minor, 2: Moderate, 3: Serious, 4: Severe, 5: Critical, 6: Maximal (untreatable); Pearson chi-square alpha value <0.01, Cramer's-V = 0.142.
* Other vehicle types not presented in table.

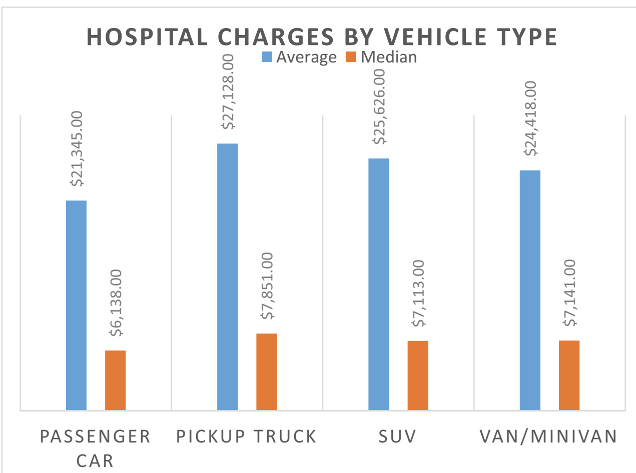


Fig. 3. Average and median hospital charges of pedestrian and pedalcyclist crash victims by vehicle type.

median and average hospital charges are presented to demonstrate the effect a few very high hospital bills can have on the calculation of a mean. So the typical pedestrian or pedalcyclist crash victim struck by a large motor vehicle could expect an additional $1,230 in median hospital charges compared to those struck by a passenger car; and an additional $4,380 in average hospital charges. Further, those struck by a pickup truck could expect to be charged the most for hospital treatment. This finding is consistent with the above analysis, which demonstrated that a greater proportion of cases involving striking pickup trucks resulted in non-minor injuries to both the head and thorax.

Fig. 4 shows that hospital charges increase by nearly a factor of three from children to senior pedestrian and pedalcyclist crash victims – a finding supported by the logistic regression presented above. One cause of higher charges is likely due to a greater likelihood of seniors suffering a more severe injury compared to other age groups, according to the maximum abbreviated injury scale (MAIS) field in the hospital data file.

3.8. Race and ethnicity

Blacks were overrepresented as victims of pedestrian and pedalcyclist crashes throughout Illinois. Outside of the City of Chicago 23% (945) of pedestrian and pedalcyclist crash victims were Black, despite Blacks making up only about 10% of the population of Illinois outside of Chicago (ACS 2019 5-Year Estimates) U.S. Census Bureau (2019). Within Chicago. 31% (1,154) of pedestrian and pedalcycle crash victims were Black, where 29.6% of the population is Black (Ibid). Statewide (all of Illinois and Chicago), 27% (2,099) of victims were Black despite representing just 14.2% of the population (Ibid).

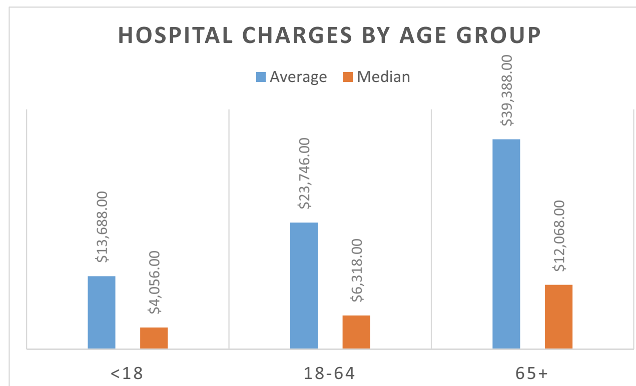


Fig. 4. Average and median hospital charges of pedestrian and pedalcyclist crash victims by age group.

The Hispanic/Latino population was underrepresented as victims of pedestrian and pedalcyclist crashes throughout Illinois. Some 20% (719) of crash victims were Hispanic/Latino within Chicago, despite making up about 29% of the population there (Ibid). And 14% (562) of pedestrian and pedalcyclist crash victims in the rest of Illinois (exclusive of Chicago) were Hispanic/Latino, where they make up 14.4% of the population (Ibid). Statewide, 16.5% (1,281) of victims were Hispanic/Latino despite representing 17.5% of the population. This finding of underrepresentation of the Hispanic/Latino population likely has several causes. One to note for this particular study is the potential for communication difficulties between a responding emergency professional and a Hispanic/Latino crash victim. If the crash victim lacks identification and their personal information is recorded incorrectly in the crash file, that case has a greater likelihood of not being linked with hospital data – from which the race/ethnicity fields are drawn.

Still, minority children were overrepresented as pedestrian and pedalcyclist crash victims. Though 27% of cases involved a Black victim, closer to 30% of cases involving children were Black. And 16.5% of all crash victims were Hispanic/Latino, yet 23.2% of all cases involving children were Hispanic/Latino. White seniors were also significantly overrepresented as victims of pedestrian and pedalcycle crashes. While Whites made up about 47% of cases, greater than 57% of crashes involving a senior was White.

4. Discussion

Three topic areas for future research are suggested: (1) A robust and nuanced understanding of the relationship between the rise of large vehicles and pedestrian/pedalcyclist fatalities and severe injuries. Clearly part of the explanation is that large vehicles carry more momentum and more severely harm human bodies compared to passenger cars. Still, another factor may be socioeconomic. Pedestrian fatalities began rising in 2009 just as the Great

Recession was taking hold, and those on the fringe of financial failure began to lose their grip on car ownership – leaving few transport choices but for walking and cycling. That same year also saw the first dip in national vehicle miles traveled for the first time since 1981, presumably leaving fewer opportunities for crashes. The right answer is likely a confluence of factors ranging from the physical (big vehicles) to the socioeconomic (more people walking during tough times) to the behavioral (feelings of security in a high-tech car that leads to speeding), as they commonly are in the social sciences. (2) Future research should investigate why Blacks are overrepresented and Hispanics/Latinos are underrepresented as pedestrian and pedalcyclist crash victims in Illinois. One possible reason for the underrepresentation of the Hispanic/Latino population could be that their crash and hospital files go unlinked because of miscommunication resulting in errors being recorded in a data field of either the crash or hospital file. Lack of state-issued identification may also contribute to the recording of erroneous patient information. Other possible explanations may simply be a distrust of police or lack of health insurance. (3) Future research should also examine the neighborhoods and communities in which cases are found to be most frequent. Findings in this area would provide powerful insights into interventions that may be tailored to narrowly target the most at-risk populations.

5. Conclusion

This paper has demonstrated the high cost of large motor vehicles on pedestrian and pedalcyclist injury severity, fatalities, and hospital charges. And once more, the most vulnerable among us seem to bear the greatest burden. Various solutions have been proposed and/or are in the works to address the mounting danger posed to pedestrians and pedalcyclists. The City of Chicago has taken aim at reducing all traffic injuries and fatalities, especially pedestrian and pedalcyclist, through their Vision Zero Chicago (VZC) Action Plan. VZC, (2017) articulates several goals: invest resources in communities equitably, foster a culture of safety, make streets safer for all users, and create safer drivers and vehicles. Accomplishing these stated VZC goals should go a long way in reducing death and injury on Chicago roadways; though some Vision Zero U.S. cities, including Chicago, may have actually seen pedestrian deaths increase since implementation (Bliss & Montgomery, 2019). At the federal level, the Government Accountability Office (GAO) reports that something as simple as reducing posted speed limits could reduce injury and death, findings also reflected by Arias et al. (2021). The GAO (2021) finds that 81% of pedestrian and 78% of cyclist fatalities occurred on a road with a posted speed limit of at least 35 miles per hour. Tiwari (2020) also finds that lowering motor-vehicle speed through enforcement has been effective at reducing pedestrian injuries and fatalities, while attempts at altering pedestrian behavior has largely not resulted in a reduction.

Other proposals are aimed directly at large motor vehicles. A bill originating in the New York State Senate would create a pedestrian rating system for motor vehicles that would be posted on the Department of Motor Vehicle's website (S7876, 2020) New York Assembly (2020). The intent there is to educate consumers of the danger posed to pedestrians and pedalcyclists by rating motor vehicles 1–5 based on the relative frequency of which they strike people. The rating system may also be used by insurance companies to charge drivers of high incidence vehicles higher premiums. Another more direct method of internalizing the external societal cost of driving large motor vehicles has been proposed by Tyndall (2021), among other economists and social scientists. Such proposals include implementing a Pigouvian tax at the federal, state, or local level equivalent to the marginal cost each example places upon society.

The widespread application of advanced automotive technology also has a role to play in harm reduction. Auto-braking and blind spot monitoring technology is not yet standard equipment on all motor vehicles sold in the United States, and the automakers' technologies do not all perform equally. One study found Subaru's EyeSight crash avoidance system with pedestrian detection reduced pedestrian-related insurance claims by 35% (HDLI, 2017). While another study of the Chevrolet Malibu, Honda Accord, Tesla Model 3, and Toyota Camry equipped with automatic emergency braking systems found their pedestrian detection systems were "significantly challenged" (AAA, 2019 p 4) in detecting pedestrians in three key metrics. Regulators should consider requiring detection technology to be standard vehicle equipment that also meets a minimum performance threshold for crash reduction.

As articulated here and in the VZC Plan, the identification of neighborhoods, roads, and intersections with a high frequency of occurrences is an important first step in the reduction of pedestrian and pedalcyclist injuries and fatalities. Arming communities with the knowledge, resources, and agency they require to address this important public health crisis is critical in achieving the goal of making our roads safe for all. Left to fester, the problem of pedestrian and pedalcyclist injuries and fatalities is certain to worsen.

Disclaimer

This research was supported by the National Center for Injury Prevention and Control of the Centers for Disease Prevention and Control (CDC) under award number CE16-1602. The content is solely the responsibility of the authors and does not necessarily represent the official views of the CDC.

Conflict of Interest

Supporting agency(ies) had no influence over the findings presented or analysis conducted. The Authors declare no other conflicting interests.

References

AAA. (2019). Automatic emergency braking with pedestrian detection. NewsRoom. AAA.com.

Arias, D., Ederer D., Rodgers M. O., Hunter M. P., & Watkins, K. E. (2021). Estimating the effect of vehicle speeds on bicycle and pedestrian safety on the Georgia arterial roadway network. *Accident Analysis & Prevention* Volume 161, 106351, ISSN 0001-4575.

Bliss, L., & Montgomery, D. (2019). What happens when a city tries to end traffic deaths? Bloomberg CityLab.

Braun, L. M., Le, H. T. K., Voulgaris, C. T., & Nethery, R. C. (2021). Healthy for whom? Equity in the spatial distribution of cycling risks in Los Angeles, CA. *Journal of Transport & Health, 23*, 101227. ISSN 2214-1405.

Coleman, H., & Mizenko, K. (2018). *Pedestrian and bicyclist data analysis (Research Note. Report No DOT HS 812 502)*. Washington, DC: National Highway Traffic Safety Administration.

Cunningham, R. M., Walton, M. A., & Carter, P. M. (2018). The major causes of death in children and adolescents in the United States. *The New England Journal of Medicine, 379*, 2468–2475.

Davis, S., & Boundy, R. (2020). *Transportation energy data book: Edition 38*. Oak Ridge National Laboratory: Energy and Transportation Science Division.

Desapriya, E., Subzwari, S., Sasges, D., Basic, A., Alidina, A., Turcotte, K., & Pike, I. (2010). Do light truck vehicles (LTV) impose greater risk of pedestrian injury than passenger cars? A meta-analysis and systematic review. *Traffic Injury Prevention, 11*(1), 48–56.

Doggett, S., Ragland, D., & Felschundneff, G. (2018). *Evaluating research on data linkage to assess underreporting of pedestrian and bicyclists injury in police crash data*. UC Berkeley: Safe Transportation Research & Education Center.

Government Accountability Office. (2021). *Pedestrians and cyclists: Better information to states and enhanced performance management could help DOT improve safety.* Report to the ranking member, Committee on Commerce, Science, and Transportation, U.S. Senate.

Governors Highway Safety Association (GHSA): Spotlight on Highway Safety. 2021. *Pedestrian Traffic Fatalities by State.*

Highway Loss Data Institute (HDLI). 2017. *Effect of Subaru EyeSight on pedestrian-related bodily injury liability claim frequencies.* Vol. 34: #39.

McGlincy, M. H. (2021). *LinkSolv*. Morrisonville, NY: Strategic Matching.

New York Assembly. New York State Senate. Transportation Committee. 2020. *An act to amend the vehicle and traffic law, in relation to creating a pedestrian safety rating system for motor vehicles*. 2019-2020 Legislative Session. S7876.

Roudsari, B.S., C.N. Mock, R. Kaufman, D. Grossman, B.Y. Henary, and J. Crandall. 2004. "Pedestrian crashes: higher injury severity and mortality rate for light truck vehicles compared with passenger vehicles." *Inj. Prev*., 10 (3) (2004), pp. 154-158.

Tiwari, G. (2020). Progress in pedestrian safety research. *International Journal of Injury Control and Safety Promotion, 27*(1), 35–43.

Tyndall, J. (2021). Pedestrian deaths and large vehicles. *Economics of Transportation, 26-27*.

U.S. Census Bureau. (2019). 2015-2019 5-year American Community Survey.

U.S. Environmental Protection Agency. (2021). *The 2020 EPA Automotive Trends Report*. EPA-420-R-21-003.

Vision Zero Chicago (VZC). (2017). *Chicago's initiative to eliminate traffic fatalities and serious injuries by 2026: Action Plan.*

Mickey Edwards, PhD, is a Visiting Researcher at the University of Illinois Springfield where he studies motor vehicle crashes and their effect on pedestrians, cyclists, and communities. Other research interests include transportation and infrastructure equity. He has worked on policy in the U.S. Senate and taught several courses at the University of Cincinnati. Previous careers include several years in photojournalism and a stint as an engineer for a large consumer goods company.

Journal of Safety Research 82 (2022) 459–462

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Erratum

Erratum to "Highway-rail grade crossings accident prediction using Zero Inflated Negative Binomial and Empirical Bayes method" [J. Saf. Res. 79 (2021) 211–236]

Jacob Mathew *, Rahim F. Benekohal

Newmark Civil Engineering Lab, 205 N Mathews Ave., University of Illinois at Urbana-Champaign, IL 61820, United States

The publisher regrets that there were some editorial discrepancies in the original article. Below you'll find the exchange between the FRA and the authors of the article. Only minor editorial changes were made to the article.

The publisher would like to apologize for any inconvenience caused.

From the United States Department of Transportation, Federal Railroad Administration

The United States Department of Transportation, Federal Railroad Administration (FRA) firmly denies the allegation by Jacob Mathew and Professor Rahim Benekohal that the "New Model for Highway-Rail Grade Crossing Accident Prediction and Severity" (New APS) adopted any of their work. FRA determined that the Mathew-Benekohal articles relied on an outdated 3-equation approach and that their small sample sizes were inadequate. For a complete list of the resources used to develop the New APS, see full report, Section 8. References, page 50, posted here: https://railroads.dot.gov/elibrary/new-model-highway-rail-grade-crossing-accident-prediction-and-severity.

Mathew and Benekohal's response to The United States Department of Transportation, Federal Railroad Administration (FRA)'s alleged errata for the article "Highway-rail grade crossings accident prediction using Zero Inflated Negative Binomial and Empirical Bayes method".

Why we say the new FRA model was adopted from the work we had done previously.

The reasons we say the new FRA model was adopted from the work we had done previously are threefold.

A. We had advocated for the Zero Inflated Negative Binomial model with Empirical Bayes Adjustment for a while. Before publishing this article in JSR, we had worked on developing the ZINB model and published the following 4 study documents that are easily available:

1. Mathew J, Benekohal. R. F., "A New Accident Prediction Model for Highway-Rail Grade Crossings Using the USDOT Formula Variables". J. Traffic and Transportation Eng [Internet]. 2020;8(1):1–13. Available from: http://www.davidpublisher.com/Public/uploads/Contribute/5f43533833a24.pdf

2. Mathew J, Benekohal RF, Medina J. C., "Accident Prediction Models using Macro and Micro Scale Analysis: Dynamic Tree and Zero Inflated Negative Binomial Models with Empirical Bayes Accident History Adjustment". 2019; Available from: https://conservancy.umn.edu/bitstream/handle/11299/201747/CTS 19-02.pdf

3. Medina J. C, Shen S, Benekohal R. F. "Micro and Macro Level Safety Analysis at Railroad Grade Crossings" [Internet]. 2016. Available from: https://www.nurailcenter.org/research/final_reports/UIUC/NURail2012-UIUC-R02_Final_Report_Benkohal.pdf

4. Medina J. C, Benekohal R. F. "Macroscopic Models for Accident Prediction at Railroad Grade Crossings". Transp Res Rec J Transp Res Board [Internet]. 2015;2476:85–93. Available from: https://journals.sagepub.com/doi/10.3141/2476-12

B. We had discussed the ZINB model with Empirical Bayes adjustment models with an FRA Engineer in change grade crossing safety research and had provided him the above 4 documents (initial version of document 1 and the final versions of the remaining three documents) prior to the conduct of the study that led to the new FRA technical report. The FRA Engineer is acknowledged in the FRA report – "The authors appreciate the insightful review comments by Federal Railroad Administration (FRA) staff Karen McClure and Francesco Bedini". After the FRA report was published, we complained to the FRA engineer that our publications are not even listed among the references, and he suggested that our publications should have been listed among the references for the FRA report. It was his sugges-

DOI of original article: https://doi.org/10.1016/j.jsr.2021.09.003

* Corresponding author.

E-mail addresses: jmathew7@illinois.edu (J. Mathew), rbenekoh@illinois.edu (R. F. Benekohal).

https://doi.org/10.1016/j.jsr.2022.07.013

tion that we send him the list of our papers to be added to the reference list of the FRA report. We did, but now FRA refuses to cite any of these publications.

C. The validation approach used by the new FRA report is very similar to the approaches we had developed and used in our publications. Those three approaches are:

a. Plotting the cumulative predicted value with respect to the field data.

b. Plotting the predicted value with respect to the observed value

c. Ranking of the crossings based on the predicted value and comparing to the field data.

The new FRA model has used the three approaches that we used with very slight modifications. The validation approach used by the FRA is given in Section 6 (page 39) of the article "New Model for Highway-Rail Grade Crossing Accident Prediction and Severity.".

"The first validation compares cumulative predicted accidents by the new model and the APS with the actual risk as measured by accident counts.
The second validation shows the predicted accidents for the new model and the APS for crossings grouped by accident count.
The third comparison examines the model results (the new model and APS) for different groupings of high-risk crossings and shows the results in a chart.".

Therefore, when the methodology and the validation approaches used are very similar to a previously published article, it is expected to cite the original articles and give proper credit to them. We do not claim that we developed statistical procedures like the Zero Inflated Negative Binomial method or the Empirical Bayes method. We, however, have been advocating using these two procedures together for the prediction of accidents at Highway-Rail Grade Crossings for a while.

In our correspondence with Ms. Karen McClure of FRA, she claimed "FRA used the normalizing constants approach developed by the U. S. DOT in 1987 to validate the New APS model." Using normalization constant is not a validation procedure but rather an attempt to adjust the predictions to changes in the conditions over time. It should not be mistaken for a validation procedure. This is explained in the report titled "FRA Collects Reliable Grade Crossing Incident Data but Needs To Update Its Accident Prediction Model and Improve Guidance for Using the Data To Focus Inspections". Page 9 of this report states that "These constants keep the accident predictions matched with current accident trends, numbers of open public grade crossings, and changes in warning devices. Each constant is a ratio of the actual number of accidents to the predicted number of accidents. The constant used in the formula depends on the type of warning device at the pertinent grade crossing—a passive warning device, a flashing light, or a gate".

In a later correspondence with Ms. McClure, she claimed "The New APS validation methods (cumulative distribution and comparison of predicted to actual accident rates) have been used by FRA in the GradeDec model since 2001 and in the US DOT Accident Prediction and Severity Model since 1987.".

<u>Regarding the use of validation approaches in the GradeDec model</u>.

"The Federal Railroad Administration developed GradeDec.NET, a highway-rail grade crossing investment analysis tool, to provide grade crossing investment decision support." GradeDec model is "a web-based application and decision support tool for the identification and evaluation of highway-rail grade crossing upgrades, separations and closures."

https://railroads.dot.gov/program-areas/highway-rail-grade-crossing/gradedecnet-crossing-evaluation-tool

https://gradedec.fra.dot.gov/

We checked the following documents associated with GradeDec model that are available in the GradeDec website to check for Ms. McClure's claims:

1. GradeDec.Net 2017 User's Manual.pdf. Available at https://railroads.dot.gov/elibrary/gradedecnet-2017-users-manual
2. GradeDec.Net 2019 Reference Manual. Available at https://railroads.dot.gov/elibrary/gradedecnet-2019-reference-manual
3. Workbook 2003. Available at https://railroads.dot.gov/elibrary/gradedec-crossing-evaluation-tool-workbook-2003

The word "validation" appears once in the GradeDec.Net 2017 User's Manual in Section 5.1, which is given below.

"This page possesses a number of features that let you easily visualize data and quickly develop probability distributions that best reflect available information and judgments on operations, future developments and social costs. These features include:

- Ease of navigation among variables
- Instant viewing of statistics and charts
- Instant ***validation*** and saving of ranges"

We couldn't find any other reference to the validation of the GradeDec model in the GradeDec documentation.

<u>Regarding the use of validation approaches in USDOT Accident Prediction and Severity model since 1987.</u>

We found the following five articles related to the USDOT Accident Prediction and Severity model that were published in 1979, 1982, 1986, 1987, and 1987 respectively.

1. Rail-highway crossing hazard prediction: research results. Available at https://rosap.ntl.bts.gov/view/dot/8584
2. Summary of the Department of Transportation Rail-Highway Crossing Accident Prediction Formulas and Resource Model. Available at https://rosap.ntl.bts.gov/view/dot/11643
3. Rail-Highway Crossing Resource Allocation Procedure. User's Guide. 2nd edition. Available at https://rosap.ntl.bts.gov/view/dot/10417
4. Rail-Highway Crossing Resource Allocation Procedure User's Guide Third Edition. Available at https://railroads.dot.gov/elibrary/rail-highway-crossing-resource-allocation-procedure-users-guide-third-edition
5. Summary of DOT Rail-Highway Crossing Resource Allocation Procedure – Revised. Available at https://rosap.ntl.bts.gov/view/dot/11385

Only in the first of the above five articles (Rail-highway crossing hazard prediction: research results), the author used a validation approach in which power factors (i.e., fraction of accidents occurring at a given fraction of the most hazardous crossings) and empirical operating characteristics (i.e., a table giving power factors, cumulative accidents at various percentages of hazardous crossings, etc.) to compare the model developed in the article to New Hampshire and Coleman-Stewart models. In none of the remaining four manuscripts, do the authors of the respective manuscripts describe a validation procedure.

After all, it is not our responsibility to prove that our approaches were used in a later study; it is the responsibility of the researchers conducting the new study to properly cite the previous study.

Regarding Sample Size

We would like to emphasize that we do not believe our sample size is small.

We used data from 10,292 crossings of which 5097 are in Illinois and 5195 are in Texas. The dataset from Illinois consists of 2755 datapoints for gates, 960 datapoints for flashing lights, and 1382 datapoints for crossbucks. The dataset from Texas consists of 3573 datapoints for gates, 346 points for flashing lights, and 1276 data points for crossbucks. We believe this a good sample size and yield a reliable model. Previous traffic safety studies that utilized the Empirical Bayes approach have used much smaller sample sizes to conduct their studies (a few of them are listed below).

1. The study by Perusad et al. used Empirical Bayes approach to study stop controlled intersections in CA. The data consisted of 1669 stop-controlled intersections

Persaud, Bhagwant, and Craig Lyon. "Empirical Bayes before–after safety studies: lessons learned from two decades of experience and future directions." Accident Analysis & Prevention 39.3 (2007): 546–555. Available at: https://web.engr.uky.edu/~rsouley/CE%20635-2021/docs/Persaud%20-%2010%20years%20of%20EB.pdf.

2. The study by Harwood et al, used Empirical Bayes to study rural two-lane highways. The data consisted of 619 segments from Minnesota and 712 from Washington.

Harwood, Douglas W., et al. Prediction of the expected safety performance of rural two-lane highways. No. FHWA-RD-99-207, MRI 4584–09, Technical Report. United States. Federal Highway Administration, 2000. Available at: https://rosap.ntl.bts.gov/view/dot/14465.

3. Perusad et al. used 197 signalized intersections to develop an EB procedure to rank intersections.

Persaud, Bhagwant, Craig Lyon, and Thu Nguyen. "Empirical Bayes procedure for ranking sites for safety investigation by potential for safety improvement." Transportation research record 1665.1 (1999): 7–12.

Additionally, the notion that data form two large states (IL and TX) is not enough for modeling railroad grade crossing accident is baseless. Different states use their own data for modeling as well as planning and management of their system and the issue of sample size is not a concern when you use your own state's data.

Furthermore, we checked the adequacy of our sample size by fitting a model using half of the data (test-model) to see if it would drastically alter the coefficients of the model based on full data (which is our ZINEBS model). In Tables 1–3 the coefficients of the test-models and ZINEBS models are shown. Table 1 is for the model for Gates, Table 2 is for the model for flashing lights, and Table 3 is for the model for crossbucks. In all the three cases, the coefficients of the test-models are similar to the coefficients of ZINEBS models.

Table 1
Coefficients for ZINB model for Gates.

	Model using all data ($n = 2755$)	Model using half the data ($n = 1377$)
Count Model		
Intercept	−2.4701	−2.2361
Total Tracks	0.2738	0.1647
Traffic Lanes	0.2838	0.2736
Angle Category (Angle > 60)	−0.6618	−0.5108
Zero Inflation Model		
Intercept	4.97211	5.86447
Total Train	−0.05382	−0.05297
Log(Aadt)	−0.48475	−0.59775

This clearly indicates that our sample size in more than adequate and is not small as the Ms. McClure claimed.

In addition, the authors would also like to point out that although using data from many states gives a large sample size, the large data is not always advantageous and may have issues because of the differences in data collection approaches used by different states. For example, a previous study by Ng and Hauer examined a pooled data from 7 states (Alabama, Michigan, Montana, North Carolina, Utah, Washington, and West Virginia) in a study of the accidents on rural two-lane roads. They concluded, "for the same amount of traffic, different states record widely discrepant numbers of accidents."; and said that data from different states must not be pooled.

Ng, Jerry CN, and Ezra Hauer. "Accidents on rural two-lane roads: differences between seven states (with discussion and closure)." Transportation research record 1238 (1989): 1–9.

Regarding 3-equation approach

We do not believe a 3-equation approach is an outdated approach, and it has its advantages. Using a 3-equation approach gives flexibility in selecting variables that is a significant contributor for crossings of a particular warning device type (gates, flashing

Table 2
Coefficients for ZINB model for Flashing Lights.

	Model using all data ($n = 960$)	Model using half the data ($n = 480$)
Count Model		
Intercept	−3.58567	−3.67268
HwySpeed	0.03604	0.05028
Zero Inflation Model		
Intercept	14.8944	8.5511
Total Train	−0.2528	−0.1442
Log(Aadt)	−1.8479	−0.9169

Table 3
Coefficients for ZINB model for Crossbucks.

	Model using all data ($n = 1382$)	Model using half the data ($n = 691$)
Count Model		
Intercept	−2.33572	−2.2176
Surface = Concrete	0.47339	0.7707
Surface = Rubber	0.03915	0.4232
Surface = Timber	−0.82825	−1.8172
Surface = Unconsolidated	−0.09580	−0.8064
Zero Inflation Model		
Intercept	7.3949	10.5639
Total Train	−0.4195	−0.7554
Log(Aadt)	−1.0705	−1.4680

lights, and crossbucks). The variables selected in the ZINEBS model were the results of classical statistical methods for variable selection. In addition, we made sure that the selected variables make sense and are the appropriate. In section 6 of our paper, we analyzed each of the variables that we identified as significant. We also showed how the variables that were significant for one device type was in fact not significant for another device type. Using a single equation model (FRA used it) not only does not give this flexibility in variable section, but also makes you to use the same variables for all devices. Furthermore, different warning devices offer different levels of protection. For this reason, the type of highway vehicle movement is expected to be different. It is therefore more appropriate to use different equations for different warning device types.

Journal of Safety Research 82 (2022) 251–260

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Examining the contribution of psychological resilience on self-reported and naturalistic driving behavior of older adults

Renée M. St. Louis [a,b,*], Sjaan Koppel [a], Lisa J. Molnar [b], Marilyn Di Stefano [c], Peteris Darzins [d], Michelle M. Porter [e], Michel Bédard [f], Nadia Mullen [f], Anita Myers [g], Shawn Marshall [h], Judith L. Charlton [a]

[a] Monash University Accident Research Centre, Monash University, Melbourne, Victoria, Australia
[b] University of Michigan Transportation Research Institute, Ann Arbor, MI, USA
[c] Department of Transport, Melbourne, Victoria, Australia
[d] Monash University Eastern Health Clinical School, Melbourne, Victoria, Australia
[e] University of Manitoba, Winnipeg, Manitoba, Canada
[f] Lakehead University, Thunder Bay, Ontario, Canada
[g] University of Waterloo, Waterloo, Ontario, Canada
[h] Ottawa Hospital Research Institute, Ottawa, Ontario, Canada

ARTICLE INFO

Article history:
Received 15 February 2021
Accepted 1 June 2022
Available online 11 June 2022

Keywords:
Psychosocial factors
Real-world driving
Older drivers
Road safety
Mobility

ABSTRACT

Introduction: This study examined the contribution of psychological resilience on self-reported driving comfort, abilities, and restrictions, and on naturalistic driving (ND) behavior of older adults at two time points, five years apart ($N = 111$; Male: 65.8%, Mean age = 86.1 years). *Method:* Participants from the Ozcandrive older driver cohort study completed a demographic questionnaire, functional assessments, psychosocial driving questionnaires, and a resilience scale. Participants' vehicles were equipped with a recording device to monitor driving behavior throughout the study. Over 1.7 million kilometers of ND data were analyzed. *Results:* There was a significant increase in resilience over time, and both self-reported and ND measures revealed reduced driving across five years. Hierarchical regression analyses using age, sex, driving exposure, functional measures, and resilience showed that adding resilience into the models at the final step resulted in statistically significant increases in the amount of variance explained for driving comfort during the day and night, perceived driving abilities, number of trips, trip distance, and proportion of night trips. *Conclusions:* This research leveraged the longitudinal nature of the Ozcandrive study to provide the first insights into the role of resilience and ND. The observed patterns of reduced driving, captured by both subjective and objective measures, are suggestive of increased levels of self-regulation. As resilience is associated with adaptive coping skills, older adults with higher resilience may be able to more effectively engage in appropriate coping behaviors with regard to driving behavior, safety, and mobility. *Practical Applications:* Effective methods of increasing resilience in the context of driving is worthy of future research as it will provide valuable information about how older drivers navigate the process of aging as it relates to driving and may assist stakeholders in developing suitable measures to support older driver safety.

1. Introduction

Population aging and road trauma data have resulted in increased attention regarding older road user safety and mobility. Despite older adults' (i.e., aged 65 years and older) reliance on personal vehicles for mobility (Newbold & Scott, 2017; Somenahalli et al., 2016), they may experience declines in abilities needed for safe driving as a result of medical conditions, the medications used to treat those conditions, or the aging process itself (Eby et al., 2018). In response to these declines and the resulting discomfort of driving in certain situations, older adults often change their driv-

* Corresponding author at: Monash University Accident Research Centre, 21 Alliance Lane, Clayton, Victoria 3800, Australia.

E-mail addresses: renee.stlouis@monash.edu (R.M. St. Louis), sjaan.koppel@monash.edu (S. Koppel), ljmolnar@umich.edu (L.J. Molnar), marilyn.distefano@roads.vic.gov.au (M. Di Stefano), peteris.darzins@monash.edu (P. Darzins), michelle.porter@umanitoba.ca (M.M. Porter), mbedard@lakeheadu.ca (M. Bédard), nmullen@lakeheadu.ca (N. Mullen), amyers@uwaterloo.ca (A. Myers), marshall@toh.ca (S. Marshall), judith.charlton@monash.edu (J.L. Charlton).

https://doi.org/10.1016/j.jsr.2022.06.001
0022-4375/

ing patterns by driving less frequently or avoiding challenging road environments, a process known as self-regulation (Molnar et al., 2015). Although more research is needed to evaluate the impact of self-regulation on actual safety outcomes, self-regulation is considered a compensatory coping strategy for older drivers (e.g., Baldock et al., 2006; Molnar et al., 2015; Tuokko et al., 2013) that enables continued mobility while also reducing the risk of crash-related injury or death.

Examining the impact of psychosocial factors on driving has greatly improved our understanding of self-regulation in older adults (see Wong et al., 2016 for a review). One consistent finding is the relationship between self-regulation and perceptions of confidence or driving comfort, specifically by avoiding particular driving situations as comfort decreases (e.g., Baldock et al., 2006; Blanchard & Myers, 2010; Charlton et al., 2006; Jouk et al., 2014, 2016; MacDonald et al., 2008; Molnar et al., 2013, 2014; Sirén & Meng, 2013). While driving comfort appears to be a significant psychosocial factor in self-regulation and driving decisions, a systematic review of psychosocial factors revealed a lack of investigation into other psychosocial factors that underlie self-regulation (Wong et al., 2016), and called for more research on other variables, such as self-efficacy, mastery, and personality traits, to be considered in studies about older driver decision making. This is particularly pertinent as psychosocial factors, specifically attitudes and perceptions, may influence driving practices more than objective driving abilities (Baldock et al., 2006; Blanchard & Myers, 2010; MacDonald et al., 2008).

Another psychosocial factor worthy of consideration in the context of older drivers is psychological resilience, conceptualized as the "process of effectively negotiating, adapting to, or managing significant sources of stress or trauma" (Windle, 2011, p. 163). Resilience is a malleable construct, evolving over time and within various contexts throughout an individual's life (Leppin et al., 2014), and reinforced through challenges faced as a result of aging. In this respect, the oldest-old (i.e., 85 years and older) have been exposed to more adverse life events, but have also had more time to learn and develop effective coping methods for managing their changing situations (Hayman et al., 2017). Within aging literature, psychological resilience is viewed as a strengths-based approach to understanding individuals' learned coping skills (Browne-Yung et al., 2017). Coupling this notion with research on driving in later-life, self-regulation can be seen as a healthy, adaptive response to declines in driving-related abilities or decreased comfort in driving abilities, which facilitates continued mobility.

Previous research into the role of resilience and driving established an association between resilience and self-reported driving comfort, abilities, and frequency in a cohort of older adults (St. Louis et al, 2020). This research showed that higher levels of resilience were associated with higher levels of self-reported daytime driving comfort, more positive perceptions of driving abilities, and more frequent driving in a variety of complex scenarios. The conclusions from this study suggested that high resilience in the context of transportation may be indicative of greater acceptance of one's declines in driving capabilities and more flexibility in planning for and utilizing alternative approaches for maintaining mobility.

Self-reported measures provide unique insight into the driving behavior of older adults and have greatly contributed to our understanding of driving decision-making; however, these results are susceptible to social desirability and recall bias (Grengs et al., 2008; Sullman & Taylor, 2010) and can lead to a misrepresentation of actual driving practices (Blanchard et al., 2010; Huebner et al., 2006; Porter et al., 2015). Global Positioning System (GPS) technology, placed unobtrusively in a driver's vehicle, allows the monitoring of driving behaviors under naturalistic conditions and provides objective measures of travel patterns (e.g., location, time of day, and speed). Naturalistic driving (ND) data, in conjunction with self-reported measures, provide a more complete picture of the driving environment and the context for specific driving behaviors. The overarching goal of this research is to extend previous findings regarding the role of resilience in the context of aging drivers by combining ND data, health and functioning information, and psychosocial measures in a longitudinal driving study.

1.1. Aims

This study had three aims: (1) evaluate resilience at two time points, five years apart; (2) determine the extent to which the previous findings regarding resilience and driving comfort, abilities and restrictions were confirmed in a subsequent year; and (3) investigate resilience in the context of real-world driving by examining GPS-derived ND data to explore the relationship between resilience and objective driving measures. Three hypotheses were assessed. Hypothesis 1 (H1): Resilience scores will significantly increase over time; (H2) Resilience will remain a significant variable in explaining the variance in self-reported driving comfort, abilities and restrictions in a five-year follow-up; and (H3) Resilience will be a significant variable in explaining the variance of driving behavior as measured by GPS data in the final year of the study.

2. Method

2.1. The Candrive/Ozcandrive study

The Candrive/Ozcandrive prospective cohort study was an international research program comprised of 1,230 participants from Canada ($n = 928$), Australia ($n = 257$), and New Zealand ($n = 45$). Its primary purpose was to develop and validate a screening tool to help clinicians identify potentially at-risk drivers. Participants were enrolled for up to eight years and completed annual in-person assessments to document physical functioning and cognitive performance, and collect information on demographics, health/functioning, driving, and self-reported crash involvement. Psychosocial data were also collected through mailed questionnaires. ND data were collected with a GPS device installed in the participants' vehicle. For a complete description of the Candrive/Ozcandrive study protocol, see Marshall et al. (2013).

2.2. Participants

The current study utilized data from Year 3 and Year 8, which allowed for an evaluation of the variables of interest at two time points that were approximately five years apart. The sample included a subset of 111 participants from the Australian cohort who had completed, in both years, an annual in-person assessment and resilience scale, and had valid ND data for both years. Participants were required to: be aged 75 years or older; have a valid driver license; drive a vehicle model year 2002 or newer; drive at least four times a week; and reside in the local regions of Melbourne. Participants were not eligible if they: planned to move out of the study area; had been diagnosed with a progressive disease that could affect driving; or had a medical contraindication to driving within the previous six months (Austroads, 2017).

2.3. Measures

The Ozcandrive database is comprised of both subjective and objective measures of functional performance and driving behavior. Data reported for the present study include: demographics; resilience; cognitive, physical and visual functioning; self-reported driving comfort, abilities, and restrictions; and naturalistic driving measures. Relevant measures are discussed below; full details are provided in the Candrive/Ozcandrive protocol paper (Marshall et al., 2013).

2.3.1. Resilience

Resilience was measured using the 14-item Resilience Scale (RS-14; Wagnild, 2009). The scale includes five underlying characteristics of resilience that serve as the conceptual foundation: self-reliance, purpose, equanimity, perseverance, and authenticity (Wagnild & Young, 1990). Participants were asked to rate their agreement on each item on a 7-point scale from strongly disagree to strongly agree. Total scores range from 14 to 98, with higher scores indicating a higher level of resilience. The RS-14 had strong internal consistency, with Cronbach's alpha coefficients of 0.94 in Year 3 and 0.92 in Year 8, the two years in which this scale was administered.

2.3.2. Physical, cognitive and vision assessments

To describe the functional and cognitive capacity of the sample, several measurements were used. The Activities of Daily Living (ADL) subsection of the Older Americans Resources and Services (OARS) Multidimensional Functional Assessment Questionnaire (Fillenbaum, 1988) was used to assess participants' ability to perform both ADLs (e.g., basic self-care tasks such as bathing and dressing) as well as instrumental activities of daily living (IADLs; slightly more complex tasks such as meal preparation and managing finances). Participants rated their ability to perform seven ADLs and seven IADLs on a 3-point scale (where 0 = unable to perform; 1 = can perform with some help; and 2 = can perform without help). The total score is the sum of all 14 items, and scores can range from 0 to 28, with higher scores indicating greater independence. A score of 26 suggests mild impairment.

Participants' cognition was measured using the Trail Making Test Part B (Trails B; Moses, 2004), a timed test of visual attention and executive functioning commonly used in older driver research. Participants were instructed to draw lines to connect circled numbers and letters in an alternating numeric and alphabetic sequence (e.g., 1-A-2-B-3-C) as quickly as possible. The score is the overall time required to complete correctly all of the connections. Scores exceeding 180 seconds indicate increased crash risk in adults aged 55 years and older (Staplin et al., 2003).

To assess physical functioning, participants completed the Rapid Pace Walk (RPW), a timed measure of motor speed, balance, and coordination (Carr et al., 2010). Participants walked 10 feet along a path marked with tape, and returned along the same path as quickly as possible. Increased crash risk is associated with a completion time greater than 10 seconds (Staplin et al., 2003).

Binocular distant visual acuity was assessed with the Snellen eye chart. Participants used their usual corrective lenses, if any were used for driving. Visual acuity scores were reported as logarithm of the minimum angle of resolution (LogMAR) scores (McGwin & Brown, 1999), where LogMAR = 0 is considered normal vision and LogMAR=+0.3 is considered reduced vision and indicates the legal limit for driving, after which corrective lenses are required to be worn as a license condition (Austroads, 2017).

2.3.3. Self-reported driving comfort, abilities and restrictions

Five self-reported measures were used to examine participants' levels of comfort while driving in various situations during the day and night, perceptions of their own driving ability, and estimates of driving frequency and avoidance.

Driving comfort: Participants rated their comfort while driving in a variety of scenarios using the 13-item daytime and 16-item nighttime Driving Comfort Scales (DCS-D and DCS-N, respectively). Self-efficacy, as defined by Social Cognitive Theory (Bandura, 1986), serves as the basis for the DCS-D and DCS-N. These scales were systematically developed with older drivers, and subjected to rigorous psychometric examination using Rasch analysis (MacDonald et al., 2008; Myers et al., 2008). Each driving scenario (e.g., "How comfortable are you driving in the daytime in heavy rain?") was rated on a 5-point scale from 0% (not at all comfortable) through 100% (completely comfortable). Possible scores range from 0% to 100%, with higher scores indicating greater driving comfort.

Perceived driving abilities: Participants rated aspects of their current driving abilities using the 15-item Perceived Driving Abilities (PDA) scale, which was also developed with older drivers and found to have strong psychometric properties (MacDonald et al., 2008; Myers et al., 2008). Participants rated their current abilities (e.g., see road signs at a distance, make quick driving decisions) on a 4-point scale (0 = poor to 3 = very good). Total scores range from 0 to 45, with higher scores indicating more positive perceptions of driving abilities.

Situational driving frequency and avoidance: Self-reported driving restrictions were assessed using the Situational Driving Frequency (SDF) and Avoidance (SDA) scales. The SDF scale asks participants how often they drive, on average, in 14 different driving scenarios (e.g., in unfamiliar areas, with passengers) on a 5-point scale from never (0) to very often (4 = four to seven days a week). Total scores range from 0 to 56, with higher scores indicating driving more often in challenging situations. On the SDA scale, older drivers are asked 'If possible, do you try and avoid any of these driving situations?' for 19 situations (e.g., rural areas at night, fog). Scores range from 0 to 19, with higher scores indicating greater avoidance (MacDonald et al., 2008; Myers et al., 2008).

2.3.4. Naturalistic driving measures

Objective driving exposure was captured using a custom-designed in-car recording device (ICRD; OttoView-CD autonomous data logging device) and software suite developed for Candrive by Persen Technologies Inc. (Winnipeg, Manitoba; Marshall, et al., 2013). The ICRD was powered via the On-Board Diagnostic (OBDII) port of each participant's primary vehicle, and collected vehicle information (e.g., time/date of trip, speed, and distance travelled). A GPS antenna enabled vehicle location information to be collected via latitude/longitude coordinates. A radio frequency identifier (RFID) system consisting of a small antenna and key chain fob that emitted RFID signals was used to identify participant trips in the event the participant vehicle was driven by someone else. Such participants also used a log book to document when a non-participant drove the vehicle. Data were collected at a rate of 1 Hz onto a Secure Digital card changed approximately every four months. When participants changed their primary vehicle, the ICRD was transferred into the new vehicle. To derive measures of interest for the current study (see Table 1), algorithms were developed and applied to the raw GPS data.

2.4. Procedure

Ethics approval was obtained from the University's Ethics Committee (CF09/3656-2009001969). Demographic, functional, and health data were collected by a Research Nurse at in-person annual assessments. Mail back questionnaires on driving were also completed. The RS-14 was administered in Years 3 and 8 as part of a separate package of mail back questionnaires. ND data were collected throughout the course of the study. The current study examined ND data from Years 3 and 8 when available. If data were missing or unavailable based on exclusion criteria (listed below), data closest to the time of administration of the RS-14 were used. Therefore, Year 2 and/or Year 7 ND data were used for participants without valid ND data in Year 3 ($n = 9$) and/or Year 8 ($n = 13$),

Table 1
Description of naturalistic driving data measures.

Driving behavior	Outcome variable	Definition
Driving Distance	Total distance	Kilometers driven per year
Driving Frequency	Total trips	Number of trips per year
Trip Distance	Mean trip distance	Mean distance per trip (ignition on/off, km)
Shorter/Longer Trips	% ≤5 km or >20 km	Proportion of trips within length categories: ≤5 km or >20 km
Night Time Driving	% Night	Proportion of total annual trips driven at night (i.e., between 1800 and 0600 hours)
Peak Hour Driving	% Peak hour	Proportion of total annual trips driven during peak traffic hours (i.e., weekday periods between 0700 and 0930 hours or between 1600 and 1800 hours)

respectively. Between Years 3 and 8, the Ozcandrive cohort had decreased from 229 to 155 participants for several reasons (deceased, health concerns, stopped driving, vehicle compatibility issues, licensing issues, or lost interest). An additional 44 participants were excluded from the analysis because they: did not complete the RS-14 in one or both years ($n = 33$), ceased driving in Year 8 ($n = 10$), or had invalid driving data for one or both years ($n = 17$). Analyses were conducted to assess differences between the sample for the current study and participants who were active in Ozcandrive at Year 3 but did not meet inclusion criteria for this study due to missing variables of interest. There were no statistically significant differences between the current sample ($N = 111$) and the active, ineligible Ozcandrive participants at Year 3 ($N = 87$) by age, sex, or resilience scores (at $p < .05$), however the groups differed on two of the seven naturalistic driving measures listed in Table 1. The current sample had driven significantly more total kilometers and more overall trips.

2.5. Analysis

Analyses were conducted using IBM SPSS Statistics version 24. To obtain summary measures for ND variables, raw data files were exported from the ICRD memory cards using the Candrive/Ozcandrive software. Based on past work (Charlton et al., 2019; Hua et al., 2018), data were subjected to similar rigorous exclusion criteria including: (1) missing data for the current variables of interest; (2) unexplained interruptions in a participant's driving (i.e., breaks in driving of two months or greater not accounted for in participant's log book); (3) data were affected by RFID fob detection issues (i.e., periods of two months or greater during which no RFID fob was detected); (4) participant drove a secondary vehicle more than 30% of their total distance driven; and (5) participants' log book entries differed significantly from driving data. Driving trips were also excluded if no RFID fob was detected for that trip, or if trip times overlapped by at least 50% with an entry in the log book. Ultimately, over 1.7 million kilometers of driving were analyzed.

Due to non-normally distributed data confirmed using the Shapiro-Wilk test, non-parametric tests were used for analyses. Hierarchical regression analyses were conducted to assess the unique proportion of variance explained by resilience in the models for self-reported and ND outcome measures for Year 8 data. Selection of independent variables was based on the theoretical framework and previous research that indicated variables of interest *a priori*.

3. Results

One hundred eleven Ozcandrive participants (Male: 65.8%, Mean age at Year 8 = 86.1 years, SD = 2.8 years) completed in-person annual assessments and the RS-14 in Years 3 and 8, and had valid ND data. Some measures remained stable over time (i.e., education level, 54.0% diploma/degree/postgraduate), while others demonstrated noteworthy shifts between Years 3 and 8, namely a higher proportion of widowed participants, and decreases in driving frequency and distance (Table 2).

Table 3 shows results from Year 3 and Year 8 for participants' resilience, functional performance, and self-reported and ND measures, as well as changes in these variables. There was a statistically significant increase in resilience scores from Year 3 (Mean = 79.8, Median = 82.0, IQR = 72.0–87.0, Range = 53–98) to Year 8 (Mean = 82.4, Median = 84.0, IQR = 77.0–89.0, Range = 38–98; $Z = -3.3$, $p < .01$). There were statistically significant declines in participants' functional performance on all measures except for Trails B. The self-reported driving measures showed significant decreases in driving comfort during the day and at night, perceived

Table 2
Demographics and self-reported driving frequency ($N = 111$).

Demographic and Driving Characteristics		Y3 % (N)	Y8 % (N)
Age group	75–79 years	31.5 (35)	0.0 (0)
	80–84 years	56.8 (63)	30.6 (34)
	85–89 years	11.7 (13)	56.8 (63)
	≥90 years	0.0 (0)	12.6 (14)
Marital status	Single (never married)	9.9 (11)	9.9 (11)
	Married/partnered	53.3 (58)	39.6 (44)
	Divorced/separated	3.6 (4)	3.6 (4)
	Widowed	34.2 (38)	46.8 (52)
Frequency of driving	Daily	41.4 (46)	38.7 (43)
	4–6 times per week	54.1 (60)	47.7 (53)
	2–3 times per week	4.5 (5)	12.6 (14)
	Once per week	0.0 (0)	0.9 (1)
Estimated kilometers driven in past year	≤5,000 km	19.8 (22)	41.4 (46)
	5,001–10,000 km	42.3 (47)	36.0 (40)
	>10,001 km	37.8 (42)	22.5 (25)

Table 3
Resilience, functional performance, self-reported, and naturalistic driving measures at Year 3 and Year 8 and change over that period (N = 111).

Measures	Year 3		Year 8		Change between Y3 & Y8	
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	Wilcoxon Z	p-value
Resilience	79.8 (10.4)	82.0 (72.0–87.0)	82.4 (9.9)	84.0 (77.0–89.0)	−3.3	<0.01
Functional performance[a]						
OARS	27.6 (0.7)	28.0 (27.0–28.0)	26.8 (1.2)	27.0 (26.0–28.0)	−6.7	<0.001
Trails B (sec)	96.9 (29.1)	96.0 (74.0–119.0)	100.2 (45.4)	94.0 (66.0–132.5)	−0.3	0.742
RPW (sec)	7.2 (1.3)	7.0 (6.0–8.0)	8.0 (1.7)	8.0 (7.0–9.0)	−4.9	<0.001
Visual Acuity LogMAR	0.04 (0.12)	0.0 (0.00–0.10)	0.12 (0.12)	0.10 (0.0–0.18)	−5.1	<0.001
Self-reported driving[b]						
DCS – D (Max = 100)	78.7 (13.8)	80.8 (71.1–86.5)	75.9 (14.5)	78.8 (67.3–86.5)	−2.5	<0.05
DCS – N (Max = 100)	71.4 (17.8)	71.9 (60.9–85.9)	66.1 (20.9)	68.8 (53.1–82.8)	−2.8	<0.01
PDA (Max = 45)	34.5 (6.2)	36.8 (30.0–39.0)	32.2 (6.9)	32.0 (27.0–38.0)	−3.8	<0.001
SDF (Max = 56)	33.9 (6.3)	34.0 (30.0–39.0)	30.4 (7.2)	32.0 (25.0–36.0)	−5.8	<0.001
SDA (Max = 19)	3.9 (3.7)	3.0 (1.0–6.0)	5.6 (4.0)	5.0 (2.0–7.0)	−5.8	<0.001
Naturalistic driving						
Total distance	9050.3 (5265.3)	8093.4 (5982.4–10675.3)	6329.5 (4755.4)	5385.2 (3257.6–7998.8)	−7.4	<0.001
Total trips	1385.7 (638.0)	1266.0 (972.0–1661.0)	1124.2 (576.7)	1051.0 (684.0–1457.0)	−5.8	<0.001
Trip distance	6.8 (3.6)	6.3 (4.4–8.0)	5.7 (3.1)	4.9 (3.6–6.9)	−5.7	<0.001
% ≤5 km	67.7 (11.5)	68.5 (61.5–74.9)	71.0 (13.3)	71.8 (62.7–80.6)	−3.8	<0.001
% >20 km	5.9 (5.5)	4.8 (1.8–7.8)	4.6 (5.4)	2.7 (1.1–6.2)	−4.6	<0.001
% Night	9.2 (7.1)	7.8 (4.1–12.4)	7.5 (7.5)	5.2 (2.5–10.6)	−4.3	<0.001
% Peak	18.2 (5.5)	17.8 (14.8–21.4)	16.9 (5.3)	16.1 (13.5–20.4)	−2.7	<0.01

Note. OARS = Older Americans Resources and Service Multidimensional Functional Assessment Questionnaire; RPW = Rapid Pace Walk; LogMAR = logarithm of the minimum angle of resolution; DCS-D = Driving Comfort Scale – Daytime; DCS-N = Driving Comfort Scale – Nighttime; PDA = Perceived Driving Abilities; SDF = Situational Driving Frequency; SDA = Situational Driving Avoidance.
[a]Data for n = 1 participant missing on each: Trails B, exceeded time limit to complete task; RPW, not recorded; Visual acuity, blindness in one eye.
[b]Data for n = 1 participant missing; did not complete self-reported driving questionnaires.

driving abilities, and situational driving frequency. The number of situations participants reported trying to avoid significantly increased. Analysis of the ND measures revealed statistically significant decreases in total distance driven, total trips, mean trip distance, and proportion of annual trips occurring during peak hour and at night, as well as proportion of trips greater than 20 km. There was a significant increase in the proportion of trips shorter than five kilometers.

Three-stage hierarchical regression analyses were conducted to determine whether resilience scores explained a statistically significant amount of variance in Year 8 for the variables of interest beyond that explained by age, sex, driving exposure, and functional abilities. All variables were treated as continuous except sex, where males served as the reference group. The initial step controlled for age, sex, and driving exposure, with the latter being operationalized using GPS-derived total kilometers driven. The OARS Multidimensional Functional Assessment Questionnaire

Table 4
Hierarchical regression analyses for self-reported driving comfort, abilities and restrictions.

Self-reported driving-related measures	DCS-D			DCS-N			PDA			SDF			SDA		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Step 1															
Age	−0.10	0.48	−0.02	0.00	0.68	0.00	−0.40	0.24	−0.16	−0.21	0.22	−0.08	0.03	0.13	0.02
Sex	−6.26	2.83	−0.21*	−13.02	3.99	−0.30**	−0.78	1.42	−0.05	−3.82	1.32	−0.25	2.37	0.75	0.28**
Km	0.00	0.00	0.23*	0.00	0.00	0.24*	0.00	0.00	0.15	0.00	0.00	0.36	0.00	0.00	−0.29**
Step 2															
Age	−0.07	0.51	−0.01	−0.08	0.71	−0.01	−0.29	0.25	−0.12	−0.18	0.24	−0.07	−0.01	0.13	−0.01
Sex	−6.19	2.89	−0.21*	−13.29	4.07	−0.30**	−0.42	1.45	−0.03	−3.73	1.34	−0.25**	2.26	0.76	0.27**
Km	0.00	0.00	0.24*	0.00	0.00	0.25*	0.00	0.00	0.13	0.00	0.00	0.36**	0.00	0.00	−0.28**
RPW	0.96	0.96	0.11	1.54	1.36	0.12	−0.16	0.48	−0.04	0.37	0.45	0.09	−0.26	0.25	−0.11
OARS	1.45	1.30	0.12	1.37	1.84	0.08	0.68	0.65	0.12	0.73	0.61	0.12	−0.62	0.34	−0.19
Step 3															
Age	−0.15	0.49	−0.03	−0.16	0.70	−0.02	−0.34	0.24	−0.13	−0.19	0.24	−0.07	0.01	0.13	0.00
Sex	−8.33	2.83	−0.28**	−15.52	4.09	−0.36**	−1.63	1.40	−0.11	−4.16	1.37	−0.27**	2.60	0.77	0.31**
Km	0.00	0.00	0.20*	0.00	0.00	0.23*	0.00	0.00	0.09	0.00	0.00	0.35**	0.00	0.00	−0.27**
RPW	1.18	0.92	0.14	1.78	1.33	0.14	−0.04	0.46	−0.01	0.42	0.45	0.10	−0.29	0.25	−0.12
OARS	1.18	1.25	0.10	1.10	1.80	0.06	0.53	0.62	0.09	0.68	0.60	0.11	−0.57	0.34	−0.17
RS-14	0.43	0.13	0.30**	0.45	0.19	0.21*	0.24	0.06	0.35**	0.09	0.06	0.12	−0.07	0.04	−0.17
	Adj. R^2	ΔR^2	*p*	Adj. R^2	ΔR^2	*p*	Adj. R^2	ΔR^2	*p*	Adj. R^2	ΔR^2	*p*	Adj. R^2	ΔR^2	*p*
Step 1	0.101		<0.01	0.159		<0.01	0.039		0.07	0.243		<0.01	0.188		<0.01
Step 2	0.097	0.013	<0.01	0.154	0.011	<0.01	0.039	0.017	0.11	0.239	0.011	<0.01	0.198	0.024	<0.01
Step 3	0.176	0.082	<0.01	0.191	0.042	<0.01	0.147	0.111	<0.01	0.246	0.013	<0.01	0.218	0.027	<0.01

Note. DCS-D = Driving Comfort Scale – Daytime; DCS-N = Driving Comfort Scale – Nighttime; PDA = Perceived Driving Abilities; SDF = Situational Driving Frequency; SDA = Situational Driving Avoidance; Km = kilometers; RPW = Rapid Pace Walk; OARS = Older Americans Resources and Service Multidimensional Functional Assessment Questionnaire; RS-14 = 14-item resilience scale.
p <.05, **p <.01.

Table 5
Hierarchical regression analyses for naturalistic driving variables.

	Naturalistic driving measures																	
	Total trips			Mean trip distance			% $\leq$ 5 km			% > 20 km			% Night			% Peak hour		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Step 1																		
Age	−13.63	13.34	−0.07	−0.03	0.09	−0.03	0.76	0.44	0.16	0.03	0.17	0.02	−0.51	0.26	−0.19	−0.01	0.19	−0.01
Sex	108.01	78.43	0.09	−1.05	0.54	−0.16	1.78	2.61	0.06	−2.00	1.02	−0.18	−3.75	1.52	−0.24*	0.09	1.09	0.01
Km	0.09	0.01	0.77**	0.00	0.00	0.49**	0.00	0.00	−0.30**	0.00	0.00	0.40**	0.00	0.00	−0.05	0.00	0.00	0.27**
Step 2																		
Age	−16.86	14.10	−0.08	0.01	0.10	0.01	0.49	0.46	0.10	0.08	0.18	0.04	−0.48	0.27	−0.18	0.02	0.20	0.01
Sex	96.94	80.34	0.08	−0.93	0.55	−0.14	0.89	2.64	0.03	−1.81	1.04	−0.16	−3.66	1.56	−0.23*	0.21	1.12	0.02
Km	0.09	0.01	0.78**	0.00	0.00	0.46**	0.00	0.00	−0.26**	0.00	0.00	0.38**	0.00	0.00	−0.06	0.00	0.00	0.26*
RPW	8.30	26.82	0.02	−0.16	0.18	−0.09	0.84	0.88	0.11	−0.05	0.35	−0.02	−0.07	0.52	−0.02	−0.19	0.37	−0.06
OARS	−16.96	36.27	−0.04	0.08	0.25	0.03	−1.14	1.19	−0.10	0.40	0.47	0.09	0.15	0.70	0.02	0.03	0.51	0.01
Step 3																		
Age	−18.30	13.90	−0.09	0.02	0.10	0.02	0.47	0.46	0.10	0.10	0.18	0.05	−0.51	0.27	−0.19	0.04	0.20	0.02
Sex	58.22	81.23	0.05	−0.63	0.55	−0.10	0.04	2.70	0.00	−1.44	1.06	−0.13	−4.51	1.57	−0.29**	0.65	1.14	0.06
Km	0.09	0.01	0.76**	0.00	0.00	0.49**	0.00	0.00	−0.28**	0.00	0.00	0.40**	0.00	0.00	−0.08	0.00	0.00	0.28*
RPW	12.32	26.46	0.04	−0.20	0.18	−0.11	0.93	0.88	0.12	−0.09	0.35	−0.03	0.02	0.51	0.01	−0.24	0.37	−0.08
OARS	−21.77	35.78	−0.05	0.12	0.24	0.05	−1.24	1.19	−0.11	0.44	0.47	0.10	0.04	0.69	0.01	0.09	0.50	0.02
RS-14	7.80	3.74	0.13*	−0.06	0.03	−0.20*	0.17	0.12	0.13	−0.07	0.05	−0.13	0.17	0.07	0.23*	−0.09	0.05	−0.16
	Adj. R^2	ΔR^2	*p*	Adj. R^2	ΔR^2	*p*	Adj. R^2	ΔR^2	*p*	Adj. R^2	ΔR^2	*p*	Adj. R^2	ΔR^2	*p*	Adj. R^2	ΔR^2	*p*
Step 1	0.574		<0.01	0.292		<0.01	0.124		<0.01	0.200		<0.01	0.062		<0.05	0.048		<0.05
Step 2	0.568	0.002	<0.01	0.288	0.010	<0.01	0.135	0.027	<0.01	0.193	0.008	<0.01	0.045	0.001	0.08	0.033	0.003	0.13
Step 3	0.582	0.017	<0.01	0.320	0.036	<0.01	0.142	0.015	<0.01	0.202	0.017	<0.01	0.085	0.047	<0.05	0.050	0.025	0.08

Note. Km = kilometers; RPW = Rapid Pace Walk; OARS = Older Americans Resources and Service Multidimensional Functional Assessment Questionnaire; RS-14 = 14-item resilience scale.
*p <.05, **p <.01.

and Rapid Pace Walk were added at Step 2 to assess changes in variance by cognitive and functional measures. Finally, resilience scores were included at Step 3. Hierarchical regression tables report the unstandardized regression coefficients, standard errors, and standardized regression coefficients for each independent variable, as well as the adjusted R-squared, R-squared change, and significance values for each step in the regression models.

Hierarchical regression analyses for self-reported driving comfort, abilities, and restrictions are reported in Table 4. The proportion of variance explained in the final models ranged from 14.7% to 24.6%. Age, RPW, and OARS were not significant at any stage of the models for the self-reported dependent variables. Sex, driving exposure, and resilience scores were statistically significant variables in the final model for driving comfort during the day and at night, explaining between 17.6% and 19.1% of the variance, respectively. The largest increase in variance, 11.1 percentage points, was observed when adding resilience scores into the model for perceived driving abilities. This was also the only dependent variable for which resilience remained the sole significant independent variable in the final model. Sex and driving exposure were significant variables in the models for situational driving frequency and avoidance.

Table 5 describes results for hierarchical regression models for summary driving characteristics and ND outcome variables suggestive of self-regulatory driving behavior (% $\leq$ 5 km, % > 20 km, % Night, % Peak hour). The proportion of variance explained was 58.2% and 32.0% for total trips and mean trip distance, respectively, with driving exposure remaining a significant variable throughout each regression step of every model. Age, sex, and functional measures were not statistically significantly associated with total trips or mean trip distance throughout any step of the models for either dependent variable. Adding resilience at Step 3 resulted in a statistically significant increase in the amount of variance explained for both total trips driven and mean trip distance. The proportion of variance explained by the regression of the independent variables on self-regulatory driving measures was lower than that found in the models for overall ND measures, ranging from 5.0% to 20.2%. Similar to earlier analyses, age and functional measures were not associated with any of the self-regulatory naturalistic outcome measures. Driving exposure was statistically significantly associated with the proportion of trips shorter than 5 km and farther than 20 km. Exposure was also a significant variable in the regression models for proportion of peak hour trips, but the final model was not significant. Both sex and resilience were significantly associated with an increase in variance explained for percent of trips driven during night time hours, but neither were significant in the models for the remaining outcome variables.

4. Discussion

This study examined a cohort of Ozcandrive older drivers at two time points to explore potential changes in functional performance, subjective and objective driving behavior, and resilience scores across a five-year period. To our knowledge this is the first study to investigate the relationship between resilience and aging driver behavior using ND data. Three hypotheses were assessed: (H1) Resilience scores will significantly increase over time; (H2) Resilience will continue to be a significant variable in explaining the variance in self-reported driving comfort, abilities and restrictions after five years; and (H3) Resilience will be a significant variable in explaining the variance of driving behavior as measured by GPS data.

The first hypothesis was supported; resilience scores significantly increased over time. This finding is consistent with previous findings that older adults tend to have higher resilience scores than younger adults (Wagnild, 2009); however, few studies have compared resilience between older and younger populations or between the oldest-old and younger old (Hayman et al., 2017). This study's focus on older adults helps to fill a significant gap in the literature as there has been scant research on psychological resilience in individuals of more advanced age (e.g., 80 years and older; Cosco et al., 2017).

The second hypothesis was partially supported. Resilience remained a significant variable in the models for driving comfort during the daytime and nighttime as well as for perceived driving abilities in Year 8, and was not a significant factor in explaining the variance in situational driving avoidance. However, unlike Year 3 where resilience was a significant variable in the model for situational driving frequency (see St. Louis et al., 2020), this was not the case for Year 8. Similar to Year 3, the relationship between resilience and perceived driving abilities remained the strongest in comparison to other self-reported measures. The addition of resilience in the final step of the model for perceived driving abilities more than doubled the amount of variance explained.

The third hypothesis was partially supported. Resilience was a significant variable in the models for overall number of trips, mean trip distance, and proportion of night trips. Higher resilience was associated with more overall trips, however, the trips were shorter in distance, suggesting that participants were perhaps staying closer to home and thus driving in more familiar areas. This finding may also indicate that participants with higher resilience were performing more trip chaining, a sequence of trips beginning from one location and returning to that location after none, one, or more intermediate stops (Golob & Hensher, 2007). Trip chaining can be considered another type of self-regulatory behavior to the extent that older adults may combine several trips into one outing to reduce the magnitude of driving challenges (Molnar et al., 2013). The association between resilience and proportion of night trips is counterintuitive to the idea that higher resilience is associated with self-regulatory behavior. It is difficult to explain this relationship without more insight into the reasons for driving at night.

As expected, participants' functional performance declined as the sample aged. However, age, OARS, and RPW were not statistically significant variables in the regression models. These findings are consistent with a previous Ozcandrive study that found changes in health between Years 1 and 5, but no association with ND outcome measures (Charlton et al., 2019). This may reflect the relative good health and functional status of the cohort at the point of recruitment and after eight years of study participation. While these decreases are indicative of age-related functional decline, the mean and median functional performance values for each measure did not exceed the standard thresholds for impairment and do not suggest increased crash risk as previously defined.

The current findings extend previous research that explored the role of psychosocial factors in relation to driving behavior and decision-making (Tuokko et al., 2013; St. Louis et al., 2020) and also provides robust, objective data over a five-year period lending support to findings of reduced driving based on self-report and short-term naturalistic driving studies (Baldock et al., 2006; Blanchard & Myers, 2010; Charlton et al., 2006; Coxon et al., 2015; Myers et al., 2011), as well as a longitudinal analysis of the full Ozcandrive cohort (Charlton et al., 2019). Continuing research in this area is certainly warranted to better understand the broad constellation of factors that play a role in driving outcomes and changes over time.

Some limitations should be noted. First, the Ozcandrive cohort is a convenience sample of older drivers who made a commitment to participate in a multi-year study, and there is likely a bias toward a healthier, more active sample. However, a study comparing the Canadian cohort to older drivers in Canada found participants were representative of the older driver population (Gagnon et al., 2016). Therefore, future research will compare the Ozcandrive cohort to a broader sample of older drivers across Australia on several self-reported measures including health, driving frequency, comfort, abilities, perceptions, restrictions, and psychological resilience.

Although it was important to control for total distance driven, it is possible the proportion of variance explained in the models was inflated by its inclusion. Despite this, while most final hierarchical regression models in our analyses were statistically significant, the R-squared values were modest. This suggests that there are other factors not examined in these analyses that play an important role in both self-reported and naturalistic driving behaviors of older adults.

Resilience was associated with several driving behaviors indicative of self-regulation, however there is a limitation in our ability to definitively interpret these results as self-regulatory behavior. The driving patterns may indeed result from self-regulation, or may simply reflect changes in preferences or lifestyle choices (Charlton et al., 2019; Molnar et al., 2013). Future research examining motivations underpinning driving decisions will provide the necessary context for better understanding driving patterns and to provide clearer interpretation of study results.

Another limitation concerns vehicle compatibility with the ICRD. Participants were required to have a vehicle that was model year 2002 or newer, which resulted in a sample of participants with vehicles that were a maximum of eight years old at study commencement. However, recent research has shown that Australian drivers aged 65 years and older were likely to be driving vehicles that were aged nine years or older (Koppel et al., 2018). Additionally, compatibility issues with newer model vehicles acquired by participants throughout the duration of the study did not allow for driving data to continue to be captured, and subsequent participation in the study was discontinued.

From the final regression models, it is clear that sex is an important factor in self-reported driving comfort, frequency, and restrictions; however due to the low number of females ($n = 38$) in the study, there was not enough power to conduct regression analyses by sex. Future research would benefit from examining sex differences in both driving behavior and resilience using a larger cohort of older adults with a larger proportion of females.

Finally, the resilience scale was administered at only two time points throughout the Ozcandrive study, limiting the years in which this variable could be examined. Although this research did address resilience over time, it would be beneficial to have at least three data points to compare across time to further understand its dynamic nature.

5. Conclusions

This study leveraged the longitudinal nature of the Ozcandrive study to provide the first insights into the role of resilience and naturalistic driving at two time points five years apart, and used independent but complementary data to better understand driving patterns and behavior of older adults. Patterns of reduced driving, captured by both subjective and objective measures, are suggestive of increased levels of self-regulation, potentially reflecting more appropriate driving decisions to reduce risk of crash-related injury or death. Self-regulatory strategies used by participants to compensate for declines in comfort or abilities have been described as: problem-focused coping strategies aimed at altering the situation via instrumental action, such as limiting driving to familiar areas; emotion-focused coping strategies that relate to gaining control over feelings about the situation, which can be seen as acceptance of the need for driving reduction or restriction and/or seeking support to maintain mobility; or a combination of both (Choi et al., 2012; Kostyniuk et al., 2000; Lazarus & Folkman, 1984). As resilience is contextual in nature, this study has demonstrated that resilience in the context of transportation is an important factor in understanding how older drivers navigate the process of aging as it relates to driving. Targeted interventions to increase resilience in the older driver population may lead to safer decision-making with regard to driving and allow stakeholders another option to engage with older adults in an effort to develop suitable measures to support older driver safety and mobility.

Funding

This research is part of a doctoral research program and was made possible through a graduate research scholarship from Monash University. Ozcandrive was funded by an Australian Research Council Linkage Grant (LP 100100078) to the Monash University in partnership with La Trobe University, Department of Transport, Victorian Government Department of Justice and Victoria Police, the Transport Accident Commission, New Zealand Transport Agency, Ottawa Hospital Research Institute and Eastern Health. The Candrive II study was funded by a Team Grant from the Canadian Institutes of Health Research (CIHR) entitled 'The CIHR Team in Driving in Older Persons (Candrive II) Research Program' (grant 90429).

Declarations of interest

None.

Acknowledgements

The authors thank Lorraine Atkinson, Ozcandrive Program Manager, for her role in managing and operationalizing the study for the Australian and New Zealand sites, as well as Jonathon Vivoda for his advice on data analysis. The authors gratefully acknowledge the invaluable contribution of the Ozcandrive team, including: Amy Allen, Louise Beasley, Russ Boag, Matthew Catchlove, Cara Dawson, Johan Davydov, Daniel Griffiths, Lei Gryffydd, Yik-Xiang Hue, Elizabeth Jacobs, Duncan Joiner, Phuong Hua, Flora Le, Peter Liu, Jason Manakis, Kevin Mascarenhas, Rachel Osborne, Emma Owen, Hiep Pham, Glenys Smith, Jarrod Verity, Zefi Vlahodimitrakou, and Chenghao Yu. The authors also thank the Ozcandrive cohort study older drivers for their dedication. Without their commitment, this research would not have been possible.

References

Austroads (2017). *Assessing fitness to drive for commercial and private vehicle drivers*. Publication No. AP-G56/17. Sydney: Austroads.

Baldock, M. R. J., Mathias, J. L., McLean, A. J., & Berndt, A. (2006). Self-regulation of driving and its relationship to driving ability among older adults. *Accident Analysis & Prevention, 38*(5), 1038–1045. https://doi.org/10.1016/j.aap.2006.04.016.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs: Prentice-Hall.

Blanchard, R. A., & Myers, A. M. (2010). Examination of driving comfort and self-regulatory practices in older adults using in-vehicle devices to assess natural driving patterns. *Accident Analysis and Prevention, 42*(4), 1213–1219. https://doi.org/10.1016/j.aap.2010.01.013.

Blanchard, R. A., Myers, A. M., & Porter, M. M. (2010). Correspondence between self-reported and objective measures of driving exposure and patterns in older drivers. *Accident Analysis & Prevention, 42*(2), 523–529. https://doi.org/10.1016/j.aap.2009.09.018.

Browne-Yung, K., Walker, R. B., & Luszcz, M. A. (2017). An examination of resilience and coping in the oldest old using life narrative method. *The Gerontologist, 57*(2), 282–291. https://doi.org/10.1093/geront/gnv137.

Carr, D. B., Schwartzberg, J. G., Manning, L., & Sempek, J. (2010). *Physician's guide to assessing and counseling older drivers* (2nd ed). Washington, DC: National Highway Traffic Safety Administration.

Charlton, J. L., Koppel, S., D'Elia, A., Hua, P., St. Louis, R., Darzins, P., ... Marshall, S. (2019). Changes in driving patterns of older Australians: Findings from the

Candrive/Ozcandrive cohort study. *Safety Science, 119*, 219–226. https://doi.org/10.1016/j.ssci.2018.11.008.
Charlton, J. L., Oxley, J., Fildes, B., Oxley, P., Newstead, S., Koppel, S., & O'Hare, M. (2006). Characteristics of older drivers who adopt self-regulatory driving behaviours. *Transportation Research Part F: Traffic Psychology and Behaviour, 9*(5), 363–373. https://doi.org/10.1016/j.trf.2006.06.006.
Choi, M., Adams, K. B., & Mezuk, B. (2012). Examining the aging process through the stress-coping framework: Application to driving cessation in later life. *Aging & Mental Health, 16*(1), 75–83. https://doi.org/10.1080/13607863.2011.583633.
Cosco, T. D., Kaushal, A., Hardy, R., Richards, M., Kuh, D., & Stafford, M. (2017). Operationalising resilience in longitudinal studies: A systematic review of methodological approaches. *Journal of Epidemiology & Community Health, 71*(1), 98–104. https://doi.org/10.1136/jech-2015-206980.
Coxon, K., Chevalier, A., Lo, S., Ivers, R., Brown, J., & Keay, L. (2015). Behind the wheel: Confidence and naturalistic measures of driving exposure among older drivers. *Transportation Research Record, 2516*(1), 35–43. https://doi.org/10.3141/2516-06.
Eby, D. W., Molnar, L. J., & St. Louis, R. M. (2018). *Perspectives and strategies for promoting safe transportation among older adults*. Elsevier. https://doi.org/10.1016/c2016-0-03275-8.
Fillenbaum, G. G. (1988). *Multidimensional functional assessment of older adults: The Duke older Americans resources and services procedures (1st ed.)*. Psychology Press. https://doi.org/10.4324/9780203771563.
Gagnon, S., Marshall, S., Kadulina, Y., Stinchcombe, A., Bédard, M., Gélinas, I., ... for the Candrive Research Team (2016). CIHR Candrive cohort comparison with Canadian household population holding valid driver's licenses. *Canadian Journal on Aging/La Revue canadienne du vieillissement, 35*(1), 99–109.
Golob, T. F., & Hensher, D. A. (2007). The trip chaining activity of Sydney residents: A cross-section assessment by age group with a focus on seniors. *Journal of Transport Geography, 15*(4), 298–312. https://doi.org/10.1016/j.jtrangeo.2006.09.005.
Grengs, J., Wang, X., & Kostyniuk, L. (2008). Using GPS data to understand driving behavior. *Journal of Urban Technology, 15*(2), 33–53. https://doi.org/10.1080/10630730802401942.
Hayman, K. J., Kerse, N., & Consedine, N. S. (2017). Resilience in context: The special case of advanced age. *Aging & Mental Health, 21*(6), 577–585. https://doi.org/10.1080/13607863.2016.1196336.
Hua, P., Charlton, J. L., Koppel, S., Griffiths, D., St Louis, R. M., Di Stefano, M., ... Marshall, S. (2018). Characteristics of low and high mileage drivers: Findings from the Ozcandrive older driver cohort study. *Journal of the Australasian College of Road Safety, 29*(4), 53. https://doi.org/10.1093/geroni/igy023.2077.
Huebner, K. D., Porter, M. M., & Marshall, S. C. (2006). Validation of an electronic device for measuring driving exposure. *Traffic Injury Prevention, 7*(1), 76–80. https://doi.org/10.1080/15389580500413067.
Jouk, A., Tuokko, H., Myers, A., Marshall, S., Man-Son-Hing, M., Porter, M., ... the Candrive II Research Team (2014). Psychosocial constructs and self-reported driving restriction in the Candrive II older adult baseline cohort. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 1–10. https://doi.org/10.1016/j.trf.2014.09.001.
Jouk, A., Sukhawathanakul, P., Tuokko, H., Myers, A., Naglie, G., Vrkljan, B., ... Bédard, M. (2016). Psychosocial constructs as possible moderators of self-reported driving restrictions. *Canadian Journal on Aging/La Revue canadienne du vieillissement, 35*(1), 32–43. https://doi.org/10.1017/s0714980816000027.
Koppel, S., Fitzharris, M., Newstead, S. V., D'Elia, A., Budd, L., & Charlton, J. L. (2018). *Development and analysis of a comprehensive data system to understand the occurrence, severity and risk factors of older road users*. Melbourne, Australia: Baseline Research Program (BRP) Committee.
Kostyniuk, L. P., Shope, J. T., & Molnar, L. J. (2000). Driving reduction/cessation among older drivers: Toward a behavioral framework. 9th International Association for Travel Behaviour Conference: Gold Coast, Queensland, Australia.
Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.
Leppin, A. L., Bora, P. R., Tilburt, J. C., Gionfriddo, M. R., Zeballos-Palacios, C., Dulohery, M. M., ... Montori, V. M. (2014). The efficacy of resiliency training programs: A systematic review and meta-analysis of randomized trials. *PloS one, 9*(10), e111420.
MacDonald, L. M., Myers, A. M., & Blanchard, R. A. (2008). Correspondence between older drivers' perceptions, abilities and driving behavior. *Topics in Geriatric Rehabilitation, 24*(3), 239–252. https://doi.org/10.1097/01.tgr.0000333756.75303.b9.
Marshall, S. C., Man-Son-Hing, M., Bedard, M., Charlton, J., Gagnon, S., Gelinas, I., ... Woolnough, A. (2013). Protocol for Candrive II/Ozcandrive, a multicentre prospective older driver cohort study. *Accident Analysis & Prevention, 61*, 245–252. https://doi.org/10.1016/j.aap.2013.02.009.
McGwin, G., Jr., & Brown, D. B. (1999). Characteristics of traffic crashes among young, middle-aged, and older drivers. *Accident Analysis & Prevention, 31*(3), 181–198. https://doi.org/10.1016/S0001-4575(98)00061-X.
Molnar, L. J., Eby, D. W., Charlton, J. L., Langford, J., Koppel, S., Marshall, S., & Man-Son-Hing, M. (2013). Driving avoidance by older adults: Is it always self-regulation? *Accident Analysis & Prevention, 57*, 96–104. https://doi.org/10.1016/j.aap.2013.07.004.
Molnar, L. J., Charlton, J. L., Eby, D. W., Langford, J., Koppel, S., Kolenic, G. E., & Marshall, S. (2014). Factors affecting self-regulatory driving practices among older adults. *Traffic Injury Prevention, 15*(3), 262–272. https://doi.org/10.1080/15389588.2013.808742.
Molnar, L. J., Eby, D. W., Zhang, L., Zanier, N., St. Louis, R. M., & Kostyniuk, L. P. (2015). *Self-regulation of driving by older adults: A synthesis of the literature and framework for future research*. Washington, DC: AAA Foundation for Traffic Safety.
Moses, J. A. (2004). Test review-Comprehensive Trail Making Test (CTMT). *Archives of Clinical Neuropsychology, 19*, 703–708. https://doi.org/10.1016/j.acn.2004.02.004.
Myers, A. M., Paradis, J. A., & Blanchard, R. A. (2008). Conceptualizing and measuring confidence in older drivers: Development of the Day and Night Driving Comfort Scales. *Archives of Physical Medicine and Rehabilitation, 89*(4), 630–640. https://doi.org/10.1016/j.apmr.2007.09.037.
Myers, A. M., Trang, A., & Crizzle, A. M. (2011). Naturalistic study of winter driving practices by older men and women: Examination of weather, road conditions, trip purposes, and comfort. *Canadian Journal on Aging/La Revue canadienne du vieillissement, 30*(4), 577–589. https://doi.org/10.1017/S0714980811000481.
Newbold, K. B., & Scott, D. M. (2017). Driving over the life course: The automobility of Canada's Millennial, generation X, baby boomer and greatest generations. *Travel Behaviour and Society, 6*, 57–63. https://doi.org/10.1016/j.tbs.2016.06.003.
Porter, M. M., Smith, G. A., Cull, A. W., Myers, A. M., Bédard, M., Gélinas, I., ... Vrkljan, B. H. (2015). Older driver estimates of driving exposure compared to in-vehicle data in the Candrive II study. *Traffic Injury Prevention, 16*(1), 24–27.
Sirén, A., & Meng, A. (2013). Older drivers' self-assessed driving skills, driving-related stress and self-regulation in traffic. *Transportation Research Part F: Traffic Psychology and Behaviour, 17*, 88–97. https://doi.org/10.1016/j.trf.2012.10.004.
Somenahalli, S., Hayashi, Y., Taylor, M., Akiyama, T., Adair, T., & Sawada, D. (2016). Accessible transportation and mobility issues of elderly—how does Australia compare with Japan? *Advances in Sustainable Urbanization, 1*(1), 31–43. https://doi.org/10.18063/JSUPP.2016.01.001.
Staplin, L., Gish, K. W., & Wagner, E. K. (2003). MaryPODS revisited: Updated crash analysis and implications for screening program implementation. *Journal of Safety Research, 34*(4), 389–397. https://doi.org/10.1016/j.jsr.2003.09.002.
St. Louis, R. M., Koppel, S., Molnar, L. J., Di Stefano, M., Darzins, P., Bédard, M., ... Charlton, J. L. (2020). The relationship between psychological resilience and older adults' self-reported driving comfort, abilities, and restrictions. *Journal of Transport & Health, 17*, 100864. https://doi.org/10.1016/j.jth.2020.100864.
Sullman, M. J., & Taylor, J. E. (2010). Social desirability and self-reported driving behaviours: Should we be worried? *Transportation Research Part F: Traffic Psychology and Behaviour, 13*(3), 215–221. https://doi.org/10.1016/j.trf.2010.04.004.
Tuokko, H., Myers, A., Jouk, A., Marshall, S., Man-Son-Hing, M., Porter, M. M., ... Naglie, G. (2013). Associations between age, gender, psychosocial and health characteristics in the Candrive II study cohort. *Accident Analysis & Prevention, 61*, 267–271. https://doi.org/10.1016/j.aap.2013.02.036.
Wagnild, G. (2009). *The Resilience Scale User's Guide for the US English version of the Resilience Scale and the 14-Item Resilience Scale (RS-14)*. Worden, MT: The Resilience Center.
Wagnild, G., & Young, H. M. (1990). Resilience among older women. *Image: The Journal of Nursing Scholarship, 22*(4), 252–255. https://doi.org/10.1111/j.1547-5069.1990.tb00224.x.
Windle, G. (2011). What is resilience? A review and concept analysis. *Reviews in Clinical Gerontology, 21*(2), 152–169. https://doi.org/10.1017/s0959259810000420.
Wong, I. Y., Smith, S. S., Sullivan, K. A., & Allan, A. C. (2016). Toward the multilevel older person's transportation and road safety model: A new perspective on the role of demographic, functional, and psychosocial factors. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 71*(1), 71–86. https://doi.org/10.1093/geronb/gbu099.

Renée M. St. Louis, PhD, is an Assistant Research Scientist in the Behavioral Sciences Group at the University of Michigan Transportation Research Institute. She is responsible for the management of a variety of research projects aimed at enhancing safe mobility throughout the lifespan, with numerous projects addressing transportation issues related to the aging driver population.

Sjaan Koppel, PhD, is Associate Professor at the Monash University Accident Research Centre. She is responsible for conducting a wide range of research projects broadly aimed at improving safe mobility and reducing transport-related injuries and deaths amongst vulnerable road users including older road users (e.g., drivers [passenger vehicle; heavy vehicle], occupants, cyclists, motorcyclists, mobility motor scooter users etc.), drivers with medical conditions and/or functional impairments, and child road users (e.g., vehicle occupants, pedestrians, cyclists etc.).

Lisa J. Molnar, PhD, is a Research Associate Professor at the University of Michigan Transportation Research Institute. Her research has focused primarily on traffic safety and driver behavior. Specific topics of interest include: older driver safety and mobility; the use of vehicle technology to improve driver safety; the travel behavior and safety of vulnerable populations, especially pedestrians and bicyclists;; behavioral effects and safety outcomes associated with advanced vehicle technologies, as well as traffic laws, policies, and programs; adolescent driving behavior; occupant protection; and the prevention of alcohol-impaired driving.

Marilyn Di Stefano, PhD, is a Senior Policy and Project Officer at Road Safety Victoria in Melbourne, Australia. Her road safety research/policy areas include human factors and human-machine interface, medical fitness to drive, disability and older/vulnerable road users.

Peteris Darzins, PhD, is Executive Clinical Director of Aged Medicine and Director of Geriatric Medicine at Eastern Health and Professor of Geriatric Medicine at Monash University.

Michelle M. Porter, PhD, is Director of the Centre on Aging at the University of Manitoba and oversees all of the activities of the Centre. In addition to being the Director, she is a professor in the Faculty of Kinesiology and Recreation Management. Her research is focused on aging, exercise physiology, neuromuscular physiology and mobility.

Michel Bédard, PhD, is a Professor in the Department of Health Sciences at Lakehead University, where he is also the Director of the Centre for Research on Safe Driving. He is also a Professor in the Human Sciences Division of the Northern Ontario School of Medicine and is the Scientific Director of the Centre for Applied Health Research at St. Joseph's Care Group. His main research focus is on aging, with a particular interest regarding automobile driving, family care giving, and mental illness.

Nadia Mullen, PhD, a Research Associate at the Centre for Research on Safe Driving. She completed her PhD (Psychology) at University of Otago, New Zealand. Her many research interests include simulator validity, increasing safe driving using applied behaviour analysis, examining the effect of drugs (e.g., cannabis) on driving performance, and examining the process of older adults' driving cessation.

Anita Myers, PhD, is Distinguished Professor Emerita at the University of Waterloo. Her research interests include health program evaluation, aging, health, and well-being. She has been involved in several projects aimed at addressing various issues focused on older adult safety and mobility. These include: evaluating and revising older driver licensing policy; projects about self-regulatory driving behavior; and developing and evaluating an intervention for drivers with dementia and their caregivers to facilitate the decision to stop driving and support the transition to non-driving.

Shawn Marshall, MD, is a Professor and physician specializing in Physical Medicine and Rehabilitation (Physiatrist). He is the Division Head of Physical Medicine and Rehabilitation at the University of Ottawa and The Ottawa Hospital where he manages both in-patients and out-patient clinics for patients with concussion to severe traumatic brain injury. Driving and disability, specifically older driver research, is one of his main interests in clinical research. His driving research interests have focused on conditional licensing as well as assessing medical fitness to drive.

Judith L. Charlton, PhD, is a Professor and former Director of the Monash University Accident Research Centre (MUARC). Her research on safe mobility includes projects focusing on individuals with vision impairment, Traumatic Brain Injury and Parkinson's Disease. These projects examine the role of vision, cognitive and physical functions in driving using state of the art driving simulators, instrumented vehicles and naturalistic driving methods.

Journal of Safety Research 82 (2022) 112–123

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Examining the influence that safety training format has on educators' perceptions of safer practices in makerspaces and integrated STEM labs

Tyler S. Love [a,*,1], Kenneth R. Roy [b], Melvin Gill [c], Mark Harrell [d]

[a] Associate Professor, Director of Career and Technology Education, Coordinator of Technology and Engineering Education, Department of the Built Environment, University of Maryland Eastern Shore, United States

[b] Director of Environmental Health & Safety, Glastonbury Public Schools, CT, Chief Safety Compliance Adviser, National Science Teaching Association, United States

[c] Technology and Engineering Education, Meade High School in Maryland, Career and Technology Education at the University of Maryland Eastern Shore, United States

[d] Kentucky Department of Education, College of Education and Human Development at the University of Louisville, United States

ARTICLE INFO

Article history:
Received 4 October 2021
Received in revised form 30 November 2021
Accepted 5 May 2022
Available online 16 May 2022

Keywords:
Technology and engineering education
Science education
STEAM
Career and Technical Education (CTE)
Libraries

ABSTRACT

Introduction: The rising popularity of makerspaces and integrated science, technology, engineering, and mathematics (STEM) education labs has increased the safety/health hazards and resulting potential risks that schools, libraries, community centers, and educators must be prepared to address. Previous studies have demonstrated that adequate safety training can enhance educators' safety perceptions and reduce accident rates. *Method:* Safety training was conducted in three different U.S. states for 48 educators working in K-12 STEM areas. Differences in the mode of delivery, length of the training, and types of hands-on activities instituted at each training site were examined in relation to the level of influence these factors had on educators' safety perceptions. A modified version of the Science Teaching Efficacy Belief Instrument (STEBI) was used, which had previously been adapted for similar safety studies and showed strong reliability measures. *Results:* The pre- and post-survey responses revealed that educators at the fully online and shortest training session did not experience significant changes in their safety perceptions. However, participants at the two face-to-face sites demonstrated significant gains in their safety perceptions. Most notably, the site that offered the longest training and integrated the most hands-on lab activities recorded the greatest gains. Additionally, correlational analyses corroborated that as the amount of hands-on activities and length of the trainings increased, there was a positive significant association with changes in educators' safety perceptions. *Conclusions:* This research helps bridge the gap between industry and K-12 STEM education research regarding better safety training practices. The findings from this study can help promote safer teaching and learning environments, while also reducing liability and the chance of a serious accident. *Practical Applications:* State departments, higher education institutions, teacher education programs, school districts, and others providing STEM safety training to K-12 educators should utilize this research to reexamine their safety training policies and practices.

1. Introduction

The growing shortage of highly-qualified graduates from science, technology, engineering, and mathematics (STEM) teacher preparation programs has resulted in an increase of alternatively licensed educators teaching in K-12 STEM education programs (Bowen, 2013; Dee & Goldhaber, 2017; Ernst & Williams, 2015; Love & Love, 2022; Volk, 2019). Additionally, an increasing number of teachers from other content areas are being tasked with teaching hazardous STEM related courses in makerspaces and integrated STEM laboratories (labs) (Ernst & Williams, 2015; Love, 2015; Love, 2022; Love & Love, 2022; Love & Maiseroulle, 2021; Reed & Ferguson, 2021; Volk, 2019). Hynes and Hynes (2018) described makerspaces as updated versions of school shops, "The wood shop of the past is now seeing new life in makerspaces that cut across various media (e.g., sewing, metalworking, woodworking, electronics) with state-of-the-art tools and resources" (p. 868). Integrated STEM labs reflect some characteristics similar to makerspaces, but are considered the full monty of interdisciplinary learning spaces. Roy and Love (2017) defined integrated STEM labs as "Collaborative spaces where the study of science, technology, engineering, and mathematics (in conjunction with other content areas) can be integrated through hands-on experiences in a pure laboratory or combined classroom laboratory setting" (pp. 6–7).

* Corresponding author at: University of Maryland Eastern Shore, 1425 Key Highway, Suite 101, Baltimore, MD 21230, United States.

E-mail addresses: tslove@umes.edu (T.S. Love), safesci@sbcglobal.net (K.R. Roy), mgill@aacps.org (M. Gill), Mark.Harrell@education.ky.gov (M. Harrell).

[1] ORCID: 0000-0002-1161-1443.

https://doi.org/10.1016/j.jsr.2022.05.003
0022-4375/

Table 1
Studies that Reported Minor and Major Accident Occurrences in K-12 STEM Courses.

Content (Grades)	Location	Minor Accidents (%)	Major Accidents (%)	Source
Science (all areas)	Texas	36	13	Fuller et al. (2001)
Chemistry (9–12)	Kentucky	74	19	Alyammahi (2015)
STEM and Makerspaces (K-12)	National	80	32	Love and Roy (2022)

Note. Accidents = Percentage of participants who reported at least one accident in this category. Minor accidents occurred within a one year span and required minor first-aid or a visit to a school nurse. Major accidents occurred within a five year span and resulted in a trip to the hospital and/or major medical attention from healthcare personnel.

Integrated STEM labs can pose increased hazards in comparison to makerspaces, including physical, chemical, biological, and other hazards (e.g., chemistry experiments, biological dissections, construction and materials processing). This requires an extensive amount of safety training and knowledge for an instructor to adequately manage.

The rising number of alternatively licensed and out of content area educators being tasked with teaching STEM in makerspaces and integrated STEM labs poses a major liability not only for the teacher, but also for the school district and administration (Love, 2013, 2014; Love & Love, 2022). One strategy to help address this concern is through meaningful and high-quality annual safety training (Love & Roy, 2017). In industry, adequate safety training has shown to reduce accident rates and save companies money (Burke et al., 2006, 2011; Cohen & Colligan, 1998; Colligan & Cohen, 2004). However, there is limited research on factors associated with accident rates and the influence of safety training on teaching and learning that occurs in makerspaces and STEM education labs (Love et al., 2021, 2022).

Teachers and administrators across the United States have identified safety as one of their top concerns within STEM education (Cannon et al., 2011; Rose et al., 2015). Additionally, concerns about the safety of design-based instructional practices were amplified after the release of the *Next Generation Science Standards* (NGSS) called for science educators to teach engineering practices (NGSS Lead States, 2013). As national STEM safety specialists have highlighted, teaching engineering practices can often require the use of hand and power tools that are commonly associated with technology and engineering (T&E) educators and T&E labs (Love, 2018; NSTA, 2020). Further contributing to the safety questions and concerns during this time was the rise in popularity of the maker movement and integrated STEM education (Love & Roy 2018; National Academies of Sciences, Engineering, and Medicine, 2019; Roy & Love, 2017). Schools and libraries soon saw an increase in potentially hazardous interdisciplinary design-based activities being conducted in non-traditional spaces, which were supervised by educators with limited to no safety training on managing hazards associated with the activities they were facilitating (Love, 2022; Love & Roy, 2017).

1.1. Increasing concerns for safety

The aforementioned changes led to an unprecedented rise in safety questions and concerns. Since many K-12 STEM laboratory accident lawsuits are settled before trial, there is limited data available regarding the occurrence of accidents in these spaces (NRC, 2006). Most states and school districts mandate the reporting of fatal or serious accidents that required care from medical personnel. However, the lack of systematic statewide and federal reporting for minor accidents involving injuries that did not require a trip to a doctor's office or the emergency room make it difficult to accurately track the occurrence of accidents in makerspaces and STEM courses (Stroud et al., 2007). Therefore, numerous studies have examined and reported data on accidents occurring in STEM classes. Gerlovich et al. (1998) reported that from 1990 to 1996, bodily injury claims sustained from science activities in Iowa's K-12 schools had increased by 49%, and the number of related lawsuits had increased 155%. Later studies specifically examined minor and major accident occurrences in various K-12 STEM education settings. When comparing these studies there is a noticeable rise in the percentage of STEM educators who reported minor and major accident occurrences over the past two decades (Table 1).

While the causes for the increases displayed in Table 1 are unknown, researchers have hypothesized that the lack of adequately prepared educators (including safety training) hired to teach in these STEM areas (Ferguson & Reed, 2019; Love, 2022; Love & Love, 2022; Love & Maiseroulle, 2021; Love & Roy, 2017; Reed & Ferguson, 2021; West et al., 2003) and the release of multiple K-12 instructional standards documents calling for more hands-on STEM learning experiences have increased the chance for accidents (Love et al., 2020a; NSTA, 2020; Stephenson et al., 2003; Stroud et al., 2007; West et al., 2003). A recent national K-12 STEM safety study corroborated these speculations, finding a strong positive correlation between the chance of an accident occurring and (a) increased hands-on instructional time in STEM courses, and (b) the comprehensiveness of STEM education safety training experiences completed by educators (Love et al., 2021). Moreover, when examining safety factors reported in K-12 STEM education safety studies from the past two decades there are some noticeable areas of improvement. Nonetheless, there remains a sustained lack of training, personal protective equipment (PPE), engineering controls, and safety practices overall (Table 2). Many of these factors have been significantly correlated with reducing the chance of an accident occurring (Love et al., 2021, 2022). Given the continued deficiency of essential safety items and practices in K-12 STEM education (Table 2) it is not surprising that the percentage of educators who reported having a minor or major accident occur in their classes has not decreased (Table 1). This reaffirms the importance of recent concerns raised about safety in K-12 STEM classes and makerspaces, and demonstrates there is still much work to be done to make these areas safer.

Given the foundational roots of T&E education and its focus on safer instruction involving hand and power tools for decades (Love, 2019), the discipline found itself in a unique position to share its valuable expertise regarding safer design-based instruction. While there has been a noticeable rise in safety questions, concerns, and litigation, the amount of research examining critical safety topics relative to K-12 makerspaces and integrated STEM labs remains limited. This is especially true regarding research pertaining to safety training, which has shown to provide positive outcomes in industry (Burke et al., 2006, 2011; Cohen & Colligan, 1998; Colligan & Cohen, 2004). Therefore, this study examined the influence that safety training has on educators who are facilitating design-based instruction in K-12 makerspaces, libraries, and integrated STEM education labs.

2. Review of literature

2.1. Safety training

2.1.1. What is required?

Safety training is required in most states under either the federal Occupational Safety and Health Administration (OSHA) stan-

Table 2
Safety Factors and Practices Reported in K-12 STEM Education Studies from 2001 to 2021.

Study	Content Area, Grade, Location	Safety Factor or Practice (%)							
		Instructor Received Safety Training within Past 5 Years	Require Signed Safety Acknowledgement Form*	Require Students to Complete Safety Tests*	Have Appropriate Eye Protection for All Students	Have Sufficient Number of GFI Outlets	Have Eye-wash in Lab	Have Fire Extinguisher in Lab	Have First-Aid Kit in Lab
Fuller et al. (2001)	Sci, K-12, TX	NR	NR	80	83	NR	83	79	66
Gerlovich et al. (2001)	Sci, 6-12, WI	37	NR	NR	80	38	82	91	65
Gerlovich and Parsa (2002)	Sci, K-12, Ntl.	40	NR	NR	81	41	67	NR	NR
Gerlovich et al. (2008)	Sci, 6-12, KY	45	77	59	86	52	84	91	NR
Alyammahi (2015)	Chem, 9-12, KY	45	85	72	75	71	89	96	82
Love and Roy (2020)	STEM, 6-12, KY	62	43	52	68	37	29	75	64
Love and Roy (2021)	STEM, 6-12, WI	36	83	95	98	57	76	98	69
Love and Roy (2022)	STEM, K-12, Ntl.	56	69	76	83	61	69	86	61

Note. * = Prior to working in the lab; Sci = Science; Chem = Chemistry; TX = Texas; WI = Wisconsin; Ntl. = National; KY = Kentucky; NR = Not reported.

dards, OSHA approved state developed Occupational Safety and Health plans, or state labor/health and safety departments that defer to specific federal OSHA standards by reference or have developed their own set of legal safety standards. Regardless, better professional safety practices, current legal safety standards, and legal precedent suggest that safety training should be a requirement for any teacher who will be working around potential safety and health hazards resulting in potential risks for themselves and their students. As explained by Love and Roy (2022), OSHA's Occupational Exposures to Hazardous Chemicals in Laboratories (also known as Laboratory Standard 29 CFR 1910.1450), Hazard Communication Standard 29 CFR 1910.1200, and additional specific OSHA legal standards (e.g., 1910 Subpart I – Personal Protective Equipment, 1910 Subpart Q – Welding, Cutting and Brazing, etc.) require employers (i.e., school districts) to provide safety training to employees (i.e., educators). These standards call for safety training upon initial hiring, when changes in work assignments present new hazards to an employee, or when there are changes in safety plans and workplace hazards (OSHA, 2020). There are also better professional safety practices established by professional educational associations (International Technology and Engineering Educators Association [ITEEA], National Science Teaching Association [NSTA], Association for Career and Technical Education [ACTE], National Science Education Leadership Association [NSELA], etc.) that recommend appropriate training of new employees and periodical updated safety training. In the event of a lawsuit, these recommended safety protocols could be presented as better professional safety practices that should have been followed for a safer teaching and learning environment (Love, 2013, 2014; Love et al., 2021). As a licensed professional, educators need to ensure they received and/or are receiving proper safety training from their district or another qualified source. If an employer (school district and administration) is not providing the appropriate training for their employees based on the criteria described above, then they could be found negligent, if not reckless.

2.1.2. Safety training efforts in STEM education

Love and Roy (2022) shed some light on the status of safety training in K-12 STEM education. Their research found that approximately 33% and 64% of educators teaching STEM courses in the United States did not receive any form of safety training during their undergraduate or coursework, respectively. Only 32% of these U.S. STEM teachers reported receiving safety training from their school district upon initial hiring, and only 56% received some form of safety training update within the past five years of the study. Further analyses revealed a similar lack in training experiences from a state level lens (Love et al., 2021; Love & Roy, 2021). Nationally, among teachers that completed training from a source outside of their district, most received training from a local source (32%), state STEM education association (18%), OSHA (18%), or a university (12%) (Love & Roy, 2022). The lack of safety training completed by STEM teachers presents potentially dangerous health and safety situations for both the educator and the students they supervise. It also can place the teacher and administration in potential legal jeopardy.

In addition to when training should be occurring and who should be providing the training, there is also literature on what content should be included in STEM education safety training. Researchers shared insight from their experiences providing safety workshops for the rising number of alternatively licensed and out of content area educators in Virginia being tasked with teaching K-12 T&E lab-based courses (Ferguson & Reed, 2019; Reed & Ferguson, 2021). They found that many alternatively licensed and out of content area educators in their state were not using tools and equipment to engage students in important design-based learning experiences because these educators did not know how to safely use the tools and equipment. Because of the limited T&E background of participants, Reed and Ferguson (2021) started by introducing teachers to the epistemology and curriculum of the field to emphasize the importance of safer hands-on learning. Their six hour face-to-face safety training focused primarily on production lab safety (woods, metals, ceramics, and composites) and included discussions, active participation, peer interactions, instructor led power and hand tool demonstrations, safety checklists to follow along during demonstrations, and a wooden toolbox construction project for educators to demonstrate safer practices. Information on lab management, organization, and maintenance activities for students was included, and participants were shown how to access a website that Reed and Ferguson (2021) developed with safety resources for the participants. The trainers concluded that while this experience appeared to improve educators' safety practices and willingness to safely use tools and machines in their instruction, the six-hour timeframe was not enough to cover the breadth and depth of safety information needed. They cautioned that safety training in this format was a compromise to the more extensive safety training educators receive when earning their degree from a higher education T&E teacher preparation program. Furthermore, they recommended these types of trainings be limited to 20 or less occupants per session and highlighted the importance of including a project that allows educators to build their safety confidence while receiving directly supervised assistance from the trainers. However, they emphasized the success of hands-on projects in such trainings is predicated by educators' understanding of the epistemological and curricular connections between T&E education and safety (Ferguson & Reed, 2019). In alignment with these recommendations, ITEEA, Flinn Scientific, CareerSafe®, and others created online safety training courses to help educators and students understand the background and importance of safer hands-on STEM teaching and learning.

Safety training for students in K-12 STEM classes traditionally covers machine nomenclature, safer operating practices, instructor demonstrations, safety tests, and directly supervised student demonstrations (DeLuca et al., 2014). Although instructors and students may be familiar with the operation of a machine, the demonstration portion is important due to differences in machine designs and features (e.g., some bandsaws might have a foot brake to help slow the blade down after shutting the power off). Studies have shown that safety training can have a significant impact on educators regardless of their years of teaching experience (Love, 2017a, 2017b, 2022). Findings from the national K-12 STEM safety study also revealed a lack of safety policies at the departmental and school district levels. Communication, consistency, and equitable enforcement of safety policies and practices are important within a school district and department (Gill et al., 2019). This includes critical practices such as requiring a district approved safety acknowledgement form and passing safety assessments be completed by students. As Gill et al. (2019) suggested, collaboration among STEM teachers in a department and school can ensure consistency in policies, safety instructional materials and assessments, addressing safety issues, and progressive disciplinary actions to develop a culture of safer habits and help reduce accidents.

3. Effectiveness of safety training

3.1. Industry connections

Numerous studies have documented the importance of safety training in both industry and K-12 STEM education contexts. Extensive literature reviews of safety training studies conducted in various industry sectors have found that safety training reduces risks from hazards in the workplace (Burke et al., 2006, 2011; Cohen & Colligan, 1998; Colligan & Cohen, 2004). Cohen and Colligan (1998) were unable to draw any strong conclusions about the influence of group size, length/frequency, and delivery method of safety training. Follow up studies investigated differences in the effectiveness of various deliver methods. Burke et al.'s (2006, 2011) extensive literature reviews found that highly engaging learner-centered methods of safety training (e.g., hands-on demonstrations) were considerably more effective than less engaging methods (e.g., lectures, slideshows) in regard to knowledge acquisition, safety performance, and reduction of negative outcomes. Burke et al. also found that although distance and electronic learning methods had monetary benefits and could reach more people, lack of participant engagement was a major issue. Advances in technology and instructors' use of technology after increased virtual teaching experiences during the COVID-19 pandemic could yield different findings. For example, Nykänen et al.'s (2020) found that in comparison to lecture-based training, virtual reality safety training had a stronger impact on construction workers' safety motivation, self-efficacy, safety-related outcome expectancies, and follow-up self-reported safety performance. While insight from industry can help inform safety training efforts in secondary education settings (Threeton & Evanoski, 2014; Threeton et al., 2021), additional research is needed to determine if these strategies would yield similar benefits among students or with educators' responsible for training and supervising students.

3.2. STEM education connections

Specific to STEM education, a few studies have investigated the influence of safety training on educators' safety perceptions, awareness, and practices. One of the most notable studies, which analyzed data from a national STEM education safety survey (Love & Roy, 2022), discovered that teachers who completed comprehensive safety training (a compilation of higher education coursework involving safety instruction, school district safety training upon initial hiring, and a safety training update session within the past five years) were 37% less likely to have a safety incident or accident occur in the STEM courses they taught (Love et al., 2021). Other studies have also documented positive benefits of safety training related to STEM education. Love (2017a, 2017b, 2022) showed that high-quality safety training can enhance teachers' self-efficacy pertaining to safer STEM instruction, and also increase their expectations for safer outcomes from students.

There have also been numerous studies conducted in response to the NGSS and concerns raised about science educators' preparation to use hand and power tools to teach engineering practices. One study compared multiple sites delivering professional development (PD) for elementary educators to learn about teaching engineering practices. The site led by T&E teacher educators had a more prominent focus on engineering safety practices and the safer use of tools and materials to develop prototypes in comparison to the sites led by science teacher educators (Grubbs et al., 2016). Furthermore, Love (2017a) found that safety training led by T&E safety specialists positively impacted elementary and middle school science educators' perceptions and awareness of safer hand and power tool use when teaching engineering practices. Educators who participated in a safety training led by T&E teacher educators reported significantly greater gains in their views toward the safer use of hand and power tools with students than participants who participated in a similar PD experience led by science teacher educators (Love, 2017a). Additional studies have revealed that high-quality safety training led by T&E safety specialists can have a positive and significant effect on the safety perceptions and awareness of female educators, as well as educators from various content areas working with engineering tools and materials in makerspaces and integrated STEM labs (Love, 2017b, 2022).

While these studies have investigated the impact of STEM education safety training according to the background of the trainers and demographics of the participants, they have not specifically examined the effectiveness of such safety training efforts according to certain factors that studies from industry have highlighted as significantly influential (e.g., length of the training, mode of delivery, and types of activities conducted during the training; Burke et al., 2006, 2011; Cohen & Colligan, 1998; Colligan & Cohen, 2004; Nykänen et al., 2020). These training factors have been shown to significantly reduce safety risks, and increase safety perceptions and self-reported safety performance. Therefore, as suggested by Love (2017a, 2017b, 2022), examining these specific factors within the context of STEM education safety training could improve the quality of training provided and potentially reduce safety risks, incidents, and accidents.

3.3. Format of STEM education safety trainings

Previous studies have demonstrated that quality STEM safety training can result in significantly positive gains in educators' safety perceptions and awareness. These studies have examined various forms of training ranging from hands-on engineering design PD lasting four-weeks and involving lab activities (Grubbs et al., 2016; Love, 2017a, 2017b), to half-day makerspace and STEM safety PD involving interactive non-lab groupwork activities (Love, 2022). Studies from industry have also found benefits among various types of safety training. However, as previously mentioned, studies from industry have specifically examined differences in safety trainings according to the mode of delivery and length of the training (Burke et al., 2006, 2011; Cohen & Colligan, 1998; Colligan & Cohen, 2004; Nykänen et al., 2020). To the authors' knowledge, there are no recent studies that have examined differences in the influence of STEM education related safety trainings based on the mode of delivery and training length.

3.4. Measuring safety perceptions

Observing STEM educators' safety practices can be time intensive, potentially dangerous, and difficult to witness authentic habits when educators know they are being observed (Love, 2017a, 2017b, 2022). Therefore, studies have found analyzing educators' safety perceptions (self-efficacy and expected outcomes) to be a reliable and more feasible measure of educators' safety practices (Love, 2017a, 2017b, 2022). As explained by Love (2017a, 2017b, 2022), self-efficacy and expected outcomes with regard to teaching and learning are based on Bandura's social learning theory. Bandura (1997) defined self-efficacy as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (p. 3). Specific to safety, Nykänen et al. (2019) defined safety-related self-efficacy as "the degree of confidence in one's ability to perform essential safety-related activities successfully" (p. 331). Moreover, safety studies in industry have found that employees who feel capable of performing particular tasks will perform them better (Bandura, 2001).

Based on Bandura's work and conclusions drawn from previous safety studies in education and industry, it is plausible to hypothesize that as one's STEM safety-related self-efficacy increases, the more confident they should be in performing essential safety duties (demonstrating proper hand and power tool usage, managing classroom safety, housekeeping practices, conducting periodic inspections, etc.). Increases in educators' self-efficacy provide important implications since self-efficacy has been linked to better teaching practices (Luft et al., 2011), higher expectations for students (Shidler, 2009), enhanced instructional quality (Holzberger et al., 2013), and greater student achievement (Çikrıkci, 2017). Furthermore, safety studies from industry have discovered that training can increase workers' self-efficacy and expected outcomes toward safety, their safety motivation, and their safety performance (Katz-Navon et al., 2007; Nykänen et al., 2018, 2019, 2020). For these reasons the authors determined that measuring changes in educators' safety perceptions offered a reliable method for analyzing the influence of STEM education related safety trainings.

3.5. Purpose and research questions

The literature indicates that safety training can positively influence educators' perceptions and awareness about safety in makerspaces and integrated STEM labs. Additionally, safety training has been correlated with reducing the chance and severity of accidents during STEM education activities (Love et al., 2021). However, the review of literature revealed a lack of research examining differences regarding the influence of K-12 STEM education safety trainings based on the training format (e.g., length of the training, mode of delivery, and instructional strategies). The overarching purpose of this study was to examine if the format of STEM education safety trainings yielded significant differences in educators' safety perceptions. The following research questions were developed to guide this study:

RQ1: Does safety training about makerspaces and integrated STEM labs have a significant difference on educators' safety self-efficacy?

RQ2: Does safety training about makerspaces and integrated STEM labs have a significant difference on educators' expected outcomes related to safety?

RQ3: Does the format of training about makerspace and integrated STEM lab safety have a significant difference on educators' safety self-efficacy?

RQ4: Does the format of training about makerspace and integrated STEM lab safety have a significant difference on educators' expected outcomes related to safety?

RQ5: To what extent is the format of makerspace and integrated STEM lab safety training associated with educators' gains in self-efficacy and expected outcomes related to safety?

4. Methodology

4.1. Safety training sites

The authors collaborated in developing the training materials used at all three sites to ensure there was consistency among the safety topics presented. However, each training site had some unique characteristics that served as the independent variables in this study. Those characteristics are described in detail below.

4.2. Site 1

This site facilitated a one-hour synchronous online training for 10 newly hired T&E teachers within a large county school system in a southern U.S. state. The training employed numerous interactive online pedagogical strategies such as responding to presenter questions and group discussions via the chat feature, verbal responses, and small group break out rooms. The training at this site did not include any instructor tool/machine demonstrations, participant tool/machine use, or project construction. It was comprised of presentations on the following safety topics as suggested by Roy and Love (2017): (1) examples of safety issues within makerspaces and STEM courses; (2) federal and state legal safety standards (e.g., OSHA); (3) liability and risk management strategies; (4)

brief overview of some biological, chemical, and physical hazards educators should look for in makerspaces and STEM labs; (5) better professional safety practices; (6) safer design considerations (e.g., engineering controls) for makerspaces and integrated STEM labs, and; (7) where to find additional safety resources.

4.3. Site 2

Site 2 provided training to two cohorts consisting of 28 educators total from a mid-Atlantic U.S. state. The cohort sizes of 16 and 12 teachers provided opportunities for more personalized instruction and met all safety criteria for the large conference style room. The training for each cohort was identical in regard to the content, format, and the safety specialist who delivered the instruction. This training consisted of one face-to-face session that lasted four-hours and included a number of interactive group activities, but did not involve any project construction, tool or machine use, or live machine demonstrations. It was comprised of presentations on the following safety topics as suggested by Roy and Love (2017): (1) introduction to makerspaces and STEM labs; (2) federal and state legal safety standards (e.g., OSHA); (3) liability and risk management strategies; (4) makerspace and STEM lab hazards (biological, chemical, physical); (5) better professional safety practices; (6) safer design considerations and exemplar makerspaces and integrated STEM labs; and (7) where to find additional safety resources. Interactive experiences such as the following were built into this training: (a) completing a mock report based on an accident scenario; (b) a think/pair/share case law activity; (c) conducting hazards analyses of makerspace and integrated STEM lab photos from participants' schools and libraries; and (d) designing a floor plan with appropriate engineering controls for a makerspace or integrated STEM lab within the physical constraints of their school or library.

4.4. Site 3

This site provided a two-day training for 10 secondary educators from a mid-West U.S. state. The training consisted of face-to-face lectures, discussions, hands-on tool and machine demonstrations, and an integrated STEM design challenge activity that utilized tools/machines/materials in a higher education fabrication lab. The lecture and discussion portions used the same presentation materials and covered safety topics identical to those described for Site 2. Participants also performed a mock lab safety inspection within the site's fabrication lab. Instructors provided safety demonstrations of various hand and power tools that are often found in makerspaces and integrated STEM labs. These items included a table saw, drill press, band saw, belt and disc sander, motorized miter saw, cordless drill, files, and other fabrication equipment. Under direct instructor supervision, participants modeled safer practices using these tools and machines to develop a solution to an integrated STEM design challenge. Their goal was to safely design, construct, and test a small catapult that launched a ping pong ball to land within targeted areas. Participants first applied some basic physics concepts to calculate the optimal launch settings for their catapult, then tested these theoretical projections by conducting multiple trials with their catapult and making any warranted adjustments.

4.5. Site safety trainers

The training instructors at each site had unique integrated STEM education safety expertise. Site 1 included an instructor who had 16 years of high school T&E teaching experience, 10 of which he served as T&E department chair. He also taught career and technology education facilities management and safety courses at an accredited southern university for nine years. This trainer served as a reviewer for journal articles and books from a national STEM educators association, published safety articles in national K-12 STEM education journals, and presented on safety topics at state and national K-12 STEM education conferences. The trainer from Sites 2 and 3 was an authorized OSHA outreach trainer for general industry. He had published numerous books and research articles on safety and served as the safety editor for publications produced by a national STEM educator association. Additionally, he served on a safety advisory board for a national science educators association. He had K-12 STEM teaching experience as well as higher education experience delivering science and T&E teacher preparation courses on facilities and lab safety. Additionally, Site 3 had a co-trainer who possessed 12 years of high school engineering technology education laboratory experience with extensive knowledge of both metalworking and woodworking equipment and processes. At the time of the training, he oversaw all public secondary education engineering technology programs in a mid-West state. Furthermore, he served as an instructor and recruiter for two higher education institutions, and previously worked for a non-profit STEM curriculum organization. Prior to each training, these three instructors met to plan the trainings, review training materials, and provide feedback. This ensured consistency in the content of the trainings, allowing this study to focus on differences in the format of each training.

4.6. Survey instrument

This study utilized a modified version of the Science Teaching Efficacy Belief Instrument, Form A (STEBI-A) (Riggs & Enochs, 1990). This instrument had previously been used in studies that examined changes in science and T&E educators' safety perceptions as a result of participating in safety trainings (Love, 2017a, 2017b, 2022). The original STEBI-A was developed to measure elementary educators' self-efficacy and expected outcomes toward teaching science. Riggs and Enochs found the instrument to have strong reliability and validity measures. It consists of 25 items measured on a five-point Likert scale (13 items examining teachers' self-efficacy, and 12 investigating their expected outcomes).

Love (2017a) modified the STEBI-A by changing all mentions of "science" to "safer use of engineering tools and materials in STEM." This modified instrument has been used to examine the influence that safety training has on science educators' perceptions toward safely teaching engineering practices involving the use of hand and power tools (Love, 2017a), differences in male and female educators' perceptions of safely using hand and power tools (2017b), and various content area educators' and librarians' perceptions about safety in makerspaces (2022). In each of the aforementioned studies the modified instrument demonstrated strong reliability measures (Love, 2017a, 2017b, 2022). For the reasons described above, Love's (2017a) modified STEBI-A was deemed the most viable instrument for this study. The instrument items can be found in Love (2017a).

4.7. Reliability and validity measures

The reliability of the survey items was tested using Crohnbach's alpha. The pre-survey (0.866) and post-survey (0.850) items demonstrated strong reliability measures. More specifically, the pre-survey (0.919) and post-survey (0.850) self-efficacy items, as well as the pre-survey (0.811) and post-survey (0.809) outcome expectancy items revealed strong reliability measures. Face validity of the instrument items was established in previous studies through a panel of national makerspace safety specialists, school district STEM supervisors, and STEM educators (Love, 2017a, 2022). The researchers conducting this study, whose unique safety expertise was described in the Site Safety Trainers section of this article, also reviewed the instrument items and found them to be appropriate for the purpose of this study. Hence, no changes were made to Love's (2017a) reliable instrument.

4.8. Administering the survey

At the start of the trainings, the pre-survey with some demographic questions was administered via online survey software. Participants were randomly assigned a number that they entered at the beginning of both surveys to protect their identity while also allowing their pre- and post-survey scores to be linked for later analyses. The randomly assigned participant number also served as encouragement for participants to provide honest and candid responses on what is often viewed as a sensitive and litigious topic. At the completion of the trainings, participants were provided time to complete the post-survey before they left the training site (Sites 2 & 3) or before signing out of the online meeting (Site 1).

4.9. Participants

The average age of the full sample was 42, with 46 (96%) identifying as White and 29 (60%) as females. Of note, Site 1 had predominantly male participants (80%), while Sites 2 and 3 were predominantly females. Approximately 22 (46%) of the participants from the overall sample had not completed any form of STEM education safety training within five years prior to participating in this study. More specifically, Site 1 had a higher percentage of participants who completed prior safety training in comparison to the other sites. Lastly, participants taught in a variety of content areas. Only 21 (44%) teachers across the three sites taught T&E where engineering hand and power tools are most commonly used. However, the majority of educators at Site 1 (60%) and Site 3 (70%) taught T&E courses. Site 2 was the most diverse in regard to participants' teaching assignments, featuring a mix of elementary educators (29%), librarians (18%), art educators (7%), and science educators (11%) who attended to learn more about safety because of their responsibilities associated with the makerspace or integrated STEM lab in their school (Table 3).

Table 3
Participant Demographics.

Characteristic	Site 1 *n* (%)	Site 2 *n* (%)	Site 3 *n* (%)
Gender			
Male	8 (80)	9 (32)	2 (20)
Female	2 (20)	19 (68)	8 (80)
Ethnicity			
White	8 (80)	28 (100)	10 (100)
Black	1 (10)	0 (0)	0 (0)
Two or more races	1 (10)	0 (0)	0 (0)
Average Age (years)	38	43	42
Completed STEM Ed Safety Training within past 5 years	5 (50)	8 (29)	13 (27)
Area Taught			
Elem. K-5	0 (0)	8 (29)	0 (0)
7-12 T&E	6 (60)	8 (29)	7 (70)
7-12 Science	1 (10)	3 (11)	1 (10)
7-12 Art	0 (0)	2 (7)	0 (0)
7-12 English	1 (10)	0 (0)	1 (10)
Librarian	0 (0)	5 (18)	1 (10)
Substitute or NC	2 (20)	2 (7)	0 (0)

Note. Area taught = Area taught at the time of the training; Elem. K-5 = elementary grades K-5; NC = Long term sub or no teaching certification.

5. Findings

5.1. Influence of safety training on safety perceptions

5.1.1. Self-efficacy (RQ1)

The first set of analyses examined each training site individually to determine if the training significantly influenced participants' self-efficacy about safety. Wilcoxon matched pairs tests were determined to be best suited for analyzing two related samples (pre- and post-survey items) with ordinal data from a nonparametric sample (Sheskin, 2011). The results revealed that the trainings at Sites 2 and 3 significantly increased participants' self-efficacy in regard to safety. The training at Site 1 did not have a significant influence on participants' safety self-efficacy (Table 4).

5.1.2. Expected outcomes (RQ2)

Similar to the first set of analyses, Wilcoxon matched pairs tests were conducted to examine changes in participants' expected outcomes for safety based on their pre- to post-survey responses. These analyses were again conducted separately for each training site. The results indicated that the workshops at Sites 2 and 3 significantly increased participants' expected outcomes for safety. The workshop at Site 1 did not have a significant influence on participants' expected outcomes for safety (Table 5).

5.1.3. Influence of safety training format

After separately analyzing the differences between pre- and post-survey scores at each training site, the researchers wanted to compare the sites to examine if the training formats had a significant influence on teachers' self-efficacy and expected outcomes related to safety. To first examine if there was a significant difference in the self-efficacy gains and expected outcome gains between training sites, Kruskal-Wallis tests were conducted. Kruskal-Wallis tests were selected due to the nonparametric and ordinal characteristics of the independent training sites (Sheskin, 2011). These analyses revealed there were statistically significant differences between sites in regard to safety self-efficacy and expected outcome gains (Table 6). Given these results, further

Table 4
Wilcoxon Matched Pairs Tests for Differences in Self-Efficacy Gains.

PD	*n*	Median	IQR	*Z*	*P*-value
Site 1					
Pre	10	49.5	16	−0.351	0.726
Post		47.5	21.25		
Site 2					
Pre	28	47.5	15.5	−4.109	<0.001*
Post		51	7		
Site 3					
Pre	10	35	10.5	−2.701	0.007*
Post		50.5	4.25		

Note. * = statistical significance at the 0.05 level.

Table 5
Wilcoxon Matched Pairs Tests for Differences in Expected Outcome Gains.

PD	*n*	Median	IQR	*Z*	*P*-value
Site 1					
Pre	10	38.5	7.75	−0.813	0.416
Post	10	41.5	6.25		
Site 2					
Pre	28	42	8.75	−3.387	<0.001*
Post	28	44.5	8.75		
Site 3					
Pre	10	40.50	10.75	−2.655	0.008*
Post	10	48.50	11.25		

Note. * = statistical significance at the 0.05 level.
PD = Professional development site.

Table 6
Kruskal-Wallis Tests for Differences between Trainings.

PD	*n*	df	Median	Mean Rank	Chi-Square	*P*-value
Self-Efficacy						
Site 1	10	2	0.5	12.60	13.975	< 0.001*
Site 2	28	2	4.5	24.66		
Site 3	10	2	14.5	35.95		
Expected Outcomes						
Site 1	10	2	0	14.75	7.897	0.019*
Site 2	28	2	2	25.30		
Site 3	10	2	6	32.00		

Note. * = statistical significance at the 0.05 level.
PD = Professional development site.

Table 7
Mann–Whitney U Tests for Differences in Self-Efficacy Gains.

PD	*n*	Median	Mean Rank	*U*	*Z*	*P*-value	*r*
Site 1	10	0.5	11.70	62.000	−2.594	0.009*	0.421
Site 2	28	4.5	22.29				
Site 1	10	0.5	6.40	9.000	−3.102	0.002*	0.694
Site 3	10	14.5	14.60				
Site 2	28	4.5	16.88	66.500	−2.443	0.015*	0.396
Site 3	10	14.5	26.85				

Note. * = statistical significance at the 0.05 level; *r* = effect size.
PD = Professional development site.

Table 8
Mann–Whitney U Tests for Differences in Expected Outcome Gains.

PD	*n*	Median	Mean Rank	*U*	*Z*	*P*-value	*R*
Site 1	10	0	13.00	75.000	−2.173	0.030*	0.353
Site 2	28	2	21.82				
Site 1	10	0	7.25	17.500	−2.480	0.013*	0.555
Site 3	10	6	13.75				
Site 2	28	2	17.98	97.500	−1.416	0.157	0.230
Site 3	10	6	23.75				

Note. * = statistical significance at the 0.05 level; *r* = effect size.
PD = Professional development site.

analyses were warranted to examine the differences between specific sites.

5.1.4. Influence on teachers' safety Self-Efficacy (RQ3)

Mann–Whitney U tests were then conducted to determine if there were significant differences in teachers' safety self-efficacy gains as a result of the training format. Mann–Whitney U tests were deemed suitable to test for significant differences among two samples with ordinal data from a nonparametric sample. This type of analysis tests for the mean difference in rank of responses between two independent groups with equal or unequal sample sizes. Additionally, the effect size of each Mann–Whitney U test was calculated using the formula provided by Pallant (2020). The Mann–Whitney U tests revealed significant differences between all trainings, with the analysis of Site 1 and Site 3 demonstrating the largest effect size (0.694), and the analysis of Site 2 and Site

3 having the lowest effect size (0.396). This indicates that trainings of greater length and with more hands-on safety experiences had a greater influence on educators' safety self-efficacy (Table 7).

5.1.5. Influence on teachers' expected outcomes (RQ4)

Mann–Whitney U tests were again conducted to determine if there were significant differences in educators' expected outcomes for safety as a result of the training format. There were significant differences when comparing Site 1 to Sites 2 and 3, with the analysis of Site 1 and Site 3 demonstrating the largest effect size (0.555). There was not a significant difference between the trainings at Sites 2 and 3, signifying that the longer and more hands-on oriented the trainings were, the greater impact they had on educators' expected outcomes for safety (Table 8).

5.1.6. Correlations between training format and safety perceptions (RQ5)

The Mann–Whitney U tests indicated there were significant differences in participants' safety self-efficacy and expected outcomes for safety when comparing the three training sites. However, the Mann–Whitney U tests do not measure association. Therefore, Spearman's rho correlation tests were conducted to examine the strength and significance of the relationship between gains in safety perceptions and the safety training format. To examine this relationship Spearman's rho tests were deemed the most appropriate measure for a nonparametric population, and also because the training format (amount of hands-on experiences and length of training) and safety perceptions were ordinal variables that could be ranked (Sheskin, 2011). Each training was assigned a rank increasing in regard to the length of the training provided and the extent of the hands-on activities involved (e.g., Site 1 received a rating of 1, Site 2 received a rating of 2, and Site 3 received a rating of 3). These analyses showed a moderate positive correlation (0.545) and significant (<0.001) association between the training format and gains in safety self-efficacy. This means that as the involvement of hands-on experiences and length of training increased, participants' self-efficacy about safety also increased. Furthermore, there was a low positive correlation (0.404) and significant (0.004) association between the training format and gains in expected outcomes toward safety. This means that as the involvement of hands-on experiences and length of training increased, participants' expectations for safer outcomes also increased (Table 9).

6. Discussion and conclusions

There are a few limitations that must be considered within the context of this study. The data were voluntarily self-reported by teachers from three U.S. states and the results may not be generalizable to every school district or state. Additionally, the length of training and number of hands-on experiences were not isolated as separate independent variables in the analyses. Therefore, no conclusions can be drawn about the influence of each factor individually. Despite these limitations, this study provided some valuable findings to enhance the safety of integrated STEM teaching and learning in K-12 makerspaces and labs.

Table 9
Spearman's rho Correlation Table of Training Format and Gains in Safety Perceptions.

Measure	r_s	*p*-value	*N*
Self-Efficacy	0.545	<0.001*	48
Expected Outcomes	0.404	0.004*	48

Note. * = statistical significance at the 0.05 level.

The Wilcoxon matched pairs tests revealed that participants' at Sites 2 and 3 had positive and significant gains in their safety self-efficacy and expected outcomes, but this was not true for Site 1 (Tables 4 and 5). According to the literature, this could potentially be attributed to a number of factors. Site 1 was the only training session that utilized an online delivery method, and the entire training for Site 1 was delivered synchronously online. Additionally, Site 1 provided the shortest training length (one hour) of all three sites. These findings are consistent with those from previous STEM education safety training studies. Prior STEM education safety training studies did not investigate the influence of an online training format, but they did find positive and significant gains in educators' safety self-efficacy and expected outcomes as a result of attending face-to-face trainings (Love, 2017a, 2017b, 2022). The findings from this study are also in alignment with the extensive literature review studies from industry, which found longer and in person trainings to be more effective than short or online safety trainings (Burke et al., 2006, 2011; Cohen & Colligan, 1998; Colligan & Cohen, 2004).

Another factor that could have had an effect on the gains at Site 1 is the percentage of participants who recently completed some form of safety training. Love and Roy (2022) found that on average, only 56% of T&E teachers across the United States had received some form of safety training within a five year span despite being correlated with significantly reducing accidents (Love et al., 2021). Site 1 yielded similar results (50%), whereas Site 2 (29%) and Site 3 (27%) were well below the national average (Table 3). The Wilcoxon matched pairs tests revealed that Site 1 had a higher pre-survey median score for safety self-efficacy in comparison to the other two sites (Table 4). This indicates that participation in other safety training experiences may have had an impact on the safety self-efficacy of Site 1 participants prior to attending the training in this study. Interestingly, participants at Site 1 were the only site to report a decrease in their safety self-efficacy as a result of the training. Their expectations or knowledge from other trainings may have had an influence.

One other interesting finding from RQ1 is that participants at Site 1 had the lowest pre-survey median score for expected outcomes of safety among all three sites (Table 5). The Wilcoxon matched pairs tests analyzed differences in median scores, but when examining the mean scores from each site, they revealed some interesting findings. In regard to expected outcomes, the mean pre-survey scores were 40.9 (Site 1), 42.11 (Site 2), and 41.4 (Site 3). The mean post-survey scores were 41.8 (Site 1), 45.11 (Site 2), and 47.2 (Site 3). This paints a clearer picture of the differences from pre- to post-survey scores at each site. When examining the means, which were not utilized for the Wilcoxon matched pairs tests, it reveals that Site 1 had the lowest change in expected outcomes and Site 3 exhibited the greatest change. The larger percentage of participants at Site 1 who had recently completed another safety training may have had an influence on gains in their safety perceptions. The literature would also suggest that the online format and training length had an impact. Since the analyses in this study did not isolate the independent variables of training length, mode of delivery, or prior safety training experiences, it is inconclusive if one of these factors had a greater influence than the others. Another interesting finding from RQ1 was related to the significant gains at Sites 2 and 3, which had a higher percentage of female participants than Site 1. These findings bear resemblance to Love's (2017b) research, which found that training on the safer use of engineering tools and materials for STEM instruction yielded greater safety self-efficacy and expected outcome gains from female educators in comparison to males.

The Mann–Whitney U analyses in RQ3 and RQ4 revealed significant increases in safety self-efficacy and expected outcomes with a large effect size between Site 1 (one-hour, fully online) and Site 3 (two days, face-to-face with demonstrations and lab activities). Although the self-efficacy gains between Site 2 (half-day, face-to-face with non-lab activity group discussions) and Site 3 were also statistically significant, the effect size was the smallest among the Mann–Whitney U analyses examining self-efficacy gains (Table 7). Moreover, there were insignificant expected outcome gains and the lowest effect size between Sites 2 and 3 (Table 8). From these findings one could conclude trainings that last longer than a half-day, and include more face-to-face hands-on lab activities, would be expected to have a greater influence on educators' STEM safety perceptions. The Spearman's rho analyses corroborated this conclusion, indicating that there were positive and significant correlations between participants' safety perceptions and the format of STEM safety trainings (length of the training, mode of delivery, and amount of hands-on lab activities included).

7. Conclusions and recommendations

The findings from this study are consistent with the literature from industry regarding factors that have been found to influence safety training. This study, along with safety research from industry (Burke et al., 2006, 2011; Cohen & Colligan, 1998; Colligan & Cohen, 2004) and STEM education (Ferguson & Reed, 2019; Love, 2022), reaffirms the notion that length of training, mode of delivery, and instructional strategies can have a significant impact on participants' perceptions related to the teaching and demonstration of safety concepts, supervision and management of proper safety practices, and safety expectations from workers or students. In light of these findings caution should be exercised when interpreting the results. Although there was consistency in the safety topics presented and discussed at each training site, the variables of training length, mode of delivery, and amount of hands-on lab activities were not isolated in the statistical analyses. Based on the findings and limitations of this study, recommendations for practitioners and future research follow.

8. Recommendations for practitioners (practical applications)

It is clear from the literature review and findings presented in this study that the format of safety trainings can have a significant influence on safety outcomes. State departments, institutions of higher education, teacher education programs, school districts, and others who are delivering safety trainings should consider the results of this study. In accordance with federal and state occupational safety and health standards, state education departments should strongly recommend safety training before alternatively certified and out of content area educators are allowed to use hazardous items in makerspaces and STEM labs. To yield greater safety benefits, it is recommended that the training be long enough to adequately cover all pertinent safety information in detail. As previous studies cautioned, although safety training is critical and required for educators working in makerspaces and STEM labs, in-service training should not be viewed as a replacement for valuable and detailed safety experiences received from higher education STEM teacher preparation programs (Ferguson & Reed, 2019; Love et al., 2021).

At a minimum, the safety topics presented at Site 3 in addition to all state and federal occupational safety and health topics applicable to hazards that educators will be exposed to, should be covered at trainings for STEM educators. Based on feedback from the site participants, training should also include an overview of methods to reduce liability (Love, 2013, 2014), instructional strategies and facility accommodations to assist students with disabilities in STEM courses (Love et al., 2020b), and approaches to work with school districts in addressing overcrowded makerspaces and integrated STEM labs (Love & Roy, 2022; West, 2016). At one site, a teacher contacted the trainer requesting additional information about safety zones and non-skid strips. Participant feedback like this is valuable to assist trainers in improving future trainings.

It is also strongly recommended that such trainings allow ample time for trainers to demonstrate the safer use of tools and equipment found in makerspaces and integrated STEM labs. Opportunities should be provided to assist educators with safely constructing a solution to an authentic integrated STEM design challenge. Additionally, educators should be required to demonstrate safer use of important tools/equipment in order to receive a certificate of completion or state approved continuing education credit. State education departments should partner with their state's occupational safety and health departments, higher education institutions, and local industries to provide training that improves safety habits in schools and the workforce.

Lastly, safety training providers should refer to OSHA resources, which provide a number of pertinent training recommendations and best practices for practitioners. Specifically, OSHA notes that accurate, credible, clear, and practical are four characteristics of effective training programs. OSHA (2021) describes those characteristics as follows:

1. *Accurate:* Training materials should be prepared by qualified individuals, updated as needed, and facilitated by appropriately qualified and experienced individuals employing appropriate training techniques and methods.
2. *Credible:* Training facilitators should have a general safety and health background or be a subject matter expert in a health or safety related field.
3. *Clear:* Training programs must not only be accurate and believable, but they must also be clear and understandable to the participant.
4. *Practical:* Training programs should present information, ideas, and skills that participants see as directly useful in their working lives. Successful transfer of learning occurs when the participant can see how information presented in a training session can be applied in the workplace. (pp. 2–3)

Safety training providers should ensure that their training reflects these characteristics.

9. Recommendations for future research

Given the similarities between findings from this study and industry safety studies, future research should investigate if factors found to be influential in industry provide similar benefits for safety training in K-12 STEM education contexts. For example, Cohen and Colligan (1998) found that feedback and positive reinforcement significantly influenced workers' safety behaviors. Parallels can be drawn between Cohen and Colligan's findings and B. F. Skinner's theory of operant conditioning, which is commonly taught in STEM teacher preparation programs. The application of this theory was noticeable at Site 3, which reported the greatest gains in safety perceptions. These gains could have potentially been attributed to the opportunities that participants at Site 3 had to demonstrate and practice safer habits while receiving feedback and positive reinforcement. Further research examining the influence of feedback and positive reinforcement during STEM education safety trainings is warranted to inform future instructional methods implemented at trainings. In addition to examining the influence of positive reinforcement, future studies should be designed to control for possible confounding variables such as recency of prior safety training, the recency of undergraduate or graduate coursework involving safety instruction, formal and informal STEM related teaching experiences, and other variables that could potentially influence one's safety perceptions and performance. Larger sample sizes in future studies would provide the potential to use parametric analyses, which could examine and control for the influence of possible cofounding variables.

Future studies should also investigate the relationship between safety perceptions and safety performance within the context of K-12 STEM teaching and learning. While Grossman and Salas (2011) found "self-efficacy has consistently shown positive relationships with the transfer of training," they cautioned that "high self-efficacy could cause individuals to feel they are adequately prepared for a challenge, and could thus reduce their motivation to prepare or put forth sufficient effort" (p. 109). Katz-Navon et al. (2007) expressed similar concerns that one could report a high general safety self-efficacy (e.g., an educator may believe that they can safely help their students) yet at the same time demonstrate low safety performance. Moreover, extensive reviews of the literature have found that only 10–20% of training content is transferred to the workplace (Grossman & Salas, 2011; Tonhäuser & Büker, 2016). One study found that 62% of training content is applied to the workplace immediately following a training; however, this fell to 44% after six months and 34% a year after the training (Saks & Belcourt, 2006). Hence, it is critical that future studies investigate the relationship between perceptions resulting from safety training and participants' safety performance following the training, especially within K-12 instructional settings that often involve unique safety factors (classroom management, percentage of students with disabilities, etc.; Love & Roy, 2022; Threeton & Evanoski, 2014). To evaluate the influence that safety trainings have on safety behaviors in the workplace, Saks and Burke (2012) suggested utilizing The Kirkpatrick Model (we suggest utilizing the latest version by Kirkpatrick & Kirkpatrick, 2016). While the study presented in this article examined educators' safety perceptions, it is strongly recommended that future studies examine changes in educators' safety knowledge (Level 2: Learning), application of safety training concepts in educators' daily instructional and/or administrative duties (Level 3: Behavior), and the longitudinal safety outcomes resulting from trainings (Level 4: Results) to gain deeper insight about the effectiveness and improvement of makerspace and STEM education safety trainings. It is imperative that future research evaluate criterion related to Levels 3 and 4 as evaluations of behavior and results have specifically been found to increase employee accountability and transfer of training to the workplace (Saks & Burke, 2012). Lastly, new and emerging instructional technologies should continue to be examined in regard to their effect on safety training and transfer of training. For example, Nykänen et al. (2020) recently found that safety training delivered via virtual reality (VR) increased worker safety. Future research should explore the applications of VR as a safety training tool and the influence it has on K-12 STEM educators' and students' safety perceptions, habits, and performance.

10. Declarations of interest

None.

Funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Alyammahi, A. R. (2015). The current status of safety in high school chemical laboratories in Kentucky (Publication No. 331). [Master's thesis, Eastern Kentucky University]. Encompass. https://encompass.eku.edu/etd/331.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman and Company: W.H.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*(1), 1–26.

Bowen, B. (2013). Measuring teacher effectiveness when comparing alternatively and traditionally licensed high school technology education teachers in North Carolina. *Journal of Technology Education, 25*(1), 82–100.

Burke, M. J., Salvador, R. O., Smith-Crowe, K., Chan-Serafin, S., Smith, A., & Sonesh, S. (2011). The dread factor: How hazards and safety training influence learning and performance. *Journal of Applied Psychology, 96*(1), 46–70.

Burke, M. J., Sarpy, S. A., Smith-Crowe, K., Chan-Serafin, S., Salvador, R. O., & Islam, G. (2006). Relative effectiveness of worker safety and health training methods. *American Journal of Public Health, 96*(2), 315–324.

Cannon, J. G., Kitchel, A., Duncan, D. W., & Arnett, S. E. (2011). Professional development needs of Idaho technology teachers: Teaching and learning. *Journal of Career and Technical Education, 26*(1), 32–47.

Çikrıkci, Ö. (2017). The effect of self-efficacy on student achievement. In E. Karadag (Ed.), *The factors effecting student achievement: Meta-analysis of empirical studies* (pp. 95–116). Springer. https://doi.org/10.1007/978-3-319-56083-0_6.

Cohen, A., & Colligan, M. J. (1998). Assessing occupational safety and health training (NIOSH Publication No. 98-145). National Institute for Occupational Safety and Health.

Colligan, M. J., & Cohen, A. (2004). The role of training in promoting workplace safety and health. In J. Barling & M. R. Frone (Eds.), *The psychology of workplace safety* (pp. 223–248). American Psychological Association. https://doi.org/10.1037/10662-011.

Dee, T. S., & Goldhaber, D. (2017). *The Hamilton Project: Understanding and addressing teacher shortages in the United States*. Brookings Institution.

DeLuca, V. W., Haynie, W. J., Love, T. S., & Roy, K. R. (2014). *Designing safer learning environments for integrative STEM education*. International Technology and Engineering Educators Association.

Ernst, J. V., & Williams, T. O. (2015). The "Who, what, and how conversation": Characteristics and responsibilities of current in-service technology and engineering educators. *The Journal of Technology Studies, 41*(1), 48–56.

Ferguson, K., & Reed, P. A. (2019). *How can career switchers and teachers without formal training be quickly prepared to teach engineering and technology education? Paper presented at the 106th Mississippi Valley Technology Teacher Education Conference*. TN: Nashville. http://www.mississippivalley.org/wp-content/uploads/2019/11/Training-Career-Switchers-Ferguson-and-Reed.pdf.

Fuller, E. J., Picucci, A. C., Collins, J. W., & Swann, P. (2001). *An analysis of laboratory safety in Texas*. Charles A. Dana Center of the University of Texas at Austin.

Gerlovich, J., McElroy, D., Parsa, R., & Kidwell, K. (2008). The status of science safety in Kentucky secondary schools. *Journal of the Kentucky Academy of Science, 69*(1), 19–28. https://doi.org/10.3101/1098-7096(2008)69[19:TSOSSI]2.0.CO;2.

Gerlovich, J. A., & Parsa, R. (2002). Surveying science safety. *The Science Teacher, 69*(7), 52–55.

Gerlovich, J. A., Whitsett, J., Lee, S., & Parsa, R. (2001). Surveying safety: How researchers addressed safety in science classrooms in Wisconsin. *The Science Teacher, 68*(4), 31–35.

Gerlovich, J. A., Wilson, E., & Parsa, R. (1998). Safety issues and Iowa science teachers. *Journal of the Iowa Academy of Sciences, 105*(4), 152–157.

Gill, M., Koperski, K., Love, T. S., & Roy, K. R. (2019). Developing a culture of safety through departmental planning. *Technology and Engineering Teacher, 79*(1), 22–25.

Grossman, R., & Salas, E. (2011). The transfer of training: What really matters: The transfer of training. *International Journal of Training and Development, 15*(2), 103–120. https://doi.org/10.1111/j.1468-2419.2011.00373.x.

Grubbs, M. E., Love, T. S., Long, D. L., & Kittrel, D. (2016). Science educators teaching engineering design: An examination across science professional development sites. *Journal of Education and Training Studies, 4*(11), 163–178. https://doi.org/10.11114/jets.v4i11.1832.

Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: A longitudinal analysis. *Journal of Educational Psychology, 105*(3), 774–786.

Hynes, M. M., & Hynes, W. J. (2018). If you build it, will they come? Student preferences for makerspace environments in higher education. *International Journal of Technology and Design Education, 28*(3), 867–883.

Katz-Navon, T., Naveh, E., & Stern, Z. (2007). Safety self-efficacy and safety performance: Potential antecedents and the moderation effect of standardization. *International Journal of Health Care Quality Assurance, 20*(7), 572–584.

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's four levels of training evaluation*. ATD Press.

Love, T. S. (2013). Addressing safety and liability in STEM education: A review of important legal issues and case law. *The Journal of Technology Studies, 39*(1), 28–41. https://doi.org/10.21061/jots.v39i1.a.3.

Love, T. S. (2014). Safety and liability in STEM education laboratories: Using case law to inform policy and practice. *Technology and Engineering Teacher [electronic edition], 73*(5), 1–13 http://www.iteea.org/File.aspx?id=86487&v=52ffd40f.

Love, T. S. (2015). Examining the demographics and preparation experiences of foundations of technology teachers. *The Journal of Technology Studies, 41*(1), 58–71. https://doi.org/10.21061/jots.v41i1.a.7.

Love, T. S. (2017a). Perceptions of teaching safer engineering practices: Comparing the influence of professional development delivered by technology and engineering, and science educators. *Science Educator, 26*(1), 21–31.

Love, T. S. (2017b). Tools and materials in primary education: Examining differences among male and female teachers' safety self-efficacy. In L. Litowitz & S. Warner

(Eds.), Technology and engineering education – Fostering the creativity of youth around the globe. Proceedings of the 34th Pupil's Attitude Toward Technology Conference, Philadelphia, PA: Millersville University. https://www.iteea.org/File.aspx?id=115739&v=21dfd7a.
Love, T. S. (2018). The T&E in STEM: A collaborative effort. *The Science Teacher, 86*(3), 8–10. https://doi.org/10.2505/4/tst18_086_03_8.
Love, T. S. (2019). STEM education safety: Temporary concern or enduring practice? Examining the progress of safety in STEM education. *Technology and Engineering Teacher, 78*(6), 15–17.
Love, T. S. (2022). Examining the influence that professional development has on educators' perceptions of integrated STEM safety in makerspaces. *Journal of Science Education and Technology, 31*(3), 289–302. https://doi.org/10.1007/s10956-022-09955-2.
Love, T. S., & Love, Z. J. (2022). The teacher recruitment crisis: Examining influential recruitment factors from a United States technology and engineering teacher preparation program. *International Journal of Technology and Design Education*. https://doi.org/10.1007/s10798-022-09727-4.
Love, T. S., & Maiseroulle, T. (2021). Are technology and engineering educator programs really declining? Reexamining the status and characteristics of programs in the United States. *Journal of Technology Education, 33*(1), 4–20. https://doi.org/10.21061/jte.v33i1.a.1.
Love, T. S., & Roy, K. R. (2017). Tools and equipment in non-traditional spaces: Safety and liability issues. *Technology and Engineering Teacher, 76*(8), 26–27.
Love, T. S., & Roy, K. R. (2018). Who should make your maker spaces? *ASEE Prism, 28*(2), 54.
Love, T. S., & Roy, K. R. (2020). *2020 national engineering education and CTE safety study data*. International Technology and Engineering Educators Association https://www.iteea.org/Resources1507/Safety/safety2020.aspx.
Love, T. S., & Roy, K. R. (2021). Key findings from Wisconsin's responses to the 2020 national T&E education safety survey. *Interface: Journal of the Wisconsin Technology Education Association, 61*(1), 22–23.
Love, T. S., & Roy, K. R. (2022). Safer engineering and CTE instruction: A national STEM education imperative. What the data tells us. *International Technology and Engineering Educators Association* https://www.iteea.org/SafetyReport.aspx.
Love, T. S., Duffy, B. C., Loesing, M. L., Roy, K. R., & West, S. S. (2020). Safety in STEM education standards and frameworks: A comparative content analysis. *Technology and Engineering Teacher, 80*(3), 34–38.
Love, T. S., Roy, K. R., & Marino, M. T. (2020). Inclusive makerspaces, fab labs, and STEM labs. *Technology and Engineering Teacher, 79*(5), 23–27.
Love, T. S., Roy, K. R., & Sirinides, P. (2021). What factors have the greatest impact on safety in Pennsylvania's T&E courses? *Technology and Engineering Education Association of Pennsylvania Journal, 69*(1), 5–22.
Love, T. S., Sirinides, P., & Roy, K. R. (2022). Examining factors associated with accidents in CTE and STEM education labs: A national safety study. *Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA*.
Luft, J. A., Firestone, J. B., Wong, S. S., Ortega, I., Adams, K., & Bang, E. (2011). Beginning secondary science teacher induction: A two-year mixed methods study. *Journal of Research in Science Teaching, 48*(10), 1199–1224.
National Academies of Sciences, Engineering, and Medicine (2019). *Science and engineering for grades 6–12: Investigation and design at the center*. The National Academies Press. https://doi.org/10.17226/25216.
National Research Council (NRC) America's lab report: Investigations in high school science. The National Academies Press. https:// 2006 doi.org/10.17226/11311.
National Science Teaching Association (NSTA) (2020). *Safety and the next generation science standards [White Paper]*. NSTA Safety Advisory Board https://static.nsta.org/pdfs/Safety%20and%20the%20Next%20Generation%20Science%20Standards_29Oct2020_FINAL.pdf.
NGSS Lead States (2013). *Next generation science standards: For states, by states*. The National Academies Press.
Nykänen, M., Puro, V., Tiikkaja, M., Kannisto, H., Lantto, E., Simpura, F., ... Teperi, A. (2020). Implementing and evaluating novel safety training methods for construction sector workers: Results of a randomized controlled trial. *Journal of Safety Research, 75*, 205–221.
Nykänen, M., Salmela-Aro, K., Tolvanen, A., & Vuori, J. (2019). Safety self-efficacy and internal locus of control as mediators of safety motivation – randomized controlled trial (RCT) study. *Safety Science, 117*, 330–338.
Nykänen, M., Sund, R., & Vuori, J. (2018). Enhancing safety competencies of young adults: A randomized field trial (RCT). *Journal of Safety Research, 67*, 45–56.
Occupational Safety and Health Administration (OSHA). (2020). Occupational safety and health standards: General industry regulations and standards (Standard No. 1910.29 CFR). https://www.osha.gov/laws-regs/regulations/standardnumber/1910.
Occupational Safety and Health Administration (OSHA). (2021). Resource for development and delivery of training to workers (OSHA 3824-05R 2021). https://www.osha.gov/sites/default/files/publications/osha3824.pdf.
Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (Seventh ed.). McGraw-Hill Education.
Reed, P. A., & Ferguson, M. K. (2021). Safety training for career and content switchers. *Technology and Engineering Teacher, 80*(7), 16–19.
Rose, M. A., Shumway, S., Carter, V., & Brown, J. (2015). Identifying characteristics of technology and engineering teachers striving for excellence using a modified Delphi. *Journal of Technology Education, 26*(2), 2–21.
Roy, K. R., & Love, T. S. (2017). Safer makerspaces, fab labs and STEM labs: A collaborative guide! National Safety Consultants, LLC.
Saks, A. M., & Belcourt, M. (2006). An investigation of training activities and transfer of training in organizations. *Human Resource Management, 45*(4), 629–648. https://doi.org/10.1002/hrm.20135.
Saks, A. M., & Burke, L. A. (2012). An investigation into the relationship between training evaluation and the transfer of training: Training evaluation. *International Journal of Training and Development, 16*(2), 118–127. https://doi.org/10.1111/j.1468-2419.2011.00397.x.
Sheskin, D. J. (2011). *Handbook of parametric and nonparametric statistical procedures* (5th ed.). Chapman and Hall.
Shidler, L. (2009). The impact of time spent coaching for teacher efficacy on student achievement. *Early Childhood Education Journal, 36*(5), 453–460.
Stephenson, A. L., West, S. S., Westerlund, J. F., & Nelson, N. C. (2003). An analysis of incident/accident reports from the Texas secondary school science safety survey, 2001. *School Science and Mathematics, 103*(6), 293–303. https://doi.org/10.1111/j.1949-8594.2003.tb18152.x.
Stroud, L. M., Stallings, C., & Korbusieski, T. J. (2007). Implementation of a science laboratory safety program in North Carolina schools. *Journal of Chemical Health & Safety, 14*(3), 20–30. https://doi.org/10.1016/j.jchas.2006.11.001.
Threeton, M. D., & Evanoski, D. C. (2014). Occupational safety and health practices: An alarming call to action. *Career and Technical Education Research, 39*(2), 119–136.
Threeton, M. D., Kwon, K., Fleck, J. A., Ketchem, R. B., & Farzam, L. (2021). An investigation of instructional practices which promote occupational safety and health. *International Journal of Occupational Safety and Ergonomics, 27*(3), 902–910.
Tonhäuser, C., & Büker, L. (2016). Determinants of transfer of training: A comprehensive literature review. *International Journal for Research in Vocational Education and Training, 3*(2), 127–165. https://doi.org/10.13152/IJRVET.3.2.4.
Volk, K. S. (2019). The demise of traditional technology and engineering education teacher preparation programs and a new direction for the profession. *Journal of Technology Education, 31*(1), 2–18.
West, S. S. (2016). Overcrowding in K-12 STEM classrooms and labs. *Technology and Engineering Teacher, 76*(4), 38–39.
West, S. S., Westerlund, J. F., Stephenson, A. L., Nelson, N. C., & Nyland, C. K. (2003). Safety in science classrooms: What research and best practice say. *The Educational Forum, 67*(2), 174–183. https://doi.org/10.1080/00131720308984555.

Tyler S. Love is an Associate Professor, Director of Graduate Studies in Career and Technology Education, and Coordinator of the undergraduate Technology and Engineering Education program at the University of Maryland Eastern Shore. He is an Authorized OSHA Outreach Trainer for General Industry, member of the National Science Teaching Association's Safety Advisory Board, and Safety Editor for the International Technology and Engineering Educators Association. He is also the recipient of 2018 CareerSafe® Safety Educator of the Year Award. His research interests include teacher preparation, safety, and liability in collaborative K-12 STEM labs, Fab Labs, and makerspaces.

Kenneth R. Roy is the Chief Safety Compliance Adviser for the National Science Teaching Association and the Director of Environmental Health and Safety for Glastonbury Public Schools in CT. Additionally, he is the Safety Compliance Officer for the National Science Education Leadership Association and General Manager/Senior Safety Consultant for the National Safety Consultants, LLC. Dr. Roy is both a nationally and internationally recognized safety specialist, accomplished author of over 12 science and STEM laboratory safety books and over 800 safety articles in professional journals. Dr. Roy also serves as an expert witness for school STEM lab accident litigation. His research interests include safety and liability in K-12 science labs, STEM labs, and makerspaces.

Melvin Gill serves as Technology and Engineering Education Department Chair at Meade High School in Maryland. He is also an Instructor of Career and Technology Education (CTE) at the University of Maryland Eastern Shore where he has taught their lab management and facilities safety course for 10 years. His research interests include practical safety applications for CTE teachers.

Mark Harrell is the Engineering Consultant at the Kentucky Department of Education and also Clinical Faculty in the College of Education and Human Development at the University of Louisville. He oversees all public secondary education engineering technology education programs in Kentucky. He taught high school engineering technology education for 12 years, served as an instructor and recruiter for two higher education institutions, and previously worked for a non-profit STEM curriculum provider. His research interests include examining safety policies and implementation at the state level to improve student and workforce safety practices.

Journal of Safety Research 82 (2022) 28–37

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Finding statistically significant high accident counts in exploration of occupational accident data

Tuula Räsänen [a,*], Arto Reiman [b], Kai Puolamäki [c,d], Rafael Savvides [c], Emilia Oikarinen [c], Eero Lantto [a]

[a] Finnish Institute of Occupational Health, Finland
[b] Industrial Engineering and Management, University of Oulu, Finland
[c] Department of Computer Science, University of Helsinki, Finland
[d] Institute for Atmospheric and Earth System Research, University of Helsinki, Finland

ARTICLE INFO

Article history:
Received 6 November 2020
Received in revised form 24 May 2021
Accepted 18 April 2022
Available online 28 April 2022

Keywords:
Occupational accident
Silent signals
Workplace
Prevention

ABSTRACT

Introduction: Finnish companies are legally required to insure their employees against occupational accidents. Insurance companies are then required to submit information about occupational accidents to the Finnish Workers' Compensation Center (TVK), which then publishes occupational accident statistics in Finland together with Statistics Finland. Our objective is to detect *silent signals,* by which we mean patterns in the data such as increased occupational accident frequencies for which there is initially only weak evidence, making their detection challenging. Detecting such patterns as early as possible is important, since there is often a cost associated with both reacting and not reacting: not reacting when an increased accident frequency is noted may lead to further accidents that could have been prevented. *Method:* In this work we use methods that allow us to detect silent signals in data sets and apply these methods in the analysis of real-world data sets related to important societal questions such as occupational accidents (using the national occupational accidents database). *Results:* The traditional approach to determining whether an effect is random is statistical significance testing. Here we formulate the described exploration workflow of contingency tables into a principled statistical testing framework that allows the user to query the significance of high accident frequencies. *Conclusions:* Our results show that we can use our iterative workflow to explore contingency tables and provide statistical guarantees for the observed frequencies. *Practical Applications:* Our method is useful in finding useful information from contingency tables constructed from accident databases, with statistical guarantees, even when we do not have a clear a priori hypothesis to test.

1. Introduction

Before undertaking preventive or corrective occupational safety actions, risks of accidents must be identified through rigorous management of information (Ross et al., 2005). Accident statistics information has been analyzed as defining different characteristics of occupational accidents by, for example, Pietilä et al. (2018), Ciarapica and Giacchetta (2009), Hovden et al. (2010), Papazoglou et al. (2015), Cruz Rios et al. (2017), and Jacinto and Guedes Soares (2008). In this paper, we present a method to find unusually high accident counts, which allows iterative exploration of data and gives a statistical guarantee for the observed counts. We call these patterns *silent signals*, "silent" because they are easy to miss with more traditional approaches, and "signals" because they may be informative about emerging patterns or changes in the data.

The digital era in which we now live provides even more possibilities for complex data gathering and analysis (Badri et al., 2018). Technological developments have made it possible to collect and analyze different kinds of data from various sources using highly developed tools and methods. However, this development trend has not eliminated the role of humans as those who determine whether the data are actually useful for accident prevention purposes (Badri et al. 2018). The concept of 'big data' is used to describe this entity of handling and processing massive data sets. For instance, Wu and Li (2019) highlight the complexity of accident database analyses and suggest applying entropy theory to be able to more deeply understand the dynamic nature of occupational safety.

Finnish workplaces are legally required to insure their employees against occupational accidents. Insurance companies are then

* Corresponding author.
E-mail address: tuula.rasanen@ttl.fi (T. Räsänen).

https://doi.org/10.1016/j.jsr.2022.04.003
0022-4375/© 2022 The Author(s). Published by the National Safety Council and Elsevier Ltd.

required to submit information about occupational accidents to the Finnish Workers' Compensation Center (TVK), which then publishes occupational accident statistics in Finland together with Statistics Finland.

Therefore, Finland's statistics and data have good coverage of blue-collar and white-collar employees' occupational accidents. In this paper we use a data set of occupational accidents in Finland from 2010–2015, available from TVK. Each accident is described by a date and 15 categorical variables (see Supplementary Material 1). Each variable consists of numerical codes that correspond to, for example, different industries, job types, accident causes, and injured body parts. We study the problem of finding accident counts that are larger than could be expected by random chance alone.

Accident frequencies can be displayed through contingency tables (or cross tabulation, pivot tables). We present a permutation testing-based statistical framework for exploring data through these tables. Contingency tables (such as Table 1) may be relatively large and contain multiple cells corresponding to accident frequencies, and multiple tables may be viewed through an iterative process. It is often the case that the analyst observes an unusually high accident frequency and does not know whether this is a significant finding or just a random effect. If the analyst can determine that it is a random effect, they can then focus on more promising hypotheses and avoid wasting resources on spurious findings or taking actions that are not based on the evidence at hand. For example, in Table 1 the two highest frequencies (405 and 149) seem to be a significant finding, while it is difficult to determine this for accidents with a lower absolute frequency (e.g., 37 or 4).

A traditional approach for determining whether an effect is random is statistical significance testing. Using, for example, a common statistical test, such as the chi-square test of independence (Agresti, 2019; Cacha, 1997), they can answer questions such as: "How unlikely is it to observe the counts in Table 1, if the variables are independent?" yielding a p-value of $\leq 10^{-16}$. The low p-value indicates that the cell values that were observed in Table 1 are extremely unlikely if the variables were independent, and thus there is evidence against independence.

However, these common statistical tests for contingency tables suffer from several shortcomings. First, they test a specific hypothesis and provide a single determination, or p-value, for the whole table. If the analyst is interested in a single accident frequency in the table, they are unable to obtain more focused answers. For example, after obtaining a low p-value using a chi-square test on Table 1, they know there is a significant finding, but cannot investigate which cells influenced this determination. If they attempt to naively test every cell in the table, they risk false discoveries due to the multiple comparisons problem (Dudoit et al., 2003). Second, most statistical tests have a specified null hypothesis that is formed before viewing the data. These are problems in practice. Answering questions such as 'what else is there in the data?' is not possible, because it would require formulating a new hypothesis that somehow takes into account what has already been observed, and then testing it on unseen data.

In practical data analyses, hypotheses are often formed after viewing the data during an iterative process, which is not in line with the assumptions made in traditional statistical testing, in which the hypothesis about the data should be formed before even observing the data at all. Therefore, there is a need for a statistical methodology that allows for testing hypotheses *during* the iterative workflow of viewing contingency tables. In this paper, we present such a methodology (initially introduced in Savvides et al., 2019) and examples of finding novel features from a data set of occupational accidents in Finland.

Table 1
Contingency table of accident frequencies for Specific physical activity and Cause of accident variables. A p-value is included in parentheses (a dot signifies a p-value equal to one). p-values with $p \le \alpha = 0.1$ are statistically significant. The p-values have been corrected for multiple testing using the minP method (see text).

Cause of accident / Specific physical activity	1100 ground level buildings/surfaces/structures	2699 other portable/mobile machines	2703 machines/chemical processes	2706 machines, other processes	2799 other fixed machines	2802 elevators/lifts/hoists/jacks etc.	2803 cranes/hoisting machines with suspended load	2811 non-lifting load transporting devices	2816 forklift trucks	2819 other handling mobile devices	2899 transport/storage systems not listed	4200 chemical/radioactive/biological substance
00 No information	4 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	11 (<0.01)
10 Operating machine	5 (.)	6 (0.032)	4 (0.51)	5 (<0.01)	7 (<0.01)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	1 (.)	12 (.)
20 Working with hand-held tools	8 (.)	1 (.)	4 (0.17)	1 (.)	2 (.)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	0 (.)	12 (0.96)
30 Driving/being on board a means of transport or handling equipment	11 (.)	3 (0.77)	1 (.)	1 (.)	0 (.)	0 (.)	1 (0.68)	0 (.)	7 (<0.01)	1 (1)	1 (.)	0 (.)
40 Handling of objects	37 (.)	7 (.)	12 (0.097)	2 (.)	8 (.)	1 (.)	0 (.)	3 (.)	1 (.)	1 (.)	0 (.)	149 (<0.01)
50 Carrying by hand	36 (.)	6 (0.14)	0 (.)	1 (.)	1 (.)	1 (.)	0 (.)	1 (.)	0 (.)	1 (.)	4 (0.045)	7 (.)
60 Movement	405 (<0.01)	2 (.)	1 (.)	3 (.)	5 (.)	3 (.)	0 (.)	2 (.)	5 (.)	2 (.)	3 (.)	16 (.)
70 Presence	6 (.)	0 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	2 (1)	0 (.)	0 (.)	23 (<0.01)
99 Other specific physical activities	9 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	14 (0.047)

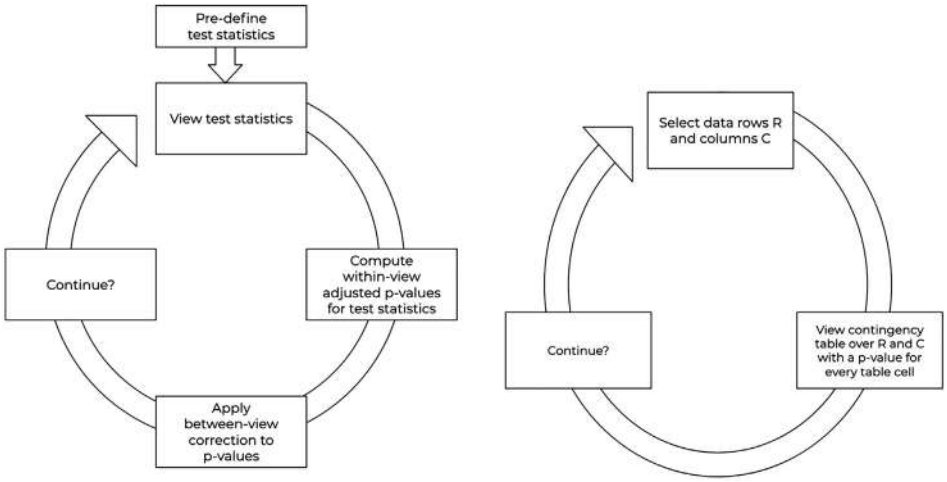


Fig. 1. Flowchart of the statistical testing procedure (left) and the practical workflow from the user's perspective (right).

The number of compensated accidents at work in Finland has been quite steady over the last 10 years. Registers show that over 126,000 occupational accidents occurred among wage earners in 2018 (Workers' Compensation Center, 2019). The vast majority (82%) of these accidents occurred at workplaces. Calculations by the Ministry of Social Affairs and Health show the annual costs of occupational accidents and injuries to be around EUR 2–2.5 billion in Finland (Rissanen & Kaseva, 2014). In addition, the human suffering of the injured person and their families and co-workers causes indirect costs that are difficult to estimate (Manuele, 2011).

Despite several measures to strengthen accident prevention and occupational safety, the statistics show a disparity between practical working life and the ambitious goal of zero accidents. Clearly, new approaches to improving occupational safety and accident prevention should be introduced. To contribute to this discussion, we introduce a new approach to large-scale occupational accident statistics categorization to achieve a more in-depth understanding of accidents for occupational accident prevention purposes.

Our objective is to detect *silent signals,* by which we mean patterns in the data such as increased occupational accident frequencies for which there may initially be only weak evidence, making their detection challenging. Detecting such patterns as early as possible is important, since there is often a cost associated with both reacting and not reacting: not reacting when an increased accident frequency is noted may lead to further accidents that could have been prevented. In this work we use methods that allow us to detect silent signals in data sets and apply these methods in the analysis of real-world data sets related to important societal questions such as occupational accidents (using the national occupational accidents database).

1.1. Motivating example

We next present an example motivating our approach. The example demonstrates how our approach works compared to a standard method. Using our approach, an analyst may ask more specific questions than standard methods.

Suppose that an analyst explores accident data using contingency tables. Table 1 displays one such contingency table (or cross tabulation, pivot table), in which each cell corresponds to an accident count in the chemical industries in Finland. If a cell appears to have a high frequency, the analyst may wish to know whether the high value is statistically significant. A standard approach, such as a chi-square test of independence, provides a *single* p-value for the whole table ($p \leq 10^{-16}$). The low p-value indicates that the table is statistically significant, and the analyst has made a "discovery." However, it is unclear which cell of the table is significant, which is especially problematic when the table is large.

In our approach, we determine whether a cell value is significantly high by computing a p-value for *every* cell in the table. The *p*-values are computed using a *permutation test*, in which the test statistic is the cell value and a p-value is computed by simulating the distribution of the test statistic under the null hypothesis. Our approach works as follows. We use a null hypothesis of independence between the variables of the table, and we simulate the null distribution by permuting each column in the data independently. This permutation scheme preserves the value distribution within each column and breaks any dependencies between columns. We permute the data multiple times and compute a contingency table on each permuted data set (Fig. 1b). This process provides a distribution of values for each cell in the table, which corresponds to the null distribution of each test statistic. A p-value can then be computed for each cell by comparing the simulated null distribution of the test statistic with its value in the original table. Finally, as we perform multiple tests, the *p*-values need to be adjusted for the multiple hypotheses problem. We adjust the *p*-values using a resampling-based adjustment procedure, called minP, which is discussed in the Methods section.

By computing a p-value for every cell, we can answer more specific questions. For example, the *p*-values in Table 1 communicate how likely it is to observe a count as high as that in the table when the 'Specific physical activity' and 'Cause of accident' variables are independent. In contrast, a standard approach, such as a chi-square test of independence, that provides one p-value for the whole table corresponds to the question: how likely is it to observe Table 1, when the 'Specific physical activity' and 'Cause of accident' variables are independent.

Another disadvantage of common statistical tests (besides not being able to test single cells), is that the analyst is unable to test more interesting hypotheses of independence. For example, how unlikely is it to observe Table 2, when the variables are *independent over most of the data, excluding a subset in which they are dependent*? In order to answer this question, we construct a permutation test, using a modified permutation scheme. Instead of permuting each column independently (as in Table 1), we now independently permute *tiles*. A tile is simply a subset of rows and columns (Fig. 1). It can act as a constraint on the permutation process, in that the rows in every tile are permuted independently to other tiles. In the previous example of Table 1, we permuted each column independently, which is equivalent to having a tile constraint over each

Table 2
An alternative hypothesis is tested in Table 1. The accident frequencies are the same as in Table 1, while the *p*-values are computed differently. In addition to independence, we use *tile constraints* to fix the relationship of the variables in the subset of the data with Specific physical activity=(10. Operating machine). As a result, all cells with Specific physical activity = 10 are insignificant, whereas the other cells that were insignificant in Table 1 are now significant, e.g., (50. Carrying by hand, 2699. other portable/mobile machines) and (20. Working with hand-held tools, 2703. machines/chemical processes).

Cause of accident Specific physical activity	1100 ground level buildings/surfaces/ structures	2699 other portable/mobile machines	2703 machines/chemical processes	2706 machines, other processes	2799 other fixed machines	2802 elevators/ lifts/ hoists/jacks etc.	2803 cranes/ hoisting machines with suspended load	2811 non-lifting load transporting devices	2816 forklift trucks	2819 other handling mobile devices	2899 transport/ storage systems not listed	4200 chemical/ radioactive/ biological substance
00 No information	4 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	11 (0.015)
10 Operating machine	5 (.)	6 (.)	4 (.)	5 (.)	7 (.)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	1 (.)	12 (.)
20 Working with hand-held tools	8 (.)	1 (.)	4 (0.098)	1 (.)	2 (0.98)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	0 (.)	12 (0.92)
30 Driving/being on board a means of transport or handling equipment	11 (.)	3 (0.55)	1 (.)	1 (.)	0 (.)	0 (.)	1 (0.67)	0 (.)	7 (<0.01)	1 (1)	1 (.)	0 (.)
40 Handling of objects	37 (.)	7 (.)	12 (0.022)	2 (.)	8 (0.7)	1 (.)	0 (.)	3 (1)	1 (.)	1 (.)	0 (.)	149 (<0.01)
50 Carrying by hand	36 (.)	6 (0.042)	0 (.)	1 (.)	1 (.)	1 (.)	0 (.)	1 (.)	0 (.)	1 (.)	4 (0.049)	7 (.)
60 Movement	405 (<0.01)	2 (.)	1 (.)	3 (.)	5 (.)	3 (.)	0 (.)	2 (.)	5 (.)	2 (.)	3 (.)	16 (.)
70 Presence	6 (.)	0 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	2 (0.99)	0 (.)	0 (.)	23 (<0.01)
99 Other specific physical activities	9 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	14 (0.023)

column and then permuting each tile independently. In this second example of Table 2, we add a tile constraint over each column, as before, and we also apply another tile constraint over a subset of rows and all columns in which 'Specific physical activity'=10. The tile constraint causes the dependencies between the columns to be preserved in this subset of the data. As a result, every permuted data set has fixed the relationship of these variables in this subset of the data, which modifies the null distribution and the computed *p*-values. In practice, this means that the contingency table for every permuted data set is identical to the original data for 'Specific physical activity'=10, and the *p*-values are insignificant. Therefore, by using modified permutation schemes we can answer questions based on what was observed before, such as "what else is there in the data that is not explained by the observed accident counts?" This is not possible with standard methods.

In this paper, we present a permutation testing-based statistical framework for exploring data through contingency tables, based on the article published by Savvides et al. (2019). The framework includes two contributions: (1) application of a powerful statistical test that computes a p-value (adjusted for multiple testing) for every cell in a contingency table, and (2) a sequential exploratory procedure that is adjusted for multiple testing.

2. Material and methods

2.1. Data

The basic reporting of occupational accidents is done by companies. They investigate each accident and fill out an accident report form for the insurance company. When compiling statistics that present the conditions under which occupational accidents occur, the TVK uses ESAW variables (European Statistics on Accidents at Work). As described in the introduction, in this study we use a data set of occupational accidents in Finland from 2010–2016. In TVK's data, each accident is described by a date and 15 categorical variables (see Supplementary Material 1). In addition to the 7 ESAW variables, TVK use their own more specific variables. Each variable consists of numerical codes that correspond to different industries, job types, accident causes, and injured body parts.

2.2. Methods

2.2.1. Overview

We study the accident data by calculating contingency tables. For example, an expert chooses the 'industry' and 'body part' variables and views a table that contains accident frequencies for every combination of industry and body part.

In this exploratory workflow, the user may observe unusually high or low accident counts, which may be true phenomena in the data or merely random artefacts. One traditional method for discarding findings that cannot be distinguished from random noise, is statistical significance testing. Hence, here we formulate the described exploration workflow of contingency tables into a principled statistical testing framework that allows the user to query the significance of high accident frequencies.

We follow an approach presented in our previous paper Savvides et al. (2019), which uses a permutation test. In this article, the authors provide a novel realization of the method for contingency tables and a new iterative correction method based on alpha investing. The test requires a test statistic and its null distribution. If the test statistic computed on the observed data is extreme compared to its null distribution, then it is significant. In our example, the test statistic corresponds to an accident frequency (a value of the cell in the contingency table) and the null distribution is defined as a model of the user's knowledge of the data as defined in Puolamäki et al. (2021). The user's knowledge is parameterized as a probability distribution over all possible data sets, and samples are drawn from this distribution to form the empirical null distribution of the test statistic.

Our null hypothesis assumes that the marginal distributions of the variables are fixed and that all possible data sets can be obtained by permuting the columns of the data sets. A sample from the null hypothesis can be obtained by computing the contingency table for such a permuted data set. Without any constraints, the columns are permuted independently and at random, which results in a null hypothesis corresponding to situations in which the data attributes are independent of each other and any relation between them is broken. During the exploration, the user's knowledge is updated, using observed contingency tables as constraints: when a pattern is observed, the permutations are constrained so that the attributes shown for the user in the permutation table are permuted together, after which all samples produce the observed contingency table. We can informally say that a test statistic is significant if it is exceptionally high compared to the user's expectations (i.e., if the test statistic has a low *p*-value).

Two aspects in the exploration process require attention. Firstly, multiple test statistics are often viewed and hence tested simultaneously. A multiple testing correction is required in order to avoid false discoveries. Secondly, we assume that the user views the data more than once (i.e., the exploration is an iterative process). If the user looks at the data enough times, they will eventually discover something significant by chance alone. This adds another level of multiple testing, which also requires a correction. We next formally describe our procedure that incorporates these two levels of multiple testing corrections to control the family-wise error rate (FWER) at a chosen level.

2.2.2. Details

In this section, we formally describe the testing procedure, as initially described in our previous paper Savvides et al. (2019). The novel contributions in this paper are the application to the domain of accident data using contingency tables, two theorems, and the use of alpha investing as an iterative correction.

Let Ω denote the *sample space* (i.e., the set of all possible data sets), and $\omega_0 \in \Omega$ the observed data set, which has been sampled from an unknown probability distribution Pr_D over Ω. As discussed above, we implicitly assume the user's knowledge is parametrized as a probability distribution of Pr_U over Ω.

Our goal is to formulate a statistical testing procedure in which Pr_U is the null distribution and the test statistic corresponds to a pattern observed in the data. Intuitively, we call the pattern significant if the test statistics (counts in contingency table) are extreme compared to Pr_U.

Test statistics. We define test statistics as functions $T_i : \Omega \rightarrow \mathcal{R}, i \in [n_T]$, which measure the 'strength' of an observed pattern and where we have used the notation $[n_T] = \{1, \dots, n_T\}$. In this paper, the observed patterns are the counts in a contingency table.

Iterative exploration. We assume that the user is shown a finite *sequence of* n_V *views* of the data. Each view V_t with $t \in [n_V]$ contains a subset of counts T_i shown in one contingency table and is defined as an index set, i.e., $V_t \subseteq [n_T]$. The idea is that in view V_t, the user observes the values of the test statistics on the observed data $T_j(\omega_0)$ for all $j \in V_t$.

Null distribution. The user's knowledge Pr_U can be updated with the use of *constraints* $C_i : \Omega \rightarrow P(\Omega)$ (where $P(\Omega)$ denotes the power set of Ω), which restrict the possible data sets to those that have a test statistic equal to the observed data set, i.e., $C_i(\Omega) = \{\omega \in \Omega : T_i(\omega) = T_i(\omega_0)\}$. We identify a set of constraints using an index set $I \subseteq [n_T]$ and we denote the set of possible data sets that satisfy a set of constraints $I \subseteq [n_T]$ as $\Omega_I = \cap_{i \in I} C_i(\Omega) = \{\omega \in \Omega : T_j(\omega) = T_j(\omega_0) \forall j \in I\}$.

In each view V_t, the null distribution is the user's *current* knowledge Pr_U, which has been updated on the basis of the test statistics I_t observed so far. We define *constrained p-values* as:

$$p_{i|I} = \frac{Pr_U(\{\omega \in \Omega_I : T_i(\omega_0) \leq T_i(\omega)\})}{Pr_U(\Omega_I)}$$

The null hypothesis that corresponds to a constrained p-value $p_{i|I}$ is that the distribution Pr_D (from which the observed data ω_0 is sampled) satisfies the following condition for any $\omega \in \Omega_I$:

$$\frac{Pr_D(\omega)}{Pr_D(\Omega_I)} = \frac{Pr_U(\omega)}{Pr_U(\Omega_I)} \tag{1}$$

The intuitive interpretation for the null hypothesis is that if true, then the conditional distribution of the data, given the constraints, is equal to the corresponding distribution assumed by the user.

Within-views correction. A view contains multiple test statistics, which are used simultaneously for testing the null hypothesis (user's knowledge). Since multiple tests are performed, a multiple testing correction is warranted. We use the *step-down minP procedure* (Westfall-Young, 1993) to compute FWER-adjusted *p*-values for the test statistics in a single view.

The minP algorithm is summarized as follows: given a vector of observed test statistics $X_0 = (x_1, \cdots, x_n)$ and a matrix of m samples of test statistic vectors from the null distribution $Y = (X_1, \cdots, X_m)$, the minP algorithm computes a vector of FWER-adjusted *p*-values $P = (p_1, \cdots, p_n)$.

An implementation of the minP algorithm in the R programming language (R Core Team, 2020) is provided in the Supplementary Material 2.

Between-views correction. The user is shown multiple views in a sequential manner and in each view, a hypothesis is tested. In addition to the within-views correction, an additional multiple testing correction is warranted for the sequence of views. If the number of views is known in advance, we can apply any multiple testing correction, such as a Bonferroni correction. However, if the number of views is not known in advance, we instead apply an online multiple testing correction, such as alpha-investing (Foster & Stine, 2008). In alpha investing, the user has an alpha wealth of total acceptable error that they may "invest" in hypotheses. If the hypothesis provides a significant result, then the alpha investment is returned and can be reused in future hypotheses.

A simple online multiple testing procedure is a generalization of the Bonferroni, called a *weighted Bonferroni correction* (Holland & Copenhaver, 1988). The weighted Bonferroni correction is summarized as follows: given a sequence of *p*-values p_t $t \in [n_V]$, multiply each p-value with a factor w_t such that $\sum_{t=1}^{\infty} 1/w_t = 1$. Then the *p*-values $\underline{p_t} = \min(1,\ w_t\ p_t)$ are adjusted for FWER.

2.3. Testing procedure

The elements described above (test statistic, null distribution, iterative exploration, within-view correction and between-view correction) are combined into a statistical testing procedure. The testing procedure consists of the following steps, for a given data set $\omega_0 \in \Omega$ sampled from Pr_D, number of views n_V and weights w_t for each view $V_t, t \in [n_V]$:

1. Set $t \leftarrow 1, V_0 \leftarrow \{\}, I_0 \leftarrow \{\}$
2. Set $I_t \leftarrow I_{t-1} \cup V_{t-1}$
3. View values of test statistics $T_j(\omega_0)$ where $j \in V_t \subseteq [n_T] \{ I_t$

4. Compute within-view adjusted p-values p_{tj} using minP algorithm and apply between-view correction to obtain final adjusted p-values $\underline{p_{tj}} = \min(1,\ w_t\ p_{tj})$
5. Set $t \leftarrow t + 1$
6. If $t \le n_V$ continue from Step 2, else terminate

The flowcharts below (Fig. 2)) visualize the statistical procedure (left) and how it translates to a workflow for the user (right). The statistical procedure is generally applicable to data exploration with any visualization, while the workflow presented here is a special case where the visualization is a contingency table. The workflow is presented in the section Case Example.

The procedure is an iterative process, as denoted by the "Continue?" block in the flowchart. The process terminates either when the user wishes to end the exploration, or when the exploration has no practical reason to continue, for example if the data are fully constrained (i.e., everything has been observed already), or the specified alpha budget is depleted when using an alpha investing method. In practice, the lack of a termination criterion means that the user is free to explore as long as there are discoveries to be made in the data and there is enough available alpha budget.

The following theorems show that the above procedure controls the family-wise error rate (FWER) at a chosen level α, both within each view and overall for the whole procedure. The theorems are a novel contribution that extends our previous work (Savvides et al., 2019).

Theorem 1 (within-view)

Let V_t be a contingency table containing test statistics T_j where $j \in V_t$, and p_j are the corresponding p-values as computed with the minP algorithm using Pr_U as a null distribution.

Then for any given constant $\alpha \in [0, 1]$ and for every $j \in V_t$ we have that $Pr(p_j \le \alpha) \le \alpha$, i.e., the probability of at least one false discovery is at most α.

Proof. Assume we have m data samples from Pr_U, denoted by ω_i. Let $X_0 = [T_j(\omega_0)], j \in V_t$ be a vector of test statistics for the observed data set, $X_i = [T_j(\omega_i)], j \in V_t$ be a vector of test statistics for data sample ω_i, and $Y = [X_1, \cdots, X_m]$ be a matrix of m test statistic vectors.

Then the p-values $p_i = MINP(X_0, Y)$ are FWER-adjusted, since the minP algorithm controls FWER.

Theorem 2 (between-views)

Let $S = (V_1, \cdots, V_{n_V})$ be a sequence of views and p_{tj} the p-values in each view, as computed with the minP algorithm using Pr_U as a null distribution and corrected with the weighted Bonferroni correction.

Then for any given constant $\alpha \in [0, 1]$, for every $t \in [n_V]$ and every $j \in V_t$ we have that $Pr(\underline{p_{tj}} \le \alpha) \le \alpha$.

Proof. We use $W_t \subseteq V_t$ to denote the views whose p-values obey the null hypothesis, according to the definition of Eq. (1), and by $S = (W_1, \cdots, W_{n_V})$ the respective sequence of views. We denote by $P_t = \min_{j \in W_t} p_{jt}$, with $P_t = 1$ if $W_t = n$ull, the minimal p-value in view $t \in [n_V]$ in which the null hypothesis is true. Since the p-values in each view have been corrected for *FWER*, we know that $P(P_t \le \alpha') \le \alpha' \forall \alpha'$.

Consider an iteration $t \in [n_V]$. The user has the option of choosing any subset of test statistics $j \in [n_T]$ to V_t. It can be shown that for all test statistics, including P_t, for which the null hypothesis is true, it holds that they are stochastically no larger than the uniform distribution. Then, P_t is multiplied by w_t, which means that the probability of a false positive at iteration $t \in [n_V]$ is therefore at most ${w_t}^{-1}\alpha$, resulting in a total false positive probability of at most α when summed over all iterations in $[n_V]$, since $\sum_{t=1}^{n_V} w_t^{-1} \leqslant 1$.

3. Case example

In this section, we demonstrate the statistical testing procedure using case studies and discuss their results. As a first case study, we focus on occupational accident data from the chemical product industry in Finland in 2010–2015. The idea is to explore the accident data using the testing procedure in order to obtain insights into unusually high accident frequencies in the chemical product industry. Focusing the analysis on one industry enables the selection of variable categories that are relevant to that industry, for example, standard variables in accident reports. Selecting only a subset of categories reduces the number of multiple hypotheses and hence improves statistical power.

Note that alternative approaches to find similar results are limited or are not typically used, to our knowledge. For example, using a standard test, such as a chi-square test of independence, we can compute a p-value for each table (Cacha, 1997). However, the test provides a *single* p-value for the whole table (as opposed to one for each table cell) and the p-value does not account for previously observed significant patterns (whereas here the user's knowledge is updated, which affects future p-values). Therefore, traditional methods are not directly comparable to our presented framework, as discussed in the Motivating example. Two general methods that can act as baselines are an approach where no corrections are performed, and an approach where the corrections are overly strict. The first approach may lead to spurious findings and no control of the error rate, which our method controls. The second approach may lead to no findings due to lack of statistical power, while our approach retains statistical power through the powerful minP correction. In addition, the correctness of the results of the case study cannot be demonstrated experimentally, since there is no "ground

A	B
A1	B1
A1	B1
A1	B2
A2	B2
A2	B3
A2	B3

	B1	B2	B3
A1	2	1	0
A2	0	1	2

A	B
A2	B1
A1	B2
A1	B3
A2	B1
A1	B3
A2	B2

	B1	B2	B3
A1	0	1	2
A2	2	1	0

A	B
A1	B1
A1	B1
A2	B3
A2	B2
A1	B3
A2	B2

	B1	B2	B3
A1	2	0	1
A2	0	2	0

Fig. 2. Illustration of permutation with *tiles* and its effect on contingency tables. *Left*: Data set D with variables A and B. A contingency table is computed from D by counting all combinations of values in variables A and B. *Centre*: Variables A and B are permuted independently. The permutation is realized through tile constraints. A tile is placed over each column and each tile is permuted independently. A contingency table is computed in the same manner. *Right*: Variables A and B are permuted independently, except for a subset of rows where B = B1. Tile constraints are placed in each column, as before, and an additional tile is placed on a subset of rows where B = B1. The result is that variables A and B retain their relationship within the subset where B = B1. This is illustrated in the contingency table in which the counts containing B1 are the same as those in the original data.

truth" to compare to. The validity of the approach is provided by the mathematical proofs in the methods section.

We now describe a case example of using the testing procedure to explore a data set of occupational accidents. The exploration consists of three iterations (i.e., three contingency tables are viewed sequentially). In each iteration, the contingency table is determined by selecting two variables (columns) and a subset of data points (rows). For each cell in the table, a FWER-adjusted p-value is computed using as a within-view correction the minP algorithm and as a between-view correction a standard Bonferroni procedure for a predetermined number of three iterations. After viewing a table, the observed accident frequencies are used to update the user's background distribution and are therefore not significant in future tables.

Iteration 1. We start by viewing a contingency table of the Specific physical activity and Cause of accident variables for the whole data. For the Cause of accident variable we view 12 out of 73 categories that are relevant to the chemical product industry (e.g., chemicals, logistics and machinery). For the Specific physical activity variable we view all nine categories (e.g., using machinery or handling objects).

We discover eight statistically significant accident frequencies in Table 3. These frequencies are unusually high compared to the current knowledge of the user, as parameterized by the null distribution.

After viewing Table 3, the user's knowledge is updated so that if Table 3 is viewed in future iterations, it does not contain significant findings. The user's knowledge is updated by modifying the null distribution through a tile constraint {R = all rows, C = (Specific physical activity, Cause of accident)} that fixes the relationship of the variables in Table 3.

Iteration 2. The table for the next iteration is determined by the user, by selecting a subset of rows and the two variables of the contingency table. The next table can be completely independent from the current one or (as in this example) it can be based on the findings of previous tables. We now focus on a subset of the data that was significant in Table 3, denoted by R1 = {Specific physical activity = 60 Movement, and Cause of accident = 1100 ground level buildings/surfaces/structures}. In subset R1, we view a contingency table of the Industry (4 digit) (using 18 out of 587 categories which, based on our knowledge, are relevant for the chemical product industry) and Working process (using all 32 categories) variables. The contingency table is presented in Table 4 (see Supplementary Material 3 for full table) and we discover one statistically significant result.

The effect of Iteration 1 on Iteration 2 has two parts. First, subset R1 was selected on the basis of the findings from Iteration 1. Second, the constraints from Iteration 1 on the null distribution may affect the *p*-values of Iteration 2. In this case, the constraints have no overlap with the data of Table 4, and as such have no effect on the *p*-values.

After viewing Table 4, the user's knowledge is updated, similarly to Iteration 1. The null distribution is updated by adding a tile constraint {R = R1, C = (Industry (4 digit), Working process)} that fixes the relationship of the variables in Table 4 for the viewed subset R1 (i.e., not for all the data). After fixing this result for subset R1, we can now test whether the result is significant for the rest of the data. We do this by using the whole data set (instead of only subset R1) to view the same variables as in Table 4.

Iteration 3. A significant result is discovered in Iteration 2, for subset R1 of the data. We now repeat the steps of Iteration 2 using all the accident reports in the chemical product industry, to investigate whether accident frequency is also significantly high for the whole data. In Table 5, we discover seven significant results (see Supplementary Material 3). However, these do not include the significant result from Iteration 2. In other words, we observe a signif-

Table 3
Contingency table of Cause of accident and Specific physical activity variables. Each cell contains the accident count for a category of each variable. A FWER-adjusted *p*-value is computed for each cell and is contained inside parentheses. Insignificant p-values ($p = 1$) are denoted by a dot. *p*-values with $p \le \alpha = 0.1$ are statistically significant. The cell in bold is the subset R1 used in Table 4.

Cause of accident / Specific physical activity	1100 ground level buildings/surfaces/structures	2699 other portable/mobile machines	2703 machines/chemical processes	2706 machines, other processes	2799 other fixed machines	2802 elevators/lifts/hoists/jacks etc.	2803 cranes/hoisting machines with suspended load	2811 non-lifting load transporting devices	2816 forklift trucks	2819 other handling mobile devices	2899 transport/storage systems not listed	4200 chemical/radioactive/biological substance
00 No information	4 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	11 (0.022)
10 Operating machine	5 (.)	6 (0.097)	4 (.)	5 (0.022)	7 (0.022)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	1 (.)	12 (.)
20 Working with hand-held tools	8 (.)	1 (.)	4 (0.5)	1 (.)	2 (.)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	0 (.)	12 (.)
30 Driving/being on board a means of transport or handling equipment	11 (.)	3 (.)	1 (.)	1 (.)	0 (.)	0 (.)	1 (.)	0 (.)	7 (0.022)	1 (.)	1 (.)	0 (.)
40 Handling of objects	37 (.)	7 (.)	12 (0.29)	2 (.)	8 (.)	1 (.)	0 (.)	3 (.)	1 (.)	1 (.)	0 (.)	149 (0.022)
50 Carrying by hand	36 (.)	6 (0.41)	0 (.)	1 (.)	1 (.)	1 (.)	0 (.)	1 (.)	0 (.)	1 (.)	4 (0.13)	7 (.)
60 Movement	**405 (0.022)**	2 (.)	1 (.)	3 (.)	5 (.)	3 (.)	0 (.)	2 (.)	5 (.)	2 (.)	3 (.)	16 (.)
70 Presence	6 (.)	0 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	2 (.)	0 (.)	0 (.)	23 (0.022)
99 Other specific physical activities	9 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	14 (0.14)

Table 4

Contingency table of Working process and Industry (4 digit) variables for the subset of the data defined by R1 (significant result from Table 3). Only a part of the table is shown here for clarity; refer to the Supplementary Material 3 for the whole table. Each cell contains the accident count for a category of each variable. A FWER-adjusted p-value is computed for each cell and is contained inside parentheses. Insignificant p-values ($p = 1$) are denoted by a dot. p-values with $p \leq \alpha = 0.1$ are statistically significant.

Working process Industry 4 digit	00 no information	11 production, manufacturing, processing	12 storing	19 other manufacturing and storing	21 excavation	22 new construction, building	23 new construction, roads, bridges, dams, ports	24 remodelling, repairing, building maintenance	25 demolition	29 other construction
2011 Manufacture of industrial gasses	2 (.)	4 (.)	2 (.)	0 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2012 Manufacture of colours and pigments	0 (.)	2 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2013 Manufacture of other non-organic basic chemicals	6 (.)	27 (.)	1 (.)	1 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2014 Manufacture of other organic basic chemicals	1 (.)	4 (.)	0 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2015 Manufacture of fertilizers and nitrogen compounds	0 (.)	3 (.)	2 (.)	0 (.)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	0 (.)
2016 Manufacture of plastic materials	1 (.)	29 (0.05)	1 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2017 Manufacture of synthetic rubber raw material	1 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2020 Manufacture of pesticides and agriculture chemicals	1 (.)	6 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2030 Manufacture of paints, printing inks and enamels	0 (.)	17 (.)	9 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)

Table 5

Contingency table of Working process and Industry (4 digit) variables for the whole data set. Only a part of the table is shown here for clarity; refer to the Supplementary Material 3 for the whole table, which contains more statistically significant cells. Each cell contains the accident count for a category of each variable. A FWER-adjusted p-value is computed for each cell and is contained inside parentheses. Insignificant p-values ($p = 1$) are denoted by a dot. p-values with $p \leq \alpha = 0.1$ are statistically significant.

Working process Industry 4 digit	00 no information	11 production, manufacturing, processing	12 storage	19 other manufacturing and storage	21 excavation	22 new construction, building	23 new construction, roads, bridges, dams, ports	24 remodelling, repairing, building maintenance	25 demolition	29 other construction
2011 Manufacture of industrial gasses	12 (.)	47 (.)	21 (.)	3 (.)	1 (.)	0 (.)	0 (.)	1 (.)	0 (.)	0 (.)
2012 Manufacture of colours and pigments	1 (.)	27 (.)	2 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2013 Manufacture of other non-organic basic chemicals	17 (.)	168 (.)	20 (.)	15 (.)	1 (.)	1 (.)	0 (.)	0 (.)	1 (.)	0 (.)
2014 Manufacture of other organic basic chemicals	5 (.)	53 (.)	8 (.)	10 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2015 Manufacture of fertilizers and nitrogen compounds	4 (.)	32 (.)	7 (.)	3 (.)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	0 (.)
2016 Manufacture of plastic materials	6 (.)	123 (.)	13 (.)	14 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2017 Manufacture of synthetic rubber raw material	2 (.)	21 (.)	2 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2020 Manufacture of pesticides and agriculture chemicals	2 (.)	25 (.)	1 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2030 Manufacture of paints, printing inks and enamels	10 (.)	209 (0.083)	50 (.)	10 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)

icantly high accident frequency in subset R1 of the data and, after taking it into account in the null distribution, the same accident frequency is not significantly high in the whole data set, even though more data are used. This suggests that the significantly high accident frequency is somehow related to the constraints added during the exploration: [{R = all, C=(Specific physical activity, Cause of accident)} + {R = R1, C=(Industry (4 digit), Working process)}]. To further illustrate this relationship, we contrast the above with a scenario in which there are no tile constraints from previous iterations (i.e., when the user has not viewed the previous two tables). In this scenario, the significantly high count in a subset (from Iteration 2) is also significant for the whole data set.

The findings obtained from the above three iteration steps are products of an exploratory data analysis. The benefit over existing approaches is that the analyst is allowed to look at the data and still be able to obtain a statistical guarantee that the observed accident counts are not due to random chance alone.

For an occupational safety analyst, the results of these three iterations in this case study propose that in the manufacture of plastic materials there may have been additional haste in production, leading to relatively many slip, trip and fall-related injuries. Now having this statistical guarantee, the analyst could start looking at other data from the industry (such as production volumes) that could explain the result according to their hypotheses. Finding larger than expected accident counts is an ubiquitous problem across safety research, for which our approach provides a practical solution.

4. Discussion

Responsive methods for accident statistics analyses have traditionally been used in safety management (Goel et al., 2017) and a selection of predictive methods have been introduced to supplement these. Predictive models developed in recent years are able to predict, for example, the number and severity of accidents at work, but silent signals that can anticipate safety situations are still poorly recognized by the commonly used analysis methods. We see that more attention should be paid to identifying silent signals and modern analytics tools in order to succeed in accident prevention. By identifying information sources that anticipate critical safety incidents and utilizing data mining, data collection and analysis can focus on relevant issues and be more cost-effective.

In practical working life, occupational accidents are often approached through uni- and bi-variate distribution analyses that show the distribution of incident characteristics in absolute numbers or percentages. In more sophisticated use, incident concentration analyses try to identify clusters of incidents with common characteristics utilizing variables similar to ours to prioritize safety measures (Kjellén and Albrechtson, 2017). Our analysis approach utilizes similar data, with the purpose of identifying silent signals from the data set of occupational accidents in Finland.

The 'traditional' way to conduct a scientific study on accident statistics data has been to form a hypothesis and then use statistical testing methods to see if the hypothesis is true (e.g., some frequencies are high). The methods presented in this article enable us to draw more fine-tuned conclusions and also perform the analysis iteratively, as the approach we present allows creating hypotheses during the analysis based on viewing contingency tables created from the data. This method would be useful for detecting 'silent signals' for informed decision making, for example, even if they concern only small portions of the data (e.g., one branch, city, company).

Previous accident analysis models suffer from the fact that it is not always obvious if the found patterns are valid in a statistical sense. The methodology presented in this paper provides a straightforward, understandable, yet powerful framework to find hidden signals and weed out random artefacts. In the examples of this paper we used raw data sets provided by TVK and only had the human expert's knowledge and intuition at hand. In principle it would have been possible to use other variables in this context (such as a person's income level, health status, etc.). However, this would have required combining different databases. As an example, Pietilä et al. (2018) similarly combined two different databases; an accident statistics database of one accident insurance company and an employee health database of an occupational health care provider.

New approaches to data analysis are needed when human capacity is not sufficient to analyze available data efficiently and reliably. Occupational safety management is facing such a challenge when it comes to utilizing fragmented information as well as large materials; this creates its own challenges for information management. In information management, information can be divided into explicit and indirect information. The collection and use of this indirect or tacit information can be of significant benefit in the prevention of accidents at work (Podgorski, 2010). Data-driven safety management, which takes advantage of more than just accident data, enables continuous improvement (Wang et al., 2018). This is what many employers strive for, as reducing accident rates with traditional analytics and data is limited.

In principle, it would be possible to use an AI method, for example, to suggest views of the data and our method to independently assess the statistical validity of the results, or augment the data set by attributes (e.g., risk indices) estimated by supervised learning models. In this article, the focus was on the 15 variables used in the TVK data. However, the TVK data also contained small verbal descriptions of every accident. This part of the data we excluded, as our focus was on statistical testing. Combining these two parts of data would be an inspiring new approach for a future study on this topic. As we have learned from the studies by, for instance, Jocelyn et al. (2016), Nanda et al. (2016), Valmuur et al. (2016) and Marucci-Wellman et al. (2017), machine learning has been successfully tested in analyzing accident descriptions. We believe that such an approach, going into the verbal data in depth, should be studied further.

5. Conclusions

The presented method is generic and can, in principle, be used to explore any data set from which one can compute contingency tables and for which contingency tables are an informative 'visualization.' Even though the examples in this paper are from quantitative measurements, there is no reason why the same approach could not be applied to qualitative data, from questionnaires, for instance. Furthermore, machine learning algorithms such as classifiers are often used to find relations in the data and to estimate unobserved variables. For example, in the case of this work accident data set we could try to estimate some of the properties using a classifier. The prediction given by a classifier could be added as a new variable to the data.

Large data sets contain a great deal of potentially useful and valuable information. Often, there is no one great question that is clear in advance; finding the useful parts requires first exploring the data. After we see something, it is then important to have some confidence in the fact that the observed patterns – in our case accident frequencies – are 'real' and not just random artefacts.

In this paper, we have proposed a method to do this on a publicly available occupational accident database. Our approach is based on iterative exploration of confidence tables. Although the underlying mathematics and algorithms require some understanding, the outcomes are easily understandable, namely contingency tables and knowledge, if any of the contingency table elements are larger than they would be expected to be by chance.

Future studies could focus on studying combined material in larger samples as they introduce an interesting possibility to gain more in-depth information. Analysis of large-scale data sets with richer information about the employer, workplace and organizational practices could provide more insight into their effects on occupational accidents.

Acknowledgements

We thank Janne Sysi-Aho and TVK for their help and insightful discussions.

This study was supported by the Academy of Finland (decisions 326280 and 326339).

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsr.2022.04.003.

References

Agresti, A. (2019). An introduction to categorical data analysis. *Wiley series in probability and statistics*. Hoboken, NJ: John Wiley & Sons.

Badri, A., Boudreau-Trudel, B., & Souissi, A. S. (2018). Occupational health and safety in the industry 4.0 era: A cause for major concern? *Safety Science, 109*, 403–411.

Cacha, C. A. (1997). *Research design and statistics for the safety and health professional*. New York, NY: Van Nostrand Reinhold.

Ciarapica, F. E., & Giacchetta, G. (2009). Classification and prediction of occupational injury risk using soft computing techniques: An Italian study. *Safety Science, 47* (1), 36–49.

Cruz Rios, F., Chong, W. K., & Grau, D. (2017). The need for detailed gender-specific occupational safety analysis. *Journal of Safety Research, 62*, 53–62.

Dudoit, S., Shaffer, J. P., & Boldrick, J. C. (2003). Multiple hypothesis testing in microarray experiments. *Statistical Science, 18*(1), 71–103.

Foster, D. P., & Stine, R. A. (2008). α-investing: A procedure for sequential control of expected false discoveries. *J Royal Statistical Soc B., 70*(2), 429–444.

Goel, P., Datta, A., & Mannan, S. (2017). Application of big data analytics in process safety and risk management. *2017 IEEE international conference on big data*.

Holland, B. S., & Copenhaver, M. D. (1988). Improved Bonferroni-type multiple testing procedures. *Psychological Bulletin, 104*(1), 145–149.

Hovden, J., Albrechtsen, E., & Herrera, I. A. (2010). Is there a need for new theories, models and approaches to occupational accident prevention? *Safety Science, 48* (8), 950–956.

Jacinto, C., & Guedes Soares, C. (2008). The added value of the new ESAW/Eurostat variables in accident analysis in the mining and quarrying industry. *Journal of Safety Research, 39*, 631–644.

Jocelyn, S., Chinniah, Y., & Ouali, M.-S. (2016). Contribution of dynamic experience feedback to the quantitative estimation of risks for preventing accidents: A proposed methodology for machinery safety. *Safety Science, 88*, 64–75.

Kjellén, U., & Albrechtson, E. (2017). *Prevention of accidents and unwanted occurrences: Theory, methods, and tools in safety management* (Second Edition). Boca Raton, FL: CRC Press.

Manuele, F.A. (2011). Accident costs. Rethinking ratios of indirect to direct costs. Professional Safety, 39-47.

Marucci-Wellman, H. R., Corns, H. L., & Lehto, M. R. (2017). Classifying injury narratives of large administrative databases for surveillance - A practical approach combining machine learning ensembles and human review. *Accident Analysis and Prevention, 98*, 359–371.

Nanda, G., Grattan, K. M., Chu, M. T., Davis, L. K., & Lehto, M. R. (2016). Bayesian decision support for coding occupational injury data. *Journal of Safety Research, 57*, 71–82.

Papazoglou, I. A., Aneziris, O., Bellamy, L., Ale, B. J., & Oh, J. I. (2015). Uncertainty assessment in the quantification of risk rates of occupational accidents. *Risk Analysis: an official publication in the Society of Risk Analysis, 35*(8), 1536–1561.

Pietilä, J., Räsänen, T., Reiman, A., Ratilainen, H., & Helander, E. (2018). Characteristics and determinants of recurrent occupational accidents. *Safety Science, 108*, 269–277.

Podgorski, D. (2010). The use of tacit knowledge in occupational safety and health management systems. International Journal of Occupational Safety and Ergonomics (JOSE) 2010, 16(3), 283–310.

Puolamäki, K., Oikarinen, E., & Henelius, A. (2021). Guided visual exploration of relations in data sets. *Journal of Machine Learning Research, 22*, 1–32.

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rissanen, M., & Kaseva, E. (2014). *Cost of lost labour input [e-publication]*. Finland: Ministry of Social Affairs and Health, Department for Occupational Safety and Health. Access method: http://stm.fi/documents/1271139/1332445/Cost+of+lost+labour+input_en.pdf/d5790088-8e3e-4d13-a5cd-56c23b67de0c.

Ross, A. J., Davies, J. B., & Plunkett, M. (2005). Reliable qualitative data for safety and risk management. *Process Safety and Environmental Protection, 83*(2), 117–121.

Savvides, R., Henelius, A., Oikarinen, E., & Puolamäki, K. (2019). Significance of patterns in data visualisations. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 1509–1517).

Valmuur, K., Marucci-Wellman, H. R., Taylor, J. A., Lehto, M. R., Corns, H. L., & Smith, G. S. (2016). Harnessing information from injury narratives in the 'big data' era: Understanding and applying machine learning for injury surveillance. *Injury Prevention, 22*(Suppl 1), i34–i42.

Wang, B., Wu, C., Huang, L., & Kang, L. (2018). Using data-driven safety decision-making to realize smart safety management in the era of big data: A theoretical perspective on basic questions and their answers. *Journal of Cleaner Production, 210*(2019), 1595–1604.

Westfall, P. H., & Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-Value adjustment*. John Wiley & Sons, Mathematics. 360 pages.

Workers' Compensation Center [Tapaturmavakuutuskeskus] (2019). Työtapaturmat 2009-2018. Tilastojulkaisu [Occupational accident statistics 2009-2018]. Workers' Compensation Center. Access method: https://indd.adobe.com/view/baa94c89-d2b1-4fa3-b10c-f421c41208a4.

Wu, D., & Li, Z. (2019). Work safety success theory based on dynamic safety entropy model. *Safety Science, 113*, 438–444.

Tuula Räsänen is a researcher at the Finnish Institute of Occupational Health. She has over 30 years' experience of occupational safety research. She completed her PhD in 2007 at Tampere University of Technology about the management of occupational safety and health information in Finnish production companies.

Arto Reiman is a research team leader at the University of Oulu in Finland. His research interests include ergonomics & human factors and occupational safety, and how they can be included in the design and development processes to improve well-being and productivity at work.

Kai Puolamäki is Associate Professor of computer science and atmospheric sciences in the Department of Computer Science and Institute for Atmospheric and Earth System Research (INAR) at the University of Helsinki. He completed his PhD in 2001 in theoretical physics. His primary interests lie in the areas of exploratory data analysis, machine learning, and related algorithms. He has a website at http://www.iki.fi/kaip/.

Rafael Savvides is a doctoral student at the Exploratory Data Analysis group at the University of Helsinki, Finland. His research interests include visual and interactive data exploration.

Emilia Oikarinen is university lecturer at Department of Computer Science at University of Helsinki, Finland. Her research interests broadly span artificial intelligence research, ranging from knowledge representation, reasoning, and optimization to explorative data analysis with applications in a wide variety of domains.

Eero Lantto M.Soc.Sci, is a researcher at the Finnish Institute of Occupational Health, working in the Occupational Safety unit. Before FIOH Eero gained experience at Eurofound on work and employment related research topics. During the two and a half years Eero has spent at FIOH he has participated in various research and development projects in different industries.

Journal of Safety Research 82 (2022) 323–328

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Go-around accidents and general aviation safety

Alex de Voogt[a,*], Hilary Kalagher[a], Brianna Santiago[a], Jonas W.B. Lang[b]

[a] Drew University, United States
[b] Business School, University of Exeter, United Kingdom

ARTICLE INFO

Article history:
Received 24 August 2021
Received in revised form 21 January 2022
Accepted 13 June 2022
Available online 22 June 2022

Keywords:
General Aviation
Fixed-wing aircraft
Landing
Go-around
Aviation training

ABSTRACT

Introduction: Changes in General Aviation (GA) accident rates, specifically in the go-around phase, are examined by comparing the number of accidents, the proportion of fatal accidents, and the proportion of certain causes of accidents over time. *Methods:* Two sets of accidents from 2000 to 2004 and from 2013 to 2017 were extracted from the National Transportation Safety Board (NTSB) online database. *Results:* Although the total number of GA accidents per landing significantly decreased over time, the proportion of fatal accidents in the go-around phase increased. Fatalities most often occurred in instrument meteorological conditions. *Conclusion:* Advances in technology and training show improvements in GA accident rates, but not for accidents in the go-around phase. *Practical Applications:* Scenario-based learning is recommended to include specific instruction concerning the timing of go-around procedures in unstable flights.

1. Introduction to General Aviation safety

General Aviation (GA) is a part of civil aviation that has exhibited the highest accident and fatality rates (Boyd, 2017; Li & Baker, 1999). This safety concern has persisted for decades and several attempts have been made to understand and improve its safety record. Understanding GA is complex as its characteristics are also highly diverse.

Accident analysis has targeted specific aspects of GA to understand its safety record. For instance, it has shown different characteristics per type of aircraft, such as sports aircraft (de Voogt & van Doorn, 2010) and helicopters (Taneja & Wiegmann, 2003); different types of operations, such as emergency medical services (Baker et al., 2006), aerial application (van Doorn, 2014; de Voogt, Uitdewilligen, & Eremenko, 2009) and instruction (Olson & Austin, 2006; Baker et al., 1996). Even within the same flight aviation rules (FAR), such as those for GA or air taxi and commuter aircraft, and only focusing on fixed-wing airplanes, research has pointed out geographical differences that greatly affect the safety of aviation operations (Thomas et al. 2000; Grabowski, Curriero, & Baker, 2002). Finally, the phase of flight is considered especially important with several studies focusing on the landing phase (e.g., Benbassat & Abramson, 2002; Benbassat, Williams, & Abramson, 2005) to allow for specific safety recommendations.

1.1. The go-around

A go-around in aviation is an aborted landing commonly instigated by a dangerous situation on the runway or an unstable approach. The procedure to move from a landing to a take-off configuration is also practiced by student pilots and generally considered a challenging emergency maneuver (Baker et al., 1996; Dehais et al., 2017; Uitdewilligen & de Voogt, 2009). Since the maneuver is meant to avoid a possible dangerous landing, accidents during the go-around phase are particularly unfortunate and suggest a lack of experience or skill on the part of the pilot.

Accidents during a go-around maneuver have been studied in cases where loss of situational awareness and possible improvements using enhanced vision systems have been suggested (Kramer, Bailey & Prinzel, 2009). In addition, the role of economic pressures (i.e., the cost of a go-around for an air carrier) on risk taking was studied experimentally (Causse et al., 2013). Both enhanced vision systems and economic pressures are less relevant for General Aviation operations; however, the accident frequency in this segment of aviation is particularly high (Boyd, 2017). Of all flight phases, landing accidents are reported as the most frequent, and landings with a high-airspeed are especially dangerous (Boyd, 2019). This suggests that go-around maneuvers, although challenging, may assuage the landing accident rate if performed correctly.

* Corresponding author.
E-mail address: adevoogt@drew.edu (A. de Voogt).

https://doi.org/10.1016/j.jsr.2022.06.008
0022-4375/

1.2. Fits

Apart from increasing insight in the characteristics of accidents, several changes have taken place in GA that may positively affect the accident and fatality rates. They range from the introduction of airplanes with emergency parachutes (Alaziz, Stolfi, & Olson, 2017) to technological advances in the field of navigation, in particular GPS. Parallel to these developments, the Federal Aviation Administration (FAA) supported initiatives to improve training curricula, known as the FAA Industry Training Standards or FITS (Craig, 2009; Summers et al., 2007) that should improve a pilot's ability to manage risk.

In 2007, the FAA started to develop FAA Industry Training Standards (FITS) for a generic commercial pilot syllabus (Craig, 2009). It used a scenario-based methodology that should improve a pilot's ability to manage risk in scenarios such as a go-around (Summers et al., 2007). A recent study on go-arounds in commercial aviation suggests that decision-making, in particular the timing of the decision, is essential and that protocols for go-around decisions during unstable flights are frequently ignored and may explain accidents in this flight phase (Blajev & Curtis, 2017). It is, however, not clear if FITS address these aspects effectively to reduce the accident rate in this flight phase.

The FITS program concentrates on scenario-based training, single pilot resource management, and learner-centered grading. While in this model flight maneuvers are still a central part of flight training, the use of real-world scenarios is used to enhance the pilot's decision-making skills. The elements of single pilot resource management have direct or indirect relevance for landing and go-around procedures since they emphasize, for instance, situational awareness, risk management, and task management (Summers et al., 2007).

An overall reduction in GA accident and/or fatality rates is difficult to determine due to the diversity of GA operations but is, on the other hand, expected in light of the developments in technology and training, as well as an increased awareness and understanding of GA accidents. In this study, we selected two sets of accidents from two different time periods for comparison. These time periods precede and follow the introduction of FITS and span an era in aviation where, for instance, GPS technology has become particularly common in all of General Aviation. We limited our data to fixed-wing GA aircraft only and concentrated on one particular flight phase, the go-around. Pilots performing a go-around are likely to benefit from the advances in technology and training. Research on go-arounds has mainly proceeded in simulators and for pilots of airliners, so that we have a reasonable understanding of the expected main causes, but not whether this insight and its possible remediation has reached GA pilots.

In this study, we expect to see a positive impact on safety in General Aviation both in the number of go-around accidents, the proportion of fatal go-around accidents, and the proportion of certain causes of go-around accidents over time as they may point to significant shifts in pilot practices. The results of this study may provide a better understanding of GA go-around accidents and also serve as a possible proxy for developments in GA safety more broadly.

We analyzed the causes and factors of 187 General Aviation go-around accidents from 2000 to the end of 2004 and compared these with 117 accidents from 2013 to the end of 2017. In both data sets the fixed-wing airplane was in a go-around phase when the accident occurred. The results may indicate whether the nature of go-around accidents in the United States has changed since the introduction of FITS, technological advances, and increasing insight in the GA safety record.

2. Method

An aviation accident is defined by the National Transportation Safety Board (NTSB) as an occasion in which the aircraft was substantially damaged or destroyed, and/or, in which occupants or people on the ground were seriously injured or died as a result of the occurrence. Accidents resulting in minor injuries and only minor damage are reported as incidents and administrated by the FAA.

All NTSB accident reports are made available online and may be accessed using the NTSB Aviation Online database using the CAROL (Case Analysis and Reporting OnLine) search query tool. Each accident has a factual report and a probable cause report that summarizes the findings of the NTSB investigator with a narrative statement, a set of findings that determines the cause and contributing factors of the incident, as well as data on the pilot, aircraft, airfield, and meteorological conditions.

United States General Aviation fixed-wing airplane accidents that took place during the go-around flight phase were extracted from the NTSB online database for the period 2000 until the end of 2004 and for 2013 to the end of 2017 (NTSB, 2020). These periods were selected to allow for changes to become visible as a result of the introduction of the FITS program, as well as technological changes in aviation. Accidents were identified using the "broad phase of flight" search tab in the database. Two cases from the first time period showed a different flight phase and were removed from the dataset. The narrative text of each accident was used to determine the reported reason for starting a go-around, the number of go-arounds attempted and, as far as possible, when the flight became unstable and when the decision to go-around was made.

The FAA (2020) provides denominator data of different kinds, including number of aircraft, number of flight hours, and number of landings. The number of landings is most relevant as denominator data for our dataset. It is noted, however, that landings are only counted for towered airports, while GA flight are often found at non-towered airports. Unfortunately, the FAA does not differentiate between General Aviation and Air Taxi landings, while in the latter two years of our dataset Commuter flights are also included in the number of landings (i.e., Flight Aviation Regulations (FAR) under Part 135 Commuter and Air Taxi). Although this still provides a reasonable comparison, some caution in the interpretation of these data is warranted.

In addition to comparing the number of landings per year, the FAA also allows for a differentiation between the number and type of engines of the aircraft. In our dataset, most airplanes had one reciprocating (piston) engine, so it is useful to provide this detail in the denominator data in case it fluctuates differently compared to other types of engines.

We used Pearson χ^2-square analysis at the significance level of 0.05 to determine the significance of relations within the datasets. In analyses in which the expected cell frequencies were less than five, a Fisher exact test was used. A logistic regression using fatal versus nonfatal as categorical outcome was used with the categorical predictors found using Pearson χ^2-square analysis to determine relative risk ratios. Unlike the proportion-testing, the logistic regression adjusted for the contributions of the other variables. A Poisson regression analysis was used to predict the number of fatal go-around accidents based on time period (early vs. later) with the natural log of the fixed-wing landings as an offset.

2.1. Risk analysis

A risk analysis of go-around accidents is part of a broader analysis of risk in General Aviation. The number of accidents for fixed-wing aircraft, the number of accidents in the landing phase, and

Journal of Safety Research 82 (2022) 323–328

the number of accidents in the go-around phase each have different characteristics. As is shown in Fig. 1 and Table 1, the percentage of fatal accidents fluctuates significantly between these three groups, and accidents in the landing phase have by far the smallest proportion of fatalities.

Fig. 1 and Table 1 show that between the two periods of study there is a drop in the number of landings. This drop is mirrored with a drop in total number of accidents. There is also a change in the percentage of fatal accidents in the second dataset, which is lower for the total number of accidents and for landing accidents, but is higher for go-around accidents.

A Poisson regression analysis was used to predict the total number of go-around fatal accidents based on time period (early vs. later) and the total General Aviation fixed-wing landings during the same time period, with the natural log of the fixed-wing landings as an offset. The analysis revealed some evidence that fatal go-around accidents were 0.915 (95% CI 0 to 1.83) times more likely to occur in the later period compared to the early period, $p = 0.05$. However, the 95% CI included zero so we cannot make an inference as to whether the rate was higher or lower for both periods. The other variable failed to reach levels of statistical significance in the analysis.

3. Results

There is a strong relation between aircraft damage and fatality in both time periods (Table 2). This relation is not unexpected, although previous studies have shown some exceptions to this seemingly obvious relation (de Voogt & van Doorn, 2006a; de Voogt, Hummel Hohl, & Kalagher, 2021).

Go-around fatal accidents were more prevalent at night in both time periods. Similarly, go-around fatal accidents were more prevalent in instrument meteorological conditions (IMC) rather than visual meteorological conditions (VMC) conditions (Table 3).

Aircraft with two engines had a significantly higher proportion of fatalities than those with one engine, while turbine engine aircraft did not report any fatalities in the two periods under study. The ratio of fatal accidents among amateur built aircraft was not significantly different from other aircraft in the dataset. (See Table 4).

In the first period, most cases reported the purpose of the flight as personal or instructional, and a Fisher exact test revealed that the proportion of fatal instructional flights was significantly lower than the proportion of fatal accidents for all others combined. However, in the second period, this was not significantly lower. Business flights, flight tests, aerial observation, ferry flights, positioning flights, and other purposes reported fewer than five fatal accidents and five or fewer nonfatal accidents in either time period.

Flight hours are not always reported in the NTSB accident reports. There were four accidents with missing data on pilot flight hours in the first time period and two in the second period, all of which were nonfatal accidents. Similarly, age was not reported in two cases, both in the first time period.

Pilot age ranged from 19 to 88 years old in the first period and from 17 to 79 years in the second period. Pilots in the United States have faced mandatory retirement at age 60, a rule that has been controversial (AMA, 2004) and may be better addressed using flight hours. There was a significantly higher proportion of over 60 pilots in the first period but not in the second, while those with more than 500 flight hours made up a significantly higher proportion of fatal accidents in both time periods (see Table 5).

From the findings in the NTSB reports, it was determined how many accidents were attributed to the pilot in command, to students, or others. Additional factors for fatal accidents included spatial disorientation, of which three occurred at night in the first time period and four were at night in the second period, and all except one in the second period occurred during IMC conditions. One night-time IMC fatal accident in the second period also reported a "somatogravic illusion," possibly exacerbated by the pilot's consumption of antihistamine.

In the first period, a total of 42 accidents occurred after more than one go-around, and 19 of these were part of go-around prac-

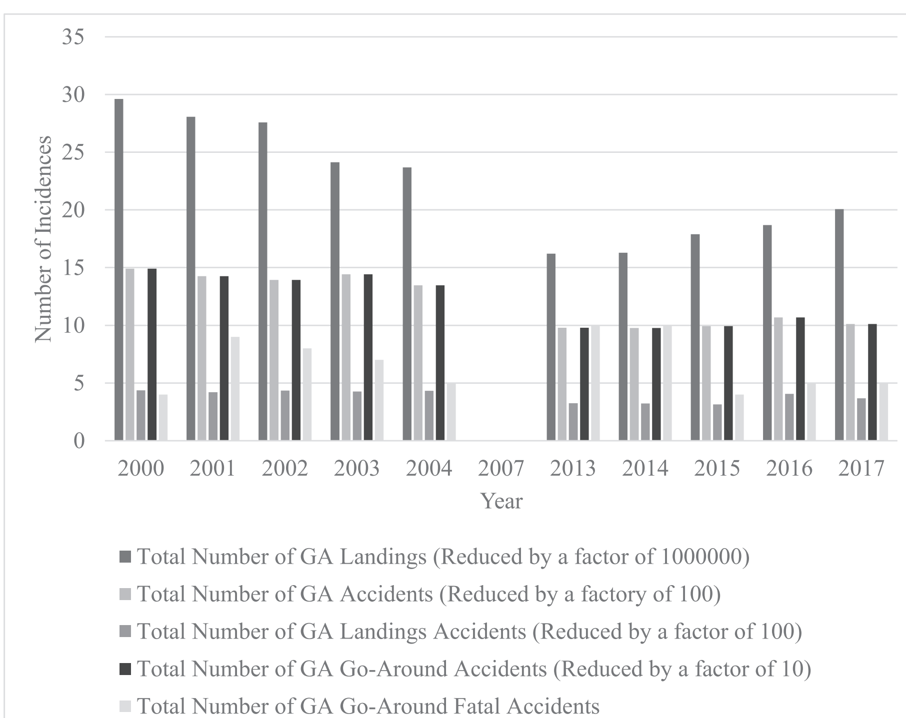


Fig. 1. Trends of GA fixed-wing accidents in landing and go-around phase.

Table 1
Overview of number of landings and accidents per year for fixed-wing aircraft in the United States.

	Total fixed-wing landings [SD]	**All go-around** accidents	**All fatal go-around** accidents	**All landing** accidents	**All** GA accidents
2000	37,914,142 [3.6]	38	4	436	1490
2001	35,011,549 [5.7]	36	9	420	1426
2002	36,321,419 [5.2]	41	8	434	1394
2003	31,959,886 [3.7]	38	7	426	1443
2004	32,171,301 [1.6]	34	5	433	1347
Total 2000–2004 (fatal)	173,378,297	187 (33/17.6%)	33 (17.6%)	2149 (41/1.9%)	7100 (1441/20.3%)
2013	24,239,819 [1.6]	29	10	324	979
2014	24,092,551 [1.6]	26	10	323	977
2015	25,884,484 [1.7]	17	4	315	993
2016 GA & Part 135	27,243,225 [1.7]	18	5	406	1068
2017 GA & Part 135	28,294,609 [1.6]	29	5	367	1012
Total 2013–2017 (fatal)	129,754,688	119 (34/28.6%)	34 (29.1%)	1735 (20/1.2%)	5029 (905/18.0%)

Table 2
Relation of fatal accidents and aircraft damage.

Fatality\damage	Destroyed aircraft	Substantially damaged	Minor damage
Fatal accidents 2000–2004	28	5	0
Nonfatal 2000–2004	9	144	1
Fatal accidents 2013–2017	21	13	0
Nonfatal 2013–2017	2	81	0

Table 3
Proportion of fatal accidents in IMC and night conditions.

Environment	2000–2004		2013–2017	
VMC (fatal)	166 (16)		102 (20)	
IMC (fatal)	21 (17)	$\chi^2 = 65.2351$, $p < 0.01$	17 (14)	$\chi^2 = 27.4026$, $p < 0.01$
Day/dusk (fatal)	157 (23)		120 (25)	
Night (fatal)	18 (10)	$\chi^2 = 19.3555$, $p < 0.01$	12 (9)	$\chi^2 = 13.6892$, $p < 0.01$

Table 4
Proportion of fatal accidents for twin-engine, turbine and amateur-built aircraft.

	2000–2004		2013–2017	
Twin engine (fatal)	25 (9)	$\chi^2 = 6.6885$, $p < 0.01$	16 (9)	$\chi^2 = 6.647$, $p < 0.01$
Turbine/turbo prop engine (fatal)	5 (0)		6 (0)	
Amateur built (fatal)	12 (3)		14 (3)	

Table 5
Purpose of flight and pilot characteristics.

	2000–2004		2013–2017	
Personal flights (fatal)	123 (22)		79 (26)	
Instructional flights (fatal)	45 (3)	$p < 0.05$	29 (5)	$p > 0.05$
Pilot age < 60 (fatal)	141 (17)		75 (18)	
Pilot age ≥ 60 (fatal)	43 (16)	$\chi^2 = 14.1638$, $p < 0.01$	42 (16)	$p > 0.05$
Total flight hours < 500 (fatal)	74 (6)		52 (7)	
Total flight hours ≥ 500 (fatal)	109 (27)	$\chi^2 = 9.1247$, $p < 0.01$	63 (27)	$\chi^2 = 5.8219$, $p < 0.02$

tice. In the second period, a total of 33 accidents occurred after more than one go-around and 8 of these were part of go-around practice. In both periods, the cases with multiple go-arounds that were not part of flying practice had a significantly higher proportion of fatalities than the remainder of flights that include practice flights and flights where the first go-around led to an accident (see Table 6).

In the first period (39 cases, 16 fatal), it could not be established from the narrative statement when the decision to go-around was made. Although nonfatal accidents had at least 29 cases in which the decision was made during touchdown and 19 cases during the landing flare, with 2 and 1 cases, respectively, for fatal accidents, this difference was not significant ($p > 0.05$).

However, in the second period, with 33 cases of which 10 fatal, it could be determined that the timing of the decision proved significant. Nonfatal accidents had at least 20 cases in which the decision was made during touchdown and 10 cases during the landing flare, with 4 and 2 cases, respectively, for fatal accidents. This showed a significantly smaller proportion of fatal accidents occurring during the touchdown and flare than in other phases of the approach ($\chi^2 = 3.8742$, $p < 0.05$).

A logistic regression was used to test possible interactions and to determine which categories significantly predicted fatality (Table 7). Accidents in IMC conditions and at night, twin and turbine engine aircraft, pilot experience, and the presence of multiple non-practice go-arounds became part of Model 1, which controlled for time period. IMC conditions, flight hours above 500 hours, and the presence of a turbine engine remained significant predictors of

Table 6
Circumstances for a go-around and cause attribution.

	2000–2004		2013–2017	
Danger on the runway (fatal)	10 (2)		11 (4)	
Go-around practice (fatal)	28 (2)		18 (2)	
Bounced landing (fatal)	8 (0)		14 (1)	
Missed approach (fatal)	64 (14)		35 (8)	
Loss-of-control (fatal)	40 (10)		79 (26)	
Weather (fatal)	26 (3)		25 (7)	
Spatial disorientation (fatal)	14 (7)		8 (8)	
Not maintaining airspeed (fatal)	56 (7)		20 (7)	
Pilot-in-command (fatal)	156 (30)		86 (7)	
Other (undetermined)	11 (7)		(8)	
More than one non-practice go-around (fatal)	23 (9)	$\chi^2 = 8.3287$, $p < 0.01$	14 (9)	$\chi^2 = 9.5723$, $p < 0.01$

fatality in this model. Flights into IMC remained a significant predictor of fatality in Model 2. There were no interactions with the time period that were significant.

4. Conclusion

While it is possible to glean trends for General Aviation safety using a risk analysis, it also illustrates that at a more granular level, in this case the specific phase of flight, the numbers may show a significantly different pattern. It confirms the need in GA accident analysis to examine specific datasets that may help to understand how safety is changing over time.

General Aviation has the highest number of accidents, for which about 20% are reported fatal (Boyd, 2017). This proportion of fatal accidents is slightly lower for those in the go-around flight phase in the earlier period we studied, but significantly higher in the later one. Although the total number of accidents concerning go-arounds in general aviation have declined between the two periods, remarkably few differences are found in the characteristics of the flight accidents involving go-arounds apart from the proportion of fatal accidents that increased. This finding is especially disappointing in light of efforts by the FAA to improve pilot training.

Go-around maneuvers are practiced regularly but result in relatively few fatal accidents for instructional flights, which is in line with other studies on student flights (Uitdewilligen & de Voogt, 2009). The more recent period also showed a significantly smaller proportion of fatal accidents occurring during the touchdown and flare. The proximity to the ground and the associated lower incidence of a fatality has also been attested in other studies (de Voogt & van Doorn, 2006b). There are also a few cases in which dangers on the runway create circumstances where a go-around is not successful. As expected, it is the pilot-in-command (rather than the student or the circumstances on the ground) who is attributed to the cause of an accident, especially a fatal accident, in the go-around flight phase.

Significant correlations between fatality and IMC as well as fatality and twin-engine aircraft are reported, but they are not necessarily specific for go-arounds as previous research indicates (e.g., Boyd, 2015, 2017). The increased complexity of twin-engine airplanes and their higher landing speeds partly explains why these

Table 7
Results of two models of logistic regression.

Statistical models	Model 1		Model 2	
	Estimate (Standard error)	Odds ratio [Confidence interval]	Estimate (Standard error)	Odds ratio [Confidence interval]
(Intercept)	−4.86* (1.99)	0.01 [−3.89; 3.90]	−3.90 (2.69)	0.02 [−5.24; 5.28]
Attempt	0.54 (0.47)	1.72 [0.80; 2.65]	0.75 (0.66)	2.13 [0.83; 3.43]
Twin engine	0.12 (0.50)	1.13 [0.15; 2.11]	0.14 (0.69)	1.15 [−0.21; 2.50]
Turbine engine	−1.52* (0.76)	0.22 [−1.27; 1.71]	−2.02 (1.06)	0.13 [−1.94; 2.21]
IMC conditions	2.72*** (0.53)	15.23 [14.19; 16.27]	3.23*** (0.77)	25.40 [23.90; 26.91]
Night conditions	0.74 (0.53)	2.09 [1.04; 3.14]	0.29 (0.78)	1.33 [−0.20; 2.86]
Flight hrs > 500	0.87* (0.39)	2.38 [1.61; 3.15]	0.67 (0.57)	1.95 [0.83; 3.07]
Time period	0.82* (0.35)	2.26 [1.58; 2.95]	−0.94 (3.95)	0.39 [−7.35; 8.13]
Period * Attempt			−0.56 (0.98)	0.57 [−1.34; 2.48]
Period * Twin engine			−0.09 (1.03)	0.91 [−1.11; 2.94]
Period * Turbine engine			1.05 (1.51)	2.87 [−0.09; 5.82]
Period * IMC conditions			−1.04 (1.08)	0.35 [−1.76; 2.47]
Period * Night conditions			0.93 (1.13)	2.53 [0.31; 4.75]
Period * Flight hrs > 500			0.37 (0.79)	1.44 [−0.11; 3.00]
Log Likelihood	−112.36		−111.23	
Deviance	224.73		222.45	
Number of observations	296		296	

*** $p < 0.001$.
* $p < 0.05$.

aircraft are also at higher risk during go-arounds (Boyd, 2019). In addition, most twin-engine planes flew in IMC conditions in this dataset.

While go-arounds are challenging procedures, experience appears inversely related to the presence of a fatal accident. It is noted that total flight experience does not necessarily mean more experience with go-arounds; at most it is more likely. Still this result is counter-intuitive if lack of training is thought to be the primary underlying cause. According to previous studies, it is not necessarily the go-around itself, but the timing that is important (Blajev & Curtis, 2017). In most cases, we were able to determine when the go-around was initiated, but it remained unclear if this was long or shortly after a flight had become unstable. This element of go-arounds is not specifically mentioned in the training protocols initiated by the FAA (i.e., FITS), but if implemented may improve the overall effect of this initiative.

Both experienced and inexperienced pilots require practice of go-around maneuvers with a focus on the timing of the go-around decision. In the case of IMC and twin-engine aircraft this practice needs to be extended to multiple different circumstances. Scenarios as taught in FITS (Summers et al., 2007) should include situations in which spatial disorientation is actively addressed, perhaps first in a simulated environment, but ultimately in an environment where the movement of and forces on the aircraft and pilot during a go-around are experienced as well. Importantly, the problem of spatial disorientation was more often reported for fatal accidents, compared to nonfatal, in both studies. This also translates to air carriers where spatial orientation has been reported as a primary concern in go-around mishaps (Dehais et al., 2017; Kramer et al., 2009). Scenario-based learning, as supported by FITS, is an important first step to achieve increased safety for both experienced and inexperience pilots.

In summary, the investments made in training curricula as well as a better understanding of problems with go-arounds in the literature have not yet shown the desired results in the accident statistics. If go-arounds are used as a proxy for the progress in GA aviation safety, the changes in both the number and the proportion of fatal accidents leaves much to be desired. At the same time, it shows that an increase in the proportion of fatalities in one specific phase of flight is contrasted with that in another, such as the landing phase. The advancements made in GA safety may only have seen their effect in certain types of accidents or phases of flight. Considering the diversity within GA, even if we only observe fixed-wing airplanes, the way forward is more likely a combination of specific and general improvements in training and regulations for which accident analyses continue to provide a guide and monitoring device over time.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Association, A. M., Committee, A. S., & Subcommittee, C. A. S. (2004). The age 60 rule. *Aviation, Space & Environmental Medicine, 75*, 708–715.
Alaziz, M., Stolfi, A., & Olson, D. M. (2017). Cirrus airframe parachute system and odds of a fatal accident in cirrus aircraft crashes. *Aerospace Medicine and Human Performance, 88*(6), 556–564.
Baker, S. P., Grabowski, J. G., Dodd, R. S., Shannahan, D. F., Lamb, M. W., & Li, G. (2006). EMS helicopter crashes: What influences fatal outcomes. *Annals of Emergency Medicine, 47*(4), 351–356.
Baker, S. P., Lamb, M. P. H., Guohua, L. D., & Dodd, R. S. (1996). Crashes of Instructional Flights. *Aviation Space and Environmental Medicine, 67*, 105–110.
Benbassat, D., & Abramson, C. I. (2002). Landing flare accident reports and pilot perception analysis. *International Journal of Aviation Psychology, 12*, 137–152.
Benbassat, D., Williams, K. W., & Abramson, C. I. (2005). General aviation leveloff, roundout, and accident rate analysis. *The International Journal of Aviation Psychology, 15*, 189–203.
Blajev, T., & Curtis, W. (2017). Go-around decision-making and execution project. Final Report to Flight Safety Foundation, March 2017. Retrieved on August 17, 2020 from https://flightsafety.org/wp-content/uploads/2017/03/Go-around-study_final.pdf.
Boyd, D. D. (2015). Causes and risk factors for fatal accidents in non-commercial twin engine piston general aviation aircraft. *Accident Analysis and Prevention, 77*, 113–119.
Boyd, D. D. (2017). A review of General Aviation Safety (1984–2017). *Aerospace Medicine and Human Performance, 88*(7), 657–664.
Boyd, D. D. (2019). Occupant injury severity in general aviation accidents involving excessive landing airspeed. *Aerospace Medicine and Human Performance, 90*(4), 355–361. https://doi.org/10.3357/AMHP.5249.2019.
Causse, M., Péran, P., Dehais, F., Caravasso, C. F., Zeffiro, T., Sabatini, U., & Pastor, J. (2013). Affective decision making under uncertainty during a plausible aviation task: An fMRI study. *NeuroImage, 71*, 19–29.
Craig, P. A. (2009). Evaluating pilots using a scenario-based methodology: A guide for instructors and examiners. *International Journal of Applied Aviation Studies, 9*(2), 155–170.
de Voogt, A., Hummel Hohl, C., & Kalagher, H. (2021). Accidents with helicopters on the ground: Fatality and operational specificity. *Aerospace Medicine and Human Performance, 92*(7), 593–596.
de Voogt, A., Uitdewilligen, S., & Eremenko, N. (2009). Safety in high-risk helicopter operations: The role of added crew in accident prevention. *Safety Science, 47*, 717–721.
de Voogt, A., & van Doorn, R. R. A. (2006a). Balloon crash damage and injuries: An analysis of 86 accidents, 2000–2004. *Aviation, Space & Environmental Medicine, 77*(5), 556–558.
de Voogt, A., & van Doorn, R. R. A. (2006b). Midair collisions in U.S. civil aviation 2000–2004: The roles of radio communication and altitude. *Aviation Space and Environmental Medicine, 77*(12), 1252–1255.
de Voogt, A., & van Doorn, R. R. A. (2010). Sports aviation accidents: Fatality and aircraft specificity. *Aviation Space and Environmental Medicine, 81*, 1033–1036.
Dehais, F., Behrend, J., Peysakhovich, V., Causse, M., & Wickens, C. D. (2017). Pilot flying and pilot monitoring's aircraft state awareness during go-around execution in aviation: A behavioral and eye tracking study. *The International Journal of Aerospace Psychology, 27*(1–2), 15–28.
FAA (Federal Aviation Administration), Extracted October 1, 2020 from https://www.faa.gov/data_research/aviation_data_statistics/general_aviation/.
Grabowski, J. G., Curriero, F. C., & Baker, S. P. (2002). Geographic patterns of pilot fatality rates in commuter and air taxi crashes. *Aviation Space and Environmental Medicine, 73*(1), 104–1020.
Kramer, L. J., Bailey, R. E., & Prinzel, L. J. III, (2009). Commercial flight crew decision making during low-visibility approach operations using fused synthetic and enhanced vision systems. *The International Journal of Aviation Psychology, 19*(2), 131–157.
Li, G., & Baker, S. P. (1999). Correlates of pilot fatality in general aviation crashes. *Aviation Space and Environmental Medicine, 70*, 305–309.
Olson, R., & Austin, J. (2006). Performance-based evaluation of flight student landings: Implications for risk management. *The International Journal of Aviation Psychology, 16*, 97–112.
NTSB (National Transportation Safety Board) Retrieved on August 15, 20202 from https://www.ntsb.gov/_layouts/ntsb.aviation/index.aspx.
Summers, M. M., Ayers, F., Connolly, T., & Robertson, C. (2007). Managing risk through scenario based training, single pilot resource management, and learner centered grading. Retrieved on August 15, 2020 from https://www.faa.gov/training_testing/training/fits/guidance/media/rm_thorugh_sbt.pdf.
Taneja, N., & Wiegmann, D. A. (2003). Analysis of injuries among pilots killed in fatal helicopter accidents. *Aviation Space and Environmental Medicine, 74*(4), 337–341.
Thomas, T. K., Bensyl, D. M., Manwaring, J. C., & Conway, G. A. (2000). Controlled flight into terrain accidents among commuter and air taxi operators in Alaska. *Aviation Space and Environmental Medicine, 71*(11), 1098–1103.
Uitdewilligen, S., & de Voogt, A. (2009). Aircraft accidents with student pilots flying solo: An analysis of 390 cases. *Aviation Space and Environmental Medicine, 80*, 803–806.
van Doorn, R. R. A. (2014). Accidents in agricultural aviation in the United States: A 28-year investigation. *Aviation Psychology and Applied Human Factors, 4*, 33–39.

Alex de Voogt is an Associate Professor at Drew University. He is an Aviation Psychologist (EAAP) and Member of the Royal Aeronautical Society (RAeS) and teaches aviation psychology at Drew University. He has a special interest in aviation safety, mostly concerning General Aviation and helicopters.

Hilary Kalagher is an Associate Professor at Drew University. She is a psychologist with a special interest in early perceptual and cognitive development. She entered the field of aviation research with a study on accidents with children in aviation.

Brianna Santiago is a BA candidate at Drew University with a special interest in aviation safety.

Jonas W. B. Lang is an associate professor in the Department of Personnel Management, Work and Organizational Psychology at Ghent University.

Journal of Safety Research 82 (2022) 469–481

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Special Report from the CDC

Health equity guiding frameworks and indices in injury: A review of the literature ☆

Natalie H. Lennon [a,b,*], Andrea E. Carmichael [a], Judith R. Qualters [a]

[a] Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, Division of Injury Prevention, Atlanta, GA 30341, USA
[b] Oak Ridge Associated Universities (ORAU), Division of Injury Prevention, Centers for Disease Control and Prevention, Atlanta, GA 30341, USA

ARTICLE INFO

Article history:
Received 1 July 2022
Accepted 1 July 2022
Available online 13 July 2022

Keywords:
Injury
Framework
Index
Health equity
Social determinants of health

ABSTRACT

Background: In early 2021, CDC released the CORE Health Equity Strategy, which resolves to integrate a comprehensive health equity approach to the work of the Agency. One priority of the Injury Center's Division of Injury Prevention is to move health equity research in injury forward. The purpose of this research is to perform an initial exploration of health equity guiding frameworks and indices to better understand which of these has been applied to injury research topics. *Methods:* A PubMed and CINAHL search of meta-analysis and systematic review articles was conducted from January 1998 through April 2022. Articles of any type and additional frameworks/indices were also identified from staff knowledge of the literature. Books were also considered, where accessible. The following areas were reviewed for each resource: population addressed, guiding framework/index, other health equity variables, gaps identified, and whether the articles addressed an injury topic. *Findings:* The PubMed/CINAHL search produced 230 articles, and an additional 29 articles and 8 books were added from previous knowledge of the literature, resulting in a total of 267 resources for review. There were 60 frameworks/indices compiled that were relevant to health equity. Out of all the resources, three reported on an injury topic and used the PROGRESS-Plus framework, the WHO Social Determinants of Health Conceptual Framework, and a social-ecological framework. *Conclusions:* This study found there were many frameworks/indices for measuring health equity; however, there were few injury-related meta-analysis and systematic review articles. Some frameworks/indices may be more appropriate than others for measuring health equity in injury topic areas, depending on which social determinants of health (SDOHs) they address. *Practical Applications:* Measuring health equity in injury and other public health research areas can help build a foundation of evidence. Moving forward, injury researchers can consider the frameworks/indices identified through this study in their health equity injury research.

1. Introduction

The Centers for Disease Control and Prevention (CDC) defines health equity as the opportunity for all persons to attain their highest level of health, regardless of social position or other socially determined conditions (CDC NCCDPHP, 2020). Health disparities explain health differences between groups that are related to economic, social, or environmental disadvantage (Department of Health and Human Services, 2020). The metrics for gauging progress towards health equity are disparities in health and disparities in the key determinants of and obstacles to health (Robert Wood Johnson Foundation, 2017). Economic and social obstacles such as poverty, unemployment, and lack of access to health care prohibit underserved communities from achieving optimal health, leading to higher rates of disease, decreased access to treatment, and premature death. These barriers often disproportionally impact people of color, rural and low-income urban communities, individuals with disabilities, and members of the LGBTQ + community (CDC NCCDPHP, 2020).

Injuries, both intentional and unintentional, are a significant public health challenge. In 2020, injuries and violence accounted for almost 280,000 deaths in the United States and represented the leading causes of death among those aged 1–44 (CDC NCIPC,

☆ The Journal of Safety Research has partnered with the Office of the Associate Director for Science, Division of Injury Prevention, National Center for Injury Prevention and Control at the CDC in Atlanta, Georgia, USA, to briefly report on some of the latest findings in the research community. This report is the 71st in a series of "Special Report from the CDC" articles on injury prevention.

* Corresponding author at: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, 4770 Buford Hwy, Atlanta, GA 30341, USA.

E-mail address: nlennon@cdc.gov (N.H. Lennon).

https://doi.org/10.1016/j.jsr.2022.07.001
0022-4375/

2020). Recent publications highlight disparities in rates of injuries by socio-demographic groups, geography, and other factors (Clemens et al., 2021; Ehlman et al., 2022; Daugherty et al., 2021; Moore et al., 2019). For example, an analysis of trends from 1999-2019 in fatal unintentional drowning among persons aged < 29 found continuing racial/ethnic disparities. Non-Hispanic (NH) American Indian/Alaskan Native (AI/AN) and NH Black persons experienced the highest rates (Clemens et al., 2021). An analysis of suicide rates from 2019 to 2020 likewise found demographic and geographic disparities with higher rates in rural counties and among men, NH AI/AN and NH White persons, and specific age groups (Ehlman et al., 2022). A study of traumatic brain injuries demonstrated geographic variability across regions of the United States with more rural states exhibiting higher rates (Daugherty et al., 2021). Moore et al.'s (2019) scoping review identified a significant body of research seeking to understand injuries and disparities and underscored the need to better measure and understand the demographic, economic, social, and environmental factors impacting risk and effective interventions to advance health equity.

In April 2021, Dr. Rochelle P. Walensky, Director of the CDC, declared racism a critical public health threat. In response, CDC released the CORE Health Equity Strategy to ensure the work of every center, division, and program remains committed to addressing health disparities in science, research, and partnerships (CDC OMHHE, 2022). Furthermore, the CORE Health Equity Strategy utilizes a comprehensive health equity approach to guide CDC and the field of public health in advancing efforts towards eliminating health disparities:

- C: Cultivate comprehensive health equity science
- O: Optimize interventions
- R: Reinforce and expand robust partnerships
- E: Enhance capacity and workforce engagement (CDC OMHHE, 2022).

To help move the work of the CORE Health Equity Strategy forward, the Injury Center's Division of Injury Prevention (DIP) aims to broaden understanding of health equity within the Division and the field of public health more broadly, including how health equity can be measured. The purpose of this research is to begin to compile and categorize health equity guiding frameworks and indices and document those that have already been applied to injury research topics. This study focused largely on meta-analysis and systematic review articles due to the high volume of individual studies that incorporate a health equity guiding framework or index. In this study, the term 'injury' is used to refer to both unintentional and intentional injury. The findings of this study can be used to assist researchers in understanding the scope and characteristics of health equity guiding frameworks and indices captured in the literature.

2. Method

The study included meta-analysis and systematic review articles published between January 1, 1998, through April 30, 2022, and also included articles and health equity guiding frameworks and indices sourced from previous knowledge of the literature. A literature search was conducted through PubMed and CINAHL (using the search terms "Health Equity"[Mesh] OR "Social Determinants of Health"[Mesh]). Books were considered, where accessible, and attempts were made to source books from university syllabi. Exclusion criteria were applied to focus the content of articles captured in the review. Articles were excluded if they were clinical-specific, were not written in English, did not have a full text available, were duplicates, or if they did not address SDOHs and/or health equity using a specific framework or measurement. For the purposes of this research, a health equity guiding framework/index was defined as a theoretical construct or standardized measure that assesses key domains and concepts associated with health equity. These key categories included: income, education, geography, gender, race/ethnicity, occupation/employment, housing, food access, energy access, childcare, transportation, environmental conditions/exposure, nativity/country of origin, power/prestige, social capital, health literacy, and access to health care.

Citations for each article and book identified in the searches were exported into an Excel file to organize for reviewing. The full text articles were split for reviewing between two reviewers (NL and AC), and data were abstracted for the following categories, when available: population addressed, health equity guiding framework/index, other notable health equity variables, gaps identified, and whether the articles addressed an injury topic. Each reviewer reviewed a sample of five articles reviewed by the other reviewer, to ensure consistency of coding.

3. Results

The database search produced 227 articles from PubMed and 3 articles from CINAHL, and an additional 29 articles and 8 books were added from previous knowledge of the literature, resulting in a total of 267 resources (articles and books) identified for review (Fig. 1). Seventeen resources were clinical-specific, eight resources were not written in English, six resources did not have full text available, and one resource was a duplicate, and thus 32 resources were excluded from the study. Out of the remaining 235 resources, upon review it was found that 160 resources did not mention a specific health equity framework or index addressing social determinants of health (SDOH) and/or health equity and were excluded from the study. In total, 75 resources were categorized as mentioning a health equity guiding framework or index (Table 1).

There were 60 unique health equity guiding frameworks/indices abstracted from the resources (Table 2). Fourteen resources mentioned more than one health equity guiding framework/index. The most commonly mentioned health equity guiding frameworks/indices were the PROGRESS/PROGRESS-Plus framework (n = 13), the World Health Organization (WHO) SDOH Conceptual Framework (n = 11), and the Social-Ecological Model (n = 9) (Table 2). The most common SDOHs addressed by the health equity guiding frameworks/indices were income, education, race/ethnicity, and occupation/employment.

Three of the identified resources reported on one or more injury topics including suicide/depression (Brown et al., 2017b), violence indicators (Armstead et al., 2021), and injury related to infant mortality (e.g., safe sleep, intimate partner violence, child abuse, and drug use during pregnancy) (Reno & Hyder, 2018) (Table 1, Table 2), and used the PROGRESS-Plus framework, the WHO SDOH Conceptual Framework, and a social ecological framework, respectively. The SDOHs analyzed in the group of three injury resources included residence, crowding, physical infrastructure, ethnicity, occupation, gender, religion, education, socioeconomic position, income, social capital, household structure, marital status, social support, and age (Brown et al., 2017b); determinants relevant to the socioeconomic and political context, socioeconomic community condition, and socioeconomic position (Armstead et al., 2021); and individual, interpersonal, organizational, community, and public policy determinants (Reno & Hyder, 2018) (data not shown). These injury articles addressed Caribbean residents (Brown et al., 2017b), African American infants (Reno & Hyder, 2018), and unspecified populations (Armstead et al., 2021) (Table 2).

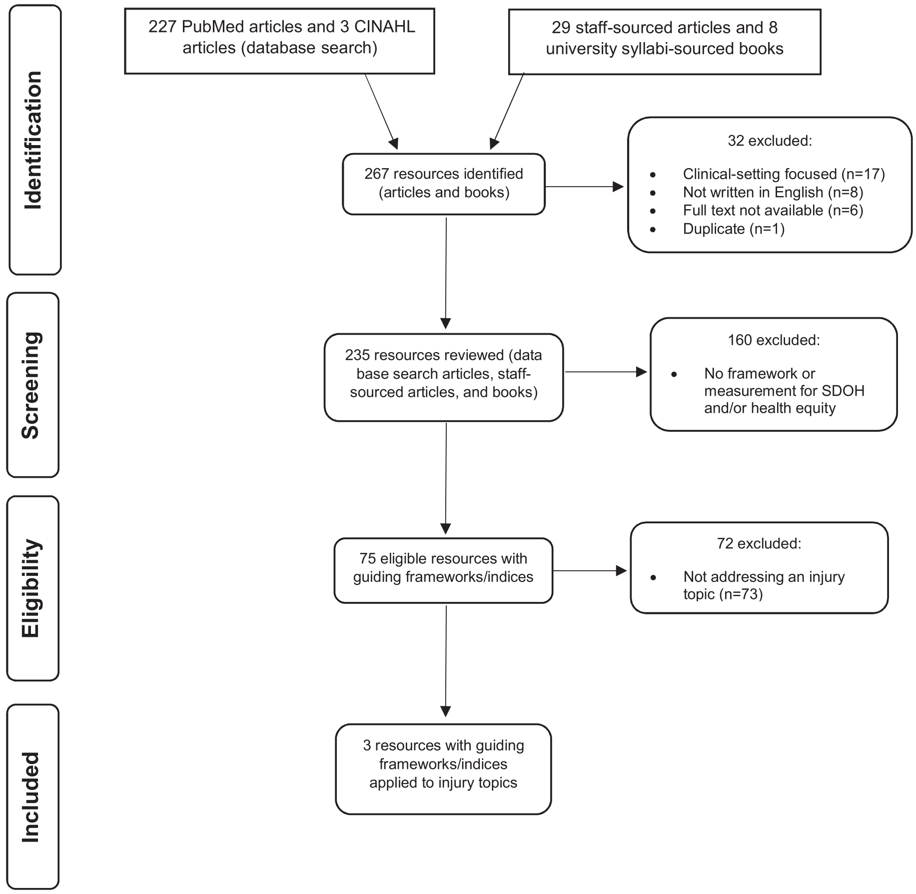


Fig. 1. Flowchart.

4. Discussion

4.1. Summary of findings

The resources examined in our research studied a wide variety of populations experiencing diverse health disparities through the application of an array of different health equity guiding frameworks and indices. Only three resources identified in our search applied health equity frameworks or indices to injury topics (Brown et al., 2017b; Armstead et al., 2021; Reno & Hyder, 2018). Our findings suggest that there remain areas of opportunity within the health equity space for injury-specific research. This includes identifying additional health equity guiding frameworks and indices suitable for specific injury topic areas from other meta-analysis, systematic reviews, and individual studies and assessing the utility and quality of well-referenced health equity guiding frameworks and indices—or their adaptations—for injury research.

4.2. Health equity guiding frameworks applied to injury

The three resources that mention the application of a specific health equity guiding framework or index to an injury topic area covered the injury topics of suicide/depression (Brown et al., 2017b), violence indicators (Armstead et al., 2021), and injury related to infant mortality (safe sleep, intimate partner violence, child abuse, and drug use during pregnancy) (Reno & Hyder, 2018), and used the PROGRESS-Plus framework, the WHO SDOH Conceptual Framework, and a variation of the Social-Ecological Model, respectively.

Depression is a risk factor for suicide (CDC NCIPC, 2021a). Brown et al conducted a systematic review that used the PROGRESS-Plus framework to examine the role of SDOHs on depression among individuals in the Caribbean (2017b). "PROGRESS" is an acronym that refers to a core set of SDOHs: place of residence, race or ethnicity, occupation, gender, religion, education, socio-economic position, and social capital. "Plus" in PROGRESS-Plus refers to other SDOHs identified based on evidence from previous literature and can include personal characteristics associated with discrimination (e.g., disability), features of relationships (e.g., bullied at school), and time-dependent relationships (e.g., hospital stays) (Cochrane, 2022). The Brown et al study identified age as an additional SDOH to analyze (2017b). Using the PROGRESS-Plus framework as a guide, Brown et al found the SDOHs that contributed most to inequalities in suicidal ideation, self-directed harm or suicide attempt, and suicide were gender, age, residence, marital status, and education (2017b).

The WHO SDOH Conceptual Framework illustrates how the structural and social determinants of health inequities (e.g., policies and social class) and intermediary determinants of SDOHs (e.g., behaviors and biological factors) work together to impact equity in health and well-being (Solar & Irwin, 2010). One review used the WHO SDOH Conceptual Framework to identify violence risk factors found in varying contexts (Armstead et al., 2021). Indicators from articles were categorized in four categories: socioeconomic and political context, socioeconomic community conditions, social and physical environments, and bridging community dynamics. The most common indicators within each of these categories were measures of income inequality (a structural determinant of health and risk factor for violence), socioeconomic

Table 1
Characteristics of eligible resources mentioning health equity guiding frameworks and indices.

Citation[1]	Population(s) Addressed	Guiding Framework/Index	Injury related?
Armstead et al., 2021[2]	Not specified	WHO Conceptual SDOH Framework	Yes (violence indicators)
Brown et al., 2017b	Caribbean residents	PROGRESS-Plus	Yes (suicide/ depression)
Reno & Hyder, 2018	African American infants	Social-Ecological Model	Yes (infant mortality-safe sleep, drugs, abuse)
Abbott & Williams, 2015	African Americans living with HIV in rural southeast	Healthy People 2020 SDOH Framework	No (HIV)
Ahmed et al., 2022	Community health workers in low- and middle-income countries	PROGRESS	No (community health worker interventions)
Allen et al., 2020	Not specified	WHO Conceptual SDOH Framework; Social-Ecological Model	No (noncommunicable diseases)
Aves et al., 2017	Not specified	PROGRESS-Plus	No (HIV)
Batista et al., 2018	Immigrants	WHO Conceptual SDOH Framework	No (health care models)
Berkman et al., 2014	Not specified	Gini coefficient	No (general concepts)
Bhojani et al., 2019	Ethnic and religious minorities in India	WHO Conceptual SDOH Framework	No (legislation)
Bowers et al., 2020	Inuit population in Canada	Dimensions of food security (Food and Agricultural Organization)	No (food security)
Braveman, 1998[2]	Low- and middle-income countries	Policy-oriented approach	No (social disparities)
Braveman, 2003[2]	Social groups with varying levels of social advantage	Conceptual framework for monitoring equity in health and healthcare	No (conceptual model for health and healthcare equity)
Brown et al., 2017a	Caribbean women	WHO Conceptual Social Determinants of Health Framework (adaptation)	No (breast cancer)
Buttazzoni et al., 2020	Urban cities	PROGRESS-Plus; Smart City 2.0 Paradigm	No (urban health)
Campos-Matos et al., 2016	Portuguese individuals	PROGRESS	No (general health)
Chandanabhumma & Narasimhan, 2020	Marginalized communities	Applied Decolonial Framework for Health Promotion	No (social justice/ health promotion)
Chandler et al., 2022	Public health students	Three Levels of Racism Framework	No (student training)
Chen et al., 2020	Not specified	Healthy People 2020 SDOH Framework; Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences (PRAPARE); WHO Conceptual SDOH Framework	No (healthcare)
Chhibber et al., 2021	Policies	PROGRESS-Plus	No (policy)
Christidis et al., 2021	Australian Aboriginal and Torres Strait Islanders	Social-Ecological Model	No (Aboriginal health/ nutrition)
Cohn & Harrison, 2022	Black women	PROGRESS-Plus	No (sexual health)
De Lima Silva et al., 2014	Elderly	Dahlgren and Whitehead model	No (mortality)
Diez Roux, 2012[2]	Racial/ethnic minorities	Fundamental Cause Model; Interaction Model; Pathways Model;	No (conceptual frameworks)
Dover & Belon, 2019	Public health workforce	Health equity measurement framework; WHO Conceptual SDOH Framework	No (health equity framework/ measurement)
Dressler et al., 2005[2]	Black Americans	Health behavior model; socioeconomic status model; psychosocial stress model; structural-constructivist model	No (theoretical models)
Driscoll et al., 2013	Not specified	Dahlgren and Whitehead model	No (circumpolar population health)
Fairfield et al., 2020[2]	Rural; low SES	Area Deprivation Index (ADI)	No (lung cancer)
Forde et al., 2019	Socially and economically disadvantaged groups	Weathering hypothesis	No (racial health disparities)
Freire et al., 2018	Older adults	Human Development Index (HDI)	No (gait performance)
Gee & Payne-Sturges, 2004[2]	Racial/ethnic minorities	Community Stress Theory; Index of Dissimilarity; Stress-Exposure Disease Framework	No (environmental health)
Ghiasvand et al., 2020	HIV infected populations	Social and demographic determinants of health-related quality of life (QoL)	No (disparities in HIV QoL)
Greenbaum et al., 2018	Human trafficking victims and their families and communities	Social-Ecological Model	No (human trafficking)
Habbab & Bhutta, 2020	Saudi adolescents	Social-Ecological Model	No (pediatric obesity)
Hu et al., 2018[2]	Urban; low SES	Area Deprivation Index (ADI)	No (readmissions)
Karger et al., 2022	Indigenous and culturally and linguistically diverse (CALD) infants	Social-Ecological Model	No (pregnancy and childbirth)
Kolahdooz et al., 2015	Indigenous Canadian populations	Integrated Life Course and Social Determinants Model of Aboriginal Health	No (SDoH)
Krieger et al., 2003[2]	Diverse race/ethnicity-gender groups	Area-based socioeconomic measures (ABSMs)	No (public health monitoring)
Lehne & Bolte, 2017	Older adults aged 50+	PROGRESS-Plus	No (physical activity)
Lorenc et al., 2014	Local policymakers, practitioners or anyone with a local-level decision-making role	Health In All Policies	No (policy)

Table 1 (*continued*)

Citation[1]	Population(s) Addressed	Guiding Framework/Index	Injury related?
Lund et al., 2018	Not specified	Social and cultural determinants of mental disorders	No (mental disorders)
Malele-Kolisa et al., 2019	Children in Africa	International Classification of Functioning, Disability and Health (ICF) model; Oral health-related quality of life (OHRQoL) conceptual framework	No (children's oral health/quality of life)
Maness & Buhi, 2016	Young females of reproductive age (aged 13–25)	Healthy People 2020 SDOH Framework	No (pregnancy)
Min et al., 2022	Asian Americans (and subgroups)	Healthy People 2030 SDOH Framework; WHO Conceptual SDOH Framework	No (cardiometabolic disease)
Mohan & Chattopadhyay, 2020	Underserved, vulnerable populations	Healthy People 2020 SDOH Framework; Public Health 3.0 Model	No (cost-effectiveness)
Morton et al., 2016	Adults with moderate to severe chronic kidney disease (CKD)	PROGRESS	No (kidney disease)
Nooh et al., 2019	Mobile pastoralist communities in Ethiopia	Equity-effectiveness model	No (disease control)
Ortiz et al., 2020	Not specified	Community-based participatory research (CBPR) conceptual model	No (community engagement)
Owusu-Addo et al., 2018	Those targeted by either conditional or unconditional cash transfers	WHO Conceptual SDOH Framework	No (cash transfers)
Payne-Sturges & Gee, 2006[2]	Racial/ethnic minorities; low SES	Framework for understanding racial/ethnic disparities in environmental health; Index of dissimilarity	No (environmental health)
Payne-Sturges et al., 2006a[2]	Racial/ethnic minorities; low SES	Stress-Exposure Disease Framework	No (environmental health)
Payne-Sturges et al., 2006b[2]	Racial/ethnic minorities	Multi-level systems approach	No (environmental health)
Pereira et al., 2019	Children w/ obesity	Social-Ecological Model	No (children's health)
Peterson et al., 2021	Public health practitioners and researchers	Health equity framework	No (health equity frameworks/ framework development)
Rajmil et al., 2020	Children in European countries	Social determinants on child health (SDCH)	No (children's health)
Restar et al., 2021	Transgender populations	Gender-based health equity framework for transgender populations	No (gender-based framework/framework development)
Salgado et al., 2020	Urban settings	Environmental determinants of health (EDoH)	No (urban health)
Schröders et al., 2015	Disadvantaged groups in Indonesia	PROGRESS	No (child/infant mortality)
Schüz et al., 2021	Not specified	PROGRESS-Plus	No (dietary nudging interventions)
Sokol et al., 2019	Children	Healthy People 2020 SDOH Framework	No (children's health)
Srivastava et al., 2022	Not specified	Area Deprivation Index (ADI); COVID-19 Community Vulnerability Index (CCVI); Healthy Places Index (HPI); Social Vulnerability Index (SVI)	No (COVID-19 vaccination)
Taggart et al., 2020	Black adolescent girls and young adult women	Life Course Approach	No (women's health)
Taylor & Lamaro Haintz, 2018	Australian refugees	Social-Ecological Model	No (healthcare services access)
The National Academies of Sciences, Engineering, and Medicine, 2016	Not specified	Danaher Framework; Rural Community Health & Well-Being Framework; WHO Conceptual SDOH Framework; The Frieden Framework; A Public Health Framework for Reducing Health Inequities	No (education for health professionals)
Tulier et al., 2019	Urban populations	Fundamental Cause Model	No (urban health inequities)
Turnbull et al., 2020	People with chronic health conditions	PROGRESS-Plus	No (chronic disease interventions)
van Daalen et al., 2021	Not specified	Health equity audits (HEA)	No (service provision)
van Hees et al., 2019	Low- and middle-income countries; vulnerable groups	Equiframe; Social, Political, Economic and Cultural (SPEC) conceptual model (WHO)	No (health care reform)
Wang et al., 2020	Pregnant women	WHO Conceptual SDOH Framework	No (maternal health)
Welch et al., 2022	Not specified	PROGRESS-Plus	No (interventions)
Wilder et al., 2021	United States adults	Healthy People 2020 SDOH Framework	No (chronic disease management)
Wilkerson, 2021[2]	Communities; racial/ethnic minorities; low SES	Systems perspective	No (social connection)
Yelton et al., 2022	Adult African American population in the United States	Healthy People 2030 SDOH Framework	No (depression/mental health)
Yiga et al., 2020	Women of reproductive age in sub-Saharan Africa	Social-Ecological Model	No (health behavior)
Zahnd & McLafferty, 2017	People with cancer	Warnecke's Model for Analysis of Population Health and Disparities	No (cancer)

[1] Citations were sourced from a PubMed/CINAHL literature search of meta-analysis and systematic review articles using the search terms "Health Equity"[Mesh] OR "Social Determinants of Health"[Mesh] and CDC Injury Center staff previous knowledge of the literature.
[2] Indicates citations sourced from previous knowledge of the health equity literature.

Table 2
Health equity guiding frameworks and indices.

Health Equity Guiding Framework/Index	Description	Scale (where applicable)	**Citations**[1]
A Public Health Framework for Reducing Health Inequities	Depicts the relationship between social inequalities and health, with a specific focus on inequities related to social, institutional, and living conditions.		The National Academies of Sciences, Engineering, and Medicine, 2016
AHRQ Social Determinants of Health Beta Data Files[3]	With funding from the Patient-Centered Outcomes Research Trust Fund, AHRQ has created multi-year (2009–2018) social determinants of health (SDOH) beta data files curated from multiple federal and other data sources; variables in the beta data files correspond to five key SDOH domains: (1) social context, (2) economic context, (3) education, (4) physical infrastructure, and (5) healthcare context.		Agency for Healthcare Research and Quality, 2021
Applied Decolonial Framework for Health Promotion	Provides guidance to public health practitioners on integrating decolonial processes into health promotion practice to achieve social justice and eliminate health inequities.		Chandanabhumma & Narasimhan, 2020
Area Deprivation Index (ADI)	Assesses a region's socioeconomic conditions to identify areas that have high levels of deprivation and may be more vulnerable to adverse health outcomes.	Geographic percentile rankings range from 1-100; higher score = more highly disadvantaged	Srivastava et al., 2022[3]; Hu et al., 2018[3]; Fairfield et al., 2020[3]
Area-based socioeconomic measures (ABSMs)	Draws on multilevel frameworks and area-based measures to characterize the cases (numerator) and catchment population (denominator), therefore enabling the calculation of rates stratified by the socioeconomic characteristics of a particular residential area.	Higher value = less advantaged	Krieger et al., 2003[3]
Community Resilience Estimate (CRE)[3]	Uses three different data sources (Community Resilience Estimates, American Community Survey, and Census Bureau's Planning Database) to provide information about the capacity of communities and neighborhoods in the United States to respond to the impacts of disasters.	CREs are categorized into three groups: 0 risks, 1–2 risks, and 3 plus risks	United States Census Bureau, 2019
Community Stress Theory	Stressors, such as issues related to inequality, can weaken the body's ability to respond to external challenges.		Gee & Payne-Sturges, 2004[3]
Community Well-Being Index[3]	Combines the Well-Being Index and the Social Determinants of Health Index into a single score to assess self-reported health-related behaviors and perceptions through five interrelated domains: (1) healthcare access, (2) food access, (3) resource access, (4) housing and transportation, and (5) economic security.	Scale ranges from 0-100; lower score = low community well-being	Sharecare, 2021
Community-based participatory research (CBPR) conceptual model	Partnership approach which involves all stakeholders (e.g., community members, organizational representatives, researchers) in the research and decision-making process with the goal of increasing knowledge and driving political and/or social change.		Ortiz et al., 2020
Conceptual framework for monitoring equity in health and healthcare	Designed to help guide the development of approaches (using existing data and simple methods) to monitoring equity in health and health care—e.g., formulating key questions, defining the social groups to be compared, and selecting health indicators and measures of disparity that are fundamental to monitoring health equity.		Braveman, 2003[3]
County Health Rankings[3]	Measures the health of nearly all counties across the United States by applying county-level measures from national and state data sources to help communities understand the health of their residents.	Lower ranking/higher percentile = healthier county	University of Wisconsin Population Health Institute, 2022
COVID-19 Community Vulnerability Index (CCVI)	Expands upon the Centers for Disease Control and Prevention's (CDC) Social Vulnerability Index (SVI) to incorporate additional variables that assess individual- and community-level vulnerability within the context of the coronavirus pandemic.	Values range from 0-1; higher value = higher level of vulnerability	Srivastava et al., 2022[3]

Table 2 (*continued*)

Health Equity Guiding Framework/Index	Description	Scale (where applicable)	**Citations**[1]
Dahlgren and Whitehead model	Maps the influence of individual (e.g., lifestyle factors) and environmental factors (e.g., community influences, living and working conditions, etc.) on health.		de Lima Silva et al., 2014; Driscoll et al., 2013
Danaher Framework	Describes how contributions from the community sector can help reduce health disparities and improve population health.		The National Academies of Sciences, Engineering, and Medicine, 2016
Dimensions of food security (Food and Agricultural Organization)	Measures the availability of food and an individual's ability to access it through the following four dimensions: (1) availability, (2) access, (3) utilization, and (4) stability.		Bowers et al., 2020
Environmental determinants of health	Environmental determinants include the physical, chemical, and biological factors external to a person and their impact on health (e.g., sanitation, exposure to toxins, climate change, pollution, etc.).		Salgado et al., 2020
Equiframe	Analyzes the inclusion of vulnerable groups and human rights in health policies through three summary indices: core concept coverage, vulnerable group coverage, and core concept quality.	Overall ranking: High = policy achieved ≥ 50 % on all 3 indices; Moderate = policy achieved ≥ 50 % on 2 indices; Low = policy achieved < 50 % on 2 or 3 indices	van Hees et al., 2019
Equity-Effectiveness Model	The effectiveness of community-level interventions decreases along a set of parameters which measures access to, and quality of, care.		Nooh et al., 2019
Expansive gender equity continuum	Expands upon previous gender equity models that define equity on a continuum from gender unequal to gender transformative, by including a broader definition of gender identity ranging from exclusive (i.e., only considers cisgender identities) to gender inclusive (i.e., considers people of all gender identities, including trans people and nonbinary individuals).		Restar et al., 2021
Framework for understanding racial/ethnic disparities in environmental health	Health disparities are partially caused by differential access to resources and exposures to hazards and can be grouped into four categories: (1) social processes, (2) environmental contaminants/exposures, (3) body burdens of environmental contaminants, and (4) health outcomes.		Payne-Sturges & Gee, 2006[3]
Fundamental Cause Model	Examines the relationship between socioeconomic inequalities and health; the ability to control disease/death is influenced by access to fundamental resources (e.g., knowledge, money, power, prestige, and beneficial social connections).		Tulier et al., 2019; Diez Roux, 2012[3]
Gini Coefficient	Statistical measure for assessing income inequality across a population.	0 = Perfect equality 1 = Perfect inequality	Berkman et al., 2014
Health-Behavior Model	Social psychological model used to understand and predict health-related behaviors among individuals and communities, particularly in the uptake of health services.		Dressler et al., 2005[3]
Health equity audits	Measures and addresses inequalities in the delivery of and access to health services, associated health outcomes, and determinants of health between different population groups.		van Daalen et al., 2021
Health equity framework	Outlines how health outcomes are influenced by complex interactions between people and their environments and centers around three foundational concepts: (1) equity at the core of health outcomes; (2) multiple, interacting spheres of influence; and a (3) historical and life-course perspective.		Peterson et al., 2021
Health equity measurement framework	Comprehensive model that describes the social determinants of health in a causal context and can be used to measure and monitor health equity; includes an expansive list of social determinants of health, such as the socioeconomic, cultural, and political context, health policy context, social stratification, social location, material and social circumstances, environment, quality of care, etc.		Dover & Belon, 2019

(*continued on next page*)

Table 2 (*continued*)

Health Equity Guiding Framework/Index	Description	Scale (where applicable)	**Citations**[1]
Health in All Policies	Health in All Policies (HiAP) is a collaborative approach that integrates and articulates health considerations into policymaking across sectors to improve the health of all communities and people. HiAP recognizes that health is created by a multitude of factors beyond healthcare and, in many cases, beyond the scope of traditional public health activities.		Lorenc et al., 2014
Healthy People (2020 and 2030)	Provides science-based, national objectives each decade dedicated to improving the health of all Americans. Healthy People 2020 developed a framework that organized the social determinants of health into five key domains: (1) Economic Stability, (2) Education, (3) Health and Health Care, (4) Neighborhood and Built Environment, and (5) Social and Community Context. Healthy People 2030 established a framework to describe the initiative's rationale and approach, including its vision, mission, foundational principles, plan of action, and overarching goals (new objectives are underway).		Wilder et al., 2021; Chen et al., 2020; Mohan & Chattopadhyay, 2020; Sokol et al., 2019; Min et al., 2022; Yelton et al., 2022; Maness & Buhi, 2016; Abbott & Williams, 2015
Healthy Places Index (HPI)	Tool used to explore the social conditions that affect health and includes indicators such as food access, job opportunities, clean air and water, single parent households, education, and others.	Value ranges from 0-100; lower score = more healthy conditions	Srivastava et al., 2022[3]
Human Development Index (HDI)	Standardized measure used to assess the extent of human development in a country through three key dimensions: (1) life expectancy, (2) education, and (3) per capita income.	Values range from 0-1; Higher score = higher lifespan, higher education level, and higher per capita income	Freire et al., 2018
Index of Dissimilarity	Measures the evenness of groups over space and can be interpreted as the percentage of a particular group who would have to move in order to integrate the two groups over the region as a whole.	Value ranges from 0-1; 0 = fully integrated environment 1 = full segregation	Payne-Sturges & Gee, 2006[3]; Gee & Payne-Sturges, 2004[3]
Integrated Life Course and Social Determinants Model of Aboriginal Health	Conceptual framework for understanding the social determinants of health that impact Aboriginal and Torres Strait Islander people as organized into three categories: (1) proximal (e.g., income, education, housing, individual health), (2) intermediate (e.g., resources, opportunities, and infrastructure), and (3) distal (e.g., colonialism, racial discrimination, natural environment, healthcare systems).		Kolahdooz et al., 2015
Interaction Model	Emphasizes the interaction between genes and their environment, such that individuals with different genotypes experience differential effects of environmental exposures and disease risk.		Diez Roux, 2012[3]
International Classification of Functioning, Disability and Health (ICF) model	In-depth classification of holistic components of functioning, disability, and health-related domains.		Malele-Kolisa et al., 2019
Life Course Approach (World Health Organization)	Applies a temporal and social perspective to analyze people's lives within social, economic, and cultural contexts across different generations to understand current patterns of health and disease.		Taggart et al., 2020
Minority Health Social Vulnerability Index (MHSVI)[3]	Expands upon the Social Vulnerability Index to include additional factors that impact COVID-19 outcomes, as organized into the following six themes: (1) Socioeconomic Status, (2) Household Composition and Disability, (3) Minority Status and Language, (4) Housing Type and Transportation, (5) Health Care Infrastructure and Access, and (6) Medical Vulnerability.	Values range from 0-1; higher value = more vulnerable	CDC & HHS OMH, 2021

Table 2 (*continued*)

Health Equity Guiding Framework/Index	Description	Scale (where applicable)	**Citations**[1]
Multi-Level Systems Approach	Focuses on individuals within broader contexts, such as within neighborhoods or communities, who may share similar characteristics and therefore may experience similar health outcomes.		Payne-Sturges et al., 2006b[3]
Pathways Model	This model aims to reduce health and social disparities in communities by connecting high-risk individuals to care and tracking the associated outcomes.		Diez Roux, 2012[3]
Policy-oriented approach[3]	Analysis of patterns and trends of social inequalities in health over time and their determinants, with a specific focus on inequalities that are commonly viewed as unjust and avoidable.		Braveman, 1998
PROGRESS/PROGRESS Plus[2]	Acronym used to identify dimensions across which health inequities may occur, specifically, place of residence; race/ethnicity /culture/ language; occupation; gender/sex; religion; education; socioeconomic status; and social capital.		Turnbull et al., 2020; Chhibber et al., 2021; Schröders et al., 2015; Lehne & Bolte, 2017; Schüz et al., 2021; Campos-Matos et al., 2016; Aves et al., 2017; Welch et al., 2022; Morton et al., 2016; Buttazzoni et al., 2020; Brown et al., 2017b; Ahmed et al., 2022; Cohn & Harrison, 2022
Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences (PRAPARE)	Risk assessment tool to help healthcare providers collect data on patients' social determinants of health to improve health, reduce costs, and ensure needs are met.		Chen et al., 2020
Psychosocial stress model	Health disparities arise from the stresses associated with institutional and interpersonal racism.		Dressler et al., 2005[3]
Public Health 3.0 Model	A model which suggests that building healthy communities requires cross-sector collaboration with various stakeholders in order to advance health and achieve health equity.		Mohan & Chattopadhyay, 2020
Rural Community Health & Well-Being Framework	Identifies key drivers (i.e., social, economic, and environmental factors) that influence health in rural communities and includes additional categories of important factors highlighted by rural residents.		The National Academies of Sciences, Engineering, and Medicine, 2016
Smart City 2.0 Paradigm	Any initiative, policy, promotion, program, or strategy developed to serve the needs of citizens.		Buttazzoni et al., 2020
Social and cultural determinants of mental disorders	Conceptual framework to understand how social determinants interact with key genetic determinants to influence mental disorders.		Lund et al., 2018
Social and demographic determinants of health-related quality of life (QoL)	An individual's overall sense of wellbeing including aspects of happiness, satisfaction of life, and physical, mental, psychological, and social perceptions.		Ghiasvand et al., 2020
Social determinants of child health (SDCH)	Examines how social determinants impact child health across time and generations through distal social factors such as poverty, material deprivation, and social inequalities.		Rajmil et al., 2020
Social Vulnerability Index (SVI)	Uses U.S. Census data to assess the extent to which communities are socially vulnerable to disasters by ranking each census tract on 15 social factors, including poverty, lack of vehicle access, and crowded housing, and grouping them into four related themes.	Values range from 0-1; higher value = higher level of vulnerability	CDC ATSDR, 2011; Srivastava et al., 2022[3]
Social, Political, Economic and Cultural (SPEC) conceptual model (WHO)	Explains social exclusions as a process rather than a state operating along different dimensions and individual, regional, and global levels.		van Hees et al., 2019
Social-Ecological Model[2]	Theory-based framework for understanding how social and structural determinants influence health and wellbeing.		Habbab & Bhutta, 2020; Yiga et al., 2020; Pereira et al., 2019; Reno & Hyder, 2018; Greenbaum et al., 2018; Christidis et al., 2021; Allen et al., 2020; Karger et al., 2022; Taylor & Lamaro Haintz, 2018
Socioeconomic Status Model	Emphasizes that race/ethnicity and socioeconomic status (SES) are related, such that certain race/ethnicity groups are disproportionately represented in lower SES groups.		Dressler et al., 2005[3]

(*continued on next page*)

Table 2 (*continued*)

Health Equity Guiding Framework/Index	Description	Scale (where applicable)	**Citations**[1]
Stress-Exposure Disease Framework	Conceptual framework that outlines the relationships between race, environmental conditions, and health.		Gee & Payne-Sturges, 2004[3]; Payne-Sturges et al., 2006a[3]
Structural-Constructivist Model	Integrates a dual perspective focused on (1) socially constructed cognitive representations within a society and (2) external factors that restrict individuals, specifically social relationships, and expectations of others (e.g., race, as a concept, is socially or culturally constructed).		Dressler et al., 2005[3]
Systems perspective	Describes communities as a set of institutions which represent various sectors, each with their own policies and decision-making processes which influence behavior and investments in the community.		Wilkerson, 2021[3]
The Frieden Framework	Five-tier pyramid for improving public health; the base of the pyramid includes (1) interventions that impact social determinants of health (e.g., poverty, education), followed by (2) interventions that benefit the general population (e.g., fluoridated water), (3) interventions that help large segments of the population (e.g., immunizations), (4) clinical interventions for the prevention of certain conditions (e.g., cardiovascular disease), and (5) health education interventions (i.e., most labor-intensive and potentially lowest impact).		The National Academies of Sciences, Engineering, and Medicine, 2016
Three Levels of Racism Framework	Theoretical framework for understanding racial health inequities and developing effective interventions to reduce inequities on three distinct levels: (1) institutionalized, (2) personally mediated, and (3) internalized.		Chandler et al., 2022
Warnecke's Model for Analysis of Population Health and Disparities	Defines factors impacting health disparities as proximal, intermediate, or distal and focuses on individual-level outcomes as they relate to specific determinants (i.e., social conditions and policies, institutional context, social context, and physical context).		Zahnd & McLafferty, 2017
Weathering Hypothesis	Proposes that cumulative exposure to social, economic, and political disadvantage leads to rapid decline in physical health.		Forde et al., 2019
WHO Conceptual SDOH Framework[2]	Outlines how social, economic, and political factors (e.g., income, education, occupation, gender, race, and ethnicity) impact an individual's socioeconomic position, which, in turn, influences their vulnerability and exposure to health conditions.		Chen et al., 2020; Bhojani et al., 2019; Dover & Belon, 2019; Wang et al., 2020; Allen et al., 2020; Armstead et al., 2021[3]; Batista et al., 2018; Owusu-Addo et al., 2018; The National Academies of Sciences, Engineering, and Medicine, 2016; Brown et al., 2017a; Min et al., 2022

[1] Citations were sourced from a PubMed/CINAHL literature search of meta-analysis and systematic review articles using the search terms "Health Equity"[Mesh] OR "Social Determinants of Health"[Mesh] and CDC Injury Center staff previous knowledge of the literature.
[2] Indicates a health equity guiding framework that was applied to an injury topic in this study.
[3] Indicates a framework, index, or article sourced from previous knowledge of the health equity literature.

disadvantage or deprivation, social disorganization, and social capital/collective efficacy, respectively.

A systematic review from 2018 used a social-ecological framework to categorize risk factors for infant mortality (specifically, mortality of African American and multiracial infants) at multiple levels (Reno & Hyder, 2018). While this study does not solely focus on injury topics, it mentions safe sleep, intimate partner violence, child abuse, and drug use while pregnant. The social-ecological framework used in this study consists of five different levels (individual, interpersonal, organizational, community, and public policy) and acknowledges the interplay that occurs between these levels. Safe sleep, drug use during pregnancy, and child abuse were all categorized as SDOHs found within the individual level, while intimate partner violence was categorized at the interpersonal level but associated with factors at the individual level (e.g., maternal drug use).

The use of health equity guiding frameworks in these studies helped to identify research gaps in health equity injury research. Namely, the studies identified the need for further examination of sociodemographic inequalities and the need to explore relationships and interactions between SDOHs as important to achieving progress towards health equity in injury, since these factors can influence an injury outcome. It is important to note that structural forces and societal constructs must also be considered in research addressing health disparities and SDOHs, since these are responsible for the inequitable distribution of SDOHs that cause health disparities and create more or less disadvantaged groups of individuals. CDC uses a four-level social ecological model to understand the interplay of risk and protective factors, inclusive of SDOHs, for various forms of injury and to guide prevention strategies, and includes levels at the individual, relationship, community, and societal levels (CDC NCIPC, 2021b). The paucity of health equity meta-analysis and systematic reviews using guiding frameworks or indices that address injury topic areas was observed in our study, suggesting areas of opportunity for injury researchers.

4.3. Limitations

This study is subject to several known limitations. For this formative study, PubMed and CINAHL were used to identify systematic review and meta-analysis articles that mention social determinants of health and health equity. Exploring other databases and broadening the article type would have produced additional results. Books were difficult to access and may also have captured more health equity guiding frameworks or indices not included in this study. Some studies may have selected certain SDOHs to focus on, but did not apply a specific name to the set of SDOHs being researched. These articles were categorized as not mentioning a health equity guiding framework or index. Lastly, the health equity guiding frameworks and indices reported here were not assessed for quality.

5. Conclusion

The findings from this study revealed that the application of health equity guiding frameworks and indices in research have been used to measure and characterize health equity within populations. Though many health equity guiding frameworks and indices were documented for measuring health equity, few were applied in meta-analyses or systematic reviews to an injury topic. However, there are large numbers of individual studies that apply health equity guiding frameworks and indices to injury topics. Some guiding frameworks/indices may be more appropriate than others for measuring health equity in injury topic areas, depending on which SDOH they address. Measuring health equity is important for understanding progress towards reducing health disparities within populations and is helpful for identifying where and for whom additional or improved programs and interventions may be necessary.

5.1. Practical Applications

Measuring health equity in injury and other public health research areas can help build a foundation of evidence. Moving forward, injury researchers can consider the health equity guiding frameworks and indices identified through this study in their health equity injury research.

Acknowledgements

We would like to acknowledge Amy Wolkin, Heather Clayton, and Martevia Bledsoe for contributing articles and health equity guiding frameworks and indices for this review.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Interest

None.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention. All materials in this report are in the public domain and may be used and copied without permission but require citation.

References

Abbott, L. S., & Williams, C. L. (2015). Influences of social determinants of health on African Americans living with HIV in the rural southeast: A qualitative meta-synthesis. *Journal of the Association of Nurses in AIDS Care, 26*(4), 340–356. https://doi.org/10.1016/j.jana.2015.03.004.

Agency for Healthcare Research and Quality (2021). Social Determinants of Health Database (Beta Version). [Available from: https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html].

Ahmed, S., Chase, L. E., Wagnild, J., Akhter, N., Sturridge, S., Clarke, A., Chowdhary, P., Mukami, D., Kasim, A., & Hampshire, K. (2022). Community health workers and health equity in low- and middle-income countries: Systematic review and recommendations for policy and practice. *International Journal for Equity in Health, 21*(1), 49. https://doi.org/10.1186/s12939-021-01615-y.

Allen, L. N., Smith, R. W., Simmons-Jones, F., Roberts, N., Honney, R., & Currie, J. (2020). Addressing social determinants of noncommunicable diseases in primary care: A systematic review. *Bulletin of the World Health Organization, 98*(11), 754–765B. https://doi.org/10.2471/BLT.19.248278.

Armstead, T. L., Wilkins, N., & Nation, M. (2021). Structural and social determinants of inequities in violence risk: A review of indicators. *Journal of Community Psychology, 49*(4), 878–906. https://doi.org/10.1002/jcop.22232.

Aves, T., Kredo, T., Welch, V., Mursleen, S., Ross, S., Zani, B., Motaze, N. V., Quinlan, L., & Mbuagbaw, L. (2017). Equity issues were not fully addressed in Cochrane human immunodeficiency virus systematic reviews. *Journal of Clinical Epidemiology, 81*, 96–100. https://doi.org/10.1016/j.jclinepi.2016.09.008.

Batista, R., Pottie, K., Bouchard, L., Ng, E., Tanuseputro, P., & Tugwell, P. (2018). Primary health care models addressing health equity for immigrants: A systematic scoping review. *Journal of Immigrant and Minority Health, 20*(1), 214–230. https://doi.org/10.1007/s10903-016-0531-y.

Berkman, L., Kawachi, I., & Glymour, M. M. (2014). *Social epidemiology*. Oxford: Oxford University Press.

Bhojani, U., Madegowda, C., Prashanth, N. S., Hebbar, P., Mirzoev, T., Karlsen, S., & Mir, G. (2019). Affirmative action, minorities, and public services in India: Charting a future research and practice agenda. *Indian Journal of Medical Ethics, 4*(4), 265–273. https://doi.org/10.20529/ijme.2019.062.

Bowers, R., Turner, G., Graham, I. D., Furgal, C., & Dubois, L. (2020). Piecing together the Labrador Inuit food security policy puzzle in Nunatsiavut, Labrador (Canada): A scoping review. *International Journal of Circumpolar Health, 79*(1), 1799676. https://doi.org/10.1080/22423982.2020.1799676.

Braveman P. (1998). Monitoring equity in health: A policy-oriented approach in low- and middle-income countries. Geneva: Department of Health Systems, World Health Organization. [Available from: https://apps.who.int/iris/handle/10665/65228].

Braveman, P. (2003). Monitoring equity in health and healthcare: A conceptual framework Available from. *Journal of Health and Population Nutrition, 21*(3), 181–192 https://www.jstor.org/stable/23499216?seq=1.

Brown, C. R., Hambleton, I. R., Hercules, S. M., Alvarado, M., Unwin, N., Murphy, M. M., Harris, E. N., Wilks, R., MacLeish, M., Sullivan, L., & Sobers-Grannum, N. (2017a). Social determinants of breast cancer in the Caribbean: A systematic review. *International Journal of Equity in Health, 16*(1), 60. https://doi.org/10.1186/s12939-017-0540-z.

Brown, C. R., Hambleton, I. R., Sobers-Grannum, N., Hercules, S. M., Unwin, N., Nigel Harris, E., Wilks, R., MacLeish, M., Sullivan, L., Murphy, M. M., & U.S. Caribbean Alliance for Health Disparities Research Group (USCAHDR) (2017b). Social determinants of depression and suicidal behaviour in the Caribbean: A systematic review. *BMC Public Health, 17*(1), 577. https://doi.org/10.1186/s12889-017-4371-z.

Buttazzoni, A., Veenhof, M., & Minaker, L. (2020). Smart city and high-tech urban interventions targeting human health: An equity-focused systematic review. *International Journal of Environmental Research in Public Health, 17*(7). https://doi.org/10.3390/ijerph17072325.

Campos-Matos, I., Russo, G., & Perelman, J. (2016). Connecting the dots on health inequalities–a systematic review on the social determinants of health in Portugal. *International Journal of Equity in Health, 15*, 26. https://doi.org/10.1186/s12939-016-0314-z.

Centers for Disease Control and Prevention, Agency for Toxic Substances and Disease Registry (2011). Social Vulnerability Index. [Available from: https://www.atsdr.cdc.gov/placeandhealth/svi/index.html].

Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion (2020). Health Equity 2020 [Available from: https://www.cdc.gov/chronicdisease/healthequity/index.htm].

Centers for Disease Control and Prevention, National Center for Injury Prevention and Control (2020). Web-based Injury Statistics Query and Reporting System (WISQARS) [online]. [Available from: http://www.cdc.gov/injury/wisqars].

Centers for Disease Control and Prevention, National Center for Injury Prevention and Control (2021). Risk and Protective Factors. [Available from: https://www.cdc.gov/suicide/factors/index.html].

Centers for Disease Control and Prevention, National Center for Injury Prevention and Control (2021). The Social-Ecological Model: A Framework for Prevention.

[Available from: https://www.cdc.gov/violenceprevention/about/social-ecologicalmodel.html].
Centers for Disease Control and Prevention, Office of Minority Health and Health Equity (2022). CDC Core Health Equity Science and Intervention Strategy. [Available from: https://www.cdc.gov/healthequity/core/index.html].
Centers for Disease Control and Prevention & United States Department of Health and Human Services Office of Minority Health (2021). Minority Health Social Vulnerability Index. [Available from: https://www.minorityhealth.hhs.gov/minority-health-svi/].
Chandanabhumma, P. P., & Narasimhan, S. (2020). Towards health equity and social justice: An applied framework of decolonization in health promotion. *Health Promotion International, 35*(4), 831–840. https://doi.org/10.1093/heapro/daz053.
Chandler, C. E., Williams, C. R., Turner, M. W., & Shanahan, M. E. (2022). Training public health students in racial justice and health equity: A systematic review. *Public Health Report, 137*, 375–385. https://doi.org/10.1177/00333549211015665.
Chen, M., Tan, X., & Padman, R. (2020). Social determinants of health in electronic health records and their impact on analysis and risk prediction: A systematic review. *Journal of the American Medical Informatics Association, 27*(11), 1764–1773. https://doi.org/10.1093/jamia/ocaa143.
Chhibber, A., Kharat, A., Kneale, D., Welch, V., Bangpan, M., & Chaiyakunapruk, N. (2021). Assessment of health equity consideration in masking/PPE policies to contain COVID-19 using PROGRESS-plus framework: A systematic review. *BMC Public Health, 21*(1), 1682. https://doi.org/10.1186/s12889-021-11688-7.
Christidis, R., Lock, M., Walker, T., Egan, M., & Browne, J. (2021). Concerns and priorities of Aboriginal and Torres Strait Islander peoples regarding food and nutrition: A systematic review of qualitative evidence. *International Journal of Equity in Health, 20*(1), 220. https://doi.org/10.1186/s12939-021-01551-x.
Clemens, T., Moreland, B., & Lee, R. (2021). Persistent racial/ethnic disparities in fatal unintentional drowning rates among persons aged ≤29 Years — United States, 1999–2019. *Morbidity and Mortality Weekly Report, 70*, 869–874. https://doi.org/10.15585/mmwr.mm7024a1.
Cohn, T., & Harrison, C. V. (2022). A systematic review exploring racial disparities, social determinants of health, and sexually transmitted infections in Black women. *Nursing for Women's Health, 26*, 128–142. https://doi.org/10.1016/j.nwh.2022.01.006.
Daugherty, J., Zhou, H., Sarmiento, K., & Waltzman, D. (2021). Differences in state traumatic brain injury–related deaths, by principal mechanism of injury, intent, and percentage of population living in rural areas — United States, 2016–2018. *Morbidity and Mortality Weekly Report, 70*, 1447–1452. https://doi.org/10.15585/mmwr.mm7041a3.
de Lima Silva, V., Cesse, E., & de Albuquerque, M. F. (2014). Social determinants of death among the elderly: A systematic literature review. *Revista Brasileira de Epidemiologia, 17*(Suppl 2), 178–193. https://doi.org/10.1590/1809-4503201400060015.
Department of Health and Human Services (2020). Healthy People 2020 [Available from: https://www.healthypeople.gov/2020].
Diez Roux, A. V. (2012). Conceptual approaches to the study of health disparities. *Annual Review of Public Health, 33*, 41–58. https://doi.org/10.1146/annurev-publhealth-031811-124534.
Dover, D. C., & Belon, A. P. (2019). Correction to: The health equity measurement framework: A comprehensive model to measure social inequities in health. *International Journal of Equity in Health, 18*(1), 58. https://doi.org/10.1186/s12939-019-0949-7.
Dressler, W., Oths, K., & Gravlee, C. (2005). Race and ethnicity in public health research: Models to explain health disparities. *Annual Reviews, 34*, 231–252. https://doi.org/10.1146/annurev.anthro.34.081804.120505.
Driscoll, D. L., Dotterrer, B., Brown, R. A., & 2nd. (2013). Assessing the social and physical determinants of circumpolar population health. *International Journal of Circumpolar Health, 72*. https://doi.org/10.3402/ijch.v72i0.21400.
Ehlman, D. C., Yard, E., Stone, D. M., Jones, C. M., & Mack, K. A. (2022). Changes in suicide rates — United States, 2019 and 2020. *Morbidity and Mortality Weekly Report, 71*, 306–312. https://doi.org/10.15585/mmwr.mm7108a5.
Fairfield, K., Black, A., Ziller, E., Murray, K., Lucas, F., Waterston, L., Korsen, N., Ineza, D., & Han, P. (2020). Area deprivation index and rurality in relation to lung cancer prevalence and mortality in a rural state. *JNCI Cancer. Spectrum, 4*(4). https://doi.org/10.1093/jncics/pkaa011.
Forde, A. T., Crookes, D. M., Suglia, S. F., & Demmer, R. T. (2019). The weathering hypothesis as an explanation for racial disparities in health: A systematic review 1-18.e3. *Annals of Epidemiology, 33*. https://doi.org/10.1016/j.annepidem.2019.02.011.
Freire, R. C., Pieruccini-Faria, F., & Montero-Odasso, M. (2018). Are Human Development Index dimensions associated with gait performance in older adults? A systematic review. *Experimental Gerontology, 102*, 59–68. https://doi.org/10.1016/j.exger.2017.12.001.
Gee, G. C., & Payne-Sturges, D. C. (2004). Environmental health disparities: A framework integrating psychosocial and environmental concepts. *Environmental Health Perspectives, 112*(17), 1645–1653. https://doi.org/10.1289/ehp.7074.
Ghiasvand, H., Higgs, P., Noroozi, M., Ghaedamini Harouni, G., Hemmat, M., Ahounbar, E., Haroni, J., Naghdi, S., Nazeri Astaneh, A., & Armoon, B. (2020). Social and demographical determinants of quality of life in people who live with HIV/AIDS infection: Evidence from a meta-analysis. *Biodemography and Social Biology, 65*(1), 57–72. https://doi.org/10.1080/19485565.2019.1587287.
Greenbaum, V. J., Titchen, K., Walker-Descartes, I., Feifer, A., Rood, C. J., & Fong, H. F. (2018). Multi-level prevention of human trafficking: The role of health care professionals. *Prevention Medicine, 114*, 164–167. https://doi.org/10.1016/j.ypmed.2018.07.006.
Habbab, R. M., & Bhutta, Z. A. (2020). Prevalence and social determinants of overweight and obesity in adolescents in Saudi Arabia: A systematic review. *Clinical Obesity, 10*(6), e12400. https://doi.org/10.1111/cob.12400.
Hu, J., Kind, A. J. H., & Nerenz, D. (2018). Area deprivation index predicts readmission risk at an urban teaching hospital. *American Journal of Medical Quality, 33*(5), 493–501. https://doi.org/10.1177/1062860617753063.
Karger, S., Bull, C., Enticott, J., & Callander, E. J. (2022). Options for improving low birthweight and prematurity birth outcomes of indigenous and culturally and linguistically diverse infants: A systematic review of the literature using the social-ecological model. *BMC Pregnancy and Childbirth, 22*(1), 3. https://doi.org/10.1186/s12884-021-04307-1.
Kolahdooz, F., Nader, F., Yi, K. J., & Sharma, S. (2015). Understanding the social determinants of health among Indigenous Canadians: Priorities for health promotion policies and actions. *Global Health Action, 8*, 27968. https://doi.org/10.3402/gha.v8.27968.
Krieger, N., Chen, J. T., Waterman, P. D., Rehkopf, D. H., & Subramanian, S. V. (2003). Race/ethnicity, gender, and monitoring socioeconomic gradients in health: A comparison of area-based socioeconomic measures–the public health disparities geocoding project. *American Journal of Public Health, 93*(10), 1655–1671. https://doi.org/10.2105/ajph.93.10.1655.
Lehne, G., & Bolte, G. (2017). Impact of universal interventions on social inequalities in physical activity among older adults: An equity-focused systematic review. *International Journal of Behavioral and Nutrition and Physical Activity, 14*(1), 20. https://doi.org/10.1186/s12966-017-0472-4.
Lorenc, T., Tyner, E. F., Petticrew, M., Duffy, S., Martineau, F. P., Phillips, G., & Lock, K. (2014). Cultures of evidence across policy sectors: Systematic review of qualitative evidence. *European Journal of Public Health, 24*(6), 1041–1047. https://doi.org/10.1093/eurpub/cku038.
Lund, C., Brooke-Sumner, C., Baingana, F., Baron, E. C., Breuer, E., Chandra, P., Haushofer, J., Herrman, H., Jordans, M., Kieling, C., Medina-Mora, M. E., Morgan, E., Omigbodun, O., Tol, W., Patel, V., & Saxena, S. (2018). Social determinants of mental disorders and the Sustainable Development Goals: A systematic review of reviews. *Lancet Psychiatry, 5*(4), 357–369. https://doi.org/10.1016/s2215-0366(18)30060-9.
Malele-Kolisa, Y., Yengopal, V., Igumbor, J., Nqcobo, C. B., & Ralephenya, T. R. D. (2019). Systematic review of factors influencing oral health-related quality of life in children in Africa. *African Journal of Primary Health Care and Family Medicine, 11*(1), e1–e12. https://doi.org/10.4102/phcfm.v11i1.1943.
Maness, S. B., & Buhi, E. R. (2016). Associations between social determinants of health and pregnancy among young people: A systematic review of research published during the past 25 years. *Public Health Report, 131*(1), 86–99. https://doi.org/10.1177/003335491613100115.
Min, L. Y., Islam, R. B., Gandrakota, N., & Shah, M. K. (2022). The social determinants of health associated with cardiometabolic diseases among Asian American subgroups: A systematic review. *BMC Health Service Research, 22*(1), 257. https://doi.org/10.1186/s12913-022-07646-7.
Mohan, G., & Chattopadhyay, S. (2020). Cost-effectiveness of leveraging social determinants of health to improve breast, cervical, and colorectal cancer screening: A systematic review. *JAMA Oncolology, 6*(9), 1434–1444. https://doi.org/10.1001/jamaoncol.2020.1460.
Moore, M., Conrick, K., Fuentes, M., Rowhani-Rahbar, A., Graves, J. M., Patil, D., Herrenkohl, M., Mills, B., Rivara, F. P., Ebel, B., & Vavilala, M. S. (2019). Research on injury disparities: A scoping review. *Health Equity, 3*(1), 504–511. https://doi.org/10.1089/heq.2019.0044.
Morton, R. L., Schlackow, I., Mihaylova, B., Staplin, N. D., Gray, A., & Cass, A. (2016). The impact of social disadvantage in moderate-to-severe chronic kidney disease: An equity-focused systematic review. *Nephrology Dialysis Transplantation, 31*(1), 46–56. https://doi.org/10.1093/ndt/gfu394.
Nooh, F., Crump, L., Hashi, A., Tschopp, R., Schelling, E., Reither, K., Hattendorf, J., Ali, S. M., Obrist, B., Utzinger, J., & Zinsstag, J. (2019). The impact of pastoralist mobility on tuberculosis control in Ethiopia: A systematic review and meta-synthesis. *Infectious Diseases of Poverty, 8*(1), 73. https://doi.org/10.1186/s40249-019-0583-z.
Ortiz, K., Nash, J., Shea, L., et al. (2020). Partnerships, Processes, and Outcomes: A Health Equity-Focused Scoping Meta-Review of Community-Engaged Scholarship. *Annual Review of Public Health, 41*, 177–199. https://doi.org/10.1146/annurev-publhealth-040119-094220.
Owusu-Addo, E., Renzaho, A. M. N., & Smith, B. J. (2018). The impact of cash transfers on social determinants of health and health inequalities in sub-Saharan Africa: A systematic review. *Health Policy and Planning, 33*(5), 675–696. https://doi.org/10.1093/heapol/czy020.
Payne-Sturges, D., & Gee, G. C. (2006). National environmental health measures for minority and low-income populations: Tracking social disparities in environmental health. *Environmental Research, 102*(2), 154–171. https://doi.org/10.1016/j.envres.2006.05.014.
Payne-Sturges, D., Gee, G. C., Crowder, K., Hurley, B. J., Lee, C., Morello-Frosch, R., Rosenbaum, A., Schulz, A., Wells, C., Woodruff, T., & Zenick, H. (2006a). Workshop summary: Connecting social and environmental factors to measure and track environmental health disparities. *Environmental Research, 102*(2), 146–153. https://doi.org/10.1016/j.envres.2005.11.001.
Payne-Sturges, D., Zenick, H., Wells, C., & Sanders, W. (2006b). We cannot do it alone: Building a multi-systems approach for assessing and eliminating

environmental health disparities. *Environmental Research, 102*(2), 141–145. https://doi.org/10.1016/j.envres.2006.01.011.
Pereira, M., Padez, C. M. P., & Nogueira, H. (2019). Describing studies on childhood obesity determinants by Socio-Ecological Model level: A scoping review to identify gaps and provide guidance for future research. *International Journal of Obesity (London), 43*(10), 1883–1890. https://doi.org/10.1038/s41366-019-0411-3.
Peterson, A., Charles, V., Yeung, D., & Coyle, K. (2021). The health equity framework: A science- and justice-based model for public health researchers and practitioners. *Health Promotion and Practice, 22*(6), 741–746. https://doi.org/10.1177/1524839920950730.
Rajmil, L., Hjern, A., Spencer, N., Taylor-Robinson, D., Gunnlaugsson, G., & Raat, H. (2020). Austerity policy and child health in European countries: A systematic literature review. *BMC Public Health, 20*(1), 564. https://doi.org/10.1186/s12889-020-08732-3.
Reno, R., & Hyder, A. (2018). The evidence base for social determinants of health as risk factors for infant mortality: A systematic scoping review. *Journal of Health Care for the Poor and Underserved, 29*(4), 1188–1208. https://doi.org/10.1353/hpu.2018.0091.
Restar, A. J., Sherwood, J., Edeza, A., Collins, C., & Operario, D. (2021). Expanding gender-based health equity framework for transgender populations. *Transgender Health, 6*(1), 1–4. https://doi.org/10.1089/trgh.2020.0026.
Robert Wood Johnson Foundation (2017). What is Health Equity? [Available from: https://www.buildhealthyplaces.org/content/uploads/2017/05/rwjf436997.pdf].
Salgado, M., Madureira, J., Mendes, A. S., Torres, A., Teixeira, J. P., & Oliveira, M. D. (2020). Environmental determinants of population health in urban settings. A systematic review. *BMC Public Health, 20*(1), 853. https://doi.org/10.1186/s12889-020-08905-0.
Schröders, J., Wall, S., Kusnanto, H., & Ng, N. (2015). Millennium development goal four and child health inequities in Indonesia: A systematic review of the literature. *PLoS One, 10*(5), e0123629. https://doi.org/10.1371/journal.pone.0123629.
Schüz, B., Meyerhof, H., Hilz, L. K., & Mata, J. (2021). Equity effects of dietary nudging field experiments: Systematic review. *Frontier Public Health, 9*, 668998. https://doi.org/10.3389/fpubh.2021.668998.
Sharecare (2021). Community Well-Being Index. [Available from: https://wellbeingindex.sharecare.com/].
Sokol, R., Austin, A., Chandler, C., Byrum, E., Bousquette, J., Lancaster, C., Doss, G., Dotson, A., Urbaeva, V., Singichetti, B., Brevard, K., Wright, S. T., Lanier, P., & Shanahan, M. (2019). Screening children for social determinants of health: A systematic review. *Pediatrics, 144*(4), e20191622. https://doi.org/10.1542/peds.2019-1622.
Solar, O., & Irwin, A. (2010). *A conceptual framework for action on the social determinants of health* Available from:. Geneva: World Health Organization https://www.who.int/publications-detail-redirect/9789241500852.
Srivastava, T., Schmidt, H., Sadecki, E., & Kornides, M. (2022). Disadvantage indices deployed to promote equitable allocation of COVID-19 vaccines in the US: A scoping review of differences and similarities in design. *JAMA, 3*(1), e214501. https://doi.org/10.1001/jamahealthforum.2021.4501.
Taggart, T., Milburn, N. G., Nyhan, K., & Ritchwood, T. D. (2020). Utilizing a life course approach to examine HIV risk for Black adolescent girls and young adult women in the United States: A systematic review of recent literature. *Ethnic Disparities, 30*(2), 277–286. https://doi.org/10.18865/ed.30.2.277.
Taylor, J., & Lamaro Haintz, G. (2018). Influence of the social determinants of health on access to healthcare services among refugees in Australia. *Australian Journal of Primary Health, 24*(1), 14–28. https://doi.org/10.1071/PY16147.
The Cochrane Collaboration. (2022). PROGRESS-Plus. [Available from: https://methods.cochrane.org/equity/projects/evidence-equity/progress-plus].
The National Academies of Sciences, Engineering, and Medicine (U.S.). (2016). A framework for educating health professionals to address the social determinants of health. [Available from: https://www.ncbi.nlm.nih.gov/books/NBK395983/].
Tulier, M. E., Reid, C., Mujahid, M. S., & Allen, A. M. (2019). "Clear action requires clear thinking": A systematic review of gentrification and health research in the United States. *Health Place, 59*, 102173. https://doi.org/10.1016/j.healthplace.2019.102173.
Turnbull, S., Cabral, C., Hay, A., & Lucas, P. J. (2020). Health equity in the effectiveness of web-based health interventions for the self-care of people with chronic health conditions: Systematic review. *Journal of Medical Internet Research, 22*(6), e17849. https://doi.org/10.2196/17849.
United States Census Bureau (2019). Community Resilience Estimates. [Available from: https://www.census.gov/programs-surveys/community-resilience-estimates.html].
University of Wisconsin Population Health Institute (2022). County Health Rankings Model. [Available from: https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model].
van Daalen, K. R., Davey, F., Norman, C., & Ford, J. A. (2021). Health equity audits: A systematic review of the effectiveness. *BMJ Open, 11*(11), e053392. https://doi.org/10.1136/bmjopen-2021-053392.
van Hees, S. G. M., O'Fallon, T., Hofker, M., Dekker, M., Polack, S., Banks, L. M., & Spaan, E. (2019). Leaving no one behind? Social inclusion of health insurance in low- and middle-income countries: A systematic review. *International Journal of Equity in Health, 18*(1), 134. https://doi.org/10.1186/s12939-019-1040-0.
Wang, E., Glazer, K. B., Howell, E. A., & Janevic, T. M. (2020). Social determinants of pregnancy-related mortality and morbidity in the United States: A systematic review. *Obstetrics & Gynecology, 135*(4), 896–915. https://doi.org/10.1097/aog.0000000000003762.
Welch, V., Dewidar, O., Tanjong Ghogomu, E., Abdisalam, S., Al Ameer, A., Barbeau, V. I., Brand, K., Kebedom, K., Benkhalti, M., Kristjansson, E., Madani, M. T., Antequera Martín, A. M., Mathew, C. M., McGowan, J., McLeod, W., Park, H. A., Petkovic, J., Riddle, A., Tugwell, P., Petticrew, M., Trawin, J., & Wells, G. A. (2022). How effects on health equity are assessed in systematic reviews of interventions. *The Cochrane Database of Systematic Reviews, 1*(1), MR000028. https://doi.org/10.1002/14651858.MR000028.pub3.
Wilkerson R. (2021). *Socially connected communities: Solutions for social isolation*. In: Moore S, ed.: Healthy Places by Design. [Available from: Socially-Connected-Communities_Solutions-for-Social-Isolation.pdf].
Yelton, B., Friedman, D. B., Noblet, S., Lohman, M. C., Arent, M. A., Macauda, M. M., Sakhuja, M., & Leith, K. H. (2022). Social determinants of health and depression among African American adults: A scoping review of current research. *International Journal of Environmental Researchin Public Health, 19*(3). https://doi.org/10.3390/ijerph19031498.
Yiga, P., Seghers, J., Ogwok, P., & Matthys, C. (2020). Determinants of dietary and physical activity behaviours among women of reproductive age in urban sub-Saharan Africa: A systematic review. *British Journal of Nutrition, 124*(8), 761–772. https://doi.org/10.1017/s0007114520001828.
Zahnd, W. E., & McLafferty, S. L. (2017). Contextual effects and cancer outcomes in the United States: A systematic review of characteristics in multilevel analyses 739-48.e3. *Annals of Epidemiology, 27*(11). https://doi.org/10.1016/j.annepidem.2017.10.002.

Natalie Lennon, MPH graduated from Emory University in 2021 with her MPH in Behavioral Sciences and Health Education. She is currently an ORISE fellow for the Division of Injury Prevention in the CDC's Injury Center, where she works on research related to nonfatal injury, health equity, and suicide prevention. Before joining CDC, Natalie completed her practicum at the Georgia Health Policy Center in the Center of Excellence for Children's Behavioral Health. Her research interests include preventing suicide and other forms of injury among disproportionally affected populations.

Andrea Carmichael, MPH graduated from George Mason University in 2018 with her MPH in Epidemiology and completed her practicum at the EPA's Office of Children's Health protection. She is currently an Associate Service Fellow for the Division of Injury Prevention in the CDC's Injury Center, where she has presented and published on various injury topics including suicide prevention, drug overdose, and nonfatal injury data. She volunteered for two COVID-19 response deployments, which involved monitoring and reporting cases among the general population and among people experiencing homelessness. Her current research interests focus on incorporating health equity into injury research.

Judy Qualters, PhD, MPH is the director of the Division of Injury Prevention (DIP) in the National Center for Injury Prevention and Control (NCIPC) at CDC. In this role, Dr. Qualters provides leadership to bridge science and practice in an effort to move the field of violence and injury prevention forward. She also leads a diverse portfolio of work that includes surveillance, data and economic analysis, information technology, policy research, evaluation, and technical assistance to state health departments.

Journal of Safety Research 82 (2022) 385–391

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

How many crashes does cellphone use contribute to? Population attributable risk of cellphone use while driving

Feng Guo [a,b,*], Danni Lu [b]

[a] Virginia Tech Transportation Institute, 3500 Transportation Research Plaza, Blacksburg, VA 24061, United States
[b] Department of Statistics, Virginia Tech, Blacksburg, VA 24061, United States

ARTICLE INFO

Article history:
Received 18 February 2022
Received in revised form 21 April 2022
Accepted 11 July 2022
Available online 25 July 2022

Keywords:
Population attributable risk
Naturalistic driving study
Case-cohort study
Distracted driving
Cellphone distraction

ABSTRACT

Background: Cellphone distraction is a major contributing factor for traffic crashes, a leading cause of death worldwide. The novel naturalistic driving study (NDS) study with continuously collected *in situ* driving videos provides an opportunity to accurately estimate the safety impact of cellphone distraction. *Methods:* We apply a case-cohort study design to the Second Strategic Highway Research Program NDS, the largest NDS up-to-date with more than 3400 participants. The data include with 842 level 1–3 crashes and 19,338 randomly selected control driving segments. We propose a partial Population Attributable Risk (PAR) estimator that provides consistent and stable estimation over time and across different driving behaviors. *Results:* The US population-adjusted PAR show that 8% of crashes (PAR = 0.08, 95 %CI: [0.06, 0.19]) can be reduced if cellphone distraction were switched to sober, alert, and attentive driving behavior. Young adults (age 20–29 years) and middle-aged drivers (age 30–64 years) each contribute 39% of the population level PAR. Within each age group, the PARs vary substantially from 18% for young adult drivers to 5% for middle-aged drivers. The contribution of cellphone visual-manual tasks to crashes is more than 4 times larger than cellphone talking and accounts for 87.5% of cellphone-related crashes (PAR = 0.07). *Conclusions:* Cellphone distraction contributes to a considerable part of crashes. Young drivers are more susceptible to the influence of cellphone distraction and visual-manual distraction accounts for the majority of cellphone-related crashes.

1. Introduction

Automobile crashes led to 33 654 fatalities and 1.8 million injuries in the Unites States in 2018 (NHTSA, 2019a). Driver behavior is a primary factor contributing to crashes, with studies showing as a critical reason in more than 90% of crashes (NHTSA, 2013). Distracted driving has been a predominant topic in driving safety studies, especially with the drastic increase in smartphone market penetration. Studies have shown that drivers spend an average of 6% of their time on a cellphone, while the percentage of time for young drivers could be as high as 11% during normal driving conditions (Dingus et al., 2016; Guo, Klauer, & et al., 2017).

There has been considerable amount of research investigating the crash risk associated with cellphone distraction using police reported crash data base or under experimental setups. Using cellphone records and a case-crossover design, Redelmeier and Tibshirani (1997) and McEvoy et al. (2005) found that cellphone engagement was associated with a four-fold increase in crash likelihood. Crash investigation can also identify crashes that are potentially related to driver distraction; for example, a study by the National Highway Traffic Safety Administration (NHTSA) indicates that 9% of fatal crashes in 2017 were reported as distraction-affected crashes (NHTSA, 2019b).

By analyzing the Fatality Analysis Reporting System, Hossain, Zhou, Das, Sun, and Hossain (2022) identified young drivers' additional risk-taking maneuvers while engaged in cellphone usage among fatal crashes, including: disregarding traffic signs and signals, speeding, and unrestrained driving. Using a simulator study, Choudhary, Gupta, and Velaga (2022) showed that drivers considered visual-manual tasks such as texting are extremely risk but tend to underestimate the risk of less risky behaviors. A simulator study showed that the mere presence of phone can be distractive for driving (Chee, Irwin, Bennett, & Carrigan, 2021). Simulator studies can provide rigorous control on the experimental environments and scenarios but lack actual crashes for risk assessment.

One issue with the traditional crash database is the lack of precise time information to catch the short, transient effect of driving behavior on crashes. The novel large-scale naturalistic driving

* Corresponding author at: Department of Statistics, Virginia Tech, Virginia Tech Transportation Institute, Blacksburg, VA 24061, United States.

E-mail address: feng.guo@vt.edu (F. Guo).

https://doi.org/10.1016/j.jsr.2022.07.005

study (NDS) method has become a major source for driver behavior evaluation through continuously collected real-life driving information (Guo, 2019). The large-scale NDS provides an opportunity to objectively observe driver behavior by continuously collecting driving data for an extended period of time through multiple sensors and cameras instrumented on participants' vehicles (Dingus et al., 2016, 2006). The video provides precise driver behavior information during both crashes and normal driving conditions. NDS has greatly expanded the understanding of risk associated with driver distraction (Farmer, Klauer, McClafferty, & Guo, 2015; Fitch, Soccolich, Guo, McClafferty, & Fang, 2013; Klauer et al., 2014, 2006).

Using data from the Second Strategic Highway Research Program (SHRP 2) NDS, Dingus et al. (2016) shows that drivers engaged in potential distracting activities among 51.9% of normal driving control segments and 68.3% of crashes. The study estimates a crash OR of 3.6 associated with overall cellphone distraction and an OR of 2.2 for cellphone talking. The ORs for cellphone visual-manual tasks are much higher, *e.g.* OR = 6.1 for texting and OR = 12.2 for dialing. In addition, studies have shown that the safety impact of distraction varies dramatically by driving experience and age (Atwood, Guo, Fitch, & Dingus, 2018; Guo, Klauer, & et al., 2017; Klauer et al., 2014; Liang & Yang, 2022). In addition to traditional epidemiological design, causal inference and advanced machine learning models has also been used in confirming the causal effects of cellphone use while driving (Lu, Guo, & Li, 2020; Lu, Tao, & et al., 2020). While substantial research has been conducted in risk assessment, there is limited research on the population attributable risk (PAR) of driver behaviors based on NDS.

Many states have imposed restrictions on cellphone use while driving (GHSA, 2019). However, there has been wide dispersed estimation of the percentage of crashes that can be prevented by eliminating cellphone distraction, varying from 3.6% to 22% (Farmer, Braitman, & Lund, 2010; Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006). Using a national survey of 1,200 US drivers, Farmer et al. (2010) estimated the population attributable risk of 22%. Using naturalistic driving data, Olson, Hanowski, Hickman, and Bocanegra (2009) stated that dialing cellphone contribute 3.31% (dialing: 2.46%, texting: 0.67%, talking: 0.18%) of safety–critical events in commercial motor vehicle operations. From a 100-car Naturalistic Driving Study, Klauer et al. (2006) concluded the population attributable risk of hand-held cellphone distraction as 3.6%. The inconsistency of PAR estimates in previous studies suggests a gap between the definition of PAR in epidemiology and its application to NDS.

The objective of this paper is to develop a rigorous PAR estimation methodology for time-variant factors with a transient effect on crashes using the NDS. The study addressed several key issues associated with large scale NDS, including the proper reference level, multi-level exposure, as well as the issue of adjustment for age group disparity between the NDS study population and the general driver population. We applied the developed method to estimate the PAR for cellphone distraction using the SHRP2 NDS, the largest NDS up to date, both methodologies and the estimated PAR by cellphone tasks contribute to the state-of-the-knowledge on the safety impacts of distraction.

The rest of this study is organized as follows. Section 2 presents the PAR estimation method in the NDS, considering the study design, the reference levels, and the multi-level exposure issues. Section 3 estimated the PAR for cellphone distraction using SHRP 2 NDS data. The summary and discussion are presented in Section 4.

2. Methodollgy

A large-scale NDS objectively collects the continuous real-world driving data at high resolution and frequency by equipping participants' vehicles with an unobtrusive data acquisition system (DAS). For example, the SHRP 2 NDS used a DAS consisting of four cameras, three-dimensional accelerometers, GPS, radar, *etc.* (Dingus et al., 2016). Driving behaviors, driving performance, environmental information, and safety outcome information were recorded from ignition-on to ignition-off at high frequency, *e.g.*, 10HZ.

Epidemiological approaches are typically used to process and analyze the massive amount of data collected by NDS. The PAR estimation proposed in this paper is based on a case-cohort study design (Guo, 2019). We focus on several key issues on multi-level exposure and partial PAR.

2.1. Estimating risk using NDS

Driver behavior information needs to be extracted by visually examining collected NDS videos. A large-scale NDS can collect millions of hours of continuous video, and it is not practical to extract all exposures. Although the NDS is a prospective cohort type study, case-based hybrid designs, such as case-cohort and case-crossover, are commonly used in evaluating the risk associated with driver behavior (Guo, 2019).

Cases, such as crashes and near crashes, can be identified using a variety of methods, including on-site identification and post-hoc processing (Hankey, Perez, & McClafferty, 2016). The post hoc processing involves screening the entire data collection for abnormal kinematic signatures, such as the high deceleration, followed by visual confirmation. The majority of severe crashes can be identified by these procedures.

Different study designs for NDS vary in how controls are selected to represent driver behavior under normal driving conditions. A control driving segment is typically a short driving segment, *e.g.*, 6-second driving segments randomly selected from the entire set of driving data (Dingus et al., 2016). The short driving segments make it feasible to extract driver behavior information from the video.

As the impact of distraction on crashes is imminent and transient, it is reasonable to assume that distraction in one driving segment is independent of the safety outcomes of other driving segments. Thus, the short driving segment samples satisfy the stable unit treatment value assumption to establish a causal relationship (Rubin, 1990).

Various designs can be used to select control samples for crash risk evaluation in NDS. The case-cohort design samples controls from the entire driving segment pool using a random sampling scheme stratified by driver (Barlow, Ichikawa, Rosner, & Izumi, 1999; Guo, 2019; Prentice, 1986). The number of controls for each driver is proportional to the driver's total driving time (Hankey et al., 2016). The case-crossover design is a matched study that samples controls from the driver who experienced the crash and also matches on other potential risk factors such as time-of-day and speed (Guo, Kim, & Klauer, 2017; Huisingh et al., 2019). The case-crossover can control for the potential confounding effects of matching factors, e.g., driver self-limit risky behavior under the match conditions, which could lead to smaller risk estimation. However, the matching mechanism limits the representation of the controls to only drivers who have experienced crashes and is not used in PAR estimation in this paper.

The randomly selected case-cohort controls provide an opportunity to evaluate both exposure prevalence and relative risk. When evaluating a specific driver distraction, the contingency table can be organized as in Table 1. Note that, there are three exposure levels in the table: behavior of interest, other behaviors, and model-driving. Model driving is sober, alert, and attentive driving, which means the driver is not engaged in any source of secondary tasks. When drivers were not engaged in the distraction of interest, they might be engaged either in other risky behaviors

Table 1
NDS contingency table.

	Behavior of interest	Other behaviors	Model-driving	Total
Crash	A	B	C	N_{crash}
Control	D	E	F	$N_{control}$
Total	$A+D$	$B+E$	$C+F$	N

$N_{crash} = A + B + C$ is total number of crashes.
$N_{control} = D + E + F$ is prespecified total number of control samples; $N = N_{crash} + N_{control}$.

or model-driving. The mixture of other behaviors could impose higher risk than the modeling driving, or even the behavior of interest.

The population exposure prevalence can be approximated using the exposure prevalence among control samples (Guo, 2019):

$$p_e^{control} = Prob(A\ control\ sample\ includes\ behavior\ of\ interest) \approx \frac{D}{N_{control}} = \frac{D}{D+E+F} \tag{1}$$

The relative risk of a risk factor can be measured by the odds ratio (OR), which represents the elevated risk when drivers engaged a risky behavior compared to a reference exposure level (Guo, 2019). The reference exposure level is a driver state with an implied risk level. Two commonly used reference levels are model-driving and all-driving. The model driving refers to the sober, alert, and attentive, i.e., an ideal driver state. The all-driving reference level includes other behaviors and model-driving. The model driving reference level generally reference a lower base risk compared to all-driving, therefore the OR with respect to model-driving is typically higher compared to all-driving. The corresponding crude OR can be conceptually illustrated in the following equations:

$$OR_{model} = \frac{A/C}{D/F}; \tag{2}$$

$$OR_{all} = \frac{A/(B+C)}{D/(E+F)}; \tag{3}$$

where OR_{model} and OR_{all} are the OR with model-driving and all-driving as the reference level respectively. In person-time cohort study, the OR estimated is a statistically consistent estimator of crash incidence density rate (IDR) under the assumption that the exposure prevalence and IDR are constant over time (Cummings, 2009; Guo, 2019; Miettinen, 1976). One driver can have multiple baselines and crashes, which will lead to correlation among these events. Instead of crude OR as in Eqs. (2) and (3), a mixed-effect logistic regression with driver-specific random effect is used for OR estimation (details can be found in Guo, 2019).

The two alternative reference levels are based on different assumptions and carry different interpretations (Guo, 2019). The all-driving reference level involves a composite of driver behaviors, which causes a reference level to depend on the exposure of interest. For example, the all-driving reference level for eating and cellphone use are different because eating is in the all-driving reference for cellphone but not when evaluate eating itself. Another issue is that driver behavior evolves over time; for example, cellphone distraction did not exist 20 years ago but now accounts for approximately 6% of total driving time. The OR based on all-driving is, therefore, potentially inconsistent across different studies and overtime. The model-driving reference level, on the other hand, is not affected by behavior of interest, which makes the risk estimations of different behaviors and studies directly comparable.

2.2. PAR for NDS

The PAR for a person-time cohort study can be calculated as:

$$PAR = \frac{p_e(R-1)}{1+p_e(R-1)} \tag{4}$$

where p_e is the prevalence of exposure; R is the relative risk measured by IDR (Cole & MacMahon, 1971; Miettinen, 1976). The PAR measures the percentage of crashes that can be prevented by eliminating a specific driving behavior. For a case-cohort NDS, the IDR R can be approximated by the OR as discussed shown in Eqs. (2) and (3). The multiple reference levels lead to alternative ways to estimate the PAR.

The reference level serves as the counterfactual in evaluating the causal effect of exposure, and therefore determines the estimand. For example, the PAR with respect to all-driving reference level is interpreted as the proportion of crashes that can be eliminated if the person-time for cellphone distraction were redistributed proportionally to other behaviors in all-driving reference level. The PAR defined on model-driving is interpreted as the proportion of crashes that can be eliminated if cellphone distraction were converted to model-driving behavior, while keeping prevalence of other distractions unchanged.

The PAR calculation for the all-driving reference level is straightforward by plugging in the prevalence in Eq. (1) and the OR in Eq. (3). However, the assumption that drivers will distribute the time spent on cellphone proportionally to other behavior will almost surely not be satisfied. In addition, the estimation will be invalid if different distractions emerge in the future, as the redistribute-proportion calculation is based on current samples. For the model-driving reference level, the OR estimation (based on Eq. (2)) is invariant to other distractions: it represents the elevated risk by cellphone use compared to an ideal driving status. However, since other behaviors category (B and E in Table 1) are excluded, the exposure prevalence estimation will be overestimated, which will lead to biased PAR estimation.

To address the aforementioned issue, we adopted the full PAR (PAR_F) and partial PAR (PAR_P) concept to accommodate multiple risk factors (Spiegelman, Hertzmark, & Wand, 2007). The objective is to provide a stable and invariant estimation of the preventable risk for cellphone use. By using model-driving as reference level, the PAR estimation will be invariant for variation in prevalence and risk of other distractions. The partial PAR approach factors in the effects of multiple exposure types thus addresses the issue of biased PAR estimation by excluding other distractions. The partial PAR evaluates the proportion of incidents that can be prevented if one or some exposure levels were unexposed while keeping other levels unmodified (Coughlin, Benichou, & Weed, 1994; Spiegelman et al., 2007). For example, the partial PAR of cellphone distraction estimates the proportion of reduced crashes if all cellphone distraction person-time were converted to model-driving person-time, while other behaviors are kept the same.

The formulation for the full and partial PAR is derived as follows. Without loss of generality, consider three exposure levels indexed by $i : i = 0, 1, 2$: model-driving ($i = 0$), distraction of interest ($i = 1$), and other distractions ($i = 2$). For the distraction status

i, let the duration of distraction be T_i , the incidence density as I_i, and the prevalence of the exposure as $p_i = \frac{T_i}{\sum T_i}$. The incident density I_i is the ratio of case frequency divided by the exposure time T_i. The expected total number of crashes can be expressed as $\Sigma_{i=0}^{2} T_i I_i$. The reduced crashes if *all distractions* (I = 1,2) were converted to model-driving is $T_1(I_1 - I_0) + T_2(I_2 - I_0)$. The full PAR can be calculated as follows:

$$\begin{aligned} PAR_{Full} &= \frac{T_1(I_1-I_0)+T_2(I_2-I_0)}{I_1T_1+I_2T_2+I_0T_0} = \frac{T_1(I_1/I_0-1)+T_2(I_2/I_0-1)}{(I_1/I_0)T_1+(I_2/I_0)T_2+T_0} \\ &= \frac{\frac{T_1}{T_0+T_1+T_2}(I_1/I_0-1)+\frac{T_2}{T_0+T_1+T_2}(I_2/I_0-1)}{\frac{T_1}{T_0+T_1+T_2}(I_1/I_0)+\frac{T_2}{T_0+T_1+T_2}(I_2/I_0)+\frac{T_0}{T_0+T_1+T_2}} \\ &= \frac{p_1(R_1-1)+p_2(R_2-1)}{R_1p_1+R_2p_2+R_0p_0} = \frac{\Sigma_{j\in Fleft\{0\}}p_j(R_j-1)}{\Sigma_{i=0}^{2}R_ip_i} \end{aligned}$$

where F is the full set of all exposure status $F = \{0,1,2\}$; R_i is the IDR of exposure i with respect to reference level 0: $R_i = I_i/I_0$. As shown in Guo (2019), the IDR can be approximated by the OR (Eq. (2)).

Using the same notation as the full PAR, the attributable crash risk of a specific driving behavior (e.g., i = 1) were converted to model-driving while other distractions were kept the same is $I_1T_1 - I_0T_1$. The corresponding partial PAR is:

$$\begin{aligned} PAR_{Partial} &= \frac{T_1(I_1 - I_0)}{I_1T_1 + I_2T_2 + I_0T_0} = \frac{T_1(I_1/I_0 - 1)}{(I_1/I_0)T_1 + (I_2/I_0)T_2 + T_0} \\ &= \frac{p_1(R_1 - 1)}{R_1p_1 + R_2p_2 + R_0p_0} = \frac{p_1(R_1 - 1)}{\Sigma_{i=0}^{2}R_ip_i} \end{aligned}$$

In summary, the PAR can be calculated as follows:

$$PAR_{Full} = \frac{\Sigma_{j\in Fleft\{0\}}p_j(R_j - 1)}{\Sigma_i R_i p_i} \quad (5)$$

$$PAR_{Partial} = \frac{\Sigma_{j\in K}p_j(R_j - 1)}{\Sigma_i R_i p_i} \quad (6)$$

where F is the full set of all exposure status with index 0 as the reference level; R_i is the IDR of exposure i with respect to reference level 0 (Eq. (2) for NDS case cohort studies); p_i is the prevalence of exposure i which can be calculated by Eq. (1); K is the subset of exposure of interest. For example, the PAR of a specific driving behavior (e.g., i = 1) converting to model-driving is:

$$PAR_P = \frac{p_1(R_1 - 1)}{\Sigma_{i=0}^{2}R_ip_i}$$

Notably, the full PAR (Eq.5) is a special case of the partial PAR (Eq.6) where all exposures are of interest. In the studies that focus only on one specific driving behavior, $K = \{1\}$. Eq. (6) provides a way to estimate the PAR by converting a specific behavior to model-driving for NDS case-cohort study.

2.3. SHRP 2 NDS case-cohort data

The SHRP 2 NDS was a large-scale NDS conducted from 2010 to 2013, with more than 3,400 participant drivers from six states: Washington, Florida, New York, North Carolina, Pennsylvania, and Indiana (Dingus et al., 2016; Hankey et al., 2016). Crashes are classified into four severity levels from Level 1, the most severe, to Level 4, low-risk tire strikes (Hankey et al., 2016). Level 1 crashes involve airbag deployment, injury, or vehicle towing due to damage; Level 2 includes crashes of minimum of $1500 worth of damage, or crashes with acceleration on any axis greater than ±1.3 g. Level 3 is defined as crashes where a vehicle makes physical contact with another object or departs the road but sustains only minimal or no damage. Level 4 crashes are tire strike low risk crashes, e.g., clipping a curb during a tight turn. In this study, we focus on the relatively severe crashes including a total of 842 levels 1, 2, and 3, crashes.

The NDS case-cohort samples include 19,338 randomly selected 6-second driving segments. The driving behaviors prior to the crashes and within the controls were coded following a rigorous video reduction protocol (Hankey et al., 2016). Cellphone distraction was aggregated into two subcategories: cellphone talking and cellphone visual-manual tasks. Cellphone talking includes primarily handheld cellphone. Cellphone visual-manual tasks include dialing, reaching, browsing, and texting, etc. (see Table 2).

The SHRP 2 NDS over-sampled young and senior drivers as they are at high crash risk but account for a small portion of the driver population compared to middle-aged drivers. To eliminate the impact of selection bias on PAR estimation, we adopt a weighting approach to estimate the PAR for the U.S. driver population:

$$PAR_{US.pop} = \Sigma_{i=1}^{4} w_i PAR^i \quad (7)$$

where w_i is the percentage of licensed drivers for i^{th} age group in the target population; PAR^i is the PAR of the i^{th} age group. The weight w_i is estimated by the percentage of licensed drivers by age group according to the Federal Highway Administration (FHWA, 2018):

$$w_i = \mathrm{n}_{i/N}$$

where n_i is the total number of licensed drivers in age group i and N is the total number of all licensed drivers.

The PAR contribution percentage of each age group is calculated as:

$$PAR\% = \frac{w_i PAR^i}{\Sigma_j w_j PAR^j}$$

The contribution percentage indicates the percentage of reduced PAR if drivers in a certain age group switch to model-driving.

3. Results

We estimated the PAR for cellphone use while driving using the SHRP 2 NDS and the partial PAR discussed above. As cellphone use prevalence and relative risk vary substantially by driver age, an age-stratified analysis was conducted based on four age groups: teenage (16–19 years old), young adult (20–29 years old), middle-aged (30–64 years old), and senior drivers (65 + years old). We used a population-weighted approach to estimate the PAR for the entire U.S. driver population.

The exposure status by age group is shown in Table 3. Notably, around 50% of the time, drivers were distracted. Cellphone use prevalence varies drastically from 1% for senior drivers to 11% for young adult drivers. The percentage of cellphone-talking and visual-manual tasks varies depending on age groups. Younger drivers tend to be involved in more visual-manual tasks than talking, while middle-aged drivers are the opposite.

3.1. Results of PAR for cellphone use while driving

The PAR of cellphone distraction was evaluated for each age group with model-driving as the reference level. As a driver might have multiple cases and controls, a mixed-effect logistic regression with driver-specific random effect was used to estimate the OR (Guo, 2019). The partial PARs of overall cellphone distraction and its subtasks were estimated using Eq.6. The 95% confidence interval (CI) was calculated using a clustered bootstrap method with

Table 2
The number of crashes and baselines.

Age	Teenage (16–19)	Young adult (20–29)	Middle-age (30–64)	Senior (65+)	Total
Crashes	221	301	157	163	842
Baselines	2608	6186	6117	4427	19,338
Total	2829	6487	6274	4590	

each driver as a cluster (Ren et al., 2010). The PAR by age group and cellphone subtasks is shown in Table 4.

The PAR varies considerably among age groups. Cellphone distraction contributes the most for young adult drivers, with a PAR of 18% (CI: 0.12–0.26). For teenage drivers, 15% (CI: 0.09–0.26) of crashes are attributed to cellphone distraction. In comparison, the decrease in crashes is smaller for middle-aged and senior drivers, with 5% and 7% PARs, respectively.

Cellphone visual-manual tasks contribute substantially more crashes than cellphone talking for all age groups. For example, for the 15% PAR for teenage drivers, 12% are contributed by cellphone visual-manual tasks and only 3% are contributed by cellphone talking. Similarly, for the 5% PAR for middle-aged drivers, 4% are due to visual-manual tasks and 1% are due to cellphone talking.

To evaluate the influence of alternative reference levels, we compared the PARs using "all-driving" condition with the partial PARs: Teenage (11% vs 15%) Young adults (12% vs 18%), middle age (4% vs 5%), and Senior (3% vs 7%). As the results shown, the PARs based on all-driving reference level are universally smaller. Since the all-driving condition includes various types of distractions, the base risk is higher than model driving condition, which leads to the low PARs based on all-driving reference level.

The population weight and the overall cellphone use PAR are shown in Table 5. Overall, 8% crashes (CI: 0.06–0.19) can be prevented if all the cellphone-distracted driving in the U.S. driver population was switched to model-driving. Among the 8% PAR for the U.S. driver population, young adults (20–29) account for 39% of the population PAR, although they only account for the 17% of the driver population. Middle-aged drivers compose 63% of the total

Table 3
Multi-level distraction status of SHRP 2 NDS.

Age	Teenage (16–19)	Young adult (20–29)	Middle-age (30–64)	Senior (65+)
Model-driving	1153* (41%)	2653 (41%)	3017 (48%)	2707 (59%)
Other distraction	1424 (50%)	3099 (48%)	2921 (47%)	1838 (40%)
Overall cellphone	252 (9%)	735 (11%)	336 (5%)	45 (1%)
Talking	80 (3%)	339 (5%)	195 (3%)	31 (1%)
Visual manual	172 (6%)	396 (6%)	141 (2%)	14 (0%)
Total	2829	6487	6274	4590

*The table includes both controls and cases.

Table 4
P artial par of cellphone distraction (model-driving as reference).

	Prevalence (p_i)*	OR_i	PAR_P*	95 %CI
Teenage (16–19)				
Talking	0.03	2.41	3%	(0.00, 0.08)
Visual manual	0.06	4.20	12%	(0.07, 0.22)
Overall cellphone			15%	(0.09, 0.26)
Young adult (20–29)				
Talking	0.05	2.03	3%	(0.00, 0.06)
Visual manual	0.06	5.90	15%	(0.10, 0.23)
Overall cellphone			18%	(0.12, 0.26)
Middle-age (30–64)				
Talking	0.03	1.53	1%	(0.00, 0.06)
Visual manual	0.02	3.00	4%	(0.00, 0.09)
Overall cellphone			5%	(0.00, 0.11)
Senior (65+)				
Talking	0.01	0.99	<1%	(0.00, 0.02)
Visual manual	<0.01	29.97	7%	(0.03, 0.80)
Overall cellphone			7%	(0.03, 0.80)

* PAR_p and p_i are defined in Eq. (6).

Table 5
P opulation level PAR and age contribution percentages.

	w_i*	PAR^i %		
		T alking	V isual M Il	O verall
Teenage	0.051	12%	9%	10%
Young adult	0.176	40%	39%	39%
Middle-aged	0.630	48%	37%	39%
Senior	0.143	<1%	15%	12%
PAR_{USPOP}				
Talking		1% (<0.01, 0.04)		
Visual manual		7% (0.04, 0.17)		
Overall cellphone		8% (0.06, 0.19)		

* wi is the percentage of licensed drivers for i^{th} age group in target population (Eq. (7)).

driver population in the U.S. and account for a similar percentage of the PAR. Teenage and senior drivers contribute to 10% and 12% of the overall PAR. Compared with talking, visual-manual tasks dominated the PAR of overall cellphone use. Out of 8% of preventable crashes due to cellphone distraction, 7% can be prevented by eliminating cellphone visual-manual tasks, and 1% can be prevented by eliminating cellphone talking.

4. Summary and discussion

Cellphones use, especially smartphones, imposes substantial cognitive and visual-manual burden on drivers and has been a major contributing factor for traffic crashes with the nearly saturated market penetration in recent years. The continuously collected large-scale NDS data provide crucial information regarding the risk associated with distraction and its culpability in crashes. This paper presents a methodology to estimate the PAR associated with risk factors with transient effects on crashes and estimates the PAR of cellphone use based on the SHRP 2 NDS.

The results show that cellphone use contributes to 8% of the crashes in the U.S. There is a substantial variation by age group, with cellphones contributing 18% to crashes for young adult drivers and only 5% for middle-aged drivers. It is of particular concern that young drivers contribute to 39% of cellphone-related crashes but only account for 17% of the driver population. Both talking and visual-manual tasks contribute to crashes, but visual-manual has a much higher PAR (4 times higher or more).

Compared to existing cellphone use PAR literatures, our overall estimation of 8% is the neither the highest or the lowest, e.g. 22% by Farmer et al. (2010), 3.31% by Olson et al. (2009), and 3.36% by Klauer et al. (2006). The high discrepancy among studies can be caused by many factors, e.g., data source, analysis method, and timeline. Note many of these studies were published a decade ago, the cellphone prevalence and the percentage of smartphones have changed dramatically. Each PAR estimation reflects the specific scenario represented by the samples thus not directly comparable. The main contribution of this study lies in it provides a tailored PAR framework for the NDS with consideration for multiple issues such as reference level and age variation.

The PARs estimated in this study should be interpreted as the proportion of crashes that can be prevented if all cellphone use were switched to sober, alert, and attentive driving. A potential alternative reference level is all-driving, which assumes that cellphone use units were proportionally converted to other driving behaviors. As discussed in Guo (2019), the all-driving reference level does vary by different secondary tasks, e.g., all-driving condition for cellphone use is different for that of adjusting in-vehicle devices tasks. The relative risk based on the all-driving is therefore not directly comparable. In addition, given the constantly evolving vehicle technology and driver behavior, the all-driving condition could also change over time. The model-driving reference level has the advantage of consistency for all risk factors and over time. One could argue that a driver may not turn all the cellphone use time to model-driving as the model-driving represents an ideal driving status. Therefore, the PAR based on model-driving status implies maximum potential benefits by eliminating cellphone use while driving.

As the risk of cellphone use while driving has been confirmed by numerous studies, to reduce cellphone distraction as well as its impacts on safety has become a major concern. Coben and Zhu (2013) advocated strict federal standards to ban cellphone use while driving. A number of studies have shown that handheld ban was associated with lower cellphone use prevalence (Carpenter & Nguyen, 2015; Rudisill & Zhu, 2017). Other technologies such as cellphone blockers also have potential to reduce cellphone use (Reagan & Cicchino, 2020). The advance in hand-free devices could mitigate the effects of visual-manual distraction. However, these technologies might encourage hand-free talking, whose risk has not been fully understood yet.

There are several limitations to the current study. The data were collected from six sites across the U.S. and might not represent the entire U.S. driver population. Antin, Stulce, Eichelberger, and Hankey (2015) conducted an extensive comparison of SHRP2 driver population with national sample. The study reveals the difference and advocates population-based adjustment to improve the representativeness. The age-adjustment approach in this study can help alleviate the representative issue. The crashes collected do not include the most-severe crashes, i.e., fatal crashes, as such the PAR may not represent the attributable risk of fatal crashes. Due to the extreme rarity of fatal crashes and high cost of NDS, it is not likely to observe sufficient numbers of fatal crashes in future NDS. Alternative study methods such as accident reconstruction, cellphone records, are needed to examine the causal effects of cellphone use and fatal crashes. Lastly, caution should be used when extrapolating the results beyond what the SHRP2 NDS data represented as the cellphone use prevalence and relative risk might change with advancements in technology such as voice-controlled user interfaces.

5. Conclusions

This study quantitatively estimated the PAR of cellphone use while driving based on the novel NDS and a partial PAR framework. The results reveal that around 8% of crashes can be attributed to cellphone use. The attributable risk of cellphone use is primarily caused by visual-manual tasks. Cellphone use while driving imposes a substantially greater hazard for young drivers, reflected in both the elevated OR and PAR. Appropriate in-vehicle cellphone user interfaces, driver education programs, regulation, and enforcement are needed to reduce cellphone-related crashes.

References

Antin, J., Stulce, K., Eichelberger, L., Hankey, J. M. (2015). Naturalistic driving study: Descriptive comparison of the study sample with national data (No. SHRP 2 Report S2-S31-RW-1).

Atwood, J., Guo, F., Fitch, G. M., & Dingus, T. A. (2018). The driver-level crash risk associated with daily cellphone use and cellphone use while driving. *Accident Analysis and Prevention, 119*, 149–154.

Barlow, W. E., Ichikawa, L., Rosner, D., & Izumi, S. (1999). Analysis of case-cohort designs. *Journal of Clinical Epidemiology, 52*(12), 1165–1172.

Carpenter, C. S., & Nguyen, H. V. (2015). Effects of a driver cellphone ban on overall, handheld, and hands-free cellphone use while driving: New evidence from Canada. *Health Economics, 24*(11), 1452–1467.

Chee, P., Irwin, J., Bennett, J. M., & Carrigan, A. J. (2021). The mere presence of a mobile phone: Does it influence driving performance? *Accident Analysis & Prevention, 159*, 106226.

Choudhary, P., Gupta, A., & Velaga, N. R. (2022). Perceived risk vs actual driving performance during distracted driving: A comparative analysis of phone use and other secondary distractions. *Transportation Research Part F: Traffic Psychology and Behaviour, 86*, 296–315.

Coben, J. H., & Zhu, M. (2013). Keeping an eye on distracted driving. *JAMA – Journal of the American Medical Association, 309*(9), 877–878.

Cole, P., & MacMahon, B. (1971). Attributable risk percent in case-control studies. *Journal of Epidemiology and Community Health, 25*(4), 242–244.

Coughlin, S. S., Benichou, J., & Weed, D. L. (1994). Attributable risk estimation in case-control studies. *Epidemiologic Reviews, 16*(1), 51–64.

Cummings, P. (2009). The relative merits of risk ratios and odds ratios. *Archives of Pediatrics and Adolescent Medicine, 163*(5), 438–445.

Dingus, T. A., Guo, F., Lee, S. E., Antin, J. F., Perez, M., Buchanan-King, M., et al. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences, 113*(10), 2636–2641.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., et al. (2006). *The 100-car naturalistic driving study (Phase II)*. Natl. Highw. Traffic Saf. Adm..

Farmer, C. M., Braitman, K. A., & Lund, A. K. (2010). Cell phone use while driving and attributable crash risk. *Traffic Injury Prevention, 11*(5), 466–470.

Farmer, C. M., Klauer, S. G., McClafferty, J. A., & Guo, F. (2015). Secondary behavior of drivers on cell phones. *Traffic Injury Prevention, 16*(8), 801–808.

FHWA (2018). Licensed drivers by age and sex (WWW Document). URL https://www.fhwa.dot.gov/ohim/onh00/bar7.htm.

Fitch, G. M., Soccolich, S. A., Guo, F., McClafferty, J., Fang, Y., Olson, et al. (2013). The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk final report.

GHSA (2019). Distracted driving laws by state updated April 2019 [WWW Document]. URL http://www.ghsa.org/state-laws/issues/Distracted-Driving.

Guo, F. (2019). Statistical methods for naturalistic driving studies. *The Annual Review of Statistics and its Applications, 6*, 309–328.

Guo, F., Kim, I., & Klauer, S. G. (2017a). Semiparametric Bayesian models for evaluating time-variant driving risk factors using naturalistic driving data and case-crossover approach. *Statistics in Medicine, 2017*, 1–15.

Guo, F., Klauer, S. G., Fang, Y., Hankey, J. M., Antin, J. F., Perez, M. A., et al. (2017). The effects of age on crash risk associated with driver distraction. *International Journal of Epidemiology, 46*(1), 258–265.

Hankey, J. M., Perez, M. A., & McClafferty, J. A. (2016). *Description of the SHRP2 naturalistic database and the crash, near-crash, and baseline data sets task report*. Blacksburg, VA: The Strategic Highway Research Program 2 Transportation Research Board of The National Academies.

Hossain, M. M., Zhou, H., Das, S., Sun, X., & Hossain, A. (2022). Young drivers and cellphone distraction: Pattern recognition from fatal crashes. *Journal of Transportation Safety & Security*.

Huisingh, C., Owsley, C., Levitan, E. B., Irvin, M. R., MacLennan, P., & McGwin, G. (2019). Distracted driving and risk of crash or near-crash involvement among older drivers using naturalistic driving data with a case-crossover study design. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 74*(4), 550–555.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near crash/crash risk: An analysis using the 100-car naturalistic driving study data*. National Highway Traffic Safety Administration.

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England Journal of Medicine, 46*(4), 600–601.

Liang, O. S., & Yang, C. C. (2022). How are different sources of distraction associated with at-fault crashes among drivers of different age gender groups? *Accident Analysis & Prevention, 165*, 106505.

Lu, D., Guo, F., & Li, F. (2020). Evaluating the causal effects of cellphone distraction on crash risk using propensity score methods. *Accident Analysis & Prevention, 143*, 105579.

Lu, D., Tao, C., Chen, J., Li, F., Guo, F., & Carin, L. (2020). Reconsidering generative objectives for counterfactual reasoning. *Advances in neural information processing systems (NeurIPS)*.

McEvoy, S. P., Stevenson, M. R., McCartt, A. T., Woodward, M., Haworth, C., Palamara, P., et al. (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: A case-crossover study. *British Medical Journal, 331*(7514), 428–430.

Miettinen, O. (1976). Estimability and estimation in case-referent studies. *American Journal of Epidemiology, 103*(2), 226–235.

NHTSA (2019a). Police-reported motor vehicle traffic crashes in 2018. Washington, DC.

NHTSA (2019b). Distracted driving in fatal crashes, 2017, Traffic Safety Facts.

NHTSA (2013). National motor vehicle crash causation survey: Report to congress. Washington, DC.

Olson, R. L., Hanowski, R. J., Hickman, J. S., & Bocanegra, J. (2009). Driver distraction in commercial vehicle operations.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika, 73*(1), 1–11.

Reagan, I. J., & Cicchino, J. B. (2020). Do not disturb while driving – Use of cellphone blockers among adult drivers. *Safety Science, 128*, 104753.

Redelmeier, D. A., & Tibshirani, R. J. (1997). Interpretation and bias in case-crossover studies. *Journal of Clinical Epidemiology, 50*(11), 1281–1287.

Ren, S., Lai, H., Tong, W., Aminzadeh, M., Hou, X., & Lai, S. (2010). Nonparametric bootstrapping for hierarchical data. *Journal of Applied Statistics, 37*(9), 1487–1498.

Rubin, D. B. (1990). Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science, 5*(4), 472–480.

Rudisill, T. M., & Zhu, M. (2017). Hand-held cell phone use while driving legislation and observed driver behavior among population sub-groups in the United States. *BMC Public Health, 17*(1), 1–10.

Spiegelman, D., Hertzmark, E., & Wand, H. C. (2007). Point and interval estimates of partial population attributable risks in cohort studies: Examples and software. *Cancer Causes and Control, 18*(5), 571–579.

Dr. Feng Guo is a professor of statistics at Virginia Tech and a lead data scientist at the Virginia Tech Transportation Institute. He holds dual Ph.D. in transportation engineering and statistics from the University of Connecticut. His major research areas include analyses of naturalistic driving studies, transportation infrastructure safety evaluation, advanced vehicle proactive safety device evaluation, and automated driving systems. He has served as the Chair of the Transportation Statistics Interest group of American Statistical Association, a member of the Transportation Research Board Committee on Statistical Methods, and a member of the Committee on Safety Data, Analysis, and Evaluation of the National Academies.

Danni Lu is an applied scientist in Safety and Insurance, Uber Technology Inc. She received a Ph.D. degree in statistics from Virginia Tech, a Ph.D. degree in transportation planning and management from Tongji University. Her work mainly focuses on transportation safety modeling, causal inference, and transportation planning and management. She was the recipient of the Boyd Harshbarger Award for outstanding academic achievement in 2016 and the Outstanding Poster Award for Summer Program on Transportation Statistics (SAMSI) in 2017.

Journal of Safety Research 82 (2022) 463–468

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Special Report from the CDC

Increased unintentional drowning deaths in 2020 by age, race/ethnicity, sex, and location, United States ☆

Briana Moreland [a,b,*], Neil Ortmann [a,c], Tessa Clemens [a]

[a] Division of Injury Prevention, National Center of Injury Prevention and Control, Centers for Disease Control and Prevention, United States
[b] Cherokee Nation Operational Solutions, United States
[c] Oak Ridge Institute for Science and Education, United States

ARTICLE INFO

Article history:
Received 21 June 2022
Accepted 22 June 2022
Available online 2 July 2022

Keywords:
Drowning
Disparities
COVID-19
Children
Young adults

ABSTRACT

Introduction: During the COVID-19 pandemic, one study in Australia showed an increase in drowning deaths in certain settings, while a study in China showed a decrease in drowning deaths. The impact of the COVID-19 pandemic on drowning deaths in the United States is unknown. Objective: To report on unintentional drowning deaths among U.S. persons aged ≤29 years by demographic characteristics and compare 2020 fatal drowning rates with rates from 2010 to 2019. *Methods:* Data from CDC WONDER were analyzed to calculate unintentional drowning death rates among persons aged ≤29 years by age group, sex, race/ethnicity, and location of drowning. These rates were compared to drowning death rates for the previous 10 years (2010–2019). *Results:* In 2020, 1.26 per 100,000 persons aged ≤29 years died from unintentional drowning, a 16.79% increase from 2019. Drowning death rates decreased 1.81% per year on average (95% CI: −3.02%, −0.59%) from 2010 to 2019. The largest increases in unintentional drowning deaths from 2019 to 2020 occurred among young adults aged 20 to 24 years (44.12%), Black or African American persons (23.73%), and males (19.55%). The location with the largest increase in drowning was natural water (26.44%). *Conclusion***:** Drowning death rates among persons aged ≤29 years significantly increased from 2019 to 2020. Further research is needed to understand the impacts of the COVID-19 pandemic on drowning and identify how drowning prevention strategies can be adapted and strengthened. *Practical applications:* Drowning remains a leading cause of injury death among persons aged ≤29 years. However, drowning is preventable. Interventions such as learning basic swimming and water safety skills, and consistent use of lifejackets on boats and among weaker swimmers in natural water, have the potential to reduce drowning deaths. Developing strategies that ensure equitable access to these interventions may prevent future drowning.

1. Introduction

Every year in the United States there are around 4,000 unintentional drowning deaths and another 8,000 nonfatal drownings that require treatment in an emergency department (CDC WISQARS). Drowning is one of the top three leading causes of unintentional injury death among persons aged ≤29 years (CDC WISQARS). More children aged 1–4 years die from drowning than from any other cause, except birth defects (CDC WISQARS). Among individuals aged ≤29 years, American Indian or Alaska Native persons and Black or African American persons have the highest drowning rates (Clemens, Moreland, & Lee, 2021). Despite a steady decrease in drowning deaths among persons aged ≤29 years overall in the past two decades, racial and ethnic disparities in drowning deaths persist (Clemens et al., 2021). The most frequent locations where drowning occurs differ by age group (Clemens et al., 2021). Infants under the age of 1 year most often drown in bathtubs, children aged 1–4 years in swimming pools, and persons aged 15 years and older in natural water (Clemens et al., 2021; Denny et al., 2021).

The COVID-19 pandemic and the associated public health response had a sudden and dramatic impact on the behaviors and lifestyles of people in the United States (Chen et al., 2021; Giuntella, Hyde, Saccardo, & Sadoff, 2021) that may have impacted

☆ The Journal of Safety Research has partnered with the Office of the Associate Director for Science, Division of Injury Prevention, National Center for Injury Prevention and Control at the CDC in Atlanta, Georgia, USA, to briefly report on some of the latest findings in the research community. This report is the 70th in a series of "Special Report from the CDC" articles on injury prevention.

* Corresponding author at: 4770 Buford Hwy, NE, Atlanta, GA 30341, United States.

E-mail address: bmoreland@cdc.gov (B. Moreland).

https://doi.org/10.1016/j.jsr.2022.06.012
0022-4375/

fatal and nonfatal injury rates. Several efforts have been made to assess the impacts of COVID-19 on general injury trends globally, with countries reporting both increases and decreases in injuries. For example, studies found significant increases in trauma patient admissions from May to June of 2020 in Italy and from March 2020 to February 2021 in the United States (Giudici et al., 2021; Moore et al., 2022) and significant increases in unintentional injury deaths from March to August 2020 in the United States (Faust et al., 2021). However, other studies found the number of injuries seen in some trauma centers significantly decreased while local COVID-19 stay-at-home orders were in place in the United States and England (Sherman et al., 2021; Sephton et al., 2021) and all-cause injury mortality significantly declined during the COVID-19 pandemic in Guangdong, China (Zheng et al., 2021).

Studies investigating the impact of the COVID-19 public health response on drowning mortality are limited. One study conducted in Guangdong, China found drowning deaths had decreased by 35% during the COVID-19 pandemic (January–June 2020) compared to the same timeframe in the prior year (Zheng et al., 2021). Another study examined unintentional coastal drowning fatalities in Australia and found the risk of drowning was 1.75 times higher during the COVID-19 restriction period (March – June 2020) compared to the previous 15 years on average (Lawes, Strasiotto, Daw, & Peden, 2021). Specifically, drowning while boating or using personal watercraft had the largest increase (88%; Lawes et al., 2021). Increases in drowning during rock fishing (60%) and unpowered watercraft activities (33%) were also reported. However, drowning related to swimming/wading (−50%) and snorkeling/diving (−67%) significantly decreased during the COVID-19 restriction period (Lawes et al., 2021).

Determining rates of drowning deaths during the COVID-19 pandemic may inform future public health strategies to prevent drowning, specifically during infrastructure disruption, and improve access to interventions for persons at the highest risk. The objective of this study is to describe drowning death rates among persons aged ≤29 years in 2020 and to compare these rates to those in the previous decade (2010–2019) by demographic characteristics and drowning locations.

2. Methods

Mortality data from the National Vital Statistics System (NVSS) were accessed through the Centers for Disease Control and Prevention's Wide-ranging Online Data for Epidemiologic Research (CDC WONDER) and analyzed to describe unintentional drowning deaths among persons aged ≤29 years from 2010 to 2020 in the United States. Through a collaboration with the National Center for Health Statistics and state and local jurisdictions, NVSS includes records of all deaths in the United States. Crude drowning death rates per 100,000 population and 95% confidence intervals from 2010 to 2020 were calculated by demographic characteristics and drowning locations. Rates were analyzed overall and by 5-year age groups (<1, 1–4, 5–9, 10–14, 15–19, 20–24, 25–29), sex (female, male), race/ethnicity (Non-Hispanic American Indian or Alaska Native [AI/AN], Non-Hispanic Asian or Pacific Islander [A/PI], Non-Hispanic Black or African American [Black], Non-Hispanic White [White], and Hispanic or Latino all races [Hispanic]), and location of drowning (bathtub, pool, natural water, watercraft, other/unspecified). Where data on ethnicity was not stated, deaths were included in the total count but were not included in the race/ethnicity counts. *International Classification of Diseases*, Tenth Revision (ICD-10) codes (W65-W74, V90, V92) were used to identify drowning-related deaths and to describe the location of drowning deaths (bathtub: W65, W66; pool W67, W68; natural water: W69, W70; other/unspecified: W73, W74; watercraft: V90, V92). Location of drowning deaths were further analyzed by sex, age group, and race and ethnicity.

The annual percent change (APC), corresponding 95% confidence intervals (CI), and significant changes in trends in crude drowning rates from 2010 to 2019 were analyzed using JoinPoint software (version 4.9.0.1). Trends were analyzed overall among persons aged ≤29 years and by age group, sex, race/ethnicity, and drowning location. The percent change from 2019 to 2020 and corresponding 95% CIs were calculated and compared to the APC from 2010 to 2019 overall and by demographic characteristics and location of drowning to describe deviations from the previously identified trends (Ingram et al., 2018).

3. Results

In 2020, 1,589 persons aged ≤29 years died from unintentional drowning (Table 1). From 2010 to 2019 unintentional drowning death rates decreased by 1.81% per year (95% CI: −3.02%, −0.59%) (Fig. 1). However, from 2019 to 2020, the rate of unintentional drowning deaths increased by 16.79% (95% CI: 8.35%, 25.24%) among persons aged ≤29 years. Rates of unintentional drowning were 3.48 times higher among males compared to females in 2020. Rates increased by 19.55% (95% CI: 9.73%, 29.37%) among males from 2019 to 2020 and decreased 2.38% per year (95% CI: −3.53%, −1.22%) from 2010 to 2019.

Unintentional drowning death rates were highest among children aged 1–4 years (2.73 per 100,000 population). Rates did not significantly change from 2010 to 2019 (APC −0.03%; 95% CI: −1.56%, 1.06%) or from 2019 to 2020 (APC 12.88%; 95% CI: −2.72%, 28.48%) for this age group (Fig. 2). However, rates increased by 44.12% (95% CI: 19.76%, 68.48%) from 2019 to 2020 among persons aged 20 to 24 years and by 28.55% (95% CI: 7.61%, 49.49%) among persons aged 25 to 29 years. From 2010 to 2019, unintentional drowning rates decreased by 3.79% per year (95% CI: −5.69%, −1.84%) among persons aged 20 to 24 years and by 2.12% per year (95% CI: −3.83%, −0.38%) among persons aged 25 to 29 years.

In 2020, for individuals aged ≤29 years, Black persons had unintentional drowning rates 1.82 times higher than White persons, while AI/AN persons had rates 1.77 times higher than White persons. From 2010 to 2019 the rate of unintentional drowning among Black persons aged ≤29 years did not significantly change (APC −1.10%; 95% CI: −2.74%, 0.57%) (Fig. 3). However, from 2019 to 2020 the rate increased by 23.73% (95% CI: 5.35%, 42.10%). Rates of unintentional drowning decreased by 2.29% per year (95% CI: −3.68%, −0.88%) among White persons aged ≤29 years from 2010 to 2019 and increased by 15.67% (95% CI: 3.48%, 27.87%) from 2019 to 2020.

The most common location for unintentional drowning deaths among persons aged ≤29 years in 2020 was natural water (47.95%) followed by swimming pools (24.17%). The location varied by age group. The age group with the highest drowning rate in bathtubs was infants aged < 1 year (0.67 per 100,000), in swimming pools was children aged 1–4 years (1.62 per 100,000), and in natural water was persons aged ≥15 years (0.86 per 100,000 persons aged 15–19 years; 0.90 per 100,000 persons aged 20–24 years; 0.83 per 100,000 persons aged 25–29 years). From 2010 to 2019 the rate of unintentional drowning among persons aged ≤29 years in natural water decreased by 2.72% per year (95% CI: −3.70%, −1.72%) and from 2019 to 2020 increased by 26.44% (95% CI: 12.96%, 39.93%) (Fig. 4). Drowning deaths in locations other than natural water among persons aged ≤29 years did not significantly change from 2010 to 2019 or from 2019 to 2020 except for other/unspecified location of drowning deaths which decreased 4.53% per year (95% CI: −7.65, −1.30) from 2010 to

Table 1
Unintentional drowning deaths and location of drowning by sex, age, and race/ethnicity among persons aged ≤29 years, National Vital Statistics System, United States 2020.

Location	Total		Bathtub		Natural Water		Pool		Watercraft		Other/Unspecified	
Characteristic	**Deaths**	**Rate[a] (95% CI)**	**Deaths**	**Rate[a] (95% CI)**	**Deaths**	**Rate[a] (95% CI)**	**Deaths**	**Rate[a] (95% CI)**	**Deaths**	**Rate[a] (95% CI)**	**Deaths**	**Rate[a] (95% CI)**
Total	1589	1.26 (1.20, 1.32)	135	0.11 (0.09, 0.13)	762	0.60 (0.56, 0.65)	384	0.30 (0.27, 0.34)	97	0.08 (0.06, 0.09)	211	0.17 (0.14, 0.19)
Sex												
Female	343	0.56 (0.50, 0.62)	78	0.13 (0.10, 0.16)	99	0.16 (0.13, 0.20)	110	0.18 (0.15, 0.21)	20	0.03 (0.02, 0.05)	36	0.06 (0.04, 0.08)
Male	1246	1.93 (1.83, 2.04)	57	0.09 (0.07, 0.11)	663	1.03, (0.95, 1.11)	274	0.43 (0.38, 0.48)	77	0.12 (0.09, 0.15)	175	0.27 (0.23, 0.31)
Age Group (years)												
<1 year	34	0.91 (0.63, 1.27)	25	0.67 (0.43, 0.99)	–	–	–	–	–	–	–	–
1–4 years	425	2.73 (2.47, 2.99)	34	0.22 (0.15, 0.31)	88	0.57 (0.45, 0.70)	252	1.62 (1.42, 1.82)	–	–	49	0.31 (0.23, 0.42)
5–9 years	117	0.58 (0.47, 0.68)	–	–	50	0.25 (0.18, 0.33)	45	0.22 (0.16, 0.30)	–	–	13	–
10–14 years	91	0.44 (0.35, 0.54)	–	–	55	0.27 (0.20, 0.34)	13	–	–	–	–	–
15–19 years	265	1.26 (1.11, 1.42)	–	–	180	0.86 (0.73, 0.98)	23	0.11 (0.07, 0.16)	16	–	37	0.18 (0.12, 0.24)
20–24 years	328	1.52 (1.35, 1.68)	27	0.13 (0.08, 0.18)	195	0.90 (0.78, 1.03)	22	0.10 (0.06, 0.15)	31	0.14 (0.10, 0.20)	53	0.25 (0.18, 0.32)
25–29 years	329	1.42 (1.26, 1.57)	27	0.12 (0.08, 0.17)	193	0.83 (0.71, 0.95)	27	0.12 (0.08, 0.17)	37	0.16 (0.11, 0.20)	45	0.19 (0.14, 0.26)
Race/Ethnicity[b]												
AI/AN	24	1.97 (1.26, 2.92)	–	–	13	–	–	–	–	–	–	–
A/PI	90	1.10 (0.89, 1.35)	–	–	50	0.61 (0.45, 0.81)	19	–	–	–	15	–
Black/African American	389	2.02 (1.82, 2.22)	30	0.16 (0.11, 0.22)	213	1.11 (0.96, 1.25)	79	0.41 (0.32, 0.51)	22	0.11 (0.07, 0.17)	45	0.23 (0.17, 0.31)
Hispanic	344	1.13 (1.01, 1.25)	25	0.08 (0.05, 0.12)	165	0.54 (0.46, 0.62)	79	0.26 (0.20, 0.32)	20	0.07 (0.04, 0.10)	55	0.18 (0.14, 0.23)
White	741	1.11 (1.03, 1.19)	76	0.11 (0.09, 0.14)	320	0.48 (0.43, 0.53)	204	0.30 (0.26, 0.35)	49	0.07 (0.05, 0.10)	92	0.14 (0.11, 0.17)

Abbreviations: AI/AN (American Indian/Alaska Native), A/PI (Asian/Pacific Islander), CI (confidence interval).
[a] Deaths per 100,000 population.
[b] Persons identified as Hispanic could be any race, persons identified as AI/AN, A/PI, Black, and White were all non-Hispanic; where ethnicity was not stated, persons were excluded from the race/ethnicity counts.
-Death counts of <10 are suppressed for confidentiality and rates based off of <20 deaths are suppressed because they are considered unreliable.

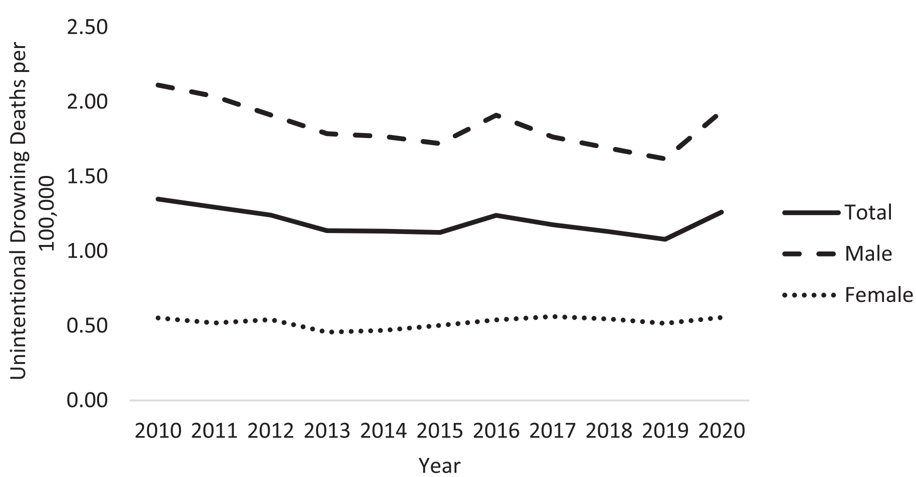


Fig. 1. Rates of unintentional drowning deaths per 100,000 persons ≤29 years overall and by sex, National Vital Statistics System, United States 2010–2020.

2019 and did not significantly change from 2019 to 2020 (APC 24.09%; 95% CI: −0.89, 49.07).

4. Discussion

Drowning deaths among persons aged ≤29 years in the United States increased by almost 17% in 2020. The largest increases were among young adults aged 20 to 29 years, Black or African American persons, males, and in natural water settings. During the initial months of the COVID-19 pandemic (spring 2020), drowning prevention organizations in the United States were concerned that stay-at-home orders would contribute to increased drowning deaths among young children who were spending more time around the home with distracted supervision (Stop Drowning

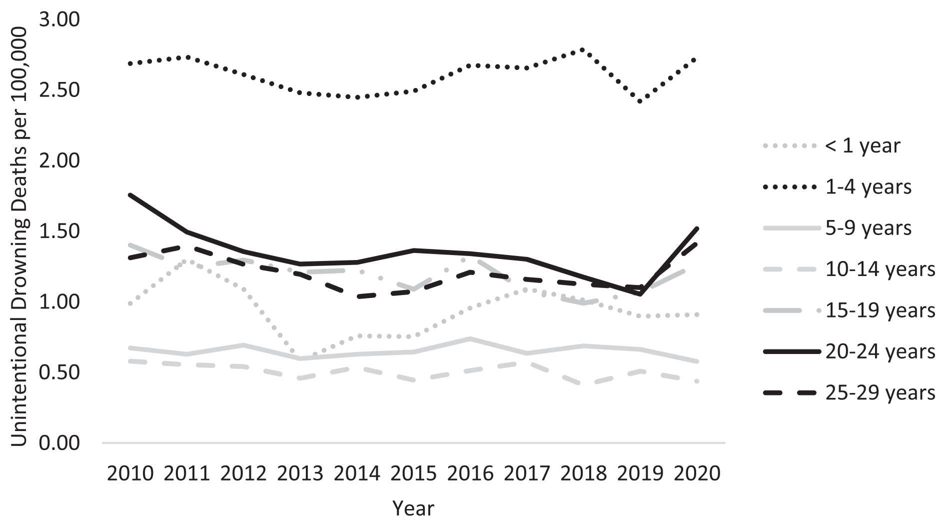
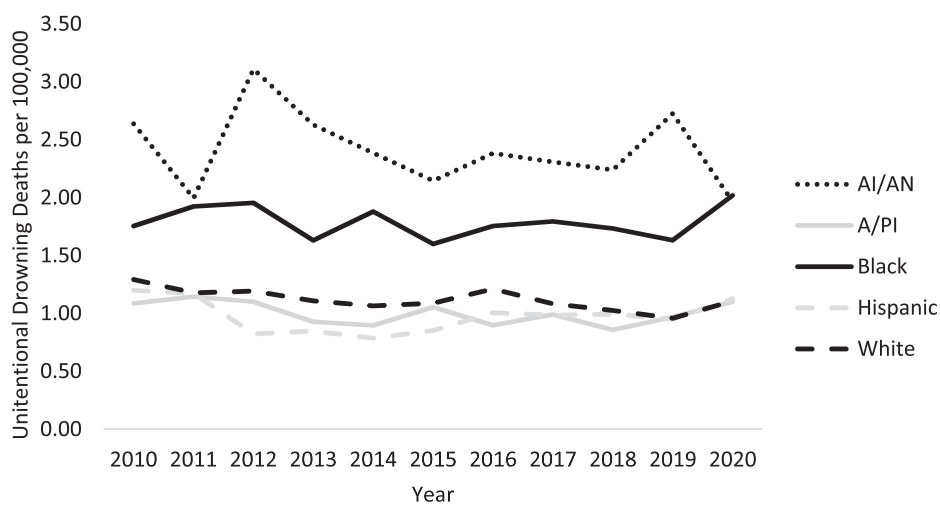


Fig. 2. Rates of unintentional drowning deaths per 100,000 persons ≤29 years by age group, National Vital Statistics System, United States 2010–2020.



Abbreviations: AI/AN (American Indian/Alaska Native), A/PI (Asian/Pacific Islander)
Persons identified as Hispanic could be any race, persons identified as AI/AN, A/PI, Black, and White were non-Hispanic

Fig. 3. Rates of unintentional drowning deaths per 100,000 persons ≤29 years by race/ethnicity, National Vital Statistics System, United States 2010–2020.

Now, 2022). During January-April 2020, a 100% increase in drowning deaths among children 1–4 years of age was reported in Florida compared to the same months of the previous year (Safe Kids, Safe States, & YMCA, 2020). Although children aged 1–4 years continued to have the highest drowning rates of all age groups, our study revealed that there were no significant increases in unintentional drowning deaths in this age group in 2020 compared with earlier years. Despite this, children 1–4 years of age remain a critical group for prioritizing drowning prevention strategies, as drowning was the leading cause of death in this age group in 2020 (CDC WISQARS).

Drowning death rates among young adults aged 20–24 years (+44%) and 25–29 years (+29%) increased significantly. The majority of deaths among these age groups occurred in natural water. A study of injury mortality during the COVID-19 pandemic in China identified lockdowns and associated avoidance of outdoor activities as drivers of a 35% decrease in drowning deaths (Zheng et al., 2021). However, our findings indicate that lockdowns and other COVID-19 related restrictions in the United States may have resulted in modified activities that increased exposure to natural water and drowning risk. People in the United States may have spent more time outdoors participating in activities where they could maintain social distancing, such as swimming and boating in natural water. Powerboat sales reached a 13-year high in 2020 in the United States, increasing 12% over the previous year (National Marine Manufacturers Association, 2020). Similar increases in boating-related drowning and boating sales contributed to increased drowning in Australia in 2020 (Lawes et al., 2021). In one study, young adults in Canada identified outdoor activities, including spending time at the beach, as an important coping mechanism during the COVID-19 pandemic (Ferguson et al., 2021). Further, COVID-19 related precautions might have resulted in decreased access to safety measures during participation in natural water recreation activities. Several lifejacket loaner

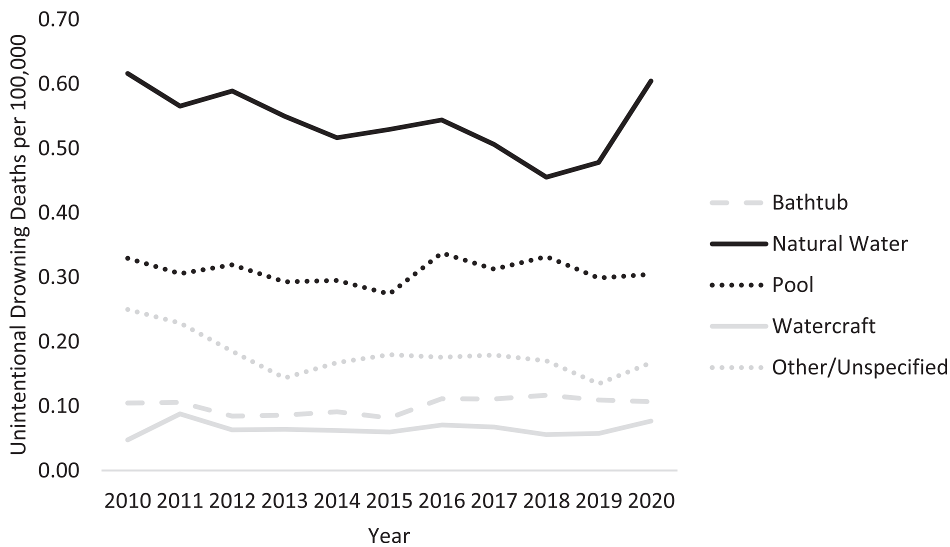


Fig. 4. Rates of unintentional drowning deaths per 100,000 persons ≤29 years by location, National Vital Statistics System, United States 2010–2020.

programs (stations that provide the public with free use of lifejackets) were closed during 2020 due to concerns of surface transfer of the SARS-CoV-2 virus through shared lifejackets that could not be adequately disinfected between uses (Washington Department of Health). Additionally, several national parks and public beaches were closed for extended periods in Deerwester and Woodyard (2020), National Park Service (2020) which might have led people to recreate in unsupervised natural water locations instead.

The increase in unintentional drowning deaths among Black persons in 2020 is particularly concerning. A recent study identified persistent racial and ethnic disparities in drowning death rates in the United States and suggested that the disparities between non-Hispanic Black or African American persons and non-Hispanic White persons had increased from 2015 to 2019 (Clemens et al., 2021). The further increases in drowning death rates among Black persons in 2020, described in this report, emphasize the urgent need to identify factors that are driving disparities in drowning death rates (Clemens et al., 2021). For example, Black children report lower swimming ability than their White peers (Irwin, Irwin, Ryan, & Drayer, 2009). One effective strategy for preventing drowning is basic swimming and water safety skills training. However, availability of swimming lessons was impacted during the COVID-19 pandemic, with many jurisdictions closing public swimming pools due to local restrictions. The impacts of these closures are unlikely to have affected 2020 drowning numbers but delays in swimming lessons may potentially impact future drowning rates. Improving equitable access to basic swimming and water safety skills training for all populations, especially those at increased risk of drowning, may reduce future drowning deaths.

This study has limitations. First, changes in drowning death rates from 2019 to 2020 should be interpreted cautiously. Year to year estimates may vary for multiple reasons and the increase in drowning deaths from one year to the next could be due to outliers in the data and not suggestive of a significant change in trend (Ingram et al., 2018). Further, we cannot determine that increases in drowning death rates were related to the COVID-19 pandemic or the associated changes in infrastructure (e.g., closing of lifejacket loaner programs) that occurred. More recent years of mortality data are needed to determine if the increases in drowning death rates in 2020 were temporary or indicative of a new trend. Second, the location of drowning deaths coded as "unspecified" could not be determined and it is possible these deaths occurred in one of the other settings described in this report (swimming pools, natural water, watercraft, or bathtubs). Third, race/ethnicity is recorded by next of kin or by observation on death certificates and may not be consistent with self-identified race/ethnicity. Persons self-identifying as American Indian or Alaska Native, Asian or Pacific Islander, or Hispanic in census data are sometimes reported as White or non-Hispanic in death data leading to potential underestimates in these racial and ethnic groups (CDC WONDER).

5. Conclusion

In 2020, children aged 1–4 years and Black and AI/AN persons aged ≤29 years had the highest fatal drowning rates. Drowning deaths were most common in natural water and swimming pools. Fatal drowning rates increased from 2019 to 2020 among persons aged ≤29 years. The largest increases occurred among persons aged 20–24 years, males, Black persons, and in drowning in natural water. Data describing risk factors and circumstances of drowning deaths and how they differ by race and ethnicity could help public health researchers adapt and strengthen drowning prevention interventions and decrease disparities in drowning deaths both before and during times of infrastructure disruption. CDC is working with partners to better understand the circumstances of drowning deaths and to improve access to effective interventions among persons at highest risk of drowning.

6. Practical applications

Drowning is preventable. Promising drowning prevention interventions include learning basic swimming and water safety skills, wearing lifejackets on boats or among weaker swimmers in natural water, close supervision of children in or near the water, and installing barriers to prevent unintended entry to water such as pool fences that completely surround swimming pools. Strategies that increase access to effective drowning prevention interventions among children and young adults at the highest risk, including during times of infrastructure disruptions such as those caused by the COVID-19 pandemic, could reduce drowning rates in the United States.

Acknowledgements

None.

Funding Source

No outside funding was used to support this work.

Previous presentations of the work

None.

Conflict of interest disclosures

None of the authors have relevant financial interests, activities, relationships, or affiliations, or other potential conflicts of interest to report.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention

References

Centers for Disease Control and Prevention (2022). *Web-based Injury Statistics Query and Reporting System (WISQARS)*. Atlanta, GA: US Department of Health and Human Services, CDC; Accessed April 26, 2022. https://www.cdc.gov/injury/wisqars/index.html.

Centers for Disease Control and Prevention (2022). *CDC WONDER*. Accessed April 26, 2022. https://wonder.cdc.gov/.

Chen, L., Li, J., Xia, T., Matthews, T. A., Tseng, T. S., Shi, L., et al. (2021). Changes of Exercise, Screen Time, Fast Food Consumption, Alcohol, and Cigarette Smoking during the COVID-19 Pandemic among Adults in the United States. *Nutrients., 13*(10), 3359.

Clemens, T., Moreland, B., & Lee, R. (2021). Persistent racial/ethnic disparities in fatal unintentional drowning rates among persons aged ≤29 years-United States, 1999–2019. *MMWR. Morbidity and Mortality Weekly Report, 70*(24), 869–874.

Deerwester, J, & Woodyard C. (2020). *Texas' Bexar joins Los Angeles, Florida counties in closing beaches, parks for July 4th weekend.* USA Today. https://www.usatoday.com/story/travel/news/2020/06/29/florida-closes-beaches-july-4th-weekend-covid-19-rates-surge/3277800001/.

Denny, S. A., Quan, L., Gilchrist, J., McCallin, T., Shenoi, R., Yusuf, S., et al. (2021). Prevention of drowning. *Pediatrics., 148*(2).

Faust, J. S., Du, C., Mayes, K. D., Li, S. X., Lin, Z., Barnett, M. L., & Krumholz, H. M. (2021). Mortality from drug overdoses, homicides, unintentional injuries, motor vehicle crashes, and suicides during the pandemic, March-August 2020. *JAMA, 326*(1), 84–86.

Ferguson, K. N., Coen, S. E., Tobin, D., Martin, G., Seabrook, J. A., & Gilliland, J. A. (2021). The mental well-being and coping strategies of Canadian adolescents during the COVID-19 pandemic: A qualitative, cross-sectional study. *Canadian Medical Association Open Access Journal., 9*(4), E1013–E1020.

Giudici, R., Lancioni, A., Gay, H., Bassi, G., Chiara, O., Mare, C., et al. (2021). Impact of the COVID-19 outbreak on severe trauma trends and healthcare system reassessment in Lombardia, Italy: An analysis from the regional trauma registry. *World Journal of Emergency Surgery, 16*(1), 1–8.

Giuntella, O., Hyde, K., Saccardo, S., & Sadoff, S. (2021). Lifestyle and mental health disruptions during COVID-19. *Proceedings of the National Academy of Sciences, 118*(9).

Ingram, D. D., Malec, D. J., Makuc, D. M., Kruszon-Moran, D., Gindi, R. M., Albert, M., et al. (2018). National Center for Health Statistics guidelines for analysis of trends. *Vital Health Statistics, 179*, 1–71.

Irwin, C. C., Irwin, R. L., Ryan, T. D., & Drayer, J. (2009). Urban minority youth swimming (in) ability in the United States and associated demographic characteristics: Toward a drowning prevention plan. *Injury Prevention, 15*(4), 234–239.

Lawes, J. C., Strasiotto, L., Daw, S., & Peden, A. E. (2021). When natural hazards intersect with public health: A preliminary exploration of the impact of bushfires and the COVID-19 pandemic on Australian coastal drowning fatalities. *International Journal of Environmental Research and Public Health., 18*(10), 5314.

Moore, A., Singleton, A., Hiatt, L., Miller, R., Phillips, S., & Leskovan, J. J. (2022). An assessment of the impact of SARS-CoV-2 on a level 1 trauma center including subgroup analysis of orthopedic injuries and mechanism of injury. *Cureus, 14*(1).

National Marine Manufacturers Association (2022). *2020 Recreational boating statistical abstract*. 2021; Accessed May 23, 2022. https://www.nmma.org/statistics/publications/statistical-abstract.

National Park Service (2020). *Park closes to support regional COVID-19 prevention efforts.* Accessed May 25, 2022. https://www.nps.gov/grsm/learn/news/park-closes-to-support-regional-covid-19-prevention-efforts.htm.

Safe Kids Worldwide, Safe States, YMCA of the USA (2020). *Florida child drownings increase 100 percent as COVID-19 pandemic evolves*. Accessed May 23, 2022. https://a3f81f198a260bc434c5-8f8af7b10ac8f957b05556159de44bb7.ssl.cf5.rackcdn.com/FL_Child_Drownings_Memo_final_04.24.2020.pdf.

Sephton, B. M., Mahapatra, P., Shenouda, M., Ferran, N., Deierl, K., Sinnett, T., et al. (2021). The effect of COVID-19 on a Major Trauma Network. An analysis of mechanism of injury pattern, referral load and operative case-mix. *Injury, 52*(3), 395–401.

Sherman, W. F., Khadra, H. S., Kale, N. N., Wu, V. J., Gladden, P. B., & Lee, O. C. (2021). How did the number and type of injuries in patients presenting to a regional level I trauma center change during the COVID-19 pandemic with a stay-at-home order?. *Clinical Orthopaedics and Related Research, 479*(2), 266–275.

State of Washington Department of Health (2021). *Guidance for lifejacket loaner stations during COVID-19*. Accessed May 25, 2022. https://www.parks.wa.gov/DocumentCenter/View/17097/Life-Jacket-Loaner-Stations-during-COVID-19_DOH-Letter_May2021.

Stop Drowning Now (2022). *Drowning rates on the rise during the pandemic*. Accessed May 23, 2022. https://www.stopdrowningnow.org/drowning-rates-on-the-rise-during-the-pandemic/.

Zheng, X. Y., Tang, S. L., Ma, S. L., Guan, W. J., Xu, X., Xu, H., et al. (2021). Trends of injury mortality during the COVID-19 period in Guangdong, China: A population-based retrospective analysis. *BMJ Open, 11*(6), e045317.

Briana Moreland MPH, is a Health Scientist and Cherokee Nation Operational Solutions contractor in the Division of Injury Prevention at CDC's National Center for Injury Prevention and Control. Her research focuses on drowning and older adult fall surveillance and prevention.

Neil Ortmann MPH, is an ORISE Fellow in the Division of Injury Prevention at CDC's National Center for Injury Prevention and Control. His research focuses on the prevention of drowning and older adult falls.

Tessa Clemens PhD, is a Health Scientist focused on drowning prevention in the Division of Injury Prevention at CDC's National Center for Injury Prevention and Control. Her research includes understanding and addressing racial and ethnic disparities in drowning rates and supporting the implementation of effective interventions among underserved populations with the highest rates of drowning.

Journal of Safety Research 82 (2022) 124–143

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Interventions to prevent and reduce work-related musculoskeletal injuries and pain among healthcare professionals. A comprehensive systematic review of the literature

Beatrice Albanesi [a,b], Michela Piredda [a,*], Marco Bravi [c], Federica Bressi [c], Raffaella Gualandi [a], Anna Marchetti [a], Gabriella Facchinetti [a], Andrea Ianni [d], Francesca Cordella [e], Loredana Zollo [e], Maria Grazia De Marinis [a]

[a] Research Unit Nursing Science, Campus Bio-Medico University, Rome, Italy
[b] Department of Public Health and Pediatrics, University of Turin, Turin, Italy
[c] Physical Medicine and Rehabilitation Unit, Campus Bio-Medico University, Rome, Italy
[d] Research Unit in Hygiene, Statistics and Public Health, Campus Bio-Medico University, Rome, Italy
[e] CREO Lab - Advanced Robotics and Human Centred Technologies, Campus Bio-Medico University, Rome, Italy

ARTICLE INFO

Article history:
Received 15 July 2021
Received in revised form 26 November 2021
Accepted 9 May 2022
Available online 18 May 2022

Keywords:
Musculoskeletal diseases
Musculoskeletal conditions
Pain
Health personnel
Systematic review

ABSTRACT

Introduction: Work-related musculoskeletal disorders (WMSDs) are among the main causes of injury and pain in healthcare professionals. Previous reviews provided a fragmented view of the interventions available for WMSDs. This review aims to provide a comprehensive description of interventions for preventing and reducing work-related musculoskeletal injuries and/or pain among healthcare professionals, and to assess the methodological quality of studies. *Methods:* A systematic literature review was performed, based on the Effective Public Health Practice Project process. A comprehensive search was conducted on six peer-reviewed databases and manually. The methodological quality of the studies included was rated as weak, moderate, or strong. The studies were organized based on the 2019 classification of the interventions by Oakman and colleagues. *Results:* Twenty-seven articles were included reporting individual ($n = 4$), task-specific ($n = 4$), work organization and job design ($n = 2$), work environment ($n = 1$), and multifactorial ($n = 16$) interventions. Overall quality rating was strong for 6 studies, moderate for 16, and weak for 5. Individual interventions such as neuromuscular and physical exercise were effective in reducing pain. Task-specific and work organization interventions could prevent certain injuries. Significant reduction of both injuries and pain resulted from multifactorial interventions, which were reported by the majority of strong ($n = 5$) and moderate ($n = 10$) quality articles. *Conclusions:* This review provides healthcare professionals with evidence-based information to plan interventions targeted towards reducing WMSDs. In particular, more efforts are needed to implement and extend effective multifactorial interventions. Moreover, studies about each professional healthcare target group are needed. *Practical Application:* Our results can guide policy-makers, healthcare managers and professionals to choose the best strategies to prevent and reduce WMSDs and to shape continuous education programs. This study prompts clinicians to develop inter-professional collaborations and to practice physical activities in order to reduce WMSDs.

1. Introduction

The World Health Organization recognizes musculoskeletal disorders (MSDs) as one of the main causes of injury worldwide, indiscriminately affecting adult and young populations, and causing significant disabilities (World Health Organization [WHO], 2019). MSDs have a gradual and multifactorial etiology characterized by high diagnostic complexity (Roquelaure, 2018). They often result in untreated lesions (WHO, 2019) and are prevalent in many work contexts (U.S. Department of Labor, 2020; Verbeek et al., 2012).

In the healthcare sector, work-related MSDs (WMSDs) represent one of the main causes of injury among professionals (European Social Statistics [EUROSTAT], 2003; European Agency

* Corresponding author at: Research Unit Nursing Science, Campus Bio-Medico di Roma University, Via Alvaro del Portillo 21, 00128 Rome, Italy.
E-mail address: m.piredda@unicampus.it (M. Piredda).

https://doi.org/10.1016/j.jsr.2022.05.004
0022-4375/

for Safety and Health at Work (2013), World Health Organization (2020)). According to the Bureau of Labor Statistics (BLS), in 2015 there were approximately 59,810 occupational injuries among healthcare workers in the United States, while, in Europe, the Danish Working Environment & Health showed that 39,000 healthcare workers experienced musculoskeletal pain (European Agency for Safety and Health at Work [EU-OSHA], 2020). The highest level of exposure to such injuries is found among nurses (37%) and healthcare assistants (46%) (EU-OSHA, 2020). Indeed, daily performing repetitive movements, such as lifting, pushing or pulling, usually necessary for positioning patients (U.S. Department of Labor, 2020), and the forward flexion required for patient lifting and handling, place the professional's spinal column in an extremely vulnerable position, meaning that even under ideal lifting conditions, the weight of any adult exceeds the lifting capacity of most healthcare professionals (Jäger et al., 2013).

The negative effects of WMSDs on healthcare professionals are not limited to injuries, but also involve other consequences that can affect their bodily health and quality of life, causing disability (Freiberg et al., 2016). WMSDs often result in acute and chronic pain, predominantly in the knees and spinal column, resulting in difficulties carrying out everyday tasks, walking difficulties, or sleep disorders, and affecting job performance; the regular use of health services and painkillers is also a direct consequence (Van Hoof et al., 2018). Additionally, WMSDs indirectly influence expenditure, increasing it both for companies and individuals; they mean a loss of working days and reduced earnings (Suni et al., 2018). In the United States alone, WMSD injuries are responsible for 34% of all workdays lost and for an increase in costs of about 33% of the total sum of workers' pay (U.S. Department of Labor, 2020). Thus, preventing WMSDs should be a priority, and the effective management of WMSDs, through a combined effort focusing on the above factors, is necessary to avoid the loss of ability to work and early involuntary exit from the healthcare sector due to WMSDs (EUROSTAT, 2003).

A number of individual causes are also associated with the onset of WMSDs (Amaro et al., 2018; Hegewald et al., 2018). First of all, socio-demographic factors such as age and gender, poor lifestyle, and lack of physical activity seem to significantly affect physical tolerance of exertion (Ngan et al., 2010; Amaro et al., 2018). Moreover, organizational constraints or incorrect policy models, such as inadequate staffing, lead to work overload and physical burden (Richardson et al., 2018; Roquelaure, 2018). Lastly, psychological pressure related to daily activities causes constant physical and emotional distress in healthcare professionals, and can influence the development of WMSDs (Roquelaure, 2018).

Oakman and colleagues (2019) modified the ergonomic framework of Macdonald and Oakman (2015) to classify interventions to reduce WMSDs into five categories: (a) individual, (b) task-specific and equipment, (c) work organization and job design, (d) workplace environment, and (e) multifactorial. The individual category includes the interventions focused on changes to an individual's working behavior, such as training, exercises, and education. The task-specific category includes interventions focused on changes to an individual's working equipment, such as workstation adjustments. The work organization and job design category include interventions such as change of working hours, overall job design, or manager training in comprehensive work risk management. The workplace environment category includes interventions focused on the physical and psychosocial environment such as general workplace culture, and job security. Finally, the multifactorial interventions category includes a combination of different types of interventions (Oakman, Clune, & Stuckey, 2019).

Several reviews have been conducted to describe interventions capable of reducing or preventing WMSDs in healthcare professionals. Some of these evaluated the effectiveness of interventions in a particular group of healthcare professionals, such as nurses or physical therapists (Anderson, & Oakman, 2016; Asuquo, Tighe, & Bradshaw, 2021; Clari et al., 2021; Milhem et al., 2016; Richardson et al., 2018). Others evaluated the effectiveness of using small aids (such as sliding sheets or walking tapes) during patient handling (Vieira & Miller, 2008; Freiberg et al., 2016; Hegewald et al., 2018), or focused on interventions for reducing disability and pain (Burdorf, Koppelaar, & Evanoff, 2013; Van Hoof et al., 2018) to improve patient safety (Hignett, 2003; Aslam et al., 2015), or identified obstacles and facilitators for the implementation of interventions to prevent WMSDs (Koppelaar et al., 2009). However, the account of interventions for reducing and preventing WMSD injuries and pain in healthcare professionals is still fragmented. Therefore, this systematic review aimed to offer a comprehensive organization and description of existing interventions for reducing and preventing WMSD injuries and/or pain among healthcare professionals (HCPs).

2. Methods

The Effective Public Health Practice Project (EPHPP) process for systematic reviews was followed (Thomas et al., 2004). The process includes seven phases: (1) question formulation; (2) literature search and retrieval; (3) identification of relevance criteria; (4) quality assessment of relevant studies; (5) data extraction and synthesis; (6) peer review; and (7) dissemination. The protocol for the review was registered in the PROSPERO database (CRD42020218598).

2.1. Question formulation

The study question was the following: 'Which interventions are effective to reduce and prevent WMSDs injuries and/or pain among healthcare professionals?' In line with this question the review had two main objectives: (a) To provide a complete description of existing interventions for preventing and reducing the WMSD injuries and/or pain among HCPs; (b) To assess the methodological quality of studies.

2.2. Literature search and retrieval

Six electronic databases were screened to identify relevant studies. A preliminary search was performed on PubMed and CINAHL databases, to identify index and key words for musculoskeletal health outcomes, healthcare professionals and nursing personnel, and patient handling. These were then appropriately combined through Boolean operators in the final search (Supplementary File 1) that was conducted on 11st November 2021 in the following databases: PubMed, the Cochrane Library Central, Excerpta Medica Database (EMBASE), Cumulative Index to Nursing and Allied Health Literature (CINAHL), Scopus and the Institute of Electrical and Electronics Engineers (IEEE) network. Additional manual searches on the key-content journals, extracted by Scimago (Falagas et al., 2008) and reference lists were also performed to identify relevant records (Supplementary File 1). The reference manager Endnote X9 software (Team, 2013) was used to gather all references and to identify duplicates. Titles and abstracts of each unique record were examined by two authors (BA and MP), then the full texts of potentially relevant studies were retrieved and assessed for eligibility. No time limits were applied. Only articles in English were included. The Updated version of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Page et al., 2021) was used for reporting the article selection process (Fig. 1).

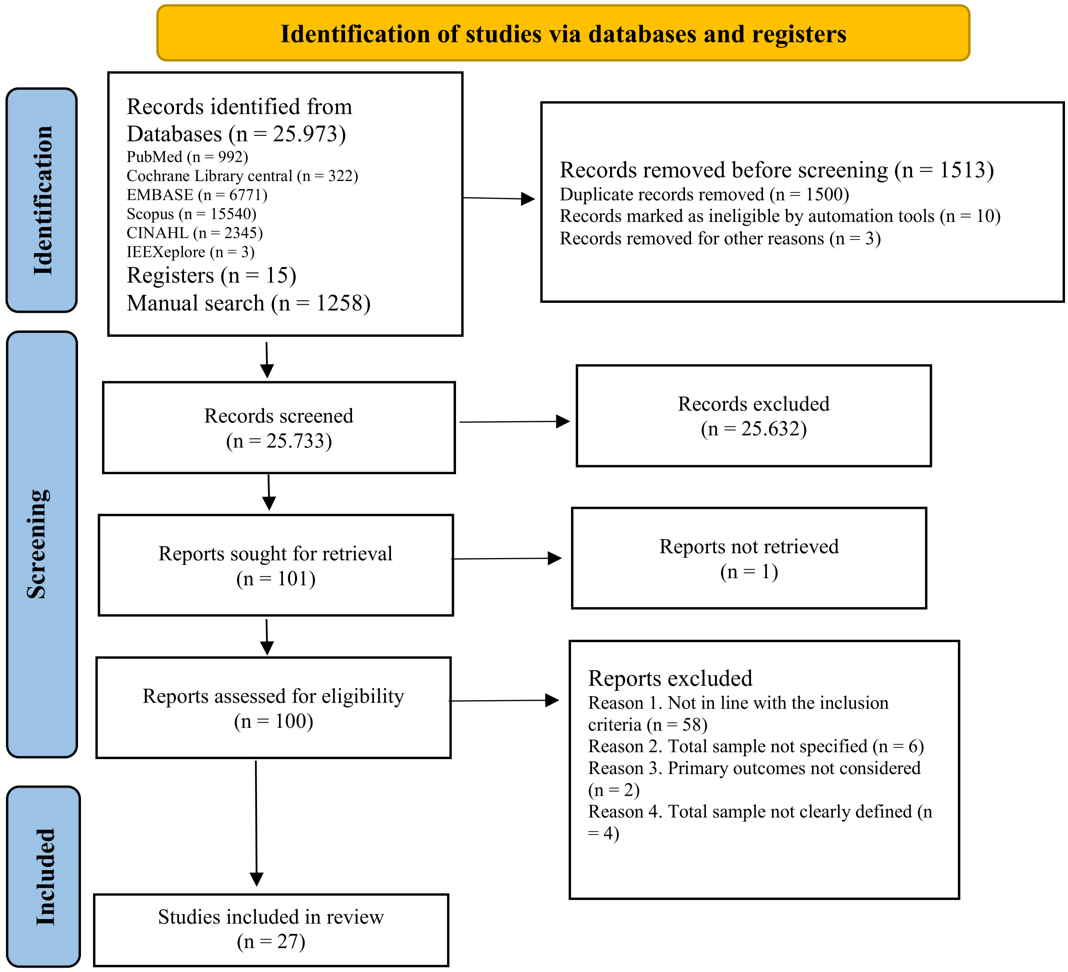


Fig. 1. PRISMA flow-chart.

2.3. Identification of relevance criteria

2.3.1. Population of interest

The population included any healthcare professionals belonging to the following disciplines: nursing (registered nurses, licensed practical nurses, nurses' aides and assistants); rehabilitation (physiotherapists or physical therapists, occupational therapists, physiotherapist or occupational therapist aides); radiology (technicians and aides); others (clerks, unit assistants and attendants). In the case of articles giving limited or unclear information about the population, the first/ corresponding author was contacted by -mail. Missing responses led to paper exclusion.

2.3.2. Interventions

To be included, studies had to investigate any intervention aimed at preventing or reducing WMSD injuries and/or pain among healthcare professionals. Both single and multifactorial interventions were considered. Interventions focused on cost saving, or delivered outside a healthcare setting, such as those conducted and tested in simulation workshops or laboratories were excluded. Studies principally considering engineering interventions (i.e., turn assist surfaces) that did not specify primary outcomes were also excluded.

2.3.3. Outcomes

Studies considering WMSD injury rates and/or pain as primary outcomes were eligible for inclusion. Injuries and pain are generally associated with other issues, such as working days lost, physicians' visits, and the use of painkillers. When these factors were reported in the articles, they were also considered as secondary outcomes. To guide the selection, the International Classification of Diseases (ICD-11) definitions were used (WHO, 2020). Musculoskeletal disorders (MSDs) were defined as 'injuries or disorders of the muscles, nerves, tendons, joints, cartilage, and spinal discs;' work-related musculoskeletal disorders (WMSDs) were defined as 'conditions in which the work environment and performance of work contribute significantly to the condition; and/or the condition is made worse or persists for longer due to work conditions;' acute pain was defined as 'pain with a duration of less than 3 months' and of chronic pain as 'an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage that persists or recurs for longer than 3 months' (WHO, 2020).

2.3.4. Study designs

Peer-reviewed articles reporting randomized and non-randomized controlled trials, as well as studies implementing pre-post evaluation, or before and after design, were included. Documents such as reviews, case reports, commentaries and dissertations; studies on biomechanical effort (in particular, regarding the evaluation of muscular effort during movement); guidelines on safety practice (handling and positioning); and studies of prevalence or retrospective studies with no intervention testing were excluded.

2.3.5. Quality assessment of relevant studies

To assess the methodological quality of each eligible study the Quality Assessment Tool for Quantitative Studies of Effective Public Health Practice Project (EPHPP) (2020) was used. This tool includes six criteria: (a) selection bias, (b) study design, (c) confounders, (d) blinding, (e) data collection method, and (f) withdrawals/dropouts, each rated as '1-strong,' '2-moderate,' or '3-weak.' Overall, studies with no weak ratings and a minimum of four strong ratings were classified as strong; those with four strong/moderate ratings and one weak rating, as moderate; and those with two or more weak ratings, as weak (Thomas et al., 2004; Armijo-Olivo et al., 2012). The integrity of interventions and the use of appropriate statistical analysis were also considered. Two reviewers (BA and MB) independently assessed each article. Any discrepancy was resolved by a third reviewer (MP).

2.3.6. Data extraction and synthesis

To address the study aims, two researchers (BA and MB) used a standardized data extraction form collecting information on: first author, year of publication, country and assigned quality, aim(s), study design, participants, setting, intervention delivered, intervention duration, outcomes and key results. The interventions were extracted and organized based on the classification by Oakman et al. (2019) (Table 1). Then a complete description and synthesis of results was performed. The methodological quality of each study was also assessed by a clear description of findings across studies.

3. Results

3.1. Study selection

A total of 27,246 records were retrieved (Fig. 1 and Supplementary file 1). After the removal of articles for specific reasons, 25,733 articles were screened. Of these, 25,632 were removed based on title and abstract, leaving 100 potentially relevant studies for a full-text screening. After full-text review, and one case of author contact to confirm that sample inclusion criteria were fulfilled, 27 papers were included. The reasons for the exclusion of 73 papers are detailed in Fig. 1 and Supplementary file 2.

Table 1
General characteristics of the included studies ($N = 27$).

Characteristics	*N* (%)
Geographical distribution	
Western Countries	**25 (92.5)**
USA	14 (51.8)
Denmark	5 (18.5)
Canada	3 (11.1)
Ireland	1 (3.7)
Finland	1 (3.7)
Sweden	1 (3.7)
Eastern Countries	**2 (7.4)**
Israel	1 (3.7)
Turkey	1 (3.7)
Type of study	
Pre-Post Design	15 (55.5)
Randomized Controlled Trial	7 (25.9)
Pilot Pre-Post Design	1 (3.7)
Retrospective Pre-Post Design	1 (3.7)
Pilot Mixed Measure Design	1 (3.7)
Longitudinal Pre-Post Design	1 (3.7)
Quasi- Experimental Design	1 (3.7)
Setting	
Hospital	19 (70.3)
Hospital & community/home services	6 (22.2)
Nursing Homes	2 (7.4)
Target population	
Healthcare professionals*	13 (48.1)
Nurses	8 (29.6)
Nurses and nurses' aides	6 (22.2)

* Nurses, nurses' aides, physiotherapists, and radiology staff.

3.2. Study characteristics

The general and specific characteristics of the included studies are presented in Tables 1 and 2 and grouped according to the intervention provided. The majority of studies were conducted in Western countries ($n = 25$; 92.5%), mostly in the United States ($n = 14$; 51.8%) and in Denmark ($n = 5$; 18.5%). Fifteen studies used a pre-post research design ($n = 15$; 55.5%), while seven (25.9%) were randomized controlled trials. Nineteen (70.3%) conducted the interventions in hospital settings only, while six (22.2%) conducted them in both hospitals and community or home services. Thirteen studies (48.1%) were conducted with a mixed sample of different healthcare professionals, principally including nurses, nurses' aides, physiotherapists, and radiology staff. Eight studies (29.6%) focused on nurses and six (22.2%) concerned nurses and nurses' aides. Most articles dealt with multifactorial interventions ($n = 15$; 55.5%) (Table 2), while five studies (18.5%) focused on individual interventions, four (14.8%) on task-specific interventions (mainly regarding manual or mechanical patient handling), two (7.4%) on work organization and job design, and one (3.7%) on work environment.

3.3. Study quality

Table 3 and supplementary file 3 show the study quality. Of the 27 studies, six (22.2%) were rated as strong, 16 (59.2%) were rated as moderate, and 5 (18.5%) were rated as weak quality, according to the EPHPP criteria. Of the five individual interventions, one (20%) obtained a strong rating, three (60%) a moderate rating, and one (20%) a weak rating. Task-specific and equipment group collected articles of moderate ($n = 2$; 50%) and weak ($n = 2$; 50%) quality. Of the two studies in the category of work organization and job design factors interventions, one received a moderate rating, and one a weak rating. In the workplace environment intervention, the only study included received a moderate rating. Multifactorial interventions were reported by 15 articles, mostly receiving either a strong ($n = 5$; 33.3%) or a moderate ($n = 9$; 60%) rating. Only one article (6.6%) received a weak rating.

3.4. Effectiveness of interventions on primary and secondary outcomes

Table 2 and Table 4 show the effectiveness of interventions for the primary and secondary outcomes. A detailed description follows, according to the classification of interventions by Oakman et al. (2019).

3.4.1. Individual interventions

These interventions consisted of physical exercise (Jakobsen et al., 2015, 2017), cognitive behavioral therapy (CBT) (Menzel & Robinson, 2006), and neuromuscular exercise (Taulaniemi et al., 2019). Pain intensity (in terms of general and widespread musculoskeletal pain) was reduced in three studies. Jakobsen et al. (2015) showed a significant reduction of musculoskeletal pain ($p = 0.01$) and low-back ($p = 0.02$) pain intensity, at baseline and 10 weeks of follow-up, for the workgroup intervention. Furthermore, the secondary analysis by Jakobsen et al. (2017) confirmed the reduction of musculoskeletal pain intensity among professionals who received the intervention at work ($p = 0.04$). Similarly, Pilates neuromuscular exercises (NME) were associated with a significant reduction of pain intensity ($p = 0.029$) and pain episodes interfer-

Table 2
Study characteristics, interventions and quality assessment ($N = 27$).

Individual (n = 4; 14.8%)							
First author, year (quality rating)	Aim	Design	Participants	Intervention	Total follow-up	Outcomes	Key Results
Menzel, and Robinson, 2006 (3)	To investigate the effectiveness of a cognitive behavioural therapy (CBT) in reducing pain intensity, pain disability, perceived stress level and to assess the association of these variables with the absence from work	Randomized Controlled Trial (RCT)	$N = 32$ Nurses Nurses' aides	A 12-week CBT intervention offered weekly for 1 h in small group sessions and homework assignments	12 weeks	*Primary* WMSDs[1] pain intensity *Secondary* Absence from work *Secondary* Absence from work Cost-related injuries	No significant differences were found for pain intensity, disability and absence from work ($p = 0.06$).
Jakobsen et al., 2015 (2)	To investigate the effect of workplace (WORK) versus home-based (HOME) physical exercise on musculoskeletal pain in back and neck/shoulders among healthcare professionals	Randomized Controlled Trial (RCT)	$N = 200$ Healthcare professionals[2]	A 10-week intervention period receiving either physical exercise (5×10 min a week), coaching sessions and ergonomic training at work or physical exercise at home.	1 year	*Primary* WMSDs[1] pain intensity *Secondary* Use of analgesics (self reported)	Pain intensity in the intervention (WORK) group decreased, compared with the control one (HOME) (-0.7, 95% CI[3] -1.0 to -0.3). A significant reduction at 10-week follow-up was showed for low-back pain intensity ($p = 0.02$) and in the use of analgesics ($p = 0.005$).
Jakobsen et al., 2017 (2)	To investigate if a WORK- versus HOME-based physical exercise intervention, pain status, frequency of patient handling activities, body mass index, age, and leisure-time activities affect musculoskeletal pain relief	Randomized Controlled Trial (RCT)	$N = 200$ Healthcare professional[2]	A 10-week intervention of either workplace of home-based physical exercise for 5x10 minutes at week, associated with an ergonomic counselling and training brief courses in patient handling and use of assistive devices	1 year	*Primary* WMSDs[1] pain intensity	The multi-adjusted analysis showed a significant effect on pain reduction in the intervention group ($p = 0.04$).
Taulaniemi et al., 2019 (1)	To investigate the effectiveness of 12-month pylates-type neuromuscular exercise (NME) on pain intensity, pain interfering with work, lumbar movement control impairments	Randomized Controlled Trial (RCT)	$N = 219$ Nurses	A supervised 6-month NME programme with twice a week classes (60 min) for 2 months associated with the next two months with one supervised work session and one DVD home session (50 min) and booklet information materials	1 year	*Primary* WMSDs[1] pain intensity Low-back pain intensity *Secondary* Work-related physical functioning (i.e. muscles strength and sense of wellbeing)	A significant decreased in the intervention group was showed in general pain intensity ($p = 0.029$), low-back pain episodes interfering with work ($p = 0.035$), and in lumbar movement control ($p = 0.046$). Work-related physical functioning significantly increased in the intervention group ($p = 0.007$).

Table 2 (*continued*)

Individual (n = 4; 14.8%)							
First author, year (quality rating)	Aim	Design	Participants	Intervention	Total follow-up	Outcomes	Key Results
Task-specific and equipment (*n* = 4; 14.8%)							
Li et al., 2004 (2)	To evaluate the effectiveness of mechanical patient lifts in reducing MSDs[1] discomfort, rates of injuries, rates of lost workday injuries, and workers' compensation costs	Pre-Post Intervention Study	$N = 36$ Nurses, nurses' aides, patient care technicians	Mechanical lifts (one portable full body and two portable stand-up sling lifts) and training on guidelines were made available to nursing personnel for transferring and lifting patients	1 year	*Primary* WMSDs[1] pain intensity WMSDs[1] rates of injury *Secondary* Absence from work Cost-related injuries	Post-intervention period showed statistically significant improvements in WMSDs[1] comfort levels ($p < 0.05$) for all the body regions (neck, back, hand, knees, anklets…). The recordable injury rates decreased in post intervention period from 10.3 to 3.8 injuries (RR[4]: 0.37, 95% CI) 0.16–0.88. The absence from work showed a decrease RR: 0.35, 95% CI 0.10–1.16). The median cost-related injuries also decreased from $484 per FTE[6] to $151 per FTE[6] post-intervention
Schoenfisch et al., 2013 (3)	To evaluate the effectiveness of a patients' lift and transfer equipment on WMSDs[1] rate of injuries and lost of workday	Pre-Post Intervention Study	$N = 11\,545$ Healthcare Professionals[5]	Implementation of a patient' lift and transfer equipment programme on patient care, following by a policy of minimal manual lift environment	13 years	*Primary* WMSDs[1] injury rates *Secondary* Absence from work	Considerable decrease in patient-handling injuries was observed following the intervention (RR[4]: 0.56, 95% CI 0.36–0.87). Patients' handling activities in hospital were twice as those in community services. The rate of days away from work associated with patient-handling MSD[1] injuries (67.5 per 100 FTE[5]) was higher than that for days away associated with nonpatient-handling MSD[1] injuries (17.1 per 100 FTE[5]).
Vieira and Brunt, 2016 (2)	To evaluate if wearing unstable shoes reduces low back pain and disability among nurses in hospital and homecare	Randomized Controlled Trial (RCT)	$N = 20$ Nurses	Wearing unstable shoes at least 36 h/week	1 year	*Primary* WMSDs[1] pain intensity	The intervention group reported a lower level of pain at weeks 4 ($p < 0.009$) and 6 ($p < 0.001$).
Alperovitch-Najenson et al., 2020 (3)	To examine the effects of sliding sheet usage on WMSDs[1] injury rates, disability, perceived workload, burnout, and job satisfaction	Interventional Prospective Repeated Measurement Study	$N = 41$ Nurses and nurses' aids	Implementation and use of a reusable tubular cylindrical sliding sheet	9 months	*Primary* WMSDs[1] pain intensity *Secondary* Job satisfaction and work motivation	After 3 and 6 months of sliding sheet usage, pain and disability decreased in the neck ($p < 0.001$); arms, shoulders, hands ($p = 0.041$); and lower back ($p < 0.001$), with an increase in job satisfaction and work motivation ($p < 0.001$).
Work organisation and job design (n = 2; 7.4%)							
Engkvist, 2006 (2)	To evaluate the use of transfer equipment, number of injuries, pain/symptoms and absence from work among nurses after the intervention of No Lifting Policy (NLS) and make a comparison with nurses at two control hospital	Cross-Sectional Pre-post intervention study	$N = 457$ Nurses	O'Shea No Lift System people/materials-handling programme	1 year	*Primary* WMSDs[1] pain intensity *Secondary* Physical tiredness Cost-related injuries	Pain intensity, especially back pain, showed a significant reduction in NLS[7] hospital ($p < 0.001$). Physical tiredness showed a significant decrease ($p < 0.01$) in NLS hospital. No significant decrease was shown for cost-related injuries.

(*continued on next page*)

Table 2 (*continued*)

First author, year (quality rating)	Aim	Design	Participants	Intervention	Total follow-up	Outcomes	Key Results
Individual (n = 4; 14.8%)							
Springer et al., 2009 (3)	To determine the effect of a lift team to reduce employee's injury rates	Pre-Post Intervention Study	*N* = not specified Nurses	Lift team help from 8 to 10 h per day	3 years	*Primary* WMSDs[1] injury rates	The risk of injuries among professionals significantly reduced ($p = 0.06$). Patients' acuity was strongly related to injuries, showing a significant double risk of injury ($p = 0.006$).
Workplace environment (n = 1; 3.7%)							
Lee et al., 2019 (2)	To evaluate the impact of the California Safe Patient Handling Legislation	Repeated Pre-post cross-Sectional Surveys	*N* = 535 Nurses	Introduction of the safe patient handling California's legislation in small (<200 beds), medium (200–399 beds) and large (≥400 or more beds)	3 years	*Primary* WMSD[1] pain intensity WMSDs[1] injury rates	A significant reduction was observed for the major WMSD[1] pain intensity (61% vs 52%) (PR[8]: 0.78 (95% CI) 0.66–0.91) specifically for low back PR[5]: 0.71 (95% CI) 0.55–0.92), neck (PR to 0.60, 95% CI 0.44–0.82), and hands/wrists (PR: 0.59, 95% CI 0.39–0.91]. For WMSD[1] injuries there was no difference (21% vs 19% (PR: 0.86, 95% CI 0.55–1.33).
Multifactorial (n = 16; 59.2%)							
Yassi et al., 2001 (1)	To compare the effectiveness of training and equipment to reduce musculoskeletal injuries, increase comfort, and reduce physical demands on staff performing patient lifts and transfers at a large acute care hospital	Randomized Controlled Trial (three arms)	*N* = 346 Nurses and nurses' aids	– Safe lifting and a no strenuous lifting (arm C); – Safe lifting by manual equipment or mechanical equipment (arm B); – Usual practice (arm A); – Intensive training in back care, patient assessment, and handling techniques;	1 year	*Primary* WMSD[1] pain intensity WMSDs[1] injury rates *Secondary* Physical tiredness Cost-related injuries	WMSD[1] injury rates did not vary significantly across the three arms ($p > 0.10$). WMSD[1] pain intensity and physical tiredness associated with patient handling tasks were reduced on both intervention arms (arm A and arm C). No reduction was shown in costs-related injuries among the three groups.
Collins et al., 2004 (2)	To conduct an intervention trial of a 'best practices' WMSD injuries prevention program designed to safely lift physically dependent nursing home residents	Pre-post intervention study	*N* = 1728 Nurses	– Best practices WMSDs injuries' prevention program; – Mechanical lifts and repositioning aids; – 45 minutes of zero lift policy, and employee training on lift usage;	6 years	*Primary* WMSDs[1] injury rates *Secondary* Absence from work	There was a significant reduction ($p = 0.05$) in WMSDs[1] injury rates. A reduction days absence from work was observed after the intervention (RR: 0.34, 95% CI 0.20–0.60).
Hartvigsen et al., 2005 (2)	To evaluate the effectiveness of an intensive educational and low-tech ergonomic intervention program aimed at reducing low back pain (LBP) among home care nurses and nurses' aids	Pre-post intervention study	*N* = 345 Nurses and nurses' aids	– Intensive educational and low-tech ergonomic; – Weekly meetings and ward training in patient transfer and lifting techniques, use of low-tech ergonomic aids;	1 year	*Primary* WMSD[1] pain intensity WMSDs[1] injury rates	No significant differences were found between the two groups for WMSD[1] pain intensity or WMSDs[1] injury rates ($p < 0.88$).
Nelson et al., 2006 (1)	To evaluate the impact of the program on injury rate, lost and modified work days, job satisfaction, self-reported unsafe patient handling acts, level of support for program and costs	Pre-post intervention study	*N* = 825 Healthcare professionals[9]	– Ergonomic Assessment Protocol; – Patient Handling Assessment Criteria and Decision Algorithms; – Peer Leader evaluation; – State-of-the-art Equipment; – After Action Reviews; – No Lift Policy;	1 year	*Primary* WMSDs[1] injury rates *Secondary* Absence of workdays Correct patient handling practice	The overall WMSDs[1] injury rates significantly decreased in the post intervention ($p = 0.036$). Absence from work decreased in post-intervention ($p = 0.79$). There was a statistically significant decrease in the number of unsafe patients' handling practices ($p = 0.027$).

Journal of Safety Research 82 (2022) 124–143

Table 2 (*continued*)

Individual (n = 4; 14.8%)							
First author, year (quality rating)	Aim	Design	Participants	Intervention	Total follow-up	Outcomes	Key Results
Zadvinskis and Salsbury, 2010 (2)	To examine the effectiveness of a multifaceted minimal-lift environment on reported equipment use, musculoskeletal injury rates, and workers' compensation costs	Pilot mixed measures study	*N* = 161 Nurses and nurses' aides	– Engineering (minimal-lift equipment); – Administrative (nursing policy); – Behavioral (peer coach program) controls;	1 year	*Primary* WMSDs[1] injury rates *Secondary* Correct patient handling practice	A reduction of WMSDs[1] injury rates was shown. A significant greater use of the floor-based lift ($p = 0.002$) and stand-assist device ($p = 0.0005$) was reported.
Black et al., 2011 (2)	To evaluate the effectiveness of a Transfer, Lifting and Repositioning program to reduce WMSDs[1]	Retrospective Pre-Post Intervention Study	*N* = 776 Healthcare professionals[10]	Education and training on anatomy, WMSDs[1] injuries, body mechanism, personal health, lifting and patient handling procedures and standardized patient handling activities	3 years	*Primary* WMSDs[1] injury rates *Secondary* Absence from work Cost-related injuries	Significant reduction ($p < 0.0001$) of WMSDs[1] injuries was shown in the intervention group. Absence from work, in terms of time-loss days/injury ($p = 0.09$) and cost-related injuries ($p = 0.013$) decreased significantly in the intervention group.
Lim et al., 2011 (2)	To evaluate repeated patient handling injuries following a multifactor ergonomic intervention program among health care workers	Quasi-experimental intervention	*N* = 1471 Healthcare professionals[11]	– Ergonomic prevention program; – Engineering and administrative controls; – Staff education on anatomy, injuries, body mechanics, personal health, lifting and patient handling procedures; – Implementation of standardized patient handling needs assessment; – Skills based learning in daily activities;	4 years	*Primary* WMSDs[1] injury rates *Secondary* Correct patient handling practice Cost-related injuries	The intervention group in medium and small size hospitals had a significant reduction of WMSDs[1] injury rates than the control group ($p = 0.001$ and $p = 0.002$). The WMSDs[1] injury rates significantly reduced in the intervention group ($p = 0.001$). For each occupational category in the intervention group there was a significant reduction of injuries than the control group ($p = 0.016$). Cost-related injuries decreased in the intervention.
Caspi et al., 2013 (2)	To evaluate unit-level changes (safety practices, supervisor and co-worker support) changes in worker behaviours, and to test the feasibility of the interventions strategies for improve WMSDs and physical activities outcomes	Pilot intervention study	*N* = 501 Healthcare professionals[12]	– Ergonomic and safety with safety features audit; – Safe-patient handling activities included 1-hour of manager training and one-to-one training in handling; – Encourage physical fitness with stretching training section and one-to-one mentoring session;	3 months	*Primary* WMSDs[1] injury rates WMSDs[1] pain intensity *Secondary* Correct patient handling practice	No changes in WMSDs[1] pain injuries were shown. A significant increase in safe patient handling practices ($p < 0.0001$) was shown.
Theis and Finkelstein (2014) (3)	To evaluate the effectiveness of a safe patient handling program (STEPS) in reducing injury due to patient transfers	Pre-post intervention study	*N* = 55 Healthcare professionals[13]	– STEPS (Safe Transfers Every Person Succeeds) – 8-hour hands-on training class and a pre-knowledge test; – Equipment ceiling lifts installation;	4 year	*Primary* WMSDs[1] injuries rates	The number of WMSDs[1]injuries rates was significantly reduced at post training compared to baseline ($p = 0.01$), but not sustained at long term.

(*continued on next page*)

Table 2 (*continued*)

Individual (n = 4; 14.8%)							
First author, year (quality rating)	Aim	Design	Participants	Intervention	Total follow-up	Outcomes	Key Results
Powell-Cope et al., 2014 (2)	To identify which components of a system-wide safe patient handling (SPH) program reduced musculoskeletal injury (MSI) due to patient handling among nurses	Longitudinal pre-post intervention	*N* = 141 Nurses	– Ceiling lifts development and other safe patient handling technologies; – Peer leader effectiveness rating by facility coordinator; – Competency assessment in the use of SPH equipment; – Peer leader training; – Achievement of program milestones at 5th data collection points; – Program support from key stakeholders; – Conduct of equipment fairs; – Incorporation of SPH into new employee orientation;	3 years	*Primary* WMSDs[1] injuries rates	The WMSDs[1] injuries rates decreased significantly ($p = 0.006$). Lastly, competency in the use of SPH equipment significantly moderated the risk association between facility complexity and WMSDs incidence rate ($p = 0.037$).
Dennerlein et al., 2017 (2)	To evaluate a safe patient handling and mobilisation programme within the context of a hospital-wide patient care and integrated safe patient equipment and practices into patient care plans	Pre-post intervention study	*N* = 1832 Healthcare professionals[14]	– Broad-based training; – Building a hospital-wide infrastructure for maintaining and servicing equipment; – Implementing a mentoring programme to sustain training efforts and dissemination of new information; – Strong communication programme with leaders, workers and clients;	3 years	*Primary* WMSDs[1] injury rates WMSDs[1] pain intensity *Secondary* Correct patient handling practice	A significant decrease in neck and shoulder injuries (RR: 0.68, 95% CI 0.46–1.00), lifting and exertion (RR: 0.73, 95% CI 0.60–0.89), and WMSDs pain intensity (RR: 0.78, 95% CI 0.62–1.00) were showed. Safe patient handling practices improved significantly ($p < 0.0001$) in the intervention hospital.
Risør et al., 2017 (1)	To evaluate an intervention to reduce WMSD injuries and lost workdays, to improve the use of patient handling equipment and improve their general health, aggressive episodes, and work-related accidents	Before-after intervention study	*N* = 937 Healthcare professionals[15]	– Development and dissemination of patient-handling guidelines; – Create guidelines for purchasing new equipment; – Purchasing new patient-handling; – Comprehensive training programme for all nursing staff in the intervention bed-wards; – Local and two days training patient handling instructors; – Project manager visiting;	1 year	*Primary* WMSDs[1] injury rates WMSDs[1] pain intensity *Secondary* Absence from work Correct patient handling practice	No changes were observed in WMSDs[1] injury rates, pain intensity and general health status and absence from work. The intervention resulted in more positive attitudes and behaviours for safe patient-handling.
Gold et al., 2018 (2)	To evaluate the effectiveness of a multi-component program (ProCare) in terms of physical exposure reduction, reduced WMSDs symptoms and injury rates	Cross-sectional intervention study	*N* = 4526 Healthcare professionals[16]	– Multi-component program included purchase of sufficient mechanical lifts for the residents in each facility – Training and protocols for lift use	4 years	*Primary* WMSDs[1] pain intensity Knee pain *Secondary* Correct patient handling practice	WMSDs[1] pain intensity and knee pain was more prevalent in nursing personnel than others were ($p = 0.0005$). Knee pain was associated with both physical and psychosocial work exposures (PRR: 1.03, 1.00–1.05), psychological job demands (PRR: 1.09, 1.02–1.17) and job strain (PRR: 1.30, 1.10–1.54). Social support was negatively associated with knee pain.
Sezgin & Esin, 2018 (1)	To evaluate effects of a PRECEDE-PROCEED Model based, nurse-delivered Ergonomic Risk Management Program (ERMP) in reducing WMSDs	Pre-post intervention study	*N* = 72 Nurses	– Health promotion programme; – Video training, interviews and training exercise;	6 months	*Primary* WMSDs[1] injury rates WMSDs[1] pain intensity	There was no difference between the intervention and control group for pain intensity and injury rates ($p > 0.05$). No significant reduction of analgesic intake ($p = 0.460$) and absence from work ($p = 0.145$) were shown.

Table 2 (*continued*)

Individual (n = 4; 14.8%)							
First author, year (quality rating)	Aim	Design	Participants	Intervention	Total follow-up	Outcomes	Key Results
Jakobsen et al., 2019 (2)	To evaluate the effect of a participatory organizational intervention for improve the use of assistive device (AD) in patient transfer	Randomized Controlled Trial	$N = 625$ Healthcare professionals[17]	– Assessment of barriers and potential solutions among participants; – Workshops and training in the assistive devices equipment; – Action plan development;	1 year	*Primary* WMSDs[1] pain intensity WMSDs[1] injury rates *Secondary* Correct patient handling practice	WMSDs[1] pain intensity, low-back pain, and WMSDs[1] injury rates did not change in the intervention group compared with the control one ($p > 0.05$). Correct patients handling practice and the general use of AD improved in the intervention group, but it was not significant ($p = 0.042$).
Coskun Beyan et al., 2020 (1)	To determine the effects of a multifaceted ergonomics intervention in reducing WMSDs in intensive care nurses	Case-control pre-post intervention study	$N = 64$ Nurses	Four ergonomic interventions were planned and assessed by the introduction of an ERGO team: – Individual training, stretching exercise and motivation meetings; – Administrative intervention; – Engineering interventions in lifts use;	2 years	*Primary* WMSDs[1] injury rates	WMSDs[1] pain intensity in the intervention group was highest for lumbar back pain (45%), knees (37%), right shoulder (36%), neck (25%), and upper right arm (24%).

*Quality rating: 1 = strong, 2 = moderate, 3 = weak; *Significant values are at $p < 0.05$.
[6]FTE: Full Time Equivalent.
[1] WMSDs: Work-related musculoskeletal disorders.
[2] Investigated healthcare professionals (HCPs): nurses, nurses' aides, attendants, physical/occupational therapists (PT/OT), clerks/unit assistants, others.
[3] CI: Confidence Interval.
[4] RR: Relative Risk.
[5] Investigated healthcare professionals (HCPs): Nurse, Nurses' aides, manager, Physiotherapist (PT), Occupational therapists (OT), PT/OT aides and radiology aides.
[7] NLS: No Lifting Policy.
[8] PR: Prevalence Ratio.
[9] Investigated healthcare professionals (HCPs): Nurses, nurses' aides, technicians and nurses' managers.
[10] Investigated healthcare professionals (HCPs): Nurses, nurses' aides, attendants, clerks, PT, unit supporter.
[11] Investigated healthcare professionals (HCPs): Nurses, patient care associate, clinical nurses.
[12] Investigated healthcare professionals (HCPs): Nurses, Nurses' aides, Therapists and service assistance.
[13] Investigated healthcare professionals (HCPs): Nurses, Nurses' aides, OT, PT, PT' assistants.
[14] Investigated healthcare professionals (HCPs): nurses, nurses' aides, attendants, physical/occupational therapists (PT/OT), clerks/unit assistants, others.
[15] Investigated healthcare professionals (HCPs): nurses, nurses' aides, attendants, physical/occupational therapists (PT/OT), clerks/unit assistants, others.
[16] PRR: Prevalence Risk Ratio.
[17] Investigated healthcare professionals (HCPs): Nurses.

Table 3
Classification of studies according to interventions and their quality rating ($N = 27$).

First author, year	Quality rating		
	Strong (1)	Moderate (2)	Weak (3)
Individual (n = 4; 14.8%)			
Menzel et al., 2004			X
Jakobsen et al., 2015		X	
Jakobsen et al., 2017		X	
Taulaniemi et al., 2019	X		
Task-specific and equipment (n = 4; 14.8%)			
Li et al., 2004		X	
Schoenfisch et al., 2013			X
Vieira and Brunt, 2016		X	
Alperovitch-Najenson et al., 2020			X
Work organisation and job design (n = 2; 7.4%)			
Engkvist, 2006		X	
Springer et al., 2009			X
Workplace environment (n = 1; 3.7%)			
Lee et al., 2019		X	
Multifactorial (n = 16; 59.2%)			
Yassi et al., 2001	X		
Collins et al., 2004		X	
Hartvigsen et al., 2005		X	
Nelson et al., 2006	X		
Zadvinskis and Salsbury, 2010		X	
Lim et al., 2011		X	
Black et al., 2011		X	
Caspi et al., 2013		X	
Theis and Finkelstein (2014)			X
Powell-Cope et al., 2014		X	
Dennerlein et al., 2017		X	
Risør et al., 2017	X		
Gold et al., 2018		X	
Sezgin and Esin, 2018	X		
Jakobsen et al., 2019		X	
Coskun Beyan et al., 2020	X		

ing with work ($p = 0.035$). In contrast, the CBT intervention reported no significant reduction in pain intensity ($p = 0.06$).

A significant reduction ($p = 0.005$) of analgesics intake and improvements in wellbeing, job satisfaction, and motivation ($p < 0.05$) were shown in work physical exercise by Jakobsen (2015, 2017). The NME intervention significantly increased ($p = 0.007$) work-related physical functioning and significantly reduced work stress ($p = 0.06$) among the intervention group.

3.4.2. Task-specific and equipment factors interventions

This group collected articles testing the use of aids, such as manual, mechanical lifting interventions. The mechanical lifting group included interventions on portable full-body or stand-up lifts (Li et al., 2004), while the studies on manual lifting evaluated the use of technical aids (i.e., sliding sheets or walk belts; Schoenfisch et al., 2013; Alperovitch-Najenson et al., 2020) and the use of unstable shoes (Vieira & Brunt, 2016). The use of portable full-body or stand-up lifts (Li et al., 2004) was effective in post-intervention, and decreased WMSDs injuries' rates, (Relative Risk [RR]: 0.37, 95% Confidence Interval [CI] 0.16–0.88). A constant use (at least 6 months) of sliding sheets significantly reduced ($p < 0.001$) low-back and neck pain (Alperovitch-Najenson et al., 2020). General pain intensity was significantly reduced among professionals wearing unstable shoes for 4 weeks ($p < 0.009$) and 6 weeks ($p < 0.001$). The 13-year patient lift and transfer program (Schoenfisch et al., 2013) showed a considerable decrease in WMSDs injuries following the intervention (RR: 0.56, 95% CI 0.36–0.87), but it was not significant.

Absence from work was reduced (RR: 0.35, 95% CI 0.10–1.16) in two studies (Li et al., 2004, Schoenfisch et al., 2013). Musculoskeletal comfort significantly improved using portable and stand-up lifts ($p < 0.05$), and cost related injuries decreased (Li et al., 2004).

3.4.3. Work organization and job design factors interventions

This group collected interventions on No Lifting Policy (NLS) (Engkvist, 2006) and the implementation of a lift team (Springer et al., 2009). NLS showed a significant reduction on back pain ($p < 0.001$) in nurses that received the intervention, compared with the professionals who did not receive it. The lift team was useful to reduce the risk of injuries among professionals ($p = 0.06$), although the frequency of lesions was strongly related to patient acuity ($p = 0.006$). A decrease ($p < 0.01$) of physical tiredness was reported at the NLS hospital compared with the control.

3.4.4. Workplace environment interventions

This group used a safe patient handling policy (Lee et al., 2019). A significant reduction was observed for pain intensity (61% vs. 52%) (Prevalence Ratio [PR]: 0.78, 95% CI 0.66–0.91, specifically for low back PR: 0.71, 95% CI 0.55–0.92, neck PR: 0.60, 95% CI 0.44–0.82, and hands/wrists PR: 0.59, 95% CI 0.39–0.91. No difference was shown for WMSD injuries (21% vs. 19%) RR: 0.86, 95% CI 0.55–1.33.

3.4.5. Multifactorial interventions

The studies included in this group evaluated combinations of different types of interventions. Eight studies (Yassi et al., 2001; Collins et al., 2004; Hartvigsen et al., 2005; Nelson et al., 2006; Black et al., 2011; Theis and Finkelstein (2014); Gold et al., 2018; Jakobsen et al., 2019) principally tested uses of mechanical aids, theoretical training, and practical in-ward activities on safe patient handling. Five studies added to the above interventions training in ergonomics that included the supervision of managers/administrative staffs (Zadvinskis & Salsbury, 2010; Lim et al., 2011; Powell-Cope et al., 2014; Risør et al. 2017), or implemented communication among work units (Dennerlein et al., 2017). Then, three studies tested mechanical aids and theoretical training, combined with physical exercises (Caspi et al., 2013; Sezgin & Esin, 2018). Only one study added a lift team (Coskun Beyan et al., 2020). Among the eight articles testing mechanical aids, theoretical training, and practical in-ward activities on safe patient handling, four reported no significant reduction in WMSD injury rates; only one study (Yassi et al., 2001) showed a reduction on WMSD pain among safe-lifting intervention' groups. Significant reduction of WMSD injuries' rates ($p = 0.05$) was found by Collins et al. (2013) and by Black et al. (2011) ($p < 0.001$) with Transfer Lifting and Repositioning Program (TLR). The STEPS program of Theis and Finkelstein (2014) also reduced the number of injuries ($p = 0.01$). The intervention of Nelson et al. (2006) showed that overall injury rates decreased significantly post-intervention (from 24.0 to 16.9; $p = 0.036$). Between the five studies testing interventions on training in ergonomics, the supervision of managers/administrative staffs and communication, four showed a reduction of WMSD injury rates. Powell-Cope et al. (2014) reported a significant decrease ($p = 0.006$) of WMSD injuries. Dennerlein et al. (2017) showed a significant decrease in neck and shoulder injuries RR: 0.68, 95% CI 0.46 –1.00, and pain RR: 0.78, 95% CI 0.62 –1.00. Zadvinskis and Salsbury (2010) showed an overall reduction in WMSD injuries of intervention group, compared to the control one. Lim et al. (2011) found fewer repeated injuries in the intervention than in the control group ($p = 0.001$ and $p = 0.002$, respectively). Only one study (Risør et al., 2017) showed no significant ($p > 0.05$) reduction of pain intensity and WMSD injuries. Studies that tested mechanical aids and theoretical training, combined with physical exercises and the addition of a lift team, observed no change in WMSD injuries and pain between intervention and

Table 4
Study effectiveness on primary and secondary outcomes proposed by the study ($N = 27$).

First author, year, (quality appraisal)	Provided intervention	Primary outcomes					Secondary outcomes				
		WMSD Injuries	General WMSD pain intensity	Low-back pain intensity	Neck/ shoulder pain intensity	Muscle strength/physical functioning or comfort	Reduction of absence from work	Reduction of analgesics intake	Wellbeing/ reduction of work stress	Increase of Safe Patients Handling activities	Reduction of cost related injuries
Individual ($n = 4$; 14.8%)											
Menzel et al., 2006 (3)	Cognitive Behavioural Therapy (CBT)										
Jakobsen et al., 2015 (2)	Workplace versus home-base physical exercise (WORK vs HOME)		√	√				√			
Jakobsen et al., 2017 (2)	Workplace versus home-base physical exercise (WORK vs HOME)		√								
Taulaniemi et al., 2019 (1)	Pylates neuromuscular exercise (NME)		√	√	√	√			√		
Task-specific and equipment ($n = 4$; 14.8%)											
Li et al., 2004 (2)	Use of mechanical patient lifts	√	√				√				√
Schoenfisch et al., 2013 (3)	Use of patients' lift and transfer equipment	√					√				
Vieira and Brunt, 2016 (2)	Wearing unstable shoes reduces		√	√							
Alperovitch-Najenson et al., 2020 (3)	Effects of sliding sheet usage		√								
Work organisation and job design ($n = 2$; 7.4%)											
Engkvist, 2006 (2)	No Lifting Policy (NLS)		√						√		√
Springer et al., 2009 (3)	Use of a lift team	√									
Workplace environment ($n = 1$; 3.7%)											
Lee et al., 2019 (2)	California Safe Patient Handling Legislation		√	√	√						
Multifactorial ($n = 16$; 59.2%)											
Yassi et al., 2001 (1)	Safe lifting and intensive training vs- Usual practice		√				√		√		√
Collins et al., 2004 (2)	Best practices training, use of mechanical lifts and no lifting policy	√					√				
Hartvigsen et al., 2005 (2)	Intensive educational, low-tech ergonomic and In-ward training on patient transfer										
Nelson et al., 2006 (1)	– Ergonomic Assessment Protocol; – Patient Handling Assessment Criteria and Decision Algorithms; – Peer Leader evaluation;	√					√			√	

(continued on next page)

Table 4 (*continued*)

First author, year, (quality appraisal)	Provided intervention	Primary outcomes					Secondary outcomes				
		WMSD Injuries	General WMSD pain intensity	Low-back pain intensity	Neck/shoulder pain intensity	Muscle strength/physical functioning or comfort	Reduction of absence from work	Reduction of analgesics intake	Wellbeing/ reduction of work stress	Increase of Safe Patients Handling activities	Reduction of cost related injuries
	– State-of-the-art Equipment; – After Action Reviews; – No Lift Policy;										
Zadvinskis and Salsbury, 2010 (2)	– Engineering (minimal-lift equipment); – Administrative (nursing policy); – Behavioral (peer coach program) controls;	√								√	
Black et al., 2011 (2)	Transfer, Lifting and Repositioning program (TLR)	√					√				√
Lim et al., 2011 (2)	– Ergonomic prevention program; – Engineering and administrative controls; – Staff education on anatomy, injuries, body mechanics, personal health, lifting and patient handling procedures; – Implementation of standardized patient handling needs assessment; – Skills based learning in daily activities;	√								√	√
Caspi et al., 2013 (2)	– - Ergonomic and safety with safety features audit; – - Safe-patient handling activities included 1-hour of manager training and one-to-one training in handling; – - Encourage physical fitness with stretching training section and one-to-one mentoring session;									√	
Theis and Finkelstein (2014) (3)	– - STEPS (Safe Transfers Every Person Succeeds) - 8-hour hands-on training class and a pre-knowledge test; - Equipment ceiling lifts installation;	√									
Powell-Cope et al., 2014 (2)	– Ceiling lifts development and other safe patient handling technologies;	√								√	

Table 4 (*continued*)

First author, year, (quality appraisal)	Provided intervention	Primary outcomes					Secondary outcomes				
		WMSD Injuries	General WMSD pain intensity	Low-back pain intensity	Neck/ shoulder pain intensity	Muscle strength/physical functioning or comfort	Reduction of absence from work	Reduction of analgesics intake	Wellbeing/ reduction of work stress	Increase of Safe Patients Handling activities	Reduction of cost related injuries
	– Peer leader effectiveness rating by facility coordinator; – Competency assessment in the use of SPH equipment; – Peer leader training; – Achievement of program milestones at 5th data collection points; – Program support from key stakeholders; – Conduct of equipment fairs; – Incorporation of SPH into new employee orientation;										
Dennerlein et al., 2017 (2)	– Broad-based training; – Building a hospital-wide infrastructure for maintaining and servicing equipment; – Implementing a mentoring programme to sustain training efforts and dissemination of new information; – Strong communication programme with leaders, workers and clients;	√	√							√	
Risør et al., 2017 (1)	– Development and dissemination of patient-handling guidelines; – Create guidelines for purchasing new equipment; – Purchasing new patient-handling; – Comprehensive training programme for all nursing staff in the intervention bed-wards; – Local and two days training patient handling instructors; – Project manager visiting;						√			√	

(*continued on next page*)

Table 4 (*continued*)

First author, year, (quality appraisal)	Provided intervention	Primary outcomes					Secondary outcomes				
		WMSD Injuries	General WMSD pain intensity	Low-back pain intensity	Neck/shoulder pain intensity	Muscle strength/physical functioning or comfort	Reduction of absence from work	Reduction of analgesics intake	Wellbeing/reduction of work stress	Increase of Safe Patients Handling activities	Reduction of cost related injuries
Gold et al., 2018 (2)	– Multi-component program included purchase of sufficient mechanical lifts for the residents in each facility – Training and protocols for lift use										
Sezgin and Esin, 2018 (1)	– Health promotion programme; – Video training, interviews and training exercise;										
Jakobsen et al., 2019 (2)	– Assessment of barriers and potential solutions among participants; – Workshops and training in the assistive devices equipment;									√	
Coskun Beyan et al., 2020 (1)	– Assessment of barriers and potential solutions among participants; – Workshops and training in the assistive devices equipment; – Action plan development Four ergonomic interventions were planned and assessed by the introduction of an ERGO team: – Individual training, stretching exercise and motivation meetings; – Administrative intervention; – Engineering interventions in lifts use;										

control groups (Caspi et al., 2013; Sezgin & Esin, 2018; Coskun Beyan et al., 2020).

Absence from work and use of medications decreased in most studies (Collins et al., 2004; Risør et al. 2017; Sezgin & Esin, 2018). Safe patient handling practices and the competency in patients handling aides increased in the intervention of Powell-Cope et al. (2014), Caspi et al. (2013), Dennerlein et al. (2017) and Zadvinskis and Salsbury (2010). Absence from work ($p = 0.09$) and cost-related injuries ($p = 0.013$) were significantly reduced only in the TLR intervention (Black et al., 2011).

4. Discussion

The main objective of this literature review was to provide a comprehensive description of existing interventions to prevent and reduce WMSD pain and/or injury rates among HCPs. The classification framework by Oakman et al. (2019) made it possible to organize and summarize the identified interventions in a structured way. Given the fragmentation of previous literature, this is an important result that could add order to the complexity of this topic.

Although several theoretical models exist, an inadequacy of the risk management strategies and tools on WMSDs seems still to be present, in line with MacDonald and Oakman (2015). In particular, the factors responsible for the development of WMSDs are often related to the casual impact of work events and psychosocial work stressor (EU-OSHA, 2020; Lang et al., 2012).

4.1. Effectiveness of interventions on WMSD injuries and pain

4.1.1. Individual interventions

Most individual interventions reduced WMSD pain. Repetitive actions or adverse postures are among the most frequent causes of MSD development (Gilchrist & Pokorná, 2021). Individual interventions such as the physical activity proposed by Jakobsen et al. (2015, 2017) or the NME of Taulaniemi et al. (2019), significantly reduced low-back pain among the HCPs studied. The positive effect of physical exercise is well known in the literature on MSDs (Airaksinen et al., 2006; Butera et al., 2019; Serra et al., 2018). Moreover, physical exercise helps in preventing new episodes of WMSD pain (Smith et al., 2019), improving strength, endurance, and neuromuscular control, as shown by two studies on rehabilitation patients by Voorn et al. (2019, 2021).

4.1.2. Task-specific and equipment factors interventions

A significant reduction in WMSD injuries was found with intervention focusing on manual and mechanical patient handing (Li et al., 2004; Schoenfisch et al., 2013; Alperovitch-Najenson et al., 2020). In particular, the use of mechanical or sliding lifts seems to reduce WMSD injuries among professionals. These results are in line with previous literature (de Cássia Pereira et al., 2016; Freiberg et al., 2016; Hegewald et al., 2018; Lietz et al., 2018), which found a reduction of the prevalence of WMSD injuries and pain among HCPs (Burdorf et al., 2013). However, several barriers to the use of the equipment have been identified including age, knowledge, availability of equipment, personal and contextual factors, and time pressures (Richardson et al., 2019). Moreover, the use of unstable shoes was effective to reduce WMSD pain intensity; perhaps unstable shoes, by strengthening muscle function, consequently ameliorate the stability of body and reduce the spinal tension (Lerebourg et al., 2020).

4.1.3. Work organization and job design factors interventions

The intervention of No Lifting Policy (NLS) proposed by (Engkvist, 2006) obtained effective results on pain. Some studies were on the application of NLS (Charney et al., 2006; Passfield et al., 2003; Harolds & Hurst, 2016; Vendittelli, Penprase & Pittiglio, 2016); however, the homogeneous application of NLS among hospitals is not easy (Vendittelli et al., 2016). From a recent study on NLS policy implementation, only a few target hospitals implemented NLS due to lack of lifting and handling equipment (Vendittelli et al., 2016). Moreover, maintaining NLS requires high costs (Vendittelli et al., 2016). The use of a lift team proposed by Springer et al. (2009) seems to reduce the risk of injuries among professionals. Possibly, by reducing the biomechanical efforts, this intervention reduces the repetitive lifting task of HCPs (Passfield et al., 2003).

4.1.4. Workplace environment interventions

The application of California Safe Patient Handling Legislation proposed by Lee et al. (2019) was effective in reducing low-back and neck pain. The dissemination of policies and guidelines on correct lifting may have increased HCPs sensitivity on correct patients handling activities and consequently reduced acute pain (da Costa & Vieira, 2010). Contextual factors, heavy workload, and temporal pressures experienced by workers have been well documented (Menzel et al., 2004; Retsas & Pinikahana, 2000; Vieira & Brunt, 2016) as casual determinants of WMSDs. Therefore, the intervention on these elements has a great potential for preventing WMSD injuries (Vendittelli et al., 2016). Moreover, the impact and control of regulations may influence professionals' work (EU-OSHA, 2020).

4.1.5. Multifactorial interventions

The majority of interventions were multifactorial (i.e., combinations of procedures considering multiple factors and acting on multiple levels; Oakman et al., 2019). This is in line with the multifaceted causes of MSDs (Zinzen et al., 2000; Soler-Font et al., 2019), whose effective management requires strategies that are integrated and comprehensive, rather than isolated from each other (Macdonald & Oakman, 2015). Generally, multifactorial interventions are defined as the set of interventions that act on multiple causes of risk of development of WMSDs (Macdonald & Oakman, 2015). Most studies in this category received a high-quality (strong or moderate) rating. This is important, as high-quality ratings suggest the feasibility and practicality of interventions to act on WMSDs (Stock et al., 2018).

Most multifactorial interventions act by combining manual or mechanical patient lifting (i.e., the implementation of lifting aids) and handling training. The most effective reduction of both WMSD injuries and pain seems to result from the interventions testing a tailored action on transferring, lifting, and repositioning patients. This could be explained by the fact that physical risk factors (also known as biomechanical risk factors, including posture-related risks, heavy lifting, or job hazards) principally depend on manual lifting efforts and insufficient handling knowledge (Garzillo et al., 2020; Hegewald et al., 2018; EU-OSHA, 2019). In particular, manual lifting is one of the main causes of injury among professionals, placing continuous stress on the ligaments of the spine, especially the lumbar ligaments (Holtermann et al., 2013). Moreover, healthcare professionals' lack of knowledge of correct manual handling or the use of aids and procedures, increases the exposure of their musculoskeletal system to lifting trauma (EU-OSHA, 2019). As noted by Garzillo et al. (2020) most professionals who have received inadequate training are at greater risk of developing WMSDs. Tailored and individual training is one of the main levels of prevention because it strengthens healthcare professionals' knowledge and enables them to develop the proper handling techniques and skills (Garzillo et al., 2020; EU-OSHA, 2019). Furthermore, training interventions act on individual work attitudes and on physical discomfort, while reducing the time lost on wrong handling techniques (Dong et al., 2019; Ruotsalainen et al., 2015). As noted by Marras et al. (2009), other effective multifactorial interventions acted on organizational constraints (e.g., implementation of communication between professionals and managers, or staffing assessment). In particular, Dennerlein et al. (2017) implemented communication channels and cooperation between staff and managers. As recognized by Yazdani and Wells (2018), a positive communication climate and organizational culture-oriented communication could be significant elements to prevent WMSDs and increase professionals' ability to work. Moreover, inadequate staffing levels were acknowledged as barriers to engaging with effective strategies, as sometimes the required approaches had to be performed alone (Richardson et al., 2019). Insufficient staffing also required professionals to work more quickly, exposing them to frequent patient handling, and reducing patient safety. In this regard, the study by Wåhlin et al. (2021) suggested that patient collaboration in handling could prevent the development of professionals' WMSDs and patients' falls.

4.2. Effectiveness of interventions on secondary outcomes

Almost all the task-specific and multifactorial (mechanical and practical in-wards activities) interventions seem to reduce workday absences and increase safe patient handling. The use of correct aids is probably related to lower musculoskeletal load, reducing work efforts of HCPs (Hegewald et al., 2018), even among those who already suffered from WMSD injuries and/or pain. The exacerbations of pain and/or muscle accidents in HCPs may be reduced by using the correct mechanical aides (Hegewald et al., 2018).

Work organization and individual interventions seem to reduce medication use and increase the sense of work well-being. The interventions may have been acting on changing the work dynamics, increasing job satisfaction, productivity, and the complicity between workers (Suni et al., 2018). Moreover, psychological stress or discomfort is causally associated with between 28% and 84% of healthcare professionals' upper back injuries. However, work attitudes and behaviors are difficult to change, as they are the product of different factors (e.g., work, family, past experiences, and beliefs), which make it challenging to provide effective interventions (Pincus et al., 2007; Leka & Cox, 2008; Van Hoof et al., 2018).

4.3. Planning targeted and effective interventions

Research evidence on the work-related causes of MSDs includes both psychosocial and physical hazards (Okaman et al., 2019). Given the primary outcome of this review on MSD injuries and pain, the multifactorial interventions resulted as the most effective. These interventions are in line with the multi-causality of WMSD injuries and pain. Among multi-factorial interventions, those that act on manual or mechanical lifting, in-ward training, and communication between leaders and workers, showed a greater effectiveness in reducing WMSD injuries and pain. In particular, the use of patients handling mechanical and manual aids is generally more effective in preventing and reducing WMSD injuries, and has lower costs (Haghewald et al., 2018). However, the use of mechanical and manual aids requires considerable training of HCPs, increasing their knowledge on handling activities. In-ward simulations, video modelling or in-ward training are possible interventions to strengthen knowledge and increase the skills of HCPs on patients' handling. Moreover, implementing the communication channels and close collaboration between managers and HCPs personnel is essential. Managers should recognize the impact of WMSDs on HCPs (Larsen et al., 2018; Richardson et al., 2018), and act on improving workers' occupational health and psychological well-being through prevention campaigns and active participation in work activities (Larsen et al., 2018).

4.4. Limitations

Some limitations should be acknowledged. It was not possible to conduct a meta-analysis due to the high heterogeneity of included studies. Moreover, although several attempts were made to obtain specific data on different professional groups, the non-targeting analysis and methodology of the studies included made it impossible. Another study limitation relates to difficulties in stratifying data according to clinical setting. In fact, most of the studies did not make a detailed analysis of the organizational characteristics, such as financial and material resources, staffing, and organizational models of individual departments. Furthermore, the type of patients, their level of autonomy, and their ability to collaborate in transfers were not analyzed. Finally, the review excluded studies from grey literature; this may mean that potentially effective interventions were omitted.

4.5. Practical application

The classification proposed by Oakman et al. (2019) and used in this study could orient healthcare managers in evaluating different categories of interventions for WMSDs. Evidence-based information from this review can guide managers to shape effective and continuous education programs, and clinicians to plan strategies for preventing and acting on WMSDs. Moreover, this study suggests the need to develop inter-professional collaborations among clinicians (such as the implementation of multi-professional lift teams). Finally, given the importance of individual characteristics in the development of WMSDs, healthcare professionals should take care of themselves by practicing physical activities, to reduce the causes associated with the development of WMSDs.

5. Conclusion

This review provides a comprehensive summary of all interventions for reducing WMSD injuries and pain. A significant reduction in WMSD injuries and pain was shown after multifactorial interventions. These interventions received the highest quality ratings, and this strengthens confidence in the effectiveness of their results. In particular, interventions combining the different elements (i.e., manual or mechanical lifting, theoretical, and in-ward training) act on primary and multiple causes of WMSDs. Multifactorial interventions also obtained effective results for other variables related to MSDs, such as workday absence and increased safe patients handling practices. More efforts should be made to implement and extend these promising interventions, to conduct more randomized controlled trials with homogeneous population groups, and to enable meta-analyses to be conducted. Moreover, some other interventions seem effective in reducing WMSD injuries and pain among HCPs. Individual interventions principally act on pain reduction. Task-specific and work organization or environment interventions seem to be effective in preventing MSD injuries, but may be affected by external and organizational elements. Therefore, more research is warranted on these interventions to better investigate their scope and effectiveness. Moreover, studies about each healthcare professional target group, their specific demographic and working characteristics, and the associated causes of the development of WMSDs, are required. Further understanding is also needed of the direct costs—in terms of professionals' health problems, absenteeism, injuries, and difficulties—associated with staff working continuity, particularly in light of an aging population and changes in healthcare needs.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

None.

Funding

This work is part of the project 'A uSer-centred Approach For improving both health care workers' and patients' conditions during patiEnt-handling MOVEments' - SAFE-MOVER, funded by a grant from the Campus Bio-Medico University of Rome.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsr.2022.05.004.

References

Airaksinen, O., Brox, J. I., Cedraschi, C., Hildebrandt, J., Klaber-Moffett, J., Kovacs, F., Mannion, A. F., Reis, S., Staal, J. B., Ursin, H., Zanoli, G., & COST B13 Working Group on Guidelines for Chronic Low Back Pain (2006). Chapter 4. European guidelines for the management of chronic nonspecific low back pain. *European Spine Journal*. https://doi.org/10.1007/s00586-006-1072-1.

Alperovitch-Najenson, D., Weiner, C., Ribak, J., & Kalichman, L. (2020). Sliding sheet use in nursing practice: An intervention study. *Workplace Health & Safety, 68*(4), 171–181. https://doi.org/10.1177/2165079919880566.

Amaro, J., Magalhães, J., Leite, M., Aguiar, B., Ponte, P., Barrocas, J., et al. (2018). Musculoskeletal injuries and absenteeism among healthcare professionals-ICD-10 characterization. *PloS one, 13*(12), e0207837.

Anderson, S. P., & Oakman, J. (2016). Allied health professionals and work-related musculoskeletal disorders: A systematic review. *Safety and Health at Work, 7*(4), 259–267. https://doi.org/10.1016/j.shaw.2016.04.001.

Armijo-Olivo, S., Stiles, C. R., Hagen, N. A., Biondo, P. D., & Cummings, G. G. (2012). Assessment of study quality for systematic reviews: A comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: Methodological research. *Journal of Evaluation in Clinical Practice, 18*(1), 12–18. https://doi.org/10.1111/j.1365-2753.2010.01516.x.

Aslam, I., Davis, S. A., Feldman, S. R., & Martin, W. E. (2015). A review of patient lifting interventions to reduce health care worker injuries. *Workplace health & safety, 63*(6), 267–276. https://doi.org/10.1177/2165079915580038.

Asuquo, E. G., Tighe, S. M., & Bradshaw, C. (2021). Interventions to reduce work-related musculoskeletal disorders among healthcare staff in nursing homes. An integrative literature review. *International Journal of Nursing Studies, 3*. https://doi.org/10.1016/j.ijnsa.2021.100033.

Black, T. R., Shah, S. M., Busch, A. J., Metcalfe, J., & Lim, H. J. (2011). Effect of transfer, lifting, and repositioning (TLR) injury prevention program on musculoskeletal injury among direct care workers. *Journal of Occupational and Environmental Hygiene, 8*(4), 226–235. https://doi.org/10.1080/15459624.2011.564110.

Burdorf, A., Koppelaar, E., & Evanoff, B. (2013). Assessment of the impact of lifting device use on low back pain and musculoskeletal injury claims among nurses. *Occupational and environmental medicine, 70*(7), 491–497. https://doi.org/10.1136/oemed-2012-101210.

Butera, K. A., Roff, S. R., Buford, T. W., & Cruz-Almeida, Y. (2019). The impact of multisite pain on functional outcomes in older adults: Biopsychosocial considerations. *Journal of Pain Research, 12*, 1115–1125. https://doi.org/10.2147/JPR.S192755.

Caspi, C. E., Dennerlein, J. T., Kenwood, C., Stoddard, A. M., Hopcia, K., Hashimoto, D., et al. (2013). Results of a pilot intervention to improve health and safety for health care workers. *Journal of Occupational and Environmental Medicine, 55*(12), 1449–1455. https://doi.org/10.1097/JOM.0b013e3182a7e65a.

Charney, W., Simmons, B., Lary, M., & Metz, S. (2006). Zero lift programs in small rural hospitals in Washington state: Reducing back injuries among health care workers. *AAOHN Journal., 54*(8), 355–358.

Clari, M., Godono, A., Garzaro, G., Voglino, G., Gualano, M. R., Migliaretti, G., et al. (2021). Prevalence of musculoskeletal disorders among perioperative nurses: A systematic review and META-analysis. *BMC Musculoskeletal Disorders, 22*(1), 226. https://doi.org/10.1186/s12891-021-04057-3.

Collins, J. W., Wolf, L., Bell, J., & Evanoff, B. (2004). An evaluation of a "best practices" musculoskeletal injury prevention program in nursing homes. *Injury Prevention, 10*(4), 206–211. https://doi.org/10.1136/ip.2004.005595.

Coskun Beyan, A., Dilek, B., & Demiral, Y. (2020). The effects of multifaceted ergonomic interventions on musculoskeletal complaints in intensive care units. *International Journal of Environmental Research and Public Health, 17*(10), 3719. https://doi.org/10.3390/ijerph17103719.

da Costa, B. R., & Vieira, E. R. (2010). Risk factors for work-related musculoskeletal disorders: A systematic review of recent longitudinal studies. *American Journal of Industrial Medicine, 53*(3), 285–323. https://doi.org/10.1002/ajim.20750.

de Cássia Pereira Fernandes, R., Pataro, S., de Carvalho, R. B., & Burdorf, A. (2016). The concurrence of musculoskeletal pain and associated work-related factors: a cross sectional study. *BMC Public Health*, 16, 628. https://doi.org/10.1186/s12889-016-3306-4

Dennerlein, J. T., O'Day, E. T., Mulloy, D. F., Somerville, J., Stoddard, A. M., Kenwood, C., et al. (2017). Lifting and exertion injuries decrease after implementation of an integrated hospital-wide safe patient handling and mobilisation programme. *Occupational and Environmental Medicine, 74*(5), 336–343. https://doi.org/10.1136/oemed-2015-103507.

Dong, H., Zhang, Q., Liu, G., Shao, T., & Xu, Y. (2019). Prevalence and associated factors of musculoskeletal disorders among Chinese healthcare professionals working in tertiary hospitals: A cross-sectional study. *BMC Musculoskeletal Disorders, 20*(1), 175. https://doi.org/10.1186/s12891-019-2557-5.

Effective Public Health Practice Project (2020). Quality Assessment Tool for Quantitative Studies Retrieved from https://www.ephpp.ca/quality-assessment-tool-for-quantitative-studies/.

Engkvist, I. L. (2006). Evaluation of an intervention comprising a no lifting policy in Australian hospitals. *Applied Ergonomics, 37*(2), 141–148. https://doi.org/10.1016/j.apergo.2005.05.008.

European Agency for Safety and Health at Work (2013). Priorities for occupational safety and health research in Europe: 2013–2020 Retrieved from https://osha.europa.eu/en/publications/reports/priorities-for-occupational-safety-and-health-research-in-europe-2013-2020.

European Agency for Safety and Health at Work (2019). Work-related musculoskeletal disorders: Prevalence, costs and demographics in the EU Retrieved from: https://osha.europa.eu/en/publication.

European Agency for Safety and Health at Work (2020). Musculoskeletal disorders in the healthcare sector Retrieved from https://osha.europa.eu/sites/default/files/publications/documents/Discussion_paper_MSDs_in_health_care_sector.pdf.

European Social Statistics. *Accidents at work and work-related health problems*. (2003). Retrieved from https://osha.europa.eu/it/publications/msds-facts-and-figures-overview-prevalence-costs-and-demographics-msds-europe/view.

Falagas, M. E., Kouranos, V. D., Arencibia-Jorge, R., & Karageorgopoulos, D. E. (2008). Comparison of SCImago journal rank indicator with journal impact factor. *FASEB Journal, 22*(8), 2623–2628. https://doi.org/10.1096/fj.08-107938.

Freiberg, A., Euler, U., Girbig, M., Nienhaus, A., Freitag, S., & Seidler, A. (2016). Does the use of small aids during patient handling activities lead to a decreased occurrence of musculoskeletal complaints and diseases? A systematic review. *International Archives of Occupational and Environmental Health, 89*(4), 547–559. https://doi.org/10.1007/s00420-015-1094-2.

Garzillo, E. M., Monaco, M., Corvino, A. R., D'Ancicco, F., Feola, D., Della Ventura, D., et al. (2020). Healthcare workers and manual patient handling: A pilot study for interdisciplinary training. *International Journal of Environmental Research and Public Health, 17*(14), 4971. https://doi.org/10.3390/ijerph17144971.

Gilchrist, A., & Pokorná, A. (2021). Prevalence of musculoskeletal low back pain among registered nurses: Results of an online survey. *Journal of Clinical Nursing, 30*(11–12), 1675–1683. https://doi.org/10.1111/jocn.15722.

Gold, J. E., Kurowski, A., Gore, R. J., ProCare Research Team & Punnett, L. (2018). Knee pain in nursing home workers after implementation of a safe resident handling program. *American Journal of Industrial Medicine, 61*(10), 849–860. https://doi.org/10.1002/ajim.22902.

Harolds, L., & Hurst, H. (2016). Preventing workplace injuries among perinatal nurses. *Nursing for women's health, 20*(1), 99–108. https://doi.org/10.1016/j.nwh.2015.12.003.

Hartvigsen, J., Lauritzen, S., Lings, S., & Lauritzen, T. (2005). Intensive education combined with low tech ergonomic intervention does not prevent low back pain in nurses. *Occupational and Environmental Medicine, 62*(1), 13–17. https://doi.org/10.1136/oem.2003.010843.

Hegewald, J., Berge, W., Heinrich, P., Staudte, R., Freiberg, A., Scharfe, J., et al. (2018). Do technical aids for patient handling prevent musculoskeletal complaints in health care workers?-A systematic review of intervention studies. *International journal of Environmental Research and Public Health, 15*(3), 476. https://doi.org/10.3390/ijerph15030476.

Hignett, S. (2003). Intervention strategies to reduce musculoskeletal injuries associated with handling patients: A systematic review. *Occupational and Environmental Medicine, 60*(9), E6. https://doi.org/10.1136/oem.60.9.e6.

Holtermann, A., Clausen, T., Jørgensen, M. B., Burdorf, A., & Andersen, L. L. (2013). Patient handling and risk for developing persistent low-back pain among female healthcare workers. *Scandinavian Journal of Work, Environment & Health, 39*(2), 164–169. https://doi.org/10.5271/sjweh.3329.

Jäger, M., Jordan, C., Theilmeier, A., Wortmann, N., Kuhn, S., Nienhaus, A., et al. (2013). Lumbar-load analysis of manual patient-handling activities for biomechanical overload prevention among healthcare workers. *The Annals of Occupational Hygiene, 57*(4), 528–544. https://doi.org/10.1093/annhyg/mes088.

Jakobsen, M. D., Aust, B., Kines, P., Madeleine, P., & Andersen, L. L. (2019). Participatory organizational intervention for improved use of assistive devices in patient transfer: A single-blinded cluster randomized controlled trial.

Scandinavian Journal of Work, Environment & Health, 45(2), 146–157. https://doi.org/10.5271/sjweh.3769.
Jakobsen, M. D., Sundstrup, E., Brandt, M., & Andersen, L. L. (2017). Factors affecting pain relief in response to physical exercise interventions among healthcare workers. *Scandinavian Journal of Medicine & Science in Sports, 27*(12), 1854–1863. https://doi.org/10.1111/sms.12802.
Jakobsen, M. D., Sundstrup, E., Brandt, M., Jay, K., Aagaard, P., & Andersen, L. L. (2015). Effect of workplace- versus home-based physical exercise on musculoskeletal pain among healthcare workers: A cluster randomized controlled trial. *Scandinavian Journal of Work, Environment & Health, 41*(2), 153–163. https://doi.org/10.5271/sjweh.3479.
Koppelaar, E., Knibbe, J. J., Miedema, H. S., & Burdorf, A. (2009). Determinants of implementation of primary preventive interventions on patient handling in healthcare: A systematic review. *Occupational and Environmental Medicine, 66*(6), 353–360. https://doi.org/10.1136/oem.2008.042481.
Lang, J., Ochsmann, E., Kraus, T., & Lang, J. W. (2012). Psychosocial work stressors as antecedents of musculoskeletal problems: A systematic review and meta-analysis of stability-adjusted longitudinal studies. *Social science & medicine, 75*(7), 1163–1174. https://doi.org/10.1016/j.socscimed.2012.04.015.
Larsen, A. K., Falkenstrøm, S., Jørgensen, M. B., et al. (2018). The role of managers in addressing employees with musculoskeletal pain: A mixed methods study. *International Archives of Occupational and Environmental Health, 91*, 361–372. https://doi.org/10.1007/s00420-017-1284-1.
Lee, S. J., Lee, J. H., & Harrison, R. (2019). Impact of California's safe patient handling legislation on musculoskeletal injury prevention among nurses. *American Journal of Industrial Medicine, 62*(1), 50–58. https://doi.org/10.1002/ajim.22923.
Leka, S., & Cox, T. (Eds.). (2008). *PRIMA-EF: Guidance on the European framework for psychosocial risk management*. Geneva: World Health Organization.
Lerebourg, L., L'Hermette, M., Menez, C., & Coquart, J. (2020). The effects of shoe type on lower limb venous status during gait or exercise: A systematic review. *PloS ONE, 15*(11), e0239787.
Li, J., Wolf, L., & Evanoff, B. (2004). Use of mechanical patient lifts decreased musculoskeletal symptoms and injuries among health care workers. *Injury Prevention, 10*(4), 212–216. https://doi.org/10.1136/ip.2003.004978.
Lietz, J., Kozak, A., & Nienhaus, A. (2018). Prevalence and occupational risk factors of musculoskeletal diseases and pain among dental professionals in Western countries: A systematic literature review and meta-analysis. *PloS ONE, 13*(12), e0208628.
Lim, H. J., Black, T. R., Shah, S. M., Sarker, S., & Metcalfe, J. (2011). Evaluating repeated patient handling injuries following the implementation of a multi-factor ergonomic intervention program among health care workers. *Journal of Safety Research, 42*(3), 185–191. https://doi.org/10.1016/j.jsr.2011.05.002.
Macdonald, W., & Oakman, J. (2015). Requirements for more effective prevention of work-related musculoskeletal disorders. *BMC Musculoskeletal Disorders, 16*(293). https://doi.org/10.1186/s12891-015-0750-8.
Marras, W. S., Cutlip, R. G., Burt, S. E., & Waters, T. R. (2009). National occupational research agenda (NORA) future directions in occupational musculoskeletal disorder health research. *Applied Ergonomics, 40*(1), 15–22. https://doi.org/10.1016/j.apergo.2008.01.018.
Menzel, N. N., & Robinson, M. E. (2006). Back pain in direct patient care providers: Early intervention with cognitive behavioral therapy. *Pain Management Nursing, 7*(2), 53–63. https://doi.org/10.1016/j.pmn.2006.02.002.
Menzel, N. N., Brooks, S. M., Bernard, T. E., & Nelson, A. (2004). The physical workload of nursing personnel: Association with musculoskeletal discomfort. *International Journal of Nursing Studies, 41*(8), 859–867. https://doi.org/10.1016/j.ijnurstu.2004.03.012.
Milhem, M., Kalichman, L., Ezra, D., & Alperovitch-Najenson, D. (2016). Work-related musculoskeletal disorders among physical therapists: A comprehensive narrative review. *International Journal of Occupational Medicine and Environmental Health, 29*(5), 735–747. https://doi.org/10.13075/ijomeh.1896.00620.
Nelson, A., Matz, M., Chen, F., Siddharthan, K., Lloyd, J., & Fragala, G. (2006). Development and evaluation of a multifaceted ergonomics program to prevent injuries associated with patient handling tasks. *International Journal of Nursing Studies, 43*(6), 717–733. https://doi.org/10.1016/j.ijnurstu.2005.09.004.
Ngan, K., Drebit, S., Siow, S., Yu, S., Keen, D., & Alamgir, H. (2010). Risks and causes of musculoskeletal injuries among health care workers. *Occupational Medicine, 60*(5), 389–394. https://doi.org/10.1093/occmed/kqq052.
Oakman, J., Macdonald, W., & Kinsman, N. (2019). Barriers to more effective prevention of work-related musculoskeletal and mental health disorders. Applied ergonomics, 75, 184–192. https://doi.org/10.1016/j.apergo.2018.10.007.
Oakman, J., Clune, S., & Stuckey, R. (2019). *Work-related musculoskeletal disorders in Australia, 2019*. Canberra: Safe Work Australia. Retrieved from https://www.safeworkaustralia.gov.au/system/files/documents/1912/work-related_musculoskeletal_disorders_in_australia_0.pdf.
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, [n71]. https://doi.org/10.1136/bmj.n71.
Passfield, J., Marshall, E., & Adams, R. (2003). "No lift" patient handling policy implementation and staff injury rates in a public hospital. *Journal of Occupational Health and Safety, 19*, 73–85.
Pincus, T., Foster, N. E., Vogel, S., Santos, R., Breen, A., & Underwood, M. (2007). Attitudes to back pain amongst musculoskeletal practitioners: A comparison of professional groups and practice settings using the ABS-mp. *Manual Therapy, 12*(2), 167–175. https://doi.org/10.1016/j.math.2006.06.005.
Powell-Cope, G., Toyinbo, P., Patel, N., Rugs, D., Elnitsky, C., Hahm, B., et al. (2014). Effects of a national safe patient handling program on nursing injury incidence rates. *The Journal of Nursing Administration, 44*(10), 525–534. https://doi.org/10.1097/NNA.0000000000000111.
Retsas, A., & Pinikahana, J. (2000). Manual handling activities and injuries among nurses: An Australian hospital study. *Journal of Advanced Nursing, 31*(4), 875–883. https://doi.org/10.1046/j.1365-2648.2000.01362.x.
Richardson, A., Gurung, G., Derrett, S., & Harcombe, H. (2019). Perspectives on preventing musculoskeletal injuries in nurses: A qualitative study. *Nursing Open, 6*(3), 915–929. https://doi.org/10.1002/nop2.272.
Richardson, A., McNoe, B., Derrett, S., & Harcombe, H. (2018). Interventions to prevent and reduce the impact of musculoskeletal injuries among nurses: A systematic review. *International Journal of Nursing Studies, 82*, 58–67. https://doi.org/10.1016/j.ijnurstu.2018.03.018.
Risør, B. W., Casper, S. D., Andersen, L. L., & Sørensen, J. (2017). A multi-component patient-handling intervention improves attitudes and behaviors for safe patient handling and reduces aggression experienced by nursing staff: A controlled before-after study. *Applied Ergonomics, 60*, 74–82. https://doi.org/10.1016/j.apergo.2016.10.011.
Roquelaure, Y. (2018). *Musculoskeletal disorders and psychosocial factors at work*. ETUI Research Paper-Report.
Ruotsalainen, J. H., Verbeek, J. H., Mariné, A., & Serra, C. (2015). Preventing occupational stress in healthcare workers. *The Cochrane Database of Systematic Reviews, 2015*(4), CD002892. https://doi.org/10.1002/14651858.CD002892.pub5.
Schoenfisch, A. L., Lipscomb, H. J., Pompeii, L. A., Myers, D. J., & Dement, J. M. (2013). Musculoskeletal injuries among hospital patient care staff before and after implementation of patient lift and transfer equipment. *Scandinavian Journal of Work Environment & Health, 39*(1), 27–36. https://doi.org/10.5271/sjweh.3288.
Serra, M., Camargo, P. R., Zaia, J. E., Tonello, M., & Quemelo, P. (2018). Effects of physical exercise on musculoskeletal disorders, stress and quality of life in workers. *International Journal of Occupational Safety and Ergonomics, 24*(1), 62–67. https://doi.org/10.1080/10803548.2016.1234132.
Sezgin, D., & Esin, M. N. (2018). Effects of a PRECEDE-PROCEED model based ergonomic risk management programme to reduce musculoskeletal symptoms of ICU nurses. *Intensive & Critical Care Nursing, 47*, 89–97. https://doi.org/10.1016/j.iccn.2018.02.007.
Smith, B. E., Hendrick, P., Bateman, M., Holden, S., Littlewood, C., Smith, T. O., et al. (2019). Musculoskeletal pain and exercise-challenging existing paradigms and introducing new. *British Journal of Sports Medicine, 53*(14), 907–912. https://doi.org/10.1136/bjsports-2017-098983.
Soler-Font, M., Ramada, J. M., van Zon, S. K. R., Almansa, J., Bültmann, U., Serra, C., et al. (2019). Multifaceted intervention for the prevention and management of musculoskeletal pain in nursing staff: Results of a cluster randomized controlled trial. *PLoS One, 14*, e0225198.
Springer, P. J., Lind, B. K., Kratt, J., Baker, E., & Clavelle, J. T. (2009). Preventing employee injury: Implementation of a lift team. *Aaohn Journal, 57*(4), 143–148.
Stock, S. R., Nicolakakis, N., Vézina, N., Vézina, M., Gilbert, L., Turcot, A., et al. (2018). Are work organization interventions effective in preventing or reducing work-related musculoskeletal disorders? A systematic review of the literature. *Scandinavian Journal of Work Environment & Health, 44*(2), 113–133. https://doi.org/10.5271/sjweh.3696.
Suni, J. H., Kolu, P., Tokola, K., Raitanen, J., Rinne, M., Taulaniemi, A., et al. (2018). Effectiveness and cost-effectiveness of neuromuscular exercise and back care counseling in female healthcare workers with recurrent non-specific low back pain: A blinded four-arm randomized controlled trial. *BMC Public Health, 18*(1), 1376. https://doi.org/10.1186/s12889-018-6293-9.
Taulaniemi, A., Kankaanpää, M., Tokola, K., Parkkari, J., & Suni, J. H. (2019). Neuromuscular exercise reduces low back pain intensity and improves physical functioning in nursing duties among female healthcare workers, secondary analysis of a randomised controlled trial. *BMC Musculoskeletal Disorders, 20*(1), 328. https://doi.org/10.1186/s12891-019-2678-x.
Team TE (2013). *EndNote [64 bit]. EndNote X9*. Philadelphia, PA: Clarivate.
Theis, J. L., & Finkelstein, M. J. (2014). Long-term effects of safe patient handling program on staff injuries. *Rehabilitation Nursing, 39*(1), 26–35. https://doi.org/10.1002/rnj.108.
Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews on Evidence-based Nursing, 1*(3), 176–184. https://doi.org/10.1111/j.1524-475X.2004.04006.x.
United States Department of Labor (2020). Occupational injuries and illnesses Retrieved from https://www.bls.gov/iif/oshwc/case/msds.htm.
Van Hoof, W., O'Sullivan, K., O'Keeffe, M., Verschueren, S., O'Sullivan, P., & Dankaerts, W. (2018). The efficacy of interventions for low back pain in nurses: A systematic review. *International Journal of Nursing Studies, 77*, 222–231. https://doi.org/10.1016/j.ijnurstu.2017.10.015.
Vendittelli, D., Penprase, B., & Pittiglio, L. (2016). Musculoskeletal injury prevention for new nurses. *Workplace Health & Safety, 64*(12), 573–585. https://doi.org/10.1177/2165079916654928.
Verbeek, J., Martimo, K. P., Karppinen, J., Kuijer, P. P., Takala, E. P., & Viikari-Juntura, E. (2012). Manual material handling advice and assistive devices for preventing and treating back pain in workers: A Cochrane Systematic Review. *Occupational and Environmental Medicine, 69*(1), 79–80. https://doi.org/10.1136/oemed-2011-100214.

Vieira, E. R., & Brunt, D. (2016). Does wearing unstable shoes reduce low back pain and disability in nurses? A randomized controlled pilot study. *Clinical Rehabilitation, 30*(2), 167–173. https://doi.org/10.1177/0269215515576812.
Vieira, E. R., & Miller, L. (2008). Facing the challenge of patient transfers: Using ceiling lifts in healthcare facilities. *HERD, 2*(1), 6–16. https://doi.org/10.1177/193758670800200102.
Voorn, E. L., Koopman, F. S., Nollet, F., & Brehm, M. A. (2021). Individualized aerobic exercise in neuromuscular diseases: A pilot study on the feasibility and preliminary effectiveness to improve physical fitness. *Physical Therapy, 101*(3), pzaa213. https://doi.org/10.1093/ptj/pzaa213.
Voorn, E. L., Koopman, F., Nollet, F., & Brehm, M. A. (2019). Aerobic exercise in adult neuromuscular rehabilitation: A survey of healthcare professionals. *Journal of Rehabilitation Medicine, 51*(7), 518–524. https://doi.org/10.2340/16501977-2567.
Wåhlin, C., Lindmark, U., Wagman, P., Johnston, V., & Rolander, B. (2021). Work and health characteristics of oral health providers who stay healthy at work – a prospective study in public dentistry. *European Journal of Physiotherapy*. https://doi.org/10.1080/21679169.2021.1905876.
World Health Organization (2019). Musculoskeletal conditions Retrieved from https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions.
World Health Organization (2020). International classification of diseases 11th revision (IDC-11) Retrieved from https://www.icd.who.int/en.
Yassi, A., Cooper, J. E., Tate, R. B., Gerlach, S., Muir, M., Trottier, J., et al. (2001). A randomized controlled trial to prevent patient lift and transfer injuries of health care workers. *Spine, 26*(16), 1739–1746. https://doi.org/10.1097/00007632-200108150-00002.
Yazdani, A., & Wells, R. (2018). Barriers for implementation of successful change to prevent musculoskeletal disorders and how to systematically address them. *Applied Ergonomics, 73*, 122–140. https://doi.org/10.1016/j.apergo.2018.05.004.
Zadvinskis, I. M., & Salsbury, S. L. (2010). Effects of a multifaceted minimal-lift environment for nursing staff: Pilot results. *Western Journal of Nursing Research, 32*(1), 47–63. https://doi.org/10.1177/0193945909342878.
Zinzen, E., Caboor, D., Verlinden, M., Cattrysse, E., Duquet, W., Van Roy, P., et al. (2000). Will the use of different prevalence rates influence the development of a primary prevention programme for low-back problems? *Ergonomics, 43*(10), 1789–1803. https://doi.org/10.1080/001401300750004195.

Beatrice Albanesi is currently a Research Fellow in nursing at the University of Turin. She was a research fellow in nursing at the University Hospital Campus Bio-Medico, with the SAFEMOVER research project, focused on developing a robotic device for improving working conditions and user subjective perspective during patient-handling movements. She obtained her PhD in Nursing Sciences and Public Health at the University of Rome Tor Vergata. In addition to her academic career, she worked as a nurse in geriatrics, oncology and internal medicine units for ten years.

Michela Piredda has more than 20 years experience as clinical nurse at the University Hospital Campus Bio-Medico mostly in oncology, orthopaedics, medicine, geriatrics and surgical wards. She attended a Master in Nursing Research at King's College, London, and a Master in Nursing Science and a PhD in Nursing at the University of Rome Tor Vergata. She has extensive experience as nursing and research teacher at undergraduate, Master and Doctoral level. Since 2019 is Associate Professor in Nursing at Campus Bio-Medico di Roma University.

Marco Bravi is currently a PhD student in Human Movement and Sport Sciences at the University of Rome "Foro Italico". He graduated in physical therapy at Catholic University of Rome and he obtained his master's degree at the Tor Vergata University of Rome. In addition he is currently the head of physiotherapists at the Campus Bio-Medico University of Rome.

Federica Bressi is currently an associate professor at the Campus Bio-Medico University of Rome. She obtained her PhD in Sciences of plasticity, organ and tissue regeneration for functional recovery at the Campus Bio-Medico University of Rome. In addition to her academic career, she is a specialist in physiatry and neurology.

Raffaella Gualandi is Assistant Director of Nursing at the University Hospital Campus Bio-Medico of Rome. She earned her Ph.D. in Management, Economics and Industrial Engineering at Politecnico di Milano. She is a lecturer in Healthcare Management and Nursing Management at the University Campus Bio-Medico of Rome.

Anna Marchetti is currently the nurse manager of the palliative care center 'Insieme nella Cura' at the University Hospital Campus Bio-Medico. She was a Research Fellow in nursing at the University Hospital Campus Bio-Medico. She obtained her PhD in Nursing Sciences and Public Health at the University of Rome Tor Vergata.

Gabriella Facchinetti, is currently responsible for the management of home care of the palliative care center "Insieme nella Cura" at University Hospital of Rome Campus Bio-Medico. She was a Research Fellow in nursing at the University Hospital Campus Bio-Medico. She developed a particular interest in applied research on older people with chronic diseases and continuity of patient care.

Andrea Ianni works in the Research Unit in Hygiene, statistics and public health (Dir. Prof. T. Petitti) at Campus Bio-Medico University, in Rome, Italy. His main research fields are: risk prevention for healthcare workers; vaccination; health education; inter-professional integration; food security. He is a member of the Quality Assurance Group of the Nursing Degree Course (Dir. Prof. M.G. De Marinis) at Campus Bio-Medico University (UCBM, Rome, Italy) and is stably involved in teaching activities in several Degree Courses at UCBM, both in the Faculty of Medicine and Surgery and in Science and Technology for Mankind and the Environment.

Francesca Cordella received the Laurea degree in Electronic Engineering and the Ph.D. in Computer and Automation Engineering both from the University of Naples Federico II. She is Assistant Professor in the Research Unit of Advanced Robotics and Human-Centred technologies at Università Campus Bio-Medico di Roma. In 2021 she received the Italian National Scientific Habilitation for Associate Professorship in Industrial Bioengineering. Her research interests are mainly in the fields of biomechanics, human-machine interfaces, assistive and collaborative robotics. She has been involved in several EU-funded and national projects in her fields of interest. She has authored/coauthored more than 50 peer-reviewed publications.

Loredana Zollo, MS 2000, PhD 2004, is Full Professor of Biomedical Engineering at Università Campus Bio-Medico di Roma. She is head of the Research Unit of Advanced Robotics and Human-Centred technologies and Director of the Master of Science in Biomedical Engineering in the same university. Her research interests are mainly in the fields of rehabilitation and assistive robotics, bio-robotics and bionics, human-machine interfaces, collaborative robotics. She has been involved, as partner and responsible, in many EU-funded and national projects in her application fields. She has authored/coauthored more than 140 peer-reviewed publications.

Maria Grazia De Marinis is Full Professor of Nursing, Director of Research Unit Nursing Science and Director of the Palliative Care Center at Campus Bio Medico University Hospital of Rome. She is President of the Italian Society of Nursing Sciences and was Vice President of the Italian Society of Medical Pedagogy; she has participated in numerous commissions and working groups for the development of guidelines for general and specialist nursing education and referent for numerous research projects, including the SAFEMOVER, focused on developing a robotic device for improving working conditions and user subjective perspective during patient-handling movements.

Journal of Safety Research 82 (2022) 233–240

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Mental health conditions and unsafe driving behaviors: A naturalistic driving study on ADHD and depression

Ou Stella Liang, Christopher C. Yang *

College of Computing and Informatics, Drexel University, Philadelphia, PA 19104, United States

ARTICLE INFO

Article history:
Received 16 September 2021
Received in revised form 2 January 2022
Accepted 26 May 2022
Available online 6 June 2022

Keywords:
ADHD
Depression
Mental health
Unsafe driving behaviors
Road safety

ABSTRACT

Introduction: Road injuries remain a persistent public health concern across the world. The task of driving is complicated by mental health conditions, which may affect drivers' executive functioning and cognitive resource allocation. This study examines whether attention-deficit/hyperactivity disorder (ADHD) and depression are associated with unsafe driving behaviors. *Method:* Generalized linear mixed models were employed to estimate the association of self-reported ADHD and depression with 18 unsafe driving behavior types found prior to at-fault crashes and near-crashes using a large-scale naturalistic driving dataset. Driver demographics, cognitive traits, environmental factors, and driver random effects were included to reduce confounding and biases. *Results:* Controlling for other covariates, people with self-reported ADHD were more likely to have performed improper braking or stopping (OR = 4.89, 95% CI 1.82–13.17) prior to an at-fault crash or near-crash, while those with self-reported depression did not have a significant association with any unsafe driving behavior. *Conclusions:* After accounting for demographic, cognitive, and environmental covariates, individuals with ADHD and depression were not prone to purposefully aggressive or reckless driving. Instead, drivers with self-reported ADHD may unintentionally execute unsafe driving behaviors in particular driving scenarios that require a high level of cognitive judgment. *Practical Applications:* These findings can inform the curriculum design of driver's education programs that help learners with mental health conditions gain practice in certain road scenarios, for example, more practice on preemptively reducing speed instead of making sudden brakes and smooth turning on curved roads for students with ADHD. Furthermore, specific advanced driver assistance systems may prove particularly helpful for drivers with ADHD, such as detection of leading objects and curve speed warning.

1. Introduction

Road injuries remain a global health burden with geographical heterogeneity – many regions continue to experience increasing incidence rates, while other parts of the world see a decrease in age-standardized mortality rate (James, 2020). According to the latest report published by the World Health Organization (WHO), annual road traffic mortalities reached 1.35 million across the world, and between 20 and 50 million more people suffer nonfatal injuries and many incur lifelong disabilities (World Health Organization, 2018). In the United States, 36,096 lives were lost and 2.74 million injured due to motor-vehicle traffic crashes during 2019 (National Highway Traffic Safety Administration, 2020).

Improving the assessment of psychophysical abilities to drive was highlighted as one of the 17 emerging issues in road safety at the Safety 2016 World Conference (Seguí-Gomez, 2016). From work commute to grocery shopping, from ride-sharing to last-mile trucking services, driving provides not only a mode of transportation but also a means of living for a lot of people. Depressive disorders and attention-deficit/hyperactivity disorder (ADHD) are common mental health conditions (MHCs) (Clarke, Schiller, & Boersma, 2019; Degenhardt, 2019; Faraone, 2021), which means that those with the symptoms have to carry on with all facets of everyday life, including driving, while receiving treatment or remaining undiagnosed. This study aims to examine how ADHD and depression may impact driving behaviors.

1.1. ADHD and road safety

A meta-analysis suggested that ADHD was likely associated with higher than normal rates of negative driving outcomes (re-

* Corresponding author at: 3675 Market St, Office 1186, Philadelphia, PA 19104, United States.

E-mail address: chris.yang@drexel.edu (C.C. Yang).

https://doi.org/10.1016/j.jsr.2022.05.014
0022-4375/

lated risk = 1.88, 95% CI [1,42, 2.50]) because of possible deficiencies in executive functioning (Jerome, Segal, & Habinski, 2006). Among adolescents and young adults, the adjusted risk of the first crash was 1.36 times higher among those with ADHD than without and did not vary by sex, licensing age, or over time (Curry, 2017). A meta-analysis of six studies suggested the positive effects of stimulant medications on driving performance (Barkley & Cox, 2007). Another study based on naturalistic driving data found that ADHD symptoms increased the crash risk by 5% for every increase in the ADHD severity score, and medication may not attenuate the risk (Aduen, 2018).

Previous studies on ADHD and driving behaviors primarily relied on self-report scales, such as the Driving Behavior Rating Scale and the Driving Behavior Questionnaire, which found more risky driving behaviors exist among ADHD groups compared to non-ADHD groups. Richards et al., using a series of psychological surveys found that both college students and adults with high ADHD symptoms reported a higher likelihood of risky and aggressive driving to express driving anger (Richards et al., 2002, 2006). Malta et al. reported that young adult drivers with self-reported high driving aggression have a significantly higher prevalence of ADHD (Malta, Blanchard, & Freidenberg, 2005). A few studies employed driving simulators to assess driving performance and found poorer steering control that may reflect poor motor control and coordination, however, the finding could not be reproduced using a large-scale study (Jerome et al., 2006).

1.2. Depression and road safety

In terms of driving outcomes, a meta-analysis of studies published between 1995 and 2015 estimated that depression increased crash risk by 2-fold (Hill, 2017). Using large-scale naturalistic driving data, researchers found that depression was not associated with increased risk of motor-vehicle crashes or collision fault, but increased risk of self-reported injuries following a collision (Aduen, Kofler, Cox, Sarver, & Lunsford, 2015). In terms of driving behaviors, multiple epidemiological studies reported mixed results on the association between depression and the risk of motor-vehicle collisions and injuries, but a more consistent association between depression and self-reported aggressive or risky driving behaviors, suggesting drivers with depressive symptoms exhibited difficulties in allocating cognitive resources to competing driving tasks (Wickens, Smart, & Mann, 2014). Using the Young Adult Driving Questionnaire, McDonald et al., reported that depressive symptoms were linked to reckless driving among adolescents and adults (McDonald, Sommers, & Fargo, 2014). One longitudinal study following a large cohort of Australian teen drivers found no significant association between depression and self-reported engagement in risky driving (Vassallo, 2008). Anxiety is frequently observed as a comorbidity with depression, and vice versa (Almeida, 2012). A study on a French cohort of drivers entering old age found that having depression, anxiety, or stress was associated with an increased crash risk despite increased driving avoidance (Turrado, 2021).

1.3. Limitations in current literature

Motor-vehicle crashes are not isolated events, rather, they happen in highly random and complex environments with a multitude of contributing factors. Contradicting findings on the effects of MHCs on driving can be attributable to methodological limitations including lack of control for driving exposure, lack of control for confounders (such as age, driving experience, and comorbidities), limited response rate or sample size, nonstandard definitions of collisions (e.g., less severe collisions were unreported), and self-report bias (Jerome et al., 2006; Kaye, Lewis, & Freeman, 2018; Wickens et al., 2014). Unsafe driving behaviors and risk factors do not generalize across different driving populations, therefore a multi-factor framework for specific risky driving behaviors is recommended (Fernandes, Job, & Hatfield, 2007).

In summary, past literature indicates that MHCs play an important role in driving performances, but there is a need to further verify these findings with naturalistic driving scenarios, in which confounding factors can be controlled. Few studies examined the association between MHCs and unsafe driving behaviors using naturalistic driving data to avoid self-report biases.

2. Materials and methods

2.1. Data source

Data from the Second Strategic Highway Research Program (SHRP 2) Naturalistic Driving Study (NDS) were used for this study (Dingus et al., 2014; Hankey, Perez, & McClafferty, 2016). The SHRP 2 is one of the largest NDS in the world, which recorded more than six million real-world driving trips by approximately 3,500 volunteer drivers across six U.S. states: Florida, Indiana, North Carolina, New York, Pennsylvania, and Washington. Participant vehicles were outfitted with a front-view camera, a rear-view camera, an in-cabin camera, a front-facing radar, and a data collection unit (DAS) that records data from the aforementioned devices as well as the vehicle's Controller Area Network (CAN bus). As a result, the SHRP 2 dataset provides a large collection of crashes, near-crashes, as well as non-eventful baselines, of which crucial data points characterizing the trips were recorded and analyzed. Human-expert annotated variables based on video recordings of the trips, as well as demographic, cognitive, behavioral, and medical information of the volunteer drivers, are known for each trip.

2.2. Human subjects protection

The SHRP 2 NDS was approved by the Institutional Review Boards of the Virginia Tech and National Academy of Sciences (NAS). Issuance of a Certificate of Confidentiality was initiated through the National Institutes of Health (NIH). Participants consented to have data collected from their main vehicle when it was driven during the study period, as well as to provide a broad set of functional assessments. Participants who were minors at the time of study provided assent to participate, and consent for participation was provided by a parent. Participants were compensated for study participation (Dingus et al., 2014).

This study is a secondary data analysis of the SHRP 2 NDS dataset. The data use of this study (Protocol Number 1902007017) was exempt by our Drexel University Institutional Review Board (IRB) because no identifiable private information of participants was obtained or analyzed.

2.3. Study population

This study included a total of 1,561 study participants whose 4,228 trips involved an at-fault crash or a near-crash. The driver selection process is presented in Fig. 1. We will elaborate on each step below.

The main criterion for identifying a crash is that the study participant vehicle came into physical contact with another object. Three crash severity levels ("severe crash," "police-reportable crash," and "minor crash") were included, but crashes labeled "low-risk tire strike" were excluded because they presented low risk to drivers. Near-crashes were circumstances that required an emergency evasive maneuver by an involved party but did not result in a crash. Further, only crashes and near-crashes that were

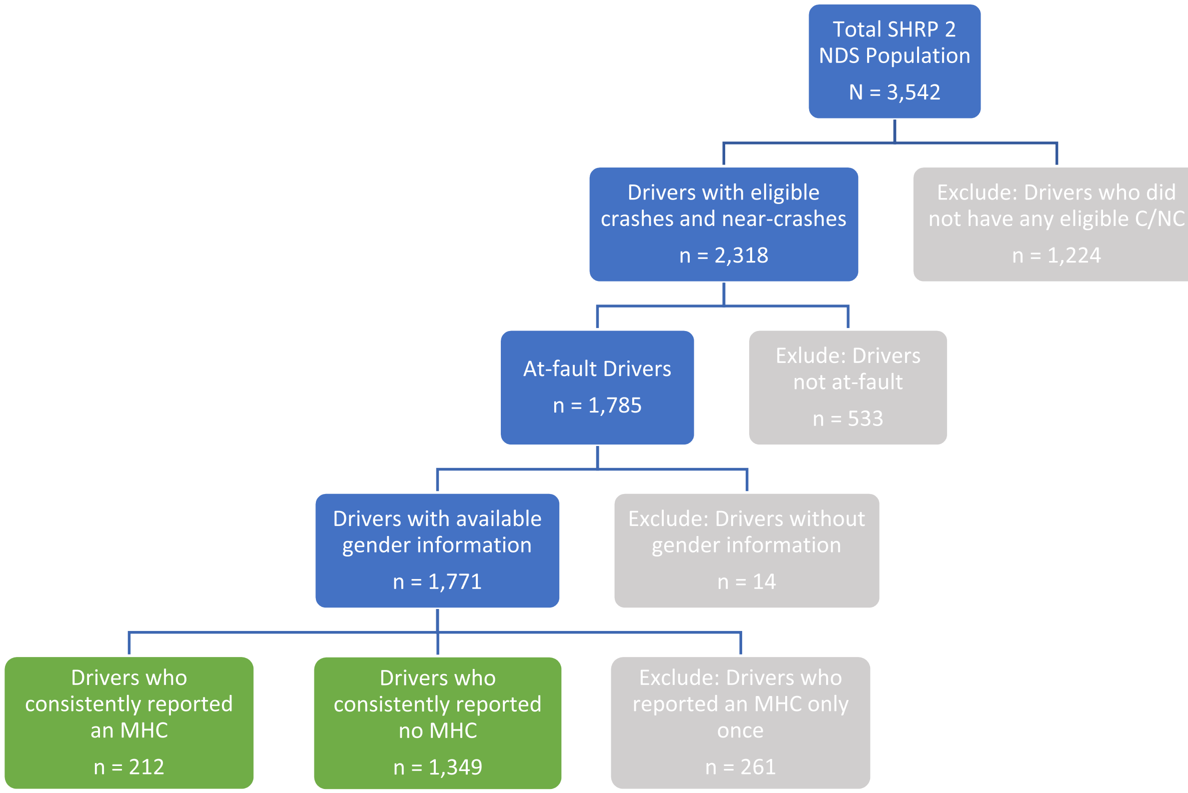


Fig. 1. Driver selection flowchart.

deemed at-fault by human experts were included because we aim to examine the unsafe driving behaviors that contributed to safety critical events. The distribution of trips by outcomes is in Table 1. Because each study participant drove multiple trips during the study and each trip had distinct environmental elements, the unit of analysis must be on the trip level, instead of the driver level.

To determine whether the study participant reported an MHC, two questionnaires were used. They are the same "medical conditions and medications" questionnaire that was dispensed at the beginning and end of the study, with questions focused on identification of conditions that could affect driving performance and safety. Study participants were asked to check all that apply among six possible MHCs, including: "ADD/ADHD/Tourette's Syndrome," "anxiety or panic attacks," "depression," "personality disorders," "psychotic disorders," and "bipolar disorder." We processed them as binary variables to allow for possible comorbidities. Because the SHRP 2 NDS lasted for an extended period, only those that consistently checked a condition at both the intake and exit surveys were included in this study as having such a condition; and only those that did not check any mental condition throughout the study were included as the control group.

We consider study participants who selected "ADD/ADHD/Tourette's Syndrome" to represent the self-reported ADHD group because Tourette's Syndrome is rare in adulthood (0.002–0.04%) (Aduen et al., 2015) and there is a significant difference ($p < 0.001$) in the mean Barkley's ADHD score between study participants who selected the answer choice "ADD/ADHD/Tourette's Syndrome" and those who did not. The self-reported ADHD and depression were treated as the main independent variables, while the other MHCs as covariates due to their limited number of responses (personality disorder, psychotic disorders, bipolar disorder) or lack of information (anxiety). Anxiety disorders alone usually are not studied with driving safety, because anxiety disorders encompass a heterogenous set of underlying concerns that may affect driving in different ways ranging from driving with exaggerated caution to hostile driving (Zinzow & Jeffirs, 2018). In this study, we do not have enough information to determine the type of anxiety or whether it is driving-related. This treatment is also seen in previous literature using the SHRP 2 dataset (Aduen et al., 2015; Aduen, 2018). The study population distribution by self-reported MHC is in Table 2. Note the numbers of study participants in each MHC category will not add to the total number of participants because some participants reported multiple MHCs. The demographics of the study population are summarized in Table 3.

Table 1
Trip outcomes ($N = 4{,}228$).

Trip Outcomes		Trips (n)
Crashes		
	Most Severe	57
	Police-reportable	78
	Minor Crash	507
Near-Crashes		3,586

Table 2
Study population ($N_{trips} = 4{,}228$, $N_{participants} = 1{,}561$).

Mental Health Condition	Trips (n)	Study Participants (n)
No MHCs	3,176	1,349
ADHD	278	66
Depression	400	113
Anxiety	342	97
Personality Disorder	0	0
Psychotic Disorders	1	1
Bipolar Disorder	31	8

Table 3
Study population demographics ($N_{participants}$ = 1,561).

		Participants n (%)
Gender		
	Females	761 (48.8)
Age Group		
	Adolescents and young adults (Age 16–24)	705 (45.2)
	Adults (Age 25–64)	521 (33.4)
	Seniors (Age 65+)	335 (21.4)
Education		
	High school	290 (18.6)
	College	869 (55.7)
	Advanced degree	402 (25.7)
Marital status		
	Not Married	1 038 (66.5)
Income level		
	Under $50,000	682 (43.7)
	$50,000 to $99,999	536 (34.3)
	Above $100,000	343 (22.0)
Annual miles		
	Under 10,000 miles	555 (35.6)
	10,000–20,000 miles	752 (48.2)
	20,000–30,000 miles	170 (10.9)
	Above 30,000 miles	84 (5.3)

2.4. Pre-crash unsafe driving behavior

We examined 60 unsafe driver behaviors qualitatively annotated from the NDS video reduction process (Virginia Tech Transportation Institute, 2015). Unsafe driver behaviors are what drivers did to cause or contribute to a crash or near-crash, for example, "exceeding speed limit" or "illegal passing." In other words, these behaviors preceded only trips that resulted in a crash or near-crash. Note this is to differentiate from risky driving behaviors as a generic term: only those of negative consequences were annotated in the SHRP 2 NDS. To emphasize on this difference, we will reference this variable as "*pre-crash* unsafe driving behaviors" in this study. These behaviors were semantically grouped into 18 types: *aggressive driving, distracted driving, driving slowly, drowsiness, following too closely, improper backing, improper braking or stopping, improper turning, inexperienced driving, lane changing error, neighbor lane conflict, no unsafe behavior, other unsafe behavior, right of way error, signal or sign violation, speeding, unfamiliar with environment or vehicle, and vehicle signaling error*. The mapping of the pre-crash unsafe driving behavior categories can be found in Appendix I.

2.5. Statistical analyses

We chose the generalized linear mixed-effect model (GLMM) to study the association between MHCs and pre-crash unsafe driving behaviors. The GLMM design allows one to control for multiple covariates, as well as the random effect of latent characteristics of individual drivers. The model is as follows:

$$g(E(y)) = X\beta + Zu + \epsilon$$

$$E(y) = P(Y = y|X, Z)$$

$$g(\cdot) = log\left(\frac{p}{1-p}\right)$$

where y is the outcome variable; $g(\cdot)$ is the logistic link function for a binomial outcome; p is the estimated probability of a positive outcome; X is a matrix of N trips and q variables; β is a $q \times 1$ vector of the fixed-effect regression coefficients; Z is a matrix of N trips and d drivers designating the driver-specific random effects; u is a $d \times 1$ vector of the random intercepts; and ϵ is the general error term not explained by the model.

To be specific, the null model has whether a driving behavior was present during the trip sample as a binary outcome, self-reported ADHD and depression as the independent variables, and the anonymous driver ID as the random effects variable. In addition to the two MHCs and driver random effect, the full model included the other MHCs, the cognitive traits, demographic characteristics, and environmental factors as control variables. These variables can be found in Table 4. Among the variables, the participant ID, gender, and MHCs must have values. For the other variables, missing values were imputed by the sample median for continuous variables to reduce the influence of outliers and sample mode for categorical variables. The numeric variables were then standardized because the variable "total trip duration" was recorded in milliseconds, thereby had a very different order of magnitude (median = 1,440,500 milliseconds, approximately 24 min) compared to the other numeric variables whose medians were between 2 and 17. The total trip duration was included to account for the time-variant risk exposure, since fatigued drivers may be more prone to some of the unsafe driving behaviors.

Take speeding, an unsafe driver behavior, as an example. The full model proved to be superior to the null model by analysis of variance (ANOVA) ($\mathcal{X}^2_{28} = 478.72$, $p < 0.001$). The Variance Inflation Factor (VIF) analysis suggested that there was high collinearity between age group and years of driving experience. Removing the age groups but preserving the years of experience further reduced the Akaike information criterion (AIC) while maintaining significantly reduced variance compared to the null model.

Finally, the p-value was adjusted following the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) to ensure the overall false discovery rate was less than 0.05, because multiple driver behaviors were tested for each MHC. The statistical analysis was performed in R version 4.0 (Fay, 2010; R Core Team, 2020).

3. Results

In the null model, self-reported ADHD saw a significant increase in the unadjusted odds ratio associated with aggressive driving (OR = 2.83, 95% CI 1.27–6.33), improper braking or stopping (OR = 5.44, 95% CI 1.92–15.40), improper turning (OR = 1.89, 95% CI 1.17–3.06), inexperienced driving (OR = 6.07, 95% CI 2.10–17.51), and speeding (OR = 2.22, 95% CI 1.42–3.48). Self-reported depression did not have a significant association with any pre-crash unsafe driving behavior.

Including cognitive traits eliminated the positive association of aggressive driving and speeding with ADHD. The association with

Table 4
Input variable groups.

	Groups	Variable Names
(a)	MHCs	ADHD, Depression, Anxiety, Psychotic Disorders, Personality Disorders, Bipolar Disorder
(b)	Cognitive Traits	Sensation Seeking Scale, Driving Knowledge Score, Risk Perception Score, Clock Drawing Assessment (for identifying signs of dementia or other neurological disorders)
(c)	Demographic Characteristics	Gender, Income Level, Education Level, Marital Status, Annual Miles, Years of Driving Experience, Age Group (removed due to high collinearity with Years of Driving Experience)
(d)	Environmental Factors (trip-specific)	Weather, Lighting, Road Surface, Road Grade, Road Alignment, Traffic Density, Total Trip Duration
(e)	Random Effect	Participant ID

inexperienced driving also disappeared after adding demographic covariates to the model. Finally, controlling for environmental factors moved the risk associated with improper turning to the boundary of significance after adjusting the *p*-value using the Benjamini-Hochberg procedure (single test $p = 0.006$, BH adjusted $p = 0.057$).

Consequently, controlling for the impact of all covariates (cognitive, demographic, and environmental), drivers with self-reported ADHD were more likely to perform improper braking or stopping (OR = 4.89, 95% CI 1.82–13.17), while those who reported depression did not have an elevated association with any pre-crash unsafe driving behavior.

The estimated associations between the MHCs and unsafe driving behaviors can be found in Fig. 2 and Tables 5 and 6.

4. Discussions

In this study, we examined the association of self-reported ADHD and depression with pre-crash unsafe driving behaviors. ADHD was significantly associated with improper braking or stopping; depression was not significantly associated with any unsafe driving behavior. Improper braking or stopping refers to sudden braking in an unsafe manner in the roadway, which may indicate that the driver was startled by a road scenario that elicited the sudden maneuver without regard to the potential impact on following vehicles. Improper turning, which approached significant association, refers to making wide turns to the extent of encroaching on neighbor lanes or road shoulders/curbs, which may suggest difficulty in maintaining stable vehicle dynamics in turning scenarios due to a lack of skills. These findings are consistent with previous research that ADHD may impact driver's executive functioning and negatively impacted their ability to master driving as a skill (Curry, 2017; Jerome et al., 2006). For drivers who report depression, this study found no significantly increased risk of any unsafe driving behaviors, which is consistent with one longitudinal study (Vassallo, 2008). Other research suggests difficulties in allocating cognitive resources to necessary or competing driving tasks both inside and outside of the vehicle among drivers with depression

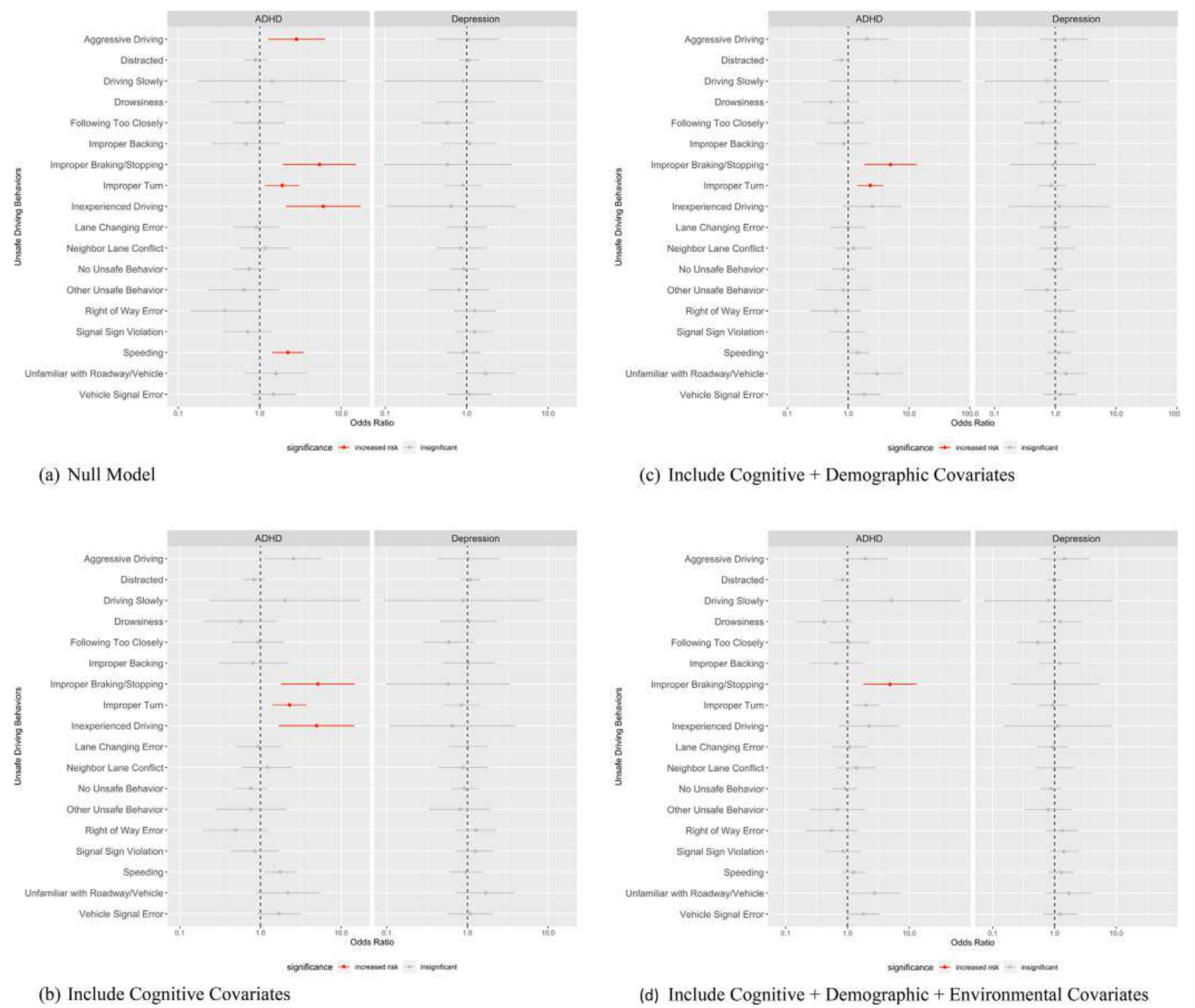


Fig. 2. Mental health conditions and pre-crash unsafe driving behaviors. (a) Null model: include only the MHCs and driver random effects; (b) Cognitive model: add cognitive traits to the null model; (c) Cognitive + Demographic model: add cognitive traits and demographic characteristics to the null model; (d) Full model: add cognitive traits, demographic characteristics, and environmental factors to the null model.

Table 5
Association of self-reported ADHD and pre-crash unsafe driving behaviors.

Behavior	ADHD					
	Null Model			Full Model		
	p-value	Adj. *p*-value	OR (95%CI)	*p*-value	Adj. *p*-value	OR (95%CI)
Aggressive Driving	0.011	0.040*	2.83 (1.27–6.33)	0.121	0.363	1.94 (0.84–4.49)
Distracted	0.463	0.641	0.89 (0.64–1.22)	0.278	0.500	0.84 (0.61–1.15)
Driving Slowly	0.747	0.827	1.42 (0.17–11.75)	0.220	0.450	5.13 (0.38–70.07)
Drowsiness	0.507	0.652	0.70 (0.24–2.01)	0.117	0.363	0.42 (0.14–1.24)
Following Too Closely	0.970	0.970	0.99 (0.47–2.05)	0.848	0.848	1.08 (0.51–2.27)
Improper Backing	0.444	0.641	0.68 (0.26–1.82)	0.408	0.566	0.66 (0.25–1.77)
Improper Braking/Stopping	0.001	0.009**	5.44 (1.92–15.40)	0.002	0.030*	4.89 (1.82–13.17)
Improper Turn	0.009	0.040*	1.89 (1.17–3.06)	0.006	0.057	2.01 (1.22–3.32)
Inexperienced Driving	0.001	0.008**	6.07 (2.10–17.51)	0.163	0.419	2.23 (0.72–6.87)
Lane Changing Error	0.781	0.827	0.91 (0.48–1.74)	0.826	0.848	1.08 (0.55–2.12)
Neighbor Lane Conflict	0.662	0.795	1.17 (0.57–2.39)	0.363	0.544	1.41 (0.67–2.96)
None	0.197	0.478	0.74 (0.47–1.17)	0.685	0.770	0.91 (0.57–1.44)
Other	0.385	0.630	0.64 (0.23–1.76)	0.479	0.616	0.69 (0.24–1.95)
Right of Way Error	0.041	0.124	0.37 (0.14–0.96)	0.225	0.450	0.55 (0.21–1.44)
Signal Sign Violation	0.324	0.583	0.71 (0.36–1.40)	0.658	0.770	0.86 (0.43–1.70)
Speeding	0.000	0.008**	2.22 (1.42–3.48)	0.318	0.520	1.25 (0.81–1.93)
Unfamiliar with Roadway/Vehicle	0.310	0.583	1.60 (0.65–3.93)	0.041	0.247	2.76 (1.04–7.29)
Vehicle Signal Error	0.212	0.478	1.48 (0.80–2.72)	0.063	0.283	1.82 (0.97–3.40)

Significance codes: <0.001 '***' [0.001–0.01) '**' [0.01–0.05) '*'.

Table 6
Association of self-reported depression and pre-crash unsafe driving behaviors.

Behavior	Depression					
	Null Model			Full Model		
	p-value	Adj. *p*-value	OR (95%CI)	*p*-value	Adj. *p*-value	OR (95%CI)
Aggressive Driving	0.939	0.972	1.04 (0.42–2.52)	0.417	0.991	1.46 (0.58–3.65)
Distracted	0.631	0.972	1.07 (0.80–1.43)	0.924	0.991	0.99 (0.74–1.31)
Driving Slowly	0.930	0.972	0.90 (0.10–8.58)	0.852	0.991	0.80 (0.07–8.71)
Drowsiness	0.972	0.972	0.99 (0.43–2.26)	0.604	0.991	1.24 (0.54–2.85)
Following Too Closely	0.149	0.972	0.58 (0.28–1.21)	0.097	0.991	0.53 (0.25–1.12)
Improper Backing	0.839	0.972	1.08 (0.50–2.34)	0.627	0.991	1.22 (0.55–2.71)
Improper Braking/Stopping	0.558	0.972	0.58 (0.10–3.56)	0.991	0.991	1.01 (0.19–5.26)
Improper Turn	0.695	0.972	0.90 (0.53–1.52)	0.810	0.991	0.93 (0.53–1.63)
Inexperienced Driving	0.635	0.972	0.64 (0.10–4.06)	0.915	0.991	1.11 (0.15–8.31)
Lane Changing Error	0.967	0.972	0.99 (0.56–1.75)	0.830	0.991	0.94 (0.52–1.68)
Neighbor Lane Conflict	0.659	0.972	0.85 (0.42–1.72)	0.985	0.991	1.01 (0.49–2.06)
None	0.722	0.972	0.93 (0.63–1.37)	0.537	0.991	0.88 (0.60–1.30)
Other	0.622	0.972	0.80 (0.34–1.91)	0.606	0.991	0.79 (0.32–1.93)
Right of Way Error	0.451	0.972	1.25 (0.70–2.26)	0.347	0.991	1.33 (0.73–2.42)
Signal Sign Violation	0.410	0.972	1.25 (0.73–2.14)	0.219	0.991	1.42 (0.81–2.49)
Speeding	0.723	0.972	0.92 (0.57–1.47)	0.306	0.991	1.28 (0.80–2.06)
Unfamiliar with Roadway/Vehicle	0.201	0.972	1.72 (0.75–3.97)	0.228	0.991	1.72 (0.71–4.13)
Vehicle Signal Error	0.818	0.972	1.08 (0.57–2.02)	0.499	0.991	1.25 (0.66–2.36)

Significance codes: <0.001 '***' [0.001–0.01) '**' [0.01–0.05) '*'.

(Wickens et al., 2014), but this study did not find such evidence (i.e., lack of association with signal or sign violation or with exhibiting unfamiliarity with road or vehicle).

Previous studies linked reckless or aggressive driving with ADHD and depression (Barkley & Cox, 2007; Jerome et al., 2006; Malta et al., 2005; Richards, Deffenbacher, Rosén, Barkley, & Rodricks, 2006). Our results did not show either condition had an increased association with the annotated behavior "aggressive driving" or "speeding" after adjusting for demographic covariates, even though the unadjusted odds ratios of aggressive driving and speeding were significantly higher among drivers with ADHD. We suspect several factors may be at play. First, the definition of aggressive driving is not standardized. Previous studies may have included several unsafe driving behaviors delineated in this study under the umbrella of aggressive driving such as improper braking and speeding. It is also possible that drivers with MHCs have more aggressive driving behaviors than those without do, when trip outcomes were not in consideration, but this study considered only behaviors that preceded an at-fault crash or near-crash. Second, previous studies measure aggressive driving by self-reported instrument results that are prone to recall bias, but the present study used video recordings of actual crash events to annotate aggressive driving behaviors. Third, as the study results demonstrated, there were significant unadjusted associations with aggressive driving and speeding among drivers with ADHD, which disappeared after accounting for demographic and environmental factors. This observation lends credence to the importance of controlling for driver and road confounders when studying driving behaviors (Barkley & Cox, 2007; Jerome et al., 2006).

This study is limited by the lack of clinical information to validate study participants' self-reported MHCs. Differences in medication use and the severity of conditions among drivers should ideally be delineated. While the self-reported ADHD condition was corroborated by the Barkley score, it is hard to gauge the validity of self-reported depression. However, the advantage of analyzing self-reported MHCs is to capture individuals with undiagnosed conditions due to lack of access to mental health services. Another limitation is that the pre-crash unsafe driving behaviors in this study were restricted to those that resulted in a crash or near-crash. It is possible that the distribution of risky driving behaviors regardless of trip outcome is different, therefore the generalizability of the study findings should be applied within the scope of analyzing the contributing factors of crashes and near-crashes.

5. Conclusions

Drivers with MHCs may face extra challenges related to driving. This study estimates the association between ADHD and depression and pre-crash unsafe driving behaviors. We showed that, after accounting for demographic, cognitive and environmental covariates, individuals with ADHD and depression were not prone to aggressive driving, contradictory to previous studies. Instead, they may unintentionally execute unsafe driving behaviors in particular driving scenarios that require a high level of cognitive judgment. These findings can provide clinicians and educators with empirical evidence of how ADHD and depression may affect patient's ability to perform safe driving.

6. Practical Applications

The study findings can inform the curriculum design of driver's education programs that help learners with MHCs gain practice in certain road scenarios (e.g., more practice on preemptively reducing speed instead of making sudden brakes as well as smooth turning on curved roads for students with ADHD). Furthermore, specific advanced driver assistance systems (ADAS) may prove particularly helpful for drivers with ADHD, such as detection of leading objects and curve speed warning.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the National Science Foundation under Grant NSF-1741306, Grant IIS-1650531, and Grant DIBBS-1443019. Any opinions and conclusions or recommendations expressed in this study are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Appendix I Mapping of unsafe driving behaviors.

Appendix I Mapping of unsafe driving behaviors. (*continued*)

Category	Unsafe Driver Behavior
Aggressive Driving	Aggressive driving, specific, directed menacing actions
	Aggressive driving, other
Distracted	Distracted
Driving Slowly	Driving slowly: below speed limit
	Driving slowly in relation to other traffic: not below speed limit
Drowsiness	Drowsy, sleepy, asleep, fatigued
Following too closely	Following too closely
Improper Backing	Improper backing, other
	Improper backing, did not see
Improper Braking/ Stopping	Sudden or improper braking
	Sudden or improper stopping on roadway
	Use of cruise control contributed to late braking
Improper Turn	Improper turn, wide left turn
	Improper turn, wide right turn
	Improper turn, other
	Improper turn, cut corner on left
	Improper turn, cut corner on right
	Making turn from wrong lane
Inexperienced Driving	Apparent general inexperience driving
	Apparent unfamiliarity with vehicle
Lane Changing Error	Cutting in, too close behind other vehicle
	Cutting in, too close in front of other vehicle
	Did not see other vehicle during lane change or merge
Neighbor Lane Conflict	Other improper or unsafe passing
	Passing on right
	Illegal passing
	Driving in other vehicle's blind zone
None	None
	No Additional Driver Behaviors
Other	Other
	Avoiding pedestrian
	Driving without lights or with insufficient lights
	Wrong side of road, not overtaking
	Avoiding other vehicle
	Avoiding animal
	Parking in improper or dangerous location
	Improper start from parked position
	Non-signed crossing violation
	Unknown
	Failure to dim headlights
Right of Way Error	Right-of-way error in relation to other vehicle or person, apparent recognition failure
	Right-of-way error in relation to other vehicle or person, apparent decision failure
	Right-of-way error in relation to other vehicle or person, other or unknown cause
Signal/Sign Violation	Signal violation, apparently did not see signal
	Stop sign violation, apparently did not see stop sign
	Stop sign violation, "rolling stop"
	Signal violation, tried to beat signal change
	Other sign (e.g., Yield) violation, apparently did not see sign
	Stop sign violation, intentionally ran stop sign at speed

(*continued on next page*)

Appendix I Mapping of unsafe driving behaviors. (*continued*)

Category	Unsafe Driver Behavior
	Signal violation, intentionally disregarded signal
	Other sign violation
	Other sign (e.g., Yield) violation, intentionally disregarded
	Disregarded officer or watchman
Speeding	Exceeded speed limit
	Exceeded safe speed but not speed limit
	Speeding or other unsafe actions in work zone
Unfamiliarity with Environment/ Vehicle	Apparent unfamiliarity with roadway
Vehicle Signal Error	Failed to signal
	Improper signal

References

Aduen, P. A. et al. (2018). ADHD, depression, and motor vehicle crashes: A prospective cohort study of continuously-monitored, real-world driving. *Journal of Psychiatric Research, 101*, 42–49.

Aduen, P. A., Kofler, M. J., Cox, D. J., Sarver, D. E., & Lunsford, E. (2015). Motor vehicle driving in high incidence psychiatric disability: Comparison of drivers with ADHD, depression, and no known psychopathology. *Journal of Psychiatric Research, 64*, 59–66.

Almeida, O. P. et al. (2012). Anxiety, depression, and comorbid anxiety and depression: Risk factors and outcome over two years. *International Psychogeriatrics, 24*, 1622–1632.

Barkley, R. A., & Cox, D. (2007). A review of driving risks and impairments associated with attention-deficit/hyperactivity disorder and the effects of stimulant medication on driving performance. *Journal of Safety Research, 38*, 113–128.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Statistical Methodology), 57*, 289–300.

Clarke, T. C., Schiller, J., & Boersma, P. *Early Release of Selected Estimates Based on Data from the 2019 National Health Interview Survey*. 2 https://www.cdc.gov/nchs/data/nhis/earlyrelease/EarlyRelease202009-508.pdf (09/20).

Curry, A. E. et al. (2017). Motor vehicle crash risk among adolescents and young adults with attention-deficit/hyperactivity disorder. *JAMA Pediatrics, 171*, 756–763.

Degenhardt, L. et al. (2019). Global patterns of opioid use and dependence: Population harms, interventions, and future action. *The Lancet Lond. Engl., 394*, 1560–1579.

Dingus, T. A. *et al. Naturalistic driving study: Technical coordination and quality control*. 22362 (Transportation Research Board, 2014). doi:10.17226/22362.

Faraone, S. V. et al. (2021). The World Federation of ADHD International Consensus Statement: 208 evidence-based conclusions about the disorder. *Neuroscience & Biobehavioral Reviews, 128*, 789–818.

Fay, M. P. (2010). Confidence intervals that match Fisher's exact or Blaker's exact tests. *Biostatistics (Oxf. England), 11*, 373–374.

Fernandes, R., Job, R. F. S., & Hatfield, J. (2007). A challenge to the assumed generalizability of prediction and countermeasure for risky driving: Different factors predict different risky driving behaviors. *Journal of Safety Research, 38*, 59–70.

Hankey, J. M., Perez, M. A. & McClafferty, J. A. (2016). *Description of the SHRP 2 naturalistic database and the crash, near-crash, and baseline data sets task report*.

Hill, L. L. et al. (2017). Depression, antidepressants and driving safety. *Injury Epidemiology, 4*, 10.

James, S. L. et al. (2020). Morbidity and mortality from road injuries: Results from the Global Burden of Disease Study 2017. *Injury Prevention, 26*, i46–i56.

Jerome, L., Segal, A., & Habinski, L. (2006). What we know about ADHD and driving risk: A literature review, meta-analysis and critique. *Journal of the American Academy of Child & Adolescent Psychiatry, 15*, 105–125.

Kaye, S.-A., Lewis, I., & Freeman, J. (2018). Comparison of self-report and objective measures of driving behavior and road safety: A systematic review. *Journal of Safety Research, 65*, 141–151.

Malta, L. S., Blanchard, E. B., & Freidenberg, B. M. (2005). Psychiatric and behavioral problems in aggressive drivers. *Behaviour Research and Therapy, 43*, 1467–1484.

McDonald, C. C., Sommers, M. S., & Fargo, J. D. (2014). Risky driving, mental health, and health-compromising behaviours: Risk clustering in late adolescents and adults. *Injury Prevention, 20*, 365–372.

National Highway Traffic Safety Administration. *Overview of Motor Vehicle Crashes in 2019*. (2020).

R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2020).

Richards, T. L., Deffenbacher, J. L., & Rosén, L. A. (2002). Driving anger and other driving-related behaviors in high and low ADHD symptom college students. *Journal of Attention Disorders, 6*, 25–38.

Richards, T. L., Deffenbacher, J. L., Rosén, L. A., Barkley, R. A., & Rodricks, T. (2006). Driving anger and driving behavior in adults with ADHD. *Journal of Attention Disorders, 10*, 54–64.

Seguí-Gomez, M. (2016). 17 Emerging issues in road safety. *Injury Prevention, 22*, A7–A8.

Turrado, J. N. et al. (2021). Chronic medical conditions and their association with crash risk and changes in driving habits: A prospective study of the GAZEL cohort. *Injury Prevention, 27*, 17–23.

Vassallo, S. et al. (2008). Risky driving among young Australian drivers II: Co-occurrence with other problem behaviours. *Accident Analysis and Prevention, 40*, 376–386.

Virginia Tech Transportation Institute. (2015). *SHRP2 Researcher dictionary for video reduction data*. 101.

Wickens, C. M., Smart, R. G., & Mann, R. E. (2014). The impact of depression on driver performance. *International Journal of Mental Health and Addiction, 12*, 524–537.

World Health Organization (2018). *Global status report on road safety 2018*. World Health Organization.

Zinzow, H. M., & Jeffirs, S. M. (2018). Driving aggression and anxiety: intersections, assessment, and interventions. *Journal of Clinical Psychology, 74*, 43–82.

Christopher C. Yang is a professor in the College of Computing and Informatics at Drexel University. He also has a courtesy appointment at the School of Biomedical Engineering, Science, and Health Systems. He is the director of the Healthcare Informatics Research Lab. His recent research includes naturalistic driving data analytics, predictive modeling of disengagement driving for injury prevention, pharmacovigilance, drug repositioning, predictive modeling of sepsis, heterogeneous network mining, distributed graph computing, health intervention through social media for substance use disorders, and social media analytics. He is the Director of Data Science Programs and the Program Director of MS in Health Informatics. He has over 300 publications in top-tier journals, conferences, and books, such as ACM Transactions on Intelligent Systems and Technology, ACM Transaction on Management Information Systems, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Computational Social Systems, PLOS One, Journal of Medical Internet Research, Artificial Intelligence in Medicine, and more. He has received over $5M research fundings from NSF, NIH, PCORI, HK RGC, etc. He is the editor-in-chief of Journal of Healthcare Informatics Research and Electronic Commerce Research and Application. He is the editor of the CRC book series on Healthcare Informatics and the founding steering committee chair of the IEEE International Conference on Healthcare Informatics. He has been the general chair of over 5 conferences and program chairs of over 10 conferences.

Ou Stella Liang is a PhD candidate in Information Science in the College of Computing and Informatics at Drexel University. She applies machine learning, statistics and qualitative methods to investigate the human factors involved in road safety and drug safety research. She has 7 years of experience analyzing healthcare data as an internal consultant and manager at the Johns Hopkins Medicine. She received a Master of Health Administration from the Johns Hopkins Bloomberg School of Public Health, and a Bachelor of Science in Applied Pharmacy from the Peking University Health Science Center.

Journal of Safety Research 82 (2022) 283–292

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Nonfatal drowning-related hospitalizations and associated healthcare expenditure in India: An analysis of nationally representative survey data

Jeetendra Yadav [a], Denny John [b,c,d], Geetha R. Menon [a], Richard C. Franklin [e,f], Amy E. Peden [e,f,g,*]

[a] ICMR-National Institute of Medical Statistics, Ansari Nagar, New Delhi 110029, India
[b] Faculty of Life and Allied Health Sciences, Ramaiah University of Applied Sciences, Bangalore - 560054, Karnataka, India
[c] Department of Public Health, Amrita Institute of Medical Sciences & Research Centre, Amrita Vishwa Vidyapeetham, Kochi 682041, Kerala, India
[d] Center for Public Health Research, MANT, Kolkata-700078, West Bengal, India
[e] College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, Queensland, Australia
[f] Royal Life Saving Society – Australia, Sydney, New South Wales, Australia
[g] School of Population Health, Faculty of Medicine, UNSW Sydney, Kensington, New South Wales, Australia

ARTICLE INFO

Article history:
Received 23 September 2021
Received in revised form 11 January 2022
Accepted 6 June 2022
Available online 15 June 2022

Keywords:
Drowning
Economics
Emergency medical systems
Injury
Public health

ABSTRACT

Background: Drowning is a global public health challenge, with significant burden in low- and middle-income countries. There are few studies exploring nonfatal drowning, including the economic and social impacts. This study aimed to quantify unintentional drowning-related hospitalization in India and associated healthcare expenditure. *Method:* Unit level data on unintentional drowning-related hospitalization were obtained from the 75th rounds of the National Sample Survey of Indian households conducted in 2018. The outcome variables were indices of health care cost such as out of pocket expenditure (OOPE), health care burden (HCB), catastrophic health expenditure (CHE), impoverishment, and hardship financing. Descriptive statistics and multivariate analysis were conducted after adjusting for inflation using the pharmaceutical price index for December 2020. The association of socio-demographic characteristics with the outcome variable was reported as relative risk with 95% CI and expenditure reported in Indian Rupees (INR) and United States dollars (USD). *Results:* 174 respondents reported drowning-related hospitalization (a crude rate of 15.91–31.34 hospitalizations per 100,000 population). Proportionately, more males (63.4%), persons aged 21–50 years (44.9%) and rural dwelling respondents (69.9%) were hospitalized. Drowning-related hospitalization costs on average INR25,421 ($345.11USD) per person per drowning incident. Costs were higher among older respondents, females, urban respondents, and longer lengths of hospital stays. About 14.4% of respondents reported hardship financing as a result of treatment costs and 9.0% of households reported pushed below the poverty line when reporting drowning-related hospitalization. *Conclusions:* Drowning can be an economically catastrophic injury, especially for those already impacted by poverty. Drowning is a significant public health problem in India. Investment in drowning prevention program will reduce hospitalization and economic burden. *Practical Applications:* This study provides support for investment in drowning prevention in India, including a need to ensure drowning prevention interventions address the determinants of health across the lifespan.

1. Introduction

Drowning has been described by the World Health Organization (WHO) as a serious and neglected public health problem (World Health Organization, 2014), the prevention of which has been recognized by a recent United Nations resolution (United Nations, 2021). The Global Burden of Disease (GBD) study estimated that unintentional drowning claimed the lives of more than 295,000 people in 2017 (Franklin et al., 2020). More recently, the WHO estimated 236,000 fatalities in 2019 (World Health Organisation, 2020). These estimates comprise fatalities due to unintentional

Abbreviations: CHE, Catastrophic Health Expenditure; CI, Confidence Intervals; GBD, Global Burden of Disease; GOI, Government of India; HCB, Health Care Burden; HCE, Household Consumption Expenditure; HH, Households; HYCE, Household Yearly Consumption Expenditure; ICD, International Classification of Diseases; INR, Indian Rupees; LMIC, Low- and Middle-Income Countries; MOSPI, Ministry of Statistics and Programme Implementation; NSS, National Sample Survey; OOPE, Out of Pocket Expenditure; PG, Poverty Gap; PGR, Poverty Gap Ratio; PHCR, Poverty Head Count Ratio; PL, Poverty line; PPI, Pharmaceutical Price Index; RR, Relative Risk; USD, United States Dollars; WHO, World Health Organization.

* Corresponding author at: Rm 323 Samuels Building, UNSW Sydney, Kensington, NSW 2052, Australia.

E-mail addresses: menongr.hq@icmr.gov.in (G.R. Menon), richard.franklin@jcu.edu.au (R.C. Franklin), a.peden@unsw.edu.au (A.E. Peden).

https://doi.org/10.1016/j.jsr.2022.06.003
0022-4375/

drowning (International Classification of Diseases [ICD] codes W65-74), excluding fatalities due to water transport, flooding, undetermined intent, and intentional drowning (Passmore, Smith, & Clapperton, 2007; Peden, Franklin, Mahony, Barnsley, & Scarr, 2017).

Drowning is defined as a process of experiencing respiratory impairment, with outcomes classified as mortality (fatal drowning) or morbidity (nonfatal drowning), which may or may not require hospitalization (van Beeck, Branche, Szpilman, Modell, & Bierens, 2005). While fatal drowning is increasingly well understood in high-income contexts (Ahlm, Saveman, & Björnstig, 2013; Clemens, 2021; Peden, Franklin, & Clemens, 2019), there is a dearth of research exploring fatal drowning in low- and middle-income countries (LMICs) (Cenderadewi, Franklin, & Devine, 2020; Miller, Alele, Emeto, & Franklin, 2019; Tyler et al., 2017). Additionally, due to lack of data and a uniform definition for nonfatal drowning, global research is limited, though proposed WHO guidelines may assist future studies (Beerman, 2018). The limited nonfatal drowning research is largely focused on patient outcomes, with a need for studies on social and economic impacts (Peden, Mahony, Barnsley, & Scarr, 2018).

The economic burden associated with drowning has been estimated in a few studies. A study from Australia estimated fatal unintentional drowning at AUD 1.24 billion annually, with the highest costs among people aged 25–34 years and for drowning in rivers, creeks, and streams (Barnsley, Peden, & Scarr, 2018). Surf-related drowning fatalities in the Great Lakes of North America were estimated to be on average USD 105 million per year, far exceeding the cost of lifeguard services and drowning prevention education programs (Houser, Arbex, & Trudeau, 2021). One study from Northern Iran explored the costs associated with both fatal and nonfatal drowning (Davoudi-Kiakalayeh, Dalal, Yousefzade-Chabok, Jansson, & Mohammadi, 2011), identifying each drowning as costing the Iranian economy 17 times the gross domestic product per capita, with males aged 10–29 years constituting the majority of the economic burden.

Low- and middle-income countries (LMICs) are overrepresented in fatal drowning data, accounting for 90% of all deaths (1, 11). India, a lower-middle-income country with a population of 1.39 billion people, is believed to experience a significant drowning burden (Gupta, Bhaumik, Roy, Panda, Peden, & Jagnoor, 2020; Suresh Kumar Shetty & Shetty, 2007). The GBD study estimated India recorded more than 88,600 drowning deaths in 2017, a rate of 4.7 per 100,000 people (Franklin et al., 2020). Drowning in natural water environments and among young children are issues of concern in India (Gupta, Zwi, & Jagnoor, 2020; Lukaszyk et al., 2019). However, there remains a lack of research exploring drowning in India, with disparate data sources posing a challenge for the reporting of nonfatal drowning (Lukaszyk, Ivers, & Jagnoor, 2018).

Given the dearth of research on nonfatal drowning in India, in particular the social and economic impacts, and a recent study indicating availability of such data for drowning (Yadav, Menon, Agarwal, & John, 2021); this study aimed to examine: Out-of-Pocket Payments, Catastrophic Health Spending, and Hardship Financing and Impoverishment due to unintentional drowning-related hospitalization in India.

2. Materials and methods

2.1. Study design

Expenditure data on unintentional drowning was utilized from the 75th round of the [Indian] National Sample Survey (NSS) conducted in 2018. The detailed study design, and the data collection methods are described in the NSS report (National Sample Survey Organisation, 2019).

Briefly, the NSS is a cross-sectional, nationally representative, large-scale survey in India that is conducted under the leadership of the Ministry of Statistics and Programme Implementation (MOSPI), Government of India (GOI) on various socio-economic aspects. The survey adopts a multi-stage sampling design to select a representative sample of households across all states and union territories in India. In the first stage of the sampling, villages/urban blocks are selected using probability proportional to size with replacement while in the second stage, households are chosen by systematic random sampling without replacement (from selected villages/urban blocks). The 2018 survey covered 113,823 households and 555,115 individuals and was collected over between July 2017 and June 2018. The NSS captures hospitalization expenditure for the last 365 days preceding the survey period.

2.2. Data and sample size

All individuals who reported hospitalization for the treatment of unintentional drowning were included in the sample. There were 174 unintentional drowning-related hospitalizations among 66,238 hospitalizations captured in the 2018 survey.

2.3. Outcome measurements

2.3.1. Out of pocket expenditure (OOPE)

Out-of-pocket expenditure was calculated as the sum of direct medical costs (i.e., hospital stay, consultation, treatment medicines and procedures, laboratory and other investigation charges), and direct non-medical costs (i.e., transportation, meals, lodging: for patients and caregivers) (Centre & Estimates, 2019; Mohanty & Kastor, 2017; Yadav, Menon, & John, 2021). The net OOPE was the excess out of pocket payment made after receipt of any kind of reimbursement. (Centre & Estimates, 2019; Mohanty & Kastor, 2017; Yadav et al., 2019, 2020, 2021a, 2021b, 2021c, 2021d). Expenditure is shown in Indian Rupees (INR) and US Dollars after conversion as per December 2020 ($1 USD = □73.66) (Exchange Rates, 2020).

2.3.2. Households yearly consumption expenditure (HYCE)

This study used yearly household consumption expenditure as a proxy variable for household income as used in numerous earlier studies (Joe & Rajpal, 2018; Kastor & Mohanty, 2018; Sangar, Dutt, & Thakur, 2019). As the NSS gives the monthly per capita consumer expenditure, this was converted to a yearly expenditure by using a factor of 12 (Yadav et al., 2021; Yadav, John, Allarakha, & Menon, 2021).

2.3.3. Health care burden (HCB)

The health care burden (HCB) for a household is defined as the percentage share of OOPE in the yearly household consumption expenditure (HCE) (Mitra, Findley, & Sambamoorthi, 2009; Sahoo & Madheswaran, 2014; Yadav et al., 2019, 2021) i.e.,

$$\text{Health Care Burden}\,(\text{HCB}) = \frac{\text{OOPE x 100}}{\text{HCE}}$$

In order to update the expenditure incurred in 2018 to December, 2020, we first adjusted the expenditure data using the pharmaceutical Price Index (PPI) for December 2020 (Gumber, 2021). The PPI usually termed as "*Inflation Rate,*" measures and examines the weighted average of prices of a basket of Manufacture of Pharmaceuticals, Medicinal Chemical and Botanical Products. (Bajwala, John, Rajasekar, & Murhekar, 2019; Journalist & Index, 2021; Bank, 2021). The adjustment equation for data captured in 2018 is as follows:

$$Expenditure\ for\ December 2020 = \frac{\text{PPI for December 2020}}{\text{Average PPI for December, 2017 to January, 2018}} * Expenditure\ for\ 2018$$

2.3.4. Catastrophic health expenditure (CHE)

A household experiences catastrophic health expenditure (CHE) if the health care expenditure exceeds a threshold of the total household consumption expenditure, in this study we used two thresholds of 10% and 20%, that is, when HCB > 10% or HCB > 20% (Ghosh, 2011; Joe, 2014; Kumar et al., 2015; Mohanty & Kastor, 2017; Selvaraj, Farooqui, & Karan, 2018; Yadav et al., 2021; Yadav et al., 2021; Yadav, Allarakha, Menon, John, & Nair, 2021) i.e.,

$$CHE = 1\ if\ HCB > 10\%\ and\ CHE = 0 HCB \leq 10\%$$

$$CHE = 1\ if\ HCB > 20\%\ and\ CHE = 0 HCB \leq 20\%$$

2.3.5. Impoverishment

Impoverishment was measured using two indices namely Poverty Head Count Ratio (PHCR) (Kumar et al., 2015; Mohanty & Kastor, 2017), which is the percentage of households that fall below the poverty line (PL) due to OOPE and the Poverty Gap Ratio (PGR) (Kumar et al., 2015; Yadav et al., 2019, 2021) measured as the average percentage deficit from the poverty line due to OOPE (Ghosh, 2011; Kumar et al., 2015; Wagstaff & Ev, 2003; Yadav et al., 2019, 2021). For the present study we used PL = 1,407INR ≈USD 26.35 in urban areas and 972 INR ≈ USD 17.36 in rural areas as per the recommendation of the Rangarajan committee in 2014, based on the 2012 data (Raveendran, 2016). Further, the PL for 2012 was adjusted for inflation using the consumer price index for December 2020 to obtain PL for December (Bajwala et al., 2019).

A household that was above the poverty line is defined as impoverished if it falls below poverty line after incurring healthcare expenditure, i.e.,

$$Poverty\ head\ count\ (PHC) = 1\ if\ HCE \geqslant PL\ \&\ (HCE - OOPE) < PL$$

*Otherwise*0, where, *HCE* is total yearly consumption expenditure of a household.

The poverty head count ratio (PHCR) is the percentage of households that fall below the poverty line due to OOPE, i.e.,

$$Poverty\ head\ count\ ratio(PHCR) = \frac{\sum_{i=1}^{N} Number\ of\ impoverished\ households_i}{Total\ households\ (N)} \times 100$$

The percentage deficit from the poverty line of an impoverished household (PHC = 1) is quantified using the poverty gap (PG), i.e.,

$$Poverty\ gap\ (PG) = PHC \times \{PL - (HCE - OOPE)\}/PL$$

The poverty gap ratio (PGR) measures the poverty gap of all impoverished households as a percentage of total households in the population.

$$Poverty\ gap\ ratio(PGR) = \frac{1}{\text{Total households (N)}} \sum_{i=1}^{N} Poverty\ gap_i \times 100$$

2.3.6. Hardship financing

Hardship financing is a condition when a household has to borrow money with interest or sell its property/assets to meet its OOPE for health care as indicated in previous studies (John & Kumar, 2017; Yadav et al., 2019, 2021). The households that met this condition were identified.

2.4. Defining predictor variables

The present study included important socioeconomic and demographic predictors such as sex, age, education, caste (social group), religion, income quintile, place of residence, types of health facility, types of wards, and duration of hospitalization as predictor variables based on the past studies and available variables in the NSS dataset (National Sample Survey Organisation, 2019; Passmore et al., 2007; Yadav et al., 2021; Yadav et al., 2021). Caste/tribe has significant relevance in the Indian context for any social and health indicator. Under *the varna* system, humans were divided into four classes. The bottommost rung of this system includes 'scheduled castes/tribes,' who were allowed to do derogatory and scavenging work only. People from these strata lack access to information and health services while the 'other backward classes' (OBC) comprise those deprived of health services primarily due to social and economic constraints. The 'other' category comprises those not comprised in any of the above categories. The description and type of measurement of each predictor variable is given in Table S1.

2.5. Analytical approach

Descriptive statistics are reported as means and standard deviations for continuous variables and percentages for categorical variables, bivariate estimates and multivariable models were performed to meet the research objectives. Descriptive analysis was used to estimate the proportion of the respondent by their socioeconomic characteristics. Crude rates of unintentional drowning-related hospitalization per 100,000 population were calculated. Additionally, crude rates of drowning-related hospitalization per 1,000 hospitalizations were also calculated. We calculated the population-based relative risk (RR) of being hospitalized for drowning by using the number of people within a particular variable (e.g., sex, age) who reported being hospitalized for drowning as the population with the outcome and the total survey population as the exposed population. Alongside the RR, the standard deviation, statistical significance *p* value ($p < 0.05$) and upper and lower 95% confidence intervals (CI) were calculated.

In the second phase of the analysis, bivariate analyses were carried out to understand the differences in mean OOPE, HCB, CHE, impoverishment, and hardship financing by socioeconomic characteristics. We applied the SVY command (Korn & Graubard, 1990) used in STATA (StataCorp, 2013) to adjust for sampling weights. In the present study, we converted the yearly consumption expenditure into an adult equivalent household consumption expenditure that has been used in various former studies (Kwesiga, Zikusooka, & Ataguba, 2015; Levine, 2012; Van Minh & Xuan, 2012; Yadav et al., 2021) (Table S2).

2.6. Ethicall approval

The data analyzed for this study are from the 75th round of the National Sample Survey, which contains anonymized data in the public domain. The authors had no access to personal identifiable data. Data available in public domain are approved for use for research purpose by MOSPI, GOI.

3. Results

A total of 174 respondents reported unintentional drowning-related hospitalization within the 12 months prior to being surveyed (Table 1). This equates to a crude rate of 31.34 hospitalization per 100,000 population. Drowning-related hospitalizations occurred at a rate of 2.263 per 1,000 unintentional hospitalizations in 2018.

3.1. Socio-economic distribution of unintentional drowning and submersion-affected individuals

Table 1 shows the socioeconomic and demographic characteristics of the respondent hospitalized due to unintentional drowning. Out of the total sample, 63.4% ($n = 107$) were males. The largest proportion of respondents hospitalized for drowning were aged 21–50 years (44.9%; $n = 65$). One-fifth of all drowning-related hospitalizations (20.2%) occurred among the 0–20 years age group. Two-thirds of all respondents hospitalized for drowning lived in rural areas (69.9%; $n = 103$).

One quarter (28.0%) of the hospitalized respondents had no formal education. One-third of those hospitalized (37.8%; $n = 73$) were from high economic status households. Almost two thirds (65.0%; $n = 110$) were treated in a private hospital, with 56.0% of respondents staying in a paid general ward. More than half of the respondents were hospitalized for five or less days (57.7%; $n = 101$). (Table 1).

Table 2 shows the relative risk (RR) of being hospitalized for unintentional drowning among those surveyed. Those aged 51 years and above had the highest risk of hospitalization due to unintentional drowning (RR = 4.56; 95 %CI: 3.06–6.80; $p < 0.01$) as compared to those aged 20 years and less. Females had a lower risk of hospitalization as compared to males (RR = 0.61; 95 %CI: 0.45–0.84; $p < 0.01$). Those who belonged to non-Hindu religions were at lower risk of hospitalization as compared to Hindus (RR = 0.57; 95 %CI: 0.38–0.86; $p = 0.01$). Drowning risk was relatively even between urban and rural residents.

Table 1
Profile of individuals who were hospitalized due to unintentional drowning in India.

Background Characteristics	%	*n*
Total	**100.0**	**174**
Age group (in years)		
0–20	20.2	36
21–50	44.9	65
51 and above	34.9	73
Unknown	0.0	0
Education		
Illiterate	28.0	47
Up to Primary	32.9	52
Middle	16.3	29
Secondary and above	22.8	42
Unknown	4.6	8
Gender		
Male	63.4	107
Female	36.6	63
Unknown	2.3	4
Marital Status		
Currently married	67.1	106
Others	32.9	64
Unknown	2.3	4
Religion		
Hindu	84.9	142
Non-Hindu	15.2	28
Unknown	2.3	4
Social group		
Scheduled Caste/Scheduled Tribes	24.9	44
Others Backwards Classes	39.7	73
Others	35.4	53
Unknown	2.3	4
Economic status of household		
Low	31.8	48
Medium	30.4	49
High	37.8	73
Unknown	2.3	4
Place of residence		
Rural	69.9	103
Urban	30.1	71
Type of health facility		
Public hospital	35.0	60
Private hospital	65.0	110
Unknown	2.3	4
Type of ward		
Free	37.1	59
Paying general	56.0	100
Paying special	6.9	15
Duration of stay in hospital		
Up to 5 days	57.7	101
6–10 days	26.4	41
11 and above	15.9	32
Health Insurance		
No	95.4	164
Yes	4.6	10

3.2. Out of pocket expenditure (OOPE), yearly household consumption expenditure (YCE) and health care burden (HCB)

Table 3 shows the average OOPE, the YCE, and HCB per episode of hospitalization care by different socioeconomic characteristics. Overall OOPE for drowning-related hospitalization amounted to INR 25,421 (345.11USD) per person with the highest OOPE seen among those aged 51 years and older (INR 31,064; 421.72USD). OOPE was highest among those with a middle school education at INR 32,982 (447.76USD). OOPE for drowning-related hospitalization was higher for females, those currently married, Hindus, and lower economic strata. As expected, those treated in private hospitals reported higher OOPE (INR 36,391; 494.04USD) as compared to those treated in public hospitals (INR 5,790; 78.60USD). Those who were treated in special paid ward reported the highest OOPE (INR 63,739; 865.31USD). A longer hospital stay resulted in higher OOPE totaling INR 65,197 (994.06USD) for those treated for 11 or more days. Health care expenditure (HCE) largely mirrored OOPE (Table 3 and Table S3).

Overall, the HCB for drowning-related hospitalization amounted to 15.7%. HCB was higher among the oldest age group (51 years and older: 18.3%), those with middle school education (25.9%), males (16.2%), those who are currently married (17.5%), lower economic strata (30.8%), rural residents (20.0%), those treated in private hospitals (19.4%), and those hospitalized for 11 or more days (43.9%).

This study also highlighted the direct medical cost, non-medical cost, cost for transport for patient, total OOPE, reimbursed by insurance, net OOPE and average number of workdays lost (Table S3).

3.3. Catastrophic health expenditure (CHE)

Table 4 provides the percentage of households that faced CHE by spending more than 10% and more than 20% of the total household's consumption expenditure on treatment for unintentional drowning. Over half of all households were pushed to CHE at the 10% threshold (53.8%). CHE was more likely: for people aged 51 years and older; when a male was hospitalized; when they lived in a rural area; they went to a private hospital; or they did not have health insurance. As duration of stay in hospital increased, more households were pushed to CHE.

3.4. Poverty effects due to OOPE

Table 4 also shows the potential for households to fall below the poverty line (poverty headcount ratio) and the average deficit from the poverty line (poverty gap ratio) due to OOPE for households, which reported seeking hospital treatment for unintentional drowning. Almost 1 in 10 households surveyed (9.0%) who

Table 2
Relative risk (RR) by background characteristics of people hospitalized for unintentional drowning in India.

Background Characteristics	RR	Std. Err.	*Level of significance* (*p*-value)	95% CI	
				Lower	Upper
Age (in years)					
0–20	Ref				
21–50	1.48	0.31	0.06	0.98	2.22
51 and above	4.56	0.93	<0.01	3.06	6.80
Education					
Illiterate	Ref				
Up to Primary	1.08	0.22	0.70	0.73	1.60
Middle	1.21	0.29	0.42	0.76	1.92
Secondary and above	0.73	0.15	0.13	0.48	1.10
Gender					
Male	Ref				
Female	0.61	0.10	<0.01	0.45	0.84
Marital Status					
Currently married	Ref				
Others	0.65	0.10	0.01	0.47	0.88
Religion					
Hindu	Ref				
Non-Hindu	0.57	0.12	0.01	0.38	0.86
Social group					
Others Backward Classes	Ref				
Scheduled Caste/Scheduled Tribes	1.26	0.24	0.23	0.87	1.83
Others	1.25	0.25	0.28	0.84	1.86
Economic status of household					
Low	Ref				
Medium	1.10	0.24	0.65	0.72	1.71
High	0.64	0.13	0.03	0.43	0.96
Place of residence					
Rural	Ref				
Urban	0.98	0.15	0.89	0.72	1.32

reported a drowning-related hospitalization fell below the poverty line. The average percentage deficit from the poverty line was 7.6% (poverty gap).

Almost 12% (11.7%) of households belonging to the middle economic strata who reported hospitalization due to unintentional drowning fell below the poverty line, compared to 8.6% of higher-income households. Low-income households also reported the greatest percentage deficit from the poverty line at 20.2%, compared to just 2.5% among middle income households and 2.7% among high income households. Households in rural areas reported a higher percentage of poverty headcount ratio (9.5%) as compared to 8.0% for urban households, and higher poverty gap deficit (10.0%) as compared to 2.4% among their urban counterparts.

Households in which members were treated in a private hospital showed a considerably higher poverty gap ratio (12.4%) than those treated in a public hospital (2.3%). Similarly, the percentage of households reporting drowning-related hospitalization who fell below the poverty line increased with duration of hospitalization (Table 4).

3.5. Hardship financing

Overall, 14.4% of respondents reported needing to sell assets or take a loan to meet the health care expenses for treatment of drowning (Table 4). The highest proportion of hardship financing was seen among the oldest age group (patients aged 51 years and over; 18.8%). Females reported a higher proportion of hardship financing (18.0%). The highest proportion of hardship financing was seen among those with the lowest education levels, at 22.0% among illiterate respondents and 18.8% among those with primary education. Similarly, those in the lower economic strata reported a higher proportion of hardship financing at 16.8% as compared to 2.0% of higher economic strata.

4. Discussion

Drowning is a significant cause of mortality and morbidity (Franklin et al., 2020; World Health Organization, 2014), particularly in LMICs such as India (Gupta et al., 2020; Lukaszyk et al., 2018; Suresh Kumar Shetty & Shetty, 2007). There is a dearth of research exploring nonfatal drowning, including the economic and social impact (Cenderadewi et al., 2020; Miller et al., 2019; Peden et al., 2018). This study aimed to quantify unintentional drowning-related hospitalization in India and examine the associated economic burden via a nationally representative survey.

The dearth of nonfatal drowning-related data in India has been previously reported. A systematic literature review of drowning in India identified a nonfatal drowning rate of 3 per 100,000 in 2014 and just 481 nonfatal drowning-related injuries reported in police data nationally in 2015, although a definition of what constitutes a nonfatal drowning was not available (Lukaszyk et al., 2018). Our rate of nonfatal drowning-related hospitalization is significantly higher than previously reported, but the incidence is lower than reported in police data, which has its own limitations (Lukaszyk et al., 2018). Further research is required to better quantify the rates of both fatal and nonfatal drowning nationally across India, including identifying the best way to gather such data at a population level, ensuring consistent definitions, and overcoming underreporting.

Males were overrepresented in nonfatal drowning-related hospitalizations in India, accounting for 63% of all cases. This is similar to many other studies globally that have identified males as being disproportionately impacted by drowning, both fatal and nonfatal (Croft & Button, 2015; Howland, Hingson, Mangione, Bell, & Bak, 1996; Peden et al., 2018). However, nonfatal drowning-related hospitalization in India among females was found to incur a higher economic burden when compared to males. Females recorded higher average out of pocket expenditure and yearly health care expenditure when compared to males. Similarly, a higher propor-

Table 3
Background Characteristic by Average Out of pocket Expenditure, yearly household consumption expenditure and Health Care Burden for treatment of unintentional drowning-related hospitalization in India.

Background Characteristics	OOPE		YCE		Health Care Burden (%)
	(in INR₹)	(in USD$)	(in INR₹)	(in USD$)	
Total	**25,421**	**345.11**	**162,352**	**2204.07**	**15.7**
Age (in years)					
0–20	16,072	218.19	190,928	2592.02	8.4
21–50	25,247	342.75	143,867	1953.12	17.5
51 and above	31,064	421.72	170,003	2307.94	18.3
Education					
Illiterate	21,974	298.32	145,803	1979.41	15.1
Up to Primary	27,422	372.28	170,943	2320.70	16.0
Middle	32,982	447.76	127,341	1728.77	25.9
Secondary and above	22,409	304.22	195,301	2651.38	11.5
Gender					
Male	23,604	320.45	146,036	1982.57	16.2
Female	29,222	396.71	190,636	2588.05	15.3
Marital Status					
Currently married	28,087	381.31	160,685	2181.44	17.5
Others	20,705	281.09	165,755	2250.27	12.5
Religion					
Hindu	27,320	370.89	167,335	2271.72	16.3
Non-Hindu	16,361	222.12	134,452	1825.31	12.2
Caste					
SC/ST	24,166	328.07	122,904	1668.53	19.7
OBC	29,552	401.19	152,454	2069.70	19.4
Others	22,338	303.26	201,233	2731.92	11.1
Economic strata					
Low	28,488	386.75	92,389	1254.26	30.8
Medium	22,344	303.34	129,493	1757.98	17.3
High	25,940	352.16	247,714	3362.94	10.5
Place of residence					
Rural	24,981	339.14	124,879	1695.34	20.0
Urban	26,442	358.97	253,159	3436.86	10.4
Type of health facility					
Public hospital	5790	78.60	111,905	1519.21	5.2
Private hospital	36,391	494.04	188,003	2552.31	19.4
Type of ward					
Free	6612	89.76	116,822	1585.96	5.7
Paying general	33,182	450.48	196,934	2673.55	16.8
Paying special	63,739	865.31	127,448	1730.22	50.0
Duration of stay in hospital					
Up to 5 days	15,214	206.54	164,279	2230.23	9.3
6–10 days	23,818	323.35	166,623	2262.06	14.3
11 and above	65,197	885.11	148,395	2014.59	43.9
Health Insurance					
No	26,247	356.32	161,212	2188.59	16.3
Yes	8415	114.24	185,507	2518.42	4.5

OOPE = *Out of pocket expenditure;* ***HCE*** = *Household Consumption Expenditure,* ***HCB =*** *Health Care Burden.*

tion of households reported hardship financing when a female was hospitalized due to drowning, when compared to males. Though gender-based drowning prevention efforts have traditionally focused on males due to a higher number of incidents, findings from this study indicate that the economic impacts of nonfatal drowning among females cannot be ignored. As such, efforts to prevent drowning among females must be given a higher priority given the devastating personal and household economic impact (Richardson, Peden, & Data, 2021; Roberts et al., 2021).

Our findings also indicate a significant nonfatal drowning burden among the oldest age group in this study, with those aged 51 years and over accounting for 35% of all nonfatal drowning-related hospitalizations reported, higher than those among young people (0–20 year olds) who often receive prominence in drowning advocacy and prevention efforts. Similarly, older respondents experienced a greater economic burden associated with nonfatal drowning-related hospitalization than younger age groups, including a higher out of pocket expenditure and a higher proportion of households falling below the poverty line. Drowning among older age groups has been described as a hidden epidemic, with few prevention efforts focused on this age group, despite a globally aging population (Clemens, Peden, & Franklin, 2021; Peden, Franklin, & Queiroga, 2018). Our findings provide further weight to the need to ensure drowning prevention efforts are addressing age-related risk factors for drowning among older people. While traditional cost/benefit analyses may place a higher value on the prevention of child drowning due to the greater years of life lost (Vos et al., 2020), our study highlights significant economic impacts associated with drowning among older people and, therefore, economic value in preventing drowning among this age group.

Although high-income respondents accounted for the highest proportion of people reporting a nonfatal drowning-related hospitalization (37.8%), people with low income reported a significantly higher proportion of health care burden for drowning-related hospitalization (30.8% for low-income people compared to 10.5% of high-income people). Additionally, drowning among low income people saw a higher proportion of households reporting a higher deficit from the poverty line due to out-of-pocket health care payments for drowning-related hospitalization, when compared to higher income households. Rural dwelling respondents were also found to represent a higher proportion of nonfatal drowning-related hospitalizations than their urban dwelling counterparts

Table 4
Percentage of households pushed to Catastrophic Health Expenditure, Percentage of households falling below the poverty line (poverty headcount ratio) and average deficit from the poverty line (poverty gap ratio) and hardship financing due to out-of-pocket expenditure in India.

Background Characteristics	Catastrophic Health Expenditure (CHE)		Poverty effects due to OOPE		Hardship financing
	10% threshold	20% threshold	Poverty headcount %	Poverty gap %	Percentage
Total	**53.8**	**25.8**	**9.0**	**7.6**	**14.4**
Age group (in years)					
0–20	36.3	16.5	8.8	4.0	13.3
21–50	60.0	26.4	6.2	4.5	12.7
51 and above	56.5	31.1	13.3	14.2	18.8
Education					
Illiterate	42.0	28.6	8.2	0.7	22.0
Up to Primary	58.6	18.3	7.8	14.6	18.8
Middle	66.8	49.2	11.9	12.2	7.5
Secondary and above	50.6	16.2	9.6	2.2	8.1
Gender					
Male	62.1	28.2	10.9	8.0	12.8
Female	38.4	21.3	5.5	7.0	18.0
Marital Status					
Currently married	58.8	29.0	7.9	9.1	14.9
Others	44.4	19.5	11.2	4.8	14.7
Religion					
Hindu	56.7	27.1	10.2	9.1	15.2
Non-Hindu	39.0	18.9	3.0	0.0	10.5
Social group					
Scheduled Caste/Scheduled Tribes	55.3	38.6	9.0	2.4	20.0
Others Backwards Classes	65.7	22.2	9.7	12.2	16.4
Others	37.9	20.3	8.2	6.1	9.0
Economic status of household					
Low	56.4	35.8	6.7	20.2	16.8
Medium	73.8	28.5	11.7	2.5	6.7
High	36.3	16.4	8.6	2.7	2.0
Place of residence					
Rural	58.2	27.5	9.5	10.0	9.1
Urban	44.0	21.8	8.0	2.4	5.0
Type of health facility					
Public hospital	21.7	10.4	2.3	0.1	1.2
Private hospital	71.3	33.9	12.4	11.5	12.4
Type of ward					
Free	25.7	9.5	2.1	0.0	1.2
Paying general	67.3	34.1	12.3	6.8	8.9
Paying special	83.5	38.3	16.9	50.6	35.6
Duration of stay in hospital					
Up to 5 days	43.3	8.1	2.2	0.5	0.8
6–10 days	65.2	41.1	13.9	2.6	6.1
11 and above	70.8	60.1	23.8	39.4	30.9
Health Insurance					
No	55.5	26.3	9.5	7.9	13.9
Yes	21.4	15.3	0.0	0.0	15.3

and to experience a greater economic burden when compared to those residing in urban areas. Higher levels of education were found to be protective against drowning-related hospitalization, with a lower relative risk of drowning-related hospitalization for those with secondary and above education when compared to illiterate respondents (RR = 0.73; 95 %CI: 0.48–1.10). Socio-economic disadvantage and rural dwelling populations are known to have a higher risk of injury due to a range of mechanisms, including drowning (Fralick, Gallinger, & Hwang, 2013; Iqbal et al., 2007; Peden & Franklin, 2021; Tyler et al., 2017). Additionally, lower maternal and paternal education levels have been found to be a risk factor for child mortality (Balaj et al., 2021). Such determinants of health are a vital consideration in the development and deployment of drowning prevention strategies to prevent death and injury and to reduce further compounding inequity among those already disadvantaged.

This study also reports drowning incidence and economic impact by severity of drowning incident, using length of hospital stay as a proxy. Almost 16% of respondents reported a hospital stay for nonfatal drowning of 11 days or longer. Unsurprisingly, the highest economic impact across a range of measures was seen for the more severe drowning cases, which required the longest length of hospital stay. While this study has contributed to addressing the gap around the impacts of drowning, there is a need for further studies to take a systematic approach to exploring the long-term impact of nonfatal drowning in both high income and low and middle income countries, including the long-term health, social, and emotional impacts to develop evidence-informed interventions (Guevarra, Peden, & Franklin, 2021; Peden et al., 2018).

Our analysis shows that a high proportion of respondents affected due to nonfatal drowning belong to lower income strata and rural areas. Most rural areas in India have an inadequate health system that are generally overcrowded (Dash & Mohanty, 2019). Many supply side factors such as shortfall of Primary Health Centres and Community Health Centres, availability of beds (currently 1 per 100,000 population), and availability of qualified medical personnel such as medical doctor, nurses, and laboratory technicians exist (Dash & Mohanty, 2019). In recent years, the Government of India has introduced Ayushman Bharat, the largest ever health insurance scheme, with an intention to provide financial protection to the bottom 40% of the population (Government of India, 2021). The National Health Policy-2017 also aims to reduce medical impoverishment and reduce inequality in health spending by 2025. In our study, we could not measure the effect of Ayushman Bharat due to data limitations, and data on health infrastructure were not collected in the NSS surveys and could not be analyzed.

4.1. Practical applications

This study identified the importance of preventing drowning, not just in terms of minimizing loss of life and injury due to drowning, but also the economic impacts, which often increase inequities among those who are already disadvantaged. The findings of this study further highlight the need to address upstream determinants of health to aid in the prevention of drowning, as well as the need to ensure that drowning prevention interventions focus on those populations at increased risk. In addition, this study provides data to support investment in drowning prevention across the age span. While such investment will prevent death and injury, it also reduces the health system burden and the broader economic and social impacts of nonfatal drowning-related hospital treatment.

4.2. Strengths and limitations of the study

The strength of our study derived from its use of national representative cross-sectional data from the National Sample Survey. The NSS followed the standardized study design. NSS covered all the states and union territories of India, which offers the generalizability to the study results. This study adds to the sparse literature reporting nonfatal drowning and economic impacts of drowning. Although this study has numerous strengths, there are some limitations as well. This study reports a cross-sectional survey, and recall bias is a major limitation for expenditure data. Similarly underreporting may be possible. The survey did not collate data on the circumstances leading to the drowning incident that resulted in hospitalization. Due to data limitations, we were unable to assess the association of economic strata to severity of injury.

5. Conclusion

Nonfatal drowning-related hospitalization causes significant economic impact in India, often compounding inequities faced by disadvantaged groups such as older people, females, low-income households, people with low levels of education, and those residing in rural areas. There is a need for further investment in drowning prevention interventions in India, including strategies aimed at those already disadvantaged. Such strategies will reduce death and injury, as well as economic and health system burden.

6. Consent to participate and ethics approval

The present study used the data from the National Sample Survey (2018). First, the NSS obtained the ethical consent from the review committee of NSS before the survey. Second, during the survey a usual written consent had been taken by the respondent once they are agreed to take part in the study. Therefore, no ethical approval is required separately for the present study.

7. Data availability statement

This study utilized the NSS 75th round data which is publicly accessible to individuals for countrywide and worldwide on http://www.mospi.gov.in/.

Conflict of interest

None declared.

Acknowledgement

This study used National Sample Survey (2018) data. The authors gratefully acknowledge members of the study field team including who were involved in mapping/listing/segmentation and main survey during data collection. The authors also acknowledge all the respondent for their active participation in this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsr.2022.06.003.

References

Ahlm, K., Saveman, B.-I., & Björnstig, U. (2013). Drowning deaths in Sweden with emphasis on the presence of alcohol and drugs - a retrospective study, 1992–2009. *BMC Public Health, 13*, 216.

Bajwala, V. R., John, D., Rajasekar, T. D., & Murhekar, M. V. (2019). Severity and costs associated with hospitalization for dengue in public and private hospitals of Surat city, Gujarat, India, 2017–2018. *Transactions of The Royal Society of Tropical Medicine and Hygiene, 113*(11), 661–669.

Balaj, M., York, H. W., Sripada, K., Besnier, E., Vonen, H. D., Aravkin, A., et al. (2021). Parental education and inequalities in child mortality: A global systematic review and meta-analysis. *The Lancet, 398*(10300), 608–620.

Barnsley, P. D., Peden, A. E., & Scarr, J. (2018). Calculating the economic burden of fatal drowning in Australia. *Journal of Safety Research, 67*, 57–63.

Beerman SB, Bierens, JJLM., Clemens, T., Meddings, D., Rahman, A., Szpilman, D. Clarification and Categorization of Non-fatal Drowning: A draft Position Statement for review and input by the global drowning community: World Health Organization; 2018. Available from: https://www.who.int/docs/default-source/documents/drowning/non-fatal-drowning-categorization.pdf?sfvrsn=44d18cc1_2 (Date accessed: 10-01-2021).

Cenderadewi, M., Franklin, R. C., & Devine, S. (2020). Socio-ecological nature of drowning in low-and middle-income countries: A review to inform health promotion approaches. *International Journal of Aquatic Research and Education, 12*(2), 6.

Clemens, T., Peden, A. E., & Franklin, R. C. (2021). Exploring a hidden epidemic: Drowning among adults aged 65 years and older. *Journal of Aging and Health, 33*(10), 828–837.

Clemens T. Persistent Racial/Ethnic Disparities in Fatal Unintentional Drowning Rates Among Persons Aged ≤29 Years—United States, 1999–201MMWR Morbidity and mortality weekly report. 2021;70.

Croft, J. L., & Button, C. (2015). Interacting factors associated with adult male drowning in New Zealand. *PLoS ONE, 10*(6), e0130545.

Cyber Journalist. All India Consumer Price Index (General) for Industrial Workers (Base= 1982=100) New Delhi: Cyber Journalist; n.d. Available from: http://cyberjournalist.org.in/manisana/aicpinew.html (Date accessed: 10-09-2021).

Dash, A., & Mohanty, S. K. (2019). Do poor people in the poorer states pay more for healthcare in India? *BMC Public Health, 19*(1), 1020.

Davoudi-Kiakalayeh, A., Dalal, K., Yousefzade-Chabok, S., Jansson, B., & Mohammadi, R. (2011). Costs related to drowning and near drowning in northern Iran (Guilan province). *Ocean and Coastal Management, 54*(3), 250–255.

Exchange Rates UK. US dollar to Indian rupee spot exchange rates for 2020; 2021. Available from: https://www.exchangerates.org.uk/USD-INR-spot-exchange-rates-history-2020.html (Date accessed: 1-09-2021).

Fralick, M., Gallinger, Z. R., & Hwang, S. W. (2013). Differences in drowning rates between rural and non-rural residents of Ontario, Canada. *International Journal of Aquatic Research and Education, 7*(4), 332–343.

Franklin, R. C., Peden, A. E., Hamilton, E. B., Bisignano, C., Castle, C. D., Dingels, Z. V., et al. (2020). The burden of unintentional drowning: Global, regional and national estimates of mortality from the Global Burden of Disease 2017 Study. *Injury Prevention, 26*(Supp 1), i83–i95.

Ghosh, S. (2011). Catastrophic payments and impoverishment due to out-of-pocket health spending. *Economic and Political Weekly, 46*(47), 63–70.

Government of India. Pradhan Mantri Jan Arogya Yojana (PM-JAY): Government of India; 2021. Available from: https://pmjay.gov.in (Date accessed: 23-09-2021).

Guevarra, J. P., Peden, A. E., & Franklin, R. C. (2021). Application of the PRECEDE–PROCEED model in the development of evidence-informed interventions for drowning prevention: A mixed-methods study protocol. *BMJ Open, 11*(7), e050688.

Gumber, A. (2021 Jun). Equity in healthcare utilisation and cost of treatment in Western India. *Journal of Social and Economic Development, 23*(1), 131–152.

Gupta, M., Bhaumik, S., Roy, S., Panda, R.K., Peden, M., Jagnoor, J. Determining child drowning mortality in the Sundarbans, India: applying the community knowledge approach. Injury Prevention. 2020:injuryprev-2020-043911.

Gupta, M., Zwi, A. B., & Jagnoor, J. (2020). Opportunities for the development of drowning interventions in West Bengal, India: A review of policy and government programs. *BMC Public Health, 20*, 1–13.

Houser, C., Arbex, M., & Trudeau, C. (2021). Short communication: Economic impact of drowning in the Great Lakes Region of North America. *Ocean & Coastal Management, 212* 105847.
Howland, J., Hingson, R., Mangione, T., Bell, N., & Bak, S. (1996). Why are most drowning victims men? Sex differences in Aquatic Skills and Behaviors. *American Journal of Public Health, 86*(1), 93–96. 769.
Iqbal, A., Shirin, T., Ahmed, T., Ahmed, S., Islam, N., Sobhan, A., et al. (2007). Childhood mortality due to drowning in rural matlab of Bangladesh: Magnitude of the problem and proposed solutions. *Journal of Health, Population, and Nutrition, 25*(3), 370.
Joe, W. (2014). Distressed financing of household out-of-pocket health care payments in India: Incidence and correlates. *Health Policy and Planning, 30*(6), 728–741.
Joe, W., & Rajpal, S. (2018). Unravelling the socioeconomic gradient in the incidence of catastrophic health care expenditure: A comment. *Health Policy and Planning, 33*(5), 699–701. https://doi.org/10.1093/heapol/czy026.
John, D., & Kumar, V. (2017). Exposure to hardship financing for healthcare among rural poor in Chhattisgarh, India. *Journal of Health Management, 19*(3), 387–400.
Kastor A, Mohanty SK. Disease-specific out-of-pocket and catastrophic health expenditure on hospitalization in India: Do Indian households face distress health financing? 2018. doi:10.1371/journal.pone.0196106.
Korn, E. L., & Graubard, B. I. (1990). Simultaneous testing of regression coefficients with complex survey data: Use of Bonferroni t statistics. *The American Statistician, 44*(4), 270–276.
Kumar, K., Singh, A., Kumar, S., Ram, F., Singh, A., Ram, U., et al. (2015). Socio-economic differentials in impoverishment effects of out-of-pocket health expenditure in China and India: evidence from WHO SAGE. *PLOS ONE, 10*(8), e0135051.
Kwesiga, B., Zikusooka, C. M., & Ataguba, J. E. (2015). Assessing catastrophic and impoverishing effects of health care payments in Uganda. *BMC Health Services Research, 15*(1), 30.
Levine, S. (2012). Exploring differences in national and international poverty estimates: Is Uganda on track to halve poverty by 2015? *Social Indicators Research, 107*(2), 331–349.
Lukaszyk, C., Ivers, R. Q., & Jagnoor, J. (2018). Systematic review of drowning in India: Assessment of burden and risk. *Injury Prevention, 24*(6), 451–458.
Lukaszyk, C., Mittal, S., Gupta, M., Das, R., Ivers, R., & Jagnoor, J. (2019). The impact and understanding of childhood drowning by a community in West Bengal, India, and the suggested preventive measures. *Acta Paediatrica, 108*(4), 731–739.
Miller, L., Alele, F. O., Emeto, T. I., & Franklin, R. C. (2019). Epidemiology, risk factors and measures for preventing drowning in Africa: A systematic review. *Medicina, 55*(10), 637.
Mitra, S., Findley, P. A., & Sambamoorthi, U. (2009). Health care expenditures of living with a disability: Total expenditures, out-of-pocket expenses, and burden, 1996 to 2004. *Archives of Physical Medicine and Rehabilitation, 90*(9), 1532–1540.
Mohanty, S. K., & Kastor, A. (2017). Out-of-pocket expenditure and catastrophic health spending on maternal care in public and private health centres in India: A comparative study of pre and post national health mission period. *Health Economics Review, 7*(1), 31.
National Health Systems Resource Centre. National Health Accounts Estimates for India (2016-17). New Delhi; 2019.
National Sample Survey Organisation (2017-18). Social consumption and health in India -NSS 75th round, 2017–18. New Delhi; 2019.
Passmore, J. W., Smith, J. O., & Clapperton, A. (2007). True burden of drowning: Compiling data to meet the new definition. *International Journal of Injury Control and Safety Promotion, 14*(1), 1–3.
Peden, A. E., & Franklin, R. C. (2021). Exploring the impact of remoteness and socio-economic status on child and adolescent injury-related mortality in Australia. *Children, 8*(1), 5.
Peden, A. E., Franklin, R. C., & Clemens, T. (2019). Exploring the burden of fatal drowning and data characteristics in three high income countries: Australia, Canada and New Zealand. *BMC Public Health, 19*(1), 794.
Peden, A., Franklin, R. C., Mahony, A., Barnsley, P., & Scarr, J. (2017). Using a retrospective cross-sectional study to analyse unintentional fatal drowning in Australia: ICD-10 coding-based methodologies verses actual deaths. *BMJ Open, 7* (e019407).
Peden, A. E., Franklin, R. C., & Queiroga, A. C. (2018). Epidemiology, risk factors and strategies for the prevention of global unintentional fatal drowning in people aged 50 years and older: A systematic review. *Injury Prevention, 24*(3), 240–247.
Peden, A. E., Mahony, A. J., Barnsley, P. D., & Scarr, J. (2018). Understanding the full burden of drowning: A retrospective, cross-sectional analysis of fatal and non-fatal drowning in Australia. *BMJ Open, 8*(11).
Raveendran, G. (2016). A review of Rangarajan Committee report on poverty estimation. *Indian Journal of Human Development, 10*(1), 85–96.
Reserve Bank of India. RBI Bulletin, 21. Wholesale Price Index. Available from: https://www.rbi.org.in/Scripts/BS_ViewBulletin.aspx?Id=19986 (Date cited: 11-12-2021).
Richardson K, Peden, AE. Another Gender Data Gap: Female Drowning in Aotearoa, New Zealand. Injury Prevention 27(6): 535-541.
Roberts, K., Thom, O., Devine, S., Leggat, P. A., Peden, A. E., & Franklin, R. C. (2021). A scoping review of female drowning: An underexplored issue in five high-income countries. *BMC Public Health, 21*(1), 1072.
Sahoo, A. K., & Madheswaran, S. (2014). Socio-economic disparities in health care seeking behaviour, health expenditure and its source of financing in Orissa: Evidence from NSSO 2004–05. *Journal of Health Management, 16*(3), 397–414.
Sangar, S., Dutt, V., & Thakur, R. (2019). Comparative assessment of economic burden of disease in relation to out of pocket expenditure. *Frontiers in Public Health, 7*, 9.
Selvaraj, S., Farooqui, H. H., & Karan, A. (2018). Quantifying the financial burden of households' out-of-pocket payments on medicines in India: A repeated cross-sectional analysis of National Sample Survey data, 1994–2014. *BMJOpen, 8*(5), e018020.
StataCorp (2013). *Stata statistical software*. College Station, TX: StataCorp LP.
Suresh Kumar Shetty, B., & Shetty, M. (2007). Epidemiology of drowning in Mangalore, a coastal Taluk of South India. *Journal of Forensic and Legal Medicine, 14*, 410–415.
Tyler, M. D., Richards, D. B., Reske-Nielsen, C., Saghafi, O., Morse, E. A., Carey, R., et al. (2017). The epidemiology of drowning in low- and middle-income countries: A systematic review. *BMC Public Health, 17*(1), 413.
Tyler, M. D., Richards, D. B., Reske-Nielsen, C., Saghafi, O., Morse, E. A., Carey, R., et al. (2017). The epidemiology of drowning in low- and middle-income countries: A systematic review. *BMC Public Health, 17*(1), 413.
United Nations. Resolution on Global Drowning Prevention: United Nations; 2021. Available from: https://www.undocs.org/en/A/75/L.76 (Date accessed: 23-09-2021).
van Beeck, E., Branche, C. M., Szpilman, D., Modell, J. H., & Bierens, J. (2005). A new definition of drowning: Towards documentation and prevention of a global public health problem. *Bulletin of the World Health Organization, 83*(11), 853–856.
Van Minh, H., & Xuan, T. B. (2012). Assessing the household financial burden associated with the chronic non-communicable diseases in a rural district of Vietnam. *Global Health Action, 5*(1), 18892.
Vos, T., Lim, S. S., Abbafati, C., Abbas, K. M., Abbasi, M., Abbasifard, M., et al. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet, 396*(10258), 1204–1222.
Wagstaff, A., & Ev, D. (2003). Catastrophe and impoverishment in paying for health care: With applications to Vietnam 1993–1998. *Health Economics, 12*(11), 921–933.
World Health Organization. Global report on drowning: Preventing a leading killer Geneva: World Health Organization; 2014.
World Health Organisation. Fact sheets - Drowning 2020. Available from: https://www.who.int/news-room/fact-sheets/detail/drowning (Date accessed: 10-09-2021).
Yadav, J., John, D., & Menon, G. (2019). Out of pocket expenditure on tuberculosis in India: Do households face hardship financing? *Indian Journal of Tuberculosis, 66* (4), 448–460.
Yadav, J., Menon, G. R., Mitra, M., Allarakha, S., & John, D. (2020). Burden and cost of communicable, maternal, perinatal and nutrition deficiency diseases in India. *Journal of Public Health*.
Yadav, J., Allarakha, S., Menon, G. R., John, D., & Nair, S. (2021). Socioeconomic impact of hospitalization expenditure for treatment of noncommunicable diseases in India: A repeated cross-sectional analysis of national sample survey data, 2004 to 2018. *Value in Health Regional Issues, 24*, 199–213.
Yadav, J., John, D., Allarakha, S., & Menon, G. R. (2021). Rising healthcare expenditure on tuberculosis: Can India achieve the End TB goal? *Tropical Medicine & International Health, 26*(10), 1256–1275.
Yadav, J., John, D., Menon, G. R., & Allarakha, S. (2021). Out-of-Pocket payments for delivery care in India: Do households face hardship financing? *Journal of Health Management, 23*(2), 197–225.
Yadav, J., Menon, G., Agarwal, A., & John, D. (2021). Burden of injuries and its associated hospitalization expenditure in India. *International Journal of Injury Control and Safety Promotion, 28*(2), 153–161.
Yadav, J., Menon, G. R., & John, D. (2020). Disease-Specific out-of-pocket payments, catastrophic health expenditure and impoverishment effects in India: An analysis of national health survey data. *Applied Health Economics and Health Policy., 19*, 769–782.

Jeetendra Yadav is currently working as Technical Officer- B at National Institute of Medical Statistics, Indian Council of Medical Research, Delhi, India. He completed Master's degree, in Economics, Population Studies and Post Graduate Diploma in Applied Statistics as well. He has received doctoral degree in Population Studies/ Demography from IIPS, Mumbai. His research interest includes demography, public health, health economics, and Epidemiology. He has experience of working on large scale health and demographic data and different statistical software. Dr. Yadav has published more than 60 research papers. He is Principal and Co-Principal investigator in various completed and ongoing research projects.

Denny John is currently working as Adjunct Professor, Faculty of Life and Allied Health Sciences, Bengaluru, India; Adjunct Faculty Amrita Institute of Medical Sciences & Research Centre, Kochi, India; and Research Methodologist, Center for Public Health Research, MANT, Kolkata. His research interests include health financing in developing countries, health technology assessment, evidence synthesis, and health economics. He is Chair, Campbell & Cochrane Economic Methods Group (CCEMG); and Co-Chair, Early Career Network, Health Technology Assessment International (HTAi).

Geetha R Menon is a trained biostatistician with a doctorate degree in Biostatistics. She is a senior scientist in Indian Council of Medical Research(ICMR), National Institute of Medical Statistics(NIMS) at New Delhi. She has more than 3 decades of experience in the field of medical statistics and public health research. She has conducted and collaborated projects of national importance in Injuries and Trauma, Non communicable diseases and Disease Burden estimation. She has been a member of National Task Force Committees of ICMR and various Advisory Committees in the Ministry of Health and Family Welfare. She is the recipient of INDUSEM 2013 Injury Researcher Award for significant contribution in injury research. Geetha has authored/ co-authored in more than 80 national and international scientific publications and has edited two books on Road Safety and Traumatic Brain Injuries. In NIMS, she undertakes statistical analysis of epidemiological and biomedical data, supports systematic reviews and is involved in capacity building and teaching programmes for young researchers

Richard C Franklin PhD, FPHAA, FARL, FACTM is a pracademic who uses evidence based approaches to develop real world solutions to improving health, safety and wellbeing focusing on drowning, health services, rural populations, agricultural safety, disasters, and road safety. He is a Professor in Public Health and the Director for the World Safety Organization Collaborating Centre - Injury Prevention and Safety Promotion at James Cook University. His research interests are wide ranging and have included epidemiological, qualitative, translational, program evaluation, product evaluation, surveillance and pure research. He is a board member of Royal Life Saving Society – Australia, Kidsafe Australia and Queensland, Farmsafe Australia and Auschem Training.

Amy Peden PhD is an injury prevention researcher and lecturer in the School of Population Health at UNSW Sydney and holds an Australian National Health and Medical Research Council Fellowship. She holds adjunct appointments with James Cook University and the George Institute for Global Health. Much of her work focuses on drowning prevention, regional and remote communities, social determinants of health and alcohol. Dr Peden is also a Senior Research Fellow with Royal Life Saving Society – Australia and is a member of the International Life Saving Federation Drowning Prevention and Public Education Commission.

Journal of Safety Research 82 (2022) 301–313

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Occupational injuries among non-standard workers in the Taiwan construction industry

Chia-Wen Liao [a,*], Tsung-Lung Chiang [b]

[a] Department of Civil Engineering and Hazard Mitigation, China University of Technology, No.56, Section 3, Hsinlong Road, Taipei 116, Taiwan
[b] Northern Occupational Safety and Health Centers of Occupational Safety and Health Administration of Ministry of Labor of Taiwan, No.439, Zhongping Rd., Xinzhuang Dist., New Taipei City 242, Taiwan

ARTICLE INFO

Article history:
Received 13 July 2021
Received in revised form 23 November 2021
Accepted 9 June 2022
Available online 21 June 2022

Keywords:
Non-standard worker
Occupational fatality
Construction safety
Pearson's chi-squared test

ABSTRACT

Introduction: Global changes in the labor force have led to an increase in non-standard employment (NSE) workers, particularly apparent in the construction industry. These workers have a higher risk of occupational injury and negative health-related outcomes. *Method:* In this study, relevant literature and the database for construction accidents are examined to identify the classification of NSE in the Taiwan construction industry. Accident reports from 2000 to 2018 are extracted from case reports of the Northern Occupational Safety and Health Center of Taiwan. Pearson's chi-squared test are then employed to analyze a total of 1,612 occupational fatality cases in the construction industry to explore the differences in occupational injuries between NSE and standard employment (SE). Further, characteristics of occupational injuries for different types of NSE in the construction industry are analyzed. *Results:* The NSE occupational injury rate for older workers over 60 years old is higher, especially for self-employed workers taking on technical work. NSE workers are more expected to suffer occupational fatalities in the small-scale, non-public, and repair projects. Occupational injuries involving self-employed and temporary agency workers are clearly regionally concentrated. Temporary agency workers involved in occupational injuries are most engaged in non-technical work and movement for worker motion with their unfamiliarity with the worksite. Most enterprises did not perform safety management on construction sites for occupational injuries involving NSE workers, especially for self-employed workers. *Conclusions:* The results show that the hazard characteristics of NSE workers are clearly different from SE workers. NSE workers face inferior job security and protection, especially for self-employed workers. *Practical Applications:* The results can be used to establish effective occupational safety management policies and programs more efficiently.

1. Introduction

Global changes in the labor force caused by an aging population and industrial transformation have led to an increase in the proportion of non-standard employment (NSE) workers. NSE is usually an informal arrangement between employee and employer or self-employment and, accordingly, does not have labor regulations and social protection (Arestis et al., 2020; Aleksynska, 2018; Ruiz et al., 2015). On the other hand, NSE can be a comprehensive condition that consists of different components: employment instability; disempowerment; vulnerability; low or insufficient wages; limited rights; and incapacity to exercise rights (Benach et al., 2014). NSE constitutes a global phenomenon, and mainly affects low and middle income regions, particularly apparent among younger workers and new workforce entrants (Lilla & Staffolani, 2012; Probst et al., 2018; Ruiz et al., 2017; Schaufeli, 2016). The NSE ratio in Taiwan across all industries increased from 2.5% in 2007 to 5.6% in 2017. In 2017, the NSE ratio in Taiwan's construction industry was 22.21%, which was much higher than other industries. Hiring NSE workers benefits organizations not only because these workers help reduce labor costs, but also because organizations can downsize or expand the business according to demands of the marketplace (Martinez et al., 2010; Sakurai et al., 2013).

Despite such economic advantages to organizations, NSE may have adverse effects on the workers. NSE workers may have a higher risk of experiencing occupational injury and negative health-related outcomes, such as poor mental health, specific health problems, and poor health status (Alfers & Rogan, 2015; Basu et al., 2016; Benavides et al., 2006; Cummings & Kreiss, 2008; Econie & Dougherty, 2019; Kawachi, 2008; Mai et al., 2019; Probst et al., 2018; Ruiz et al., 2017; Sakurai et al., 2013).

* Corresponding author.
E-mail address: cwliao@cute.edu.tw (C.-W. Liao).

https://doi.org/10.1016/j.jsr.2022.06.006
0022-4375/

A possible cause is, with their limited job tenure, NSE workers are unfamiliar with the work environment and the employer's safety issues and procedures (Sakurai et al., 2013). They may be more reluctant to admit to deficiencies in required job-related safety knowledge (Cummings & Kreiss, 2008), and organizations are less willing to invest in safety-related training and equipment for NSE workers (Aronsson, 1999). Another may be that in order to increase the chance of obtaining a new contract, NSE workers are extrinsically motivated to work hard. Such pressure might lead to risk-taking behaviors (Sakurai et al., 2013). Additionally, the types of jobs outsourced to NSE workers are often more hazardous in nature (Rousseau & Libuser, 1997; Thébaud-Mony, 1999). NSE workers were disproportionately employed in higher numbers within the highest injury rate occupations (Pierce et al., 2013), and thus they were likely to work under worse working conditions than standard employment (SE) workers (Ruiz et al., 2017).

In addition, because NSE workers worry about losing their current job and some uncertainty over the job's future, there is often job insecurity (Sakurai et al., 2013). And the stressor of job insecurity is more salient and detrimental. Job insecurity may bring cognitive, emotional, and attitudinal reactions, which lead to weakened employee well-being from decreased job satisfaction, organizational commitment, trust in management, and job involvement to more negative mental health outcomes and turnover intentions. As job insecurity increases, employee safety knowledge and motivation to comply with safety policies and procedures decreases (Probst & Brubaker, 2001). Individuals violate more safety policies, and their risk taking behaviors at work increase (Probst, 2002; Størseth, 2006). As a result, job insecurity has harmful impacts on employee safety attitudes, behaviors, and outcomes (Grunberg et al., 1996; Probst & Brubaker, 2001; Probst et al., 2018; Probst, 2002), and may produce a devastating "snow ball" effect that negatively impacts working life, housing quality, and poverty (Ruiz et al., 2017).

The occupational safety protection of workers is much lower for NSE when compared to SE. In Taiwan, the Occupational Safety and Health Act No. 1 was amended in 2014 to extend the protection of the Act from laborers to all workers. According to the amendment of Article 2 of the Act, "the term 'workers' referred to in this Act means laborers, self-employed workers, or other people engaged in work and directed or supervised by the responsible people in workplaces." Additionally, according to the new Article 51, Paragraph 2, "People engaged in work directed or supervised by the responsible people in workplaces as described in Article 2 subparagraph 1, when performing labor work at business entities' workplaces, are equally subject to this Act as laborers employed by said enterprise." However, not only do business entities lack awareness of this concept, in practical terms, it is also difficult to confirm the relationship between workers and business entities.

In Taiwan, the Occupational Injuries Ratio per Thousandth fell from 4.439 in 2005 to 2.773 in 2017 across all industries, a fall of 37.5%. The Occupational Injuries Ratio per Thousandth in the construction industry fell from 12.968 in 2005 to 10.036 in 2017, a fall of only 22.6%. This shows that the occupational accident risk in the construction industry is much higher than in other industries because the characteristics of the construction industry make it difficult to reduce occupational accidents. NSE workers accounted for half of the occupational injuries in the construction industry in the Northern Region of Taiwan. In addition, the proportion of occupational injuries involving NSE workers in the construction industry has been on the rise in recent years, as shown in Fig. 1. In particular, temporary agency work and self-employment have seen large increases, as shown in Fig. 2. Therefore, with the future trend of increasing NSE in the construction industry, how to put forward relevant safety management strategies is an issue that the government and businesses must urgently address. Nevertheless, less has been discussed about the differences in occupational injuries between NSE and SE.

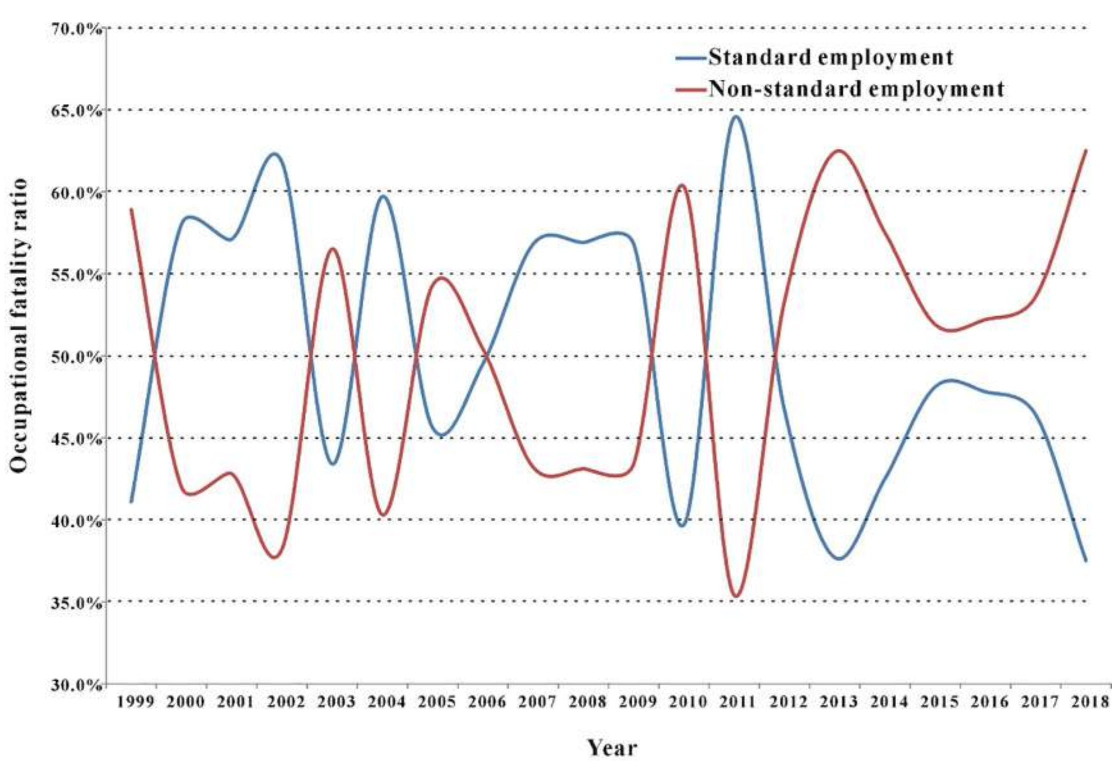


Fig. 1. Occupational fatality ratio for NSE and SE in the Taiwan construction industry.

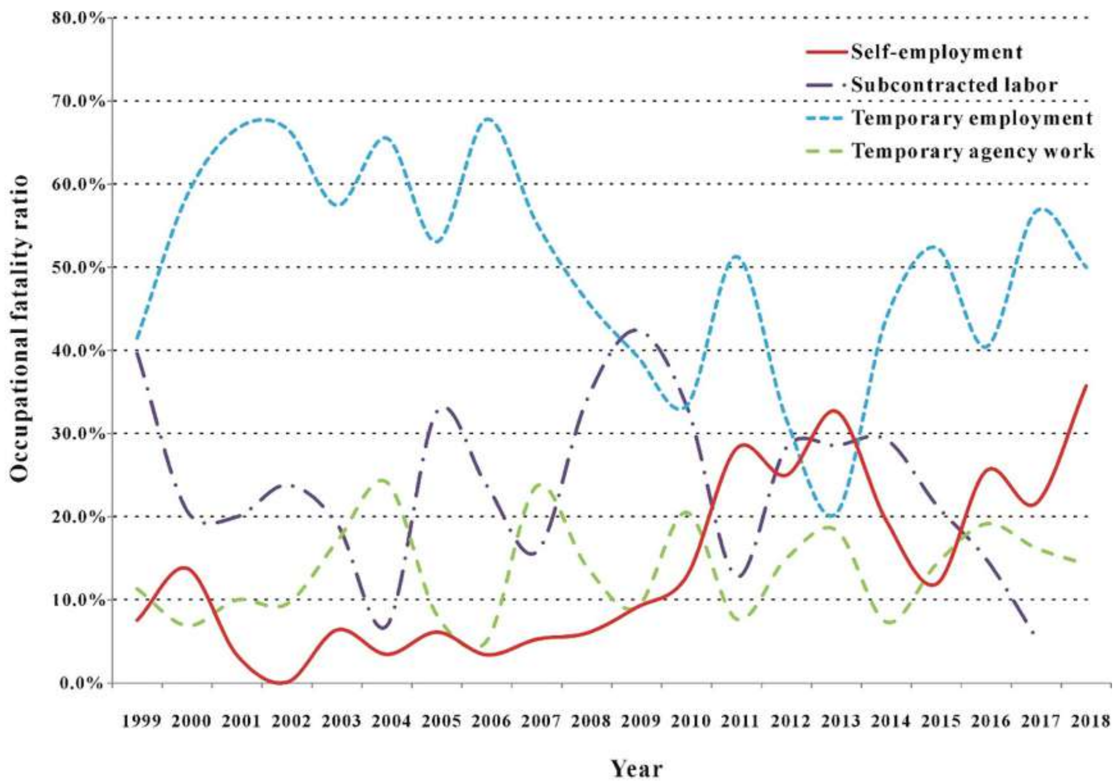


Fig. 2. Occupational fatality ratio for different types of NSE in the Taiwan construction industry.

In the present study, the relevant literature is reviewed to identify the classification of NSE in the Taiwan construction industry. A total of 1,612 occupational fatality cases in the construction industry are then analyzed to explore the differences in occupational injuries between NSE and SE. Further, we analyze the characteristics of occupational injuries for different types of NSE in the construction industry. The occupational injury characteristics of NSE are used to propose measures to improve safety management.

2. NSE classification

NSE is defined as a non-regulated placement in the labor market (International Labour Organization [ILO], 2002, 2012), that can take different forms. ILO divides NSE into four types: (1) temporary employment; (2) part-time work; (3) temporary agency work and multi-party employment; and (4) disguised employment and dependent self-employment.

Temporary employment whereby workers are engaged for a specific period of time includes fixed-term, project- or task-based contracts, as well as seasonal or casual work, including day labor. In part-time employment, the normal hours of work are fewer than those of comparable full-time workers. When workers are not directly employed by the company to which they provide their services, their employment falls under contractual arrangements involving multiple parties. In most countries, an employment relationship normally exists between the agency and the worker, whereas a commercial contract binds the agency and the user firm. According to the ILO, disguised employment can involve masking the identity of the employer by hiring the workers through a third party, or by engaging the worker in a civil, commercial, or cooperative contract instead of an employment contract and at the same time directing and monitoring the working activity in a way that is incompatible with the worker's independent status. Thus, the worker is purposefully misclassified as an independent, self-employed worker.

Referring to the ILO's NSE classification, this study adjusts the classification according to the characteristics of Taiwan's construction industry and data limitations. NSE workers in the construction industry are classified into self-employment, temporary employment, subcontracted labor, and temporary agency work. Self-employment includes both independent self-employment and dependent self-employment. Among them, subcontracted labor is labor employed through subcontractors and supervised by the general contractor. Companies that process with supplied materials are referred to as subcontractors, but in reality, their operations are almost entirely dependent on the general contractor. In other words, their operations follow the operations of the general contractor. Such companies do not have the ability to operate independently, and their work is entirely dependent on the general contractor. Even the work arrangements of the company's employees are dominated by the general contractor. These workers are nominally employed by the company, but actually managed by the general contractor. Therefore, the occupational safety facilities are basically the responsibility of the general contractor.

3. Materials and methods

This study identifies SE or NSE in the database of occupational injuries according to the victim's identity. Pearson's chi-squared test was then employed to compare several categorized risk factors between SE and NSE. Then, we identify the relationship between different types of NSE workers and various explanatory factors. The study collects relevant survey statistics on NSE in Taiwan to determine the significant relationship between employment type and occupational injury factors.

Table 1
Description of occupational fatalities in the Taiwan construction industry.

Factor	*N*	Item	*n*	%
Individual factors				
Whether the worker is non-standard worker	1612	N	806	50.0
		Y	806	50.0
Worker age	1533	Under 29	210	13.7
		30–39	344	22.4
		40–49	441	28.8
		50–59	334	21.8
		Over 60	204	13.3
Daily salary	1422	Under NT$1,000	260	18.3
		NT$1,000–1,499	482	33.9
		NT$1,500–1,999	352	24.8
		NT$2,000–2,499	194	13.6
		Over NT$2, 500	134	9.4
Whether the worker is in the possession of labor insurance status	1304	N	682	52.3
		Y	622	47.7
Job type	1512	Technical	1144	75.7
		Non-technical	368	24.3
Safety training	1409	N	1035	73.5
		Y	374	26.5
Project factors				
Project contract amount	1405	Under NT$1,000,000	315	22.4
		NT$1,000,000–10,000,000	235	16.7
		NT$10,000,000–100,000,000	268	19.1
		NT$100,000,000–500,000,000	254	18.1
		Over NT$500,000,000	333	23.7
Project type	1413	Civil engineering project	431	30.5
		Building project	608	43.0
		Repair project	374	26.5
Whether the project is public	1608	N	1052	65.4
		Y	556	34.6
Location of the project site	1612	Taipei City	211	13.1
		Others	66	4.1
		Yilan & Hualien County	180	11.2
		Taoyuan City	417	25.9
		New Taipei City	504	31.3
		Hsinchu County	234	14.5
Accident factors				
Worker motion	1340	Movement	354	26.4
		Others	29	2.2
		Applying force hard without moving	477	35.6
		Applying force slightly without moving	357	26.6
		Stand still	123	9.2
Source of injury	1440	Equipment	465	32.3
		Environment	651	45.2
		Temporary facilities	324	22.5
Accident type	1612	Unknown	18	1.1
		Traffic accident	53	3.3
		Others	15	0.9
		Falling object	82	5.1
		Collapse	233	14.5
		Caught in between, and clamped	36	2.2
		Struck by	48	3.0
		Fall on the same level	26	1.6
		Electric shock	133	8.3
		Drowning	20	1.2
		Contacting hazardous materials and/or extreme temperatures	16	1.0
		Fall	904	56.1
		Struck against	10	0.6
		Explosion, fire	18	1.1
Compensation	932	Under NT$2,000,000	287	30.8
		NT$2,000,000–4,000,000	198	21.2
		NT$4,000,000–6,000,000	268	28.8
		Over NT$6,000,000	179	19.2
Management factors				
Whether the original enterprise established a consultative organization	989	N	378	38.2
		Y	611	61.8

Table 1 (*continued*)

Factor	*N*	Item	*n*	%
Whether the original enterprise notified the contractor of potential hazards	1044	N	398	38.1
		Y	646	61.9
Whether the original enterprise established labor safety and health staff	1484	N	489	33.0
		Y	995	67.0
Whether the original enterprise provided labors with safety and health education and training	1484	N	763	51.4
		Y	721	48.6
Whether the original enterprise implemented self-inspection	1484	N	752	50.7
		Y	732	49.3
Whether the accident contractor provided labors with safety and health education and training	1482	N	1081	72.9
		Y	401	27.1
Whether the accident contractor prepared safety and health work rules	1482	N	1165	78.6
		Y	317	21.4
Whether the accident contractor established labor safety and health staff	1482	N	922	62.2
		Y	560	37.8
Whether the accident contractor implemented self-inspection	1481	N	1182	79.8
		Y	299	20.2

Note: 1 USD ≈28.14 NTD (10/25/2021).

3.1. Materials

This study analyzed 1,612 accident reports of fatal occupational injuries in the construction industry during the period 2000 to 2018. All accident reports were extracted from case reports at the Northern Occupational Safety and Health Center of Taiwan. Relevant factors are derived from the accident reports. In addition to general factors, several factors related to the regulations for safety and health management are included in the management factors. Each accident report was reviewed several times to itemize detailed information on each factor. Each factor is divided into several levels by the classification schemes mostly provided by the Ministry of Labor of Taiwan, as shown in Table 1.

3.2. Background information survey

The present study collects information on some denominators for risk factors from relevant labor and engineering statistics to facilitate the discussion of the research results. Five types of information related to the attributes in the present study are identified: (a) the proportion of NSE in the construction industry; (b) the proportion of workers in different age groups in the construction industry; (c) the proportion of technical workers across all industries; (d) the ratio of different levels of project contract amount for public works; and (e) the proportion of building permits in each county and city (Table 2–4). The first three categories are based on the Yearbook of Manpower Survey Statistics published by the Directorate-General of Budget, Accounting and Statistics, Executive Yuan, Taiwan. The ratio of different levels of project contract amount for public works is based on the 2018 bidding data of the Public Construction Commission, Executive Yuan, Taiwan,. The proportion of building permits in each county and city is based on a housing condition survey released by the Construction and Planning Agency, Ministry of the Interior, Taiwan.

Due to the differences between the information on some denominators and the risk factors in this study, in comparative analysis, the different definitions of risk factors may cause differences in ratios. For example, the Directorate General of Budget, Accounting and Statistics of Taiwan categorizes NSE workers only as part-time, temporary, or dispatched workers. This is different from the classification in the present study: self-employment,

Table 2
Age distribution of workers for the Taiwan construction industry (2000–2018).

Factor	Item	%
Worker age	Under 29	14.7%
	30–39	22.7%
	40–49	28.2%
	50–59	27.0%
	Over 60	7.4%

Table 3
Distribution of project contract amount for the Taiwan public works (2018).

Factor	Item	%
Project contract amount	Under NT$1,000,000	56.3%
	NT$1,000,000–10,000,000	6.9%
	NT$10,000,000–100,000,000	36.1%
	NT$100,000,000–500,000,000	0.5%
	Over NT$500,000,000	0.2%

Note: 1 USD ≈28.14 NTD (10/25/2021).

Table 4
Distribution of building permits in the Northern Region of Taiwan (2017).

Factor	Item	%
Location of the project site	Taipei City	4.5%
	Others	1.5%
	Yilan & Hualien County	31.8%
	Taoyuan City	31.8%
	New Taipei City	8.7%
	Hsinchu County	21.6%

temporary employment, subcontracted labor, and temporary agency work. However, based on the results of the Pearson's chi-squared test discussed later, the data can still be used as a basis of comparison to make useful conclusions.

3.3. Methods

Pearson's chi-squared test was employed to compare several categorized risk factors.

(Bunner et al., 2021; Eijkelenboom et al., 2020; Boyd et al., 2021; Kwon et al., 2019). Pearson's chi-squared test is also known as the chi-squared test for goodness-of-fit or chi-squared test for independence. A test of goodness-of-fit estimates whether an observed frequency distribution differs from a theoretical distribution. A test of independence assesses whether paired observations on two variables, expressed in the rows and columns of a contingency table, are independent of each other. In the present study, Pearson's chi-squared test was employed for the independence test. The null hypotheses are that each variable is independent of SE/NSE and that each variable is independent of NSE types. For a test of independence, a chi-squared probability of less than or equal to 0.05 is justification for rejecting the null hypothesis that the row variable is independent of the column variable. The software used to perform the statistical calculation in this study was IBM SPSS Statistics Version 19.

4. Results and discussion

4.1. Statistics on risk factors for occupational injuries in Taiwan construction industry

According to Table 1, SE and NSE each accounted for around 50% of occupational injuries. According to the survey carried out by the Directorate General of Budget, Accounting and Statistics of Taiwan during 2000–2018, the proportions of NSE and SE workers in the construction industry are 22.2% and 77.79%, respectively. Although the definition of NSE for the previous survey differs somewhat from this article, it is clear that SE accounts for the majority. However, NSE accounts for half of occupational injuries, showing that the NSE occupational injury rate in the construction industry is higher than that of SE.

When worker age is considered, the proportion of occupational injuries involving older workers (over 60 years old) in the construction industry was 13.3% (Table 1). According to Table 2, the proportion of older workers (over 60 years old) in the construction industry is only 7.4%, showing that the rate of occupational injuries among older workers (over 60 years old) is higher than that of other ages.

The ratio of occupational injuries in the construction industry involving workers with a daily salary of between NT$1,000 and NT$2,000 was over half (58.7%). Generally speaking, most workers in the construction industry have a daily salary of between NT $1,000 and NT$2,000.

For *whether the worker is in the possession of labor insurance status*, more than half (52.3%) of occupational injuries in the construction industry occurred among workers without labor insurance. The results show that a large proportion of workers in the construction industry may be uninsured. Therefore, when these workers suffer from occupational accidents, there is a high probability that they will not be compensated because they are not covered by labor insurance.

Regarding job type, the proportion of occupational injuries in the construction industry involving technical workers was as high as 75.7% (Table 1). According to the survey carried out by the Directorate General of Budget, Accounting and Statistics of Taiwan during 2000–2018, the proportions of technical and non-technical workers across all industries are 33.8% and 66.2%, respectively. The results show that the incidence of occupational accidents among technical workers in the construction industry is much higher than that of non-technical workers. This also shows that due to the more traditional skills used in the construction industry and the high-risk nature of the work environment, technical workers often have to take risks when working. As a result, the occupational injury rate involving technical workers is relatively high.

For *project contract amount*, projects with a contract amount of more than NT$500 million accounted for 23.7% of occupational injuries (Table 1), which is higher than projects of other amounts. The proportion of public construction projects with a construction budget of NT$500 million or more was only 0.2% in 2018 (Table 3), showing that the risk of occupational accidents in large projects is much higher than that in smaller projects.

Considering project type, building projects accounted for 43% of occupational injuries, showing that building projects involve a higher risk. For *whether the project is public*, 65.4% of occupational injuries occurred in non-public construction projects. For *location of the project site*, New Taipei City accounted for 31.3% of occupational injuries (Table 1). The 2017 building permits survey shows that New Taipei City accounts for only 8.7% of permits (Table 4), showing that New Taipei City has a high ratio of occupational injuries.

4.2. Pearson's chi-squared test between NSE and SE

Among the occupational injury factors, except for *location of the project site*, *whether the original enterprise established a consultative organization*, and *whether the original enterprise notified the contractor of potential hazards*, there are significant differences between NSE and SE for the remaining risk factors, as shown in Table 5. In other words, Pearson's chi-squared test results showed no significant difference between NSE and SE for *location of the project site, whether the original enterprise established a consultative organization,* and *whether the original enterprise notified the contractor of potential hazards*.

The age distribution of workers involved in occupational injuries was the same for both NSE and SE, with the highest number of accidents in the construction industry in the age range 40–49. However, the proportion of occupational injuries involving older workers (aged 60 and over) for NSE (15.6%) was higher than for SE (10.7%). Comparing all age groups, as shown in Table 6, the proportion of occupational injuries involving NSE workers increases with age. Conversely, the proportion of occupational injuries involving SE workers decreases with age. This finding may be related to the higher proportion of NSE workers with increasing age. However, Table 5 did not find such a phenomenon. Obviously, the proportion of occupational injuries involving older NSE workers is relatively high. This phenomenon may be attributed to the declining motor response of the elderly and engagement in high-risk work.

The daily salary distribution for workers involved in occupational injuries is the same for both groups, with a daily salary of between NT$1,000 and NT$2,000 accounting for the greatest number of occupational accidents. In particular, the proportion of NSE workers involved in occupational injuries with a daily salary NT $2,000 or above is higher than SE workers with a daily salary of NT$2,000 or above. The proportion of NSE workers involved in occupational injuries with a daily salary NT$2,500 or above is about three times higher than SE workers with a daily salary of NT$2,500 or above. The results show that a higher proportion of NSE workers involved in occupational injuries are in the high-paid category. This phenomenon may be because NSE workers tend to work in high-paying, high-risk work.

Table 5
Results of Pearson's chi-squared test between NSE and SE.

Factor	Item	SE (%)	NSE (%)	Chi-square test	*P* value
Individual factors					
Worker age	Under 29	14.6%	12.9%	9.623	0.047*
	30–39	24.2%	20.8%		
	40–49	28.7%	28.8%		
	50–59	21.7%	21.8%		
	Over 60	10.7%	15.6%		
Daily salary	Under NT$1,000	20.6%	16.0%	46.788	0.000**
	NT$1,000–1,499	36.8%	31.1%		
	NT$1,500–1,999	26.5%	23.1%		
	NT$2,000–2,499	11.3%	15.9%		
	Over NT$2, 500	4.7%	14.0%		
Whether the worker is in the possession of labor insurance status	N	40.0%	63.3%	70.656	0.000**
	Y	60.0%	36.7%		
Job type	Technical	81.4%	70.0%	26.898	0.000**
	Non-technical	18.6%	30.0%		
Safety training	N	58.3%	87.2%	150.695	0.000**
	Y	41.7%	12.8%		
Project factors					
Project contract amount	Under NT$1,000,000	21.4%	23.4%	10.841	0.028*
	NT$1,000,000–10,000,000	16.0%	17.4%		
	NT$10,000,000–100,000,000	17.9%	20.2%		
	NT$100,000,000–500,000,000	17.2%	18.9%		
	Over NT$500,000,000	27.5%	20.1%		
Project type	Civil engineering project	37.6%	23.3%	34.017	0.000**
	Building project	38.9%	47.2%		
	Repair project	23.5%	29.4%		
Whether the project is public	N	58.8%	72.1%	31.657	0.000**
	Y	41.2%	27.9%		
Location of the project site	Taipei City	12.7%	13.5%	6.159	0.291
	Others	4.7%	3.5%		
	Yilan & Hualien County	12.0%	10.3%		
	Taoyuan City	27.0%	24.7%		
	New Taipei City	30.4%	32.1%		
	Hsinchu County	13.2%	15.9%		
Accident factors					
Worker motion	Movement	24.9%	27.8%	15.218	0.004**
	Others	2.2%	2.2%		
	Applying force hard without moving	32.1%	38.8%		
	Applying force slightly without moving	29.5%	24.0%		
	Stand still	11.3%	7.2%		
Source of injury	Equipment	38.5%	26.2%	33.154	0.000**
	Environment	44.0%	46.4%		
	Temporary facilities	17.5%	27.4%		
Accident type	Unknown	2.1%	0.1%	46.309	0.000**
	Traffic accident	4.1%	2.5%		
	Others	1.2%	0.5%		
	Falling object	5.7%	4.5%		
	Collapse	14.1%	14.8%		
	Caught in between, and clamped	2.9%	1.6%		
	Struck by	3.6%	2.4%		
	Fall on the same level	1.1%	2.1%		
	Electric shock	8.2%	8.3%		
	Drowning	1.6%	0.9%		
	Contacting hazardous materials and/or extreme temperatures	1.5%	0.5%		
	Fall	52.0%	60.2%		
	Struck against	0.5%	0.7%		
	Explosion, fire	1.4%	0.9%		
Compensation	Under NT$2,000,000	29.2%	32.3%	21.425	0.000**
	NT$2,000,000–4,000,000	16.1%	26.3%		
	NT$4,000,000–6,000,000	31.8%	25.7%		
	Over NT$6,000,000	22.8%	15.6%		
Management factors					
Whether the original enterprise established a consultative organization	N	35.7%	40.6%	2.539	0.111
	Y	64.3%	59.4%		
Whether the original enterprise notified the contractor of potential hazards	N	35.7%	40.4%	2.435	0.119
	Y	64.3%	59.6%		
Whether the original enterprise established labor safety and health staff	N	28.1%	37.4%	14.366	0.000**
	Y	71.9%	62.6%		
Whether the original enterprise provided labors with safety and health education and training	N	45.1%	57.2%	21.464	0.000**
	Y	54.9%	42.8%		

(*continued on next page*)

Table 5 (*continued*)

Factor	Item	SE (%)	NSE (%)	Chi-square test	*P* value
Whether the original enterprise implemented self-inspection	N	45.3%	55.6%	15.839	0.000**
	Y	54.7%	44.4%		
Whether the accident contractor provided labors with safety and health education and training	N	63.1%	82.1%	67.251	0.000**
	Y	36.9%	17.9%		
Whether the accident contractor prepared safety and health work rules	N	70.0%	86.6%	60.777	0.000**
	Y	30.0%	13.4%		
Whether the accident contractor established labor safety and health staff	N	53.0%	70.7%	49.449	0.000**
	Y	47.0%	29.3%		
Whether the accident contractor implemented self-inspection	N	72.6%	86.5%	44.098	0.000**
	Y	27.4%	13.5%		

Note: 1 USD ≈28.14 NTD (10/25/2021).
2. * Indicates a significance level of 0.05.
** Indicates a significance level of 0.01.

Table 6
Worker age distribution of occupational fatalities between NSE and SE.

Factor	Item	SE (%)	NSE (%)	Subtotal (%)
Worker age	Under 29	50.5%	49.5%	100.0%
	30–39	51.2%	48.8%	100.0%
	40–49	47.4%	52.6%	100.0%
	50–59	47.3%	52.7%	100.0%
	Over 60	38.2%	61.8%	100.0%
	Total	47.4%	52.6%	100.0%

Table 7
Compensation of occupational fatalities in the Taiwan construction industry for NSE and SE workers.

Factor		Mean	SD
Compensation	SE	NT$4,579,973	NT$2,680,727
	NSE	NT$3,946,116	NT$3,719,821

Note: 1 USD ≈28.14 NTD (10/25/2021).

SE workers and NSE workers have the opposite distribution trends on *whether the worker is in the possession of labor insurance status*. The majority of SE workers who had an occupational injury were covered by labor insurance (60%). Conversely the majority of NSE workers involved in occupational injuries were not covered by labor insurance (63.3%).

The most prevalent job type of workers involved in occupational injuries was technical worker. In particular, among NSE workers the proportion of occupational injuries involving non-technical workers (30%) was much higher than among SE workers (18.6%). We find that the incidence of occupational accidents among technical workers in the construction industry is much higher than that of non-technical workers. This result may be caused by the higher number of technical workers in the construction industry. However, it may be that the safety threats faced by technical workers in the construction industry are higher than for other industries.

Most workers involved in occupational injuries had not received safety training. In particular, the proportion of NSE workers involved in occupational injuries who had not received safety training was 87.2%, much higher than the proportion of SE workers who had not received safety training (58.3%).

The distribution of *project contract amount* for occupational injuries was roughly the same in both the NSE and SE groups (more injuries for less than NT$1 million and more than NT$500 million). It is worth noting that the proportion of SE workers involved in occupational injuries working on projects with a contract amount of more than NT$500 million (27.5%) was much higher than for NSE workers (20.1%).

The proportion of occupational injuries occurring on building projects was the highest for both groups. It is worth noting that nearly half (47.2%) of NSE workers involved in occupational injuries were working on building projects, higher than the ratio of SE workers involved in occupational injuries that were working on building projects (38.9%).

A larger proportion of occupational injuries occurred in non-public construction projects (private sector construction projects) for both NSE and SE. In particular, the proportion of NSE workers involved in occupational injuries on non-public construction projects (72.1%) was much higher than the proportion of SE workers involved in occupational injuries on non-public construction projects (58.8%). This may be because NSE workers are commonly employed on private sector construction projects. It may also be because contractual requirements for public construction projects make it difficult to employ NSE workers or provide more protection for NSE workers.

The distribution of worker motion for workers involved in occupational injuries was largely the same for both groups. The highest proportion of accidents occurred when *applying force hard without moving*. The second highest ratio was in the category *movement* for NSE and *applying force slightly without moving* for SE.

Almost half of all occupational injuries were caused by the *environment*. It is worth noting that of the occupational injuries involving SE workers, the cause *equipment* (38.5%) was much higher than for NSE workers (26.2%). Conversely, for occupational injuries involving NSE workers, the cause *temporary facilities* (27.4%) was much higher than for SE workers (17.5%).

The accident type for more than half of occupational injuries in both groups was *fall*. However, the proportion of occupational injuries involving falls for NSE workers (60.2%) was about 10% higher than the proportion of occupational injuries involving falls for SE workers (52%). In addition, the ratio of *fall on the same level* for occupational injuries involving NSE workers is about twice as high as occupational injuries involving SE workers. Based on the aforementioned worker age results, it may be because there are more older NSE workers.

There was a difference in the *compensation* obtained depending on whether workers were NSE or not. For SE workers involved in occupational injuries, the largest proportion (31.8%) received NT $4 million–NT$6 million (31.8%). However, for NSE workers involved in occupational injuries, the largest proportion received less than NT$2 million. The average compensation for SE workers and NSE workers was NT$4.58 million and 3.95 million respectively, a difference of NT$630,000, as shown in Table 7. The standard deviation for SE and NSE workers was NT$2.68 million and NT$3.72 million, respectively. The results showed that not only did NSE workers obtain lower average compensation, but there were also large differences in the amount of compensation received.

Regardless of whether workers were NSE or not, most of the original enterprises involved in occupational injuries had established labor safety and health staff. It is worth noting that in cases of NSE occupational injuries, the proportion of original enterprises that had not established labor safety and health staff (37.4%) was more than for SE enterprises (28.1%).

SE workers and NSE workers have the opposite distribution trends on *whether the original enterprise provided labors with safety and health education and training*. In occupational injuries involving SE workers, *the original enterprise provided laborers with safety and health education and training* in the majority of cases (54.9%). In contrast, in occupational injuries involving NSE workers, in most cases the original enterprise *did not provide laborers with safety and health education and training* (57.2%).

SE workers and NSE workers have the opposite distribution trends on *whether the original enterprise implemented self-inspection*. In occupational injuries involving SE workers, *the original enterprise implemented self-inspection* in the majority of cases (54.7%). In contrast, in occupational injuries involving NSE workers, in most cases *the original enterprise did not implement self-inspection (55.6%).* The results are consistent with *whether the original enterprise provided labors with safety and health education and training.*

In terms of safety management of the original enterprise, for occupational injuries involving NSE workers, there was a higher ratio for *the original enterprise did not establish labor safety and health staff*. Additionally, in the majority of cases *the original enterprise did not provide laborers with safety and health education and training* and *implement self-inspection.*

Most of *the accident contractors did not provide laborers with safety and health education and training*. It is worth noting for occupational injuries involving NSE workers, the ratio for the *accident contractors did not provide laborers with safety and health education and training* (82.1%) was higher than for SE workers (63.1%).

Most of the *accident contractors did not prepare safety and health work rules*. For occupational injuries involving NSE workers, the ratio for the *accident contractors did not prepare safety and health work rules* (86.6%) was higher than for SE workers (70%).

Most *accident contractors did not establish labor safety and health staff*. For occupational injuries involving NSE workers, the ratio for *the accident contractors did not establish labor safety and health staff* (70.7%) was higher than for SE workers (53%).

The majority of accident contractors *did not implement self-inspection*. For occupational injuries involving NSE workers, the ratio for the accident contractors did not implement self-inspection (86.5%) was higher than for SE workers (72.6%).

In terms of the safety management of the accident contractor, for occupational injuries involving NSE workers, the ratios for *the accident contractor did not provide laborers with safety and health education and training, prepare safety and health work rules, establish labor safety and health staff*, and *implement self-inspection were higher*. Most accident contractors that employed NSE workers did not pay attention to regulations and site safety, greatly increasing the risk of harm to NSE workers.

4.3. Pearson's chi-squared test between different types of NSE

After removing SE data, there were 806 data records for NSE workers. Self-employment accounted for 13.8%, subcontracted labor for 24.2%, temporary employment for 48.5%, and temporary agency work for 13.5%, as shown in Table 8. Overall, temporary employment accounted for the largest number, almost 50%. Pearson's chi-squared test was then employed to compare several categorized risk factors between different types of NSE. In the majority of occupational injuries involving NSE, except for *whether the worker is in the possession of labor insurance status, whether the project is public,* and *accident type*, there were significant differences between different types of NSE for the other attributes, as shown in Table 9.

Among the four types of NSE, except for self-employment, the distribution of worker age is very similar, with the largest number aged 40–49, showing a normal distribution. However, for self-employment, the results showed an increased ratio of occupational injuries with older age. As workers get older, they have more experience and therefore, have more opportunities to be self-employed workers. Older self-employed workers often take on higher-risk technical work. Declining motor response with age is likely to result in accidents in high-risk workplaces.

The daily salary ratios of subcontracted labor and temporary employment workers involved in occupational injuries shows a normal distribution. The salary for self-employed workers involved in occupational injuries is concentrated in two groups: low to medium salary (daily salary NT$1,000–NT$1,499, accounting for 37.1%) and high salary (daily salary more than NT$2,500, accounting for 22.9%). The daily salary for temporary agency work is concentrated in the low to medium salary group (daily salary less than NT$1,500, accounting for 78.7%). The results show that a higher proportion of self-employed workers involved in occupational injuries have a high salary. Possibly because self-employed workers are well paid, they take more safety risks than other types of NSE.

Most NSE are not in the possession of labor insurance status, possibly resulting in the non-significant effect of *whether the worker is in the possession of labor insurance status* on different types of NSE. Except for temporary agency work, most of the workers involved in occupational injuries for other types of NSE are technical workers. This result shows that temporary agency workers are most engaged in non-technical work.

The ratio of workers involved in occupational injuries receiving education and training is low for all types of NSE. Of these, only 6.5% of self-employed workers involved in occupational injuries had received safety and health education and training. Only 9.7% of workers in temporary employment involved in occupational injuries had received safety and health education and training.

There is a clear difference in the distribution of *project contract amount* for occupational injuries among the four types of NSE. For workers in self-employment and temporary employment, the largest number of occupational injuries were in projects of less than NT$1 million (34.4% and 32.2%, respectively). For workers in temporary agency work, the highest number of occupational injuries were projects over NT$500 million (31.8%) and the lowest were projects under NT$1 million (8.4%).

Table 8
NSE distribution of occupational fatalities in the Taiwan construction industry.

Factor	*N*	Item	*n*	%
NSE type	806	Self-employment	111	13.8
		Subcontracted labor	195	24.2
		Temporary employment	391	48.5
		Temporary agency work	109	13.5

Table 9
Results of Pearson's chi-squared test between different types of NSE.

Factor	Item	NSE (%)				Chi-square test	*P* value
		Self-employment	Subcontracted labor	Temporary employment	Temporary agency work		
Individual factors							
Worker age	Under 29	1.8%	16.9%	13.8%	13.8%	83.492	0.000**
	30–39	13.5%	27.7%	19.7%	20.2%		
	40–49	21.6%	29.2%	31.7%	24.8%		
	50–59	23.4%	16.4%	24.6%	20.2%		
	Over 60	39.6%	9.7%	10.2%	21.1%		
Daily salary	Under NT$1,000	14.3%	9.8%	14.0%	35.2%	73.459	0.000**
	NT$1,000–1,499	37.1%	29.4%	27.9%	43.5%		
	NT$1,500–1,999	11.4%	30.9%	23.3%	12.0%		
	NT$2,000–2,499	14.3%	18.0%	17.6%	6.5%		
	Over NT$2, 500	22.9%	11.9%	17.3%	2.8%		
Whether the worker is in the possession of labor insurance status	N	52.7%	64.5%	63.0%	71.0%	6.054	0.109
	Y	47.3%	35.5%	37.0%	29.0%		
Job type	Technical	76.8%	87.0%	68.9%	37.1%	82.297	0.000**
	Non-technical	23.2%	13.0%	31.1%	62.9%		
Safety training	N	93.5%	81.4%	90.3%	84.0%	11.740	0.008**
	Y	6.5%	18.6%	9.7%	16.0%		
Project factors							
Project contract amount	Under NT$1,000,000	34.4%	10.1%	32.2%	8.4%	78.939	0.000**
	NT$1,000,000–10,000,000	27.8%	16.2%	17.3%	11.2%		
	NT$10,000,000–100,000,000	14.4%	25.7%	17.3%	25.2%		
	NT$100,000,000–500,000,000	6.7%	24.0%	18.1%	23.4%		
	Over NT$500,000,000	16.7%	24.0%	15.2%	31.8%		
Project type	Civil engineering project	18.6%	23.0%	25.1%	22.1%	53.899	0.000**
	Building project	31.4%	61.0%	39.7%	61.1%		
	Repair project	50.0%	16.0%	35.2%	16.8%		
Whether the project is public	N	78.4%	73.2%	69.7%	72.5%	3.445	0.328
	Y	21.6%	26.8%	30.3%	27.5%		
Location of the project site	Taipei City	25.2%	15.4%	9.5%	12.8%	39.552	0.001**
	Others	4.5%	2.1%	4.3%	1.8%		
	Yilan & Hualien County	9.9%	8.7%	12.5%	5.5%		
	Taoyuan City	18.9%	25.6%	28.1%	16.5%		
	New Taipei City	27.9%	33.3%	30.9%	38.5%		
	Hsinchu County	13.5%	14.9%	14.6%	24.8%		
Accident factors							
Worker motion	Movement	26.0%	23.1%	27.8%	37.9%	21.952	0.038*
	Others	4.1%		2.5%	3.2%		
	Applying force hard without moving	35.6%	40.8%	42.1%	25.3%		
	Applying force slightly without moving	26.0%	25.4%	22.2%	26.3%		
	Stand still	8.2%	10.7%	5.3%	7.4%		
Source of injury	Equipment	34.4%	23.5%	27.6%	18.2%	23.037	0.001**
	Environment	35.5%	39.7%	51.0%	52.5%		
	Temporary facilities	30.1%	36.9%	21.4%	29.3%		
Accident type	Unknown			0.3%		50.378	0.269
	Traffic accident	2.7%	1.5%	3.1%	1.8%		
	Others	0.9%	0.5%	0.6%	0.9%		
	Falling object	4.5%	4.6%	3.1%	9.2%		
	Collapse	6.3%	16.9%	16.6%	12.8%		
	Caught in between, and clamped	2.7%	2.6%	0.5%	2.8%		
	Struck by	1.8%	1.5%	3.3%	0.9%		
	Fall on the same level	4.5%	2.6%	1.3%	1.8%		
	Electric shock	10.8%	8.7%	7.9%	6.4%		
	Drowning	0.9%		1.3%	0.9%		
	Contacting hazardous materials and/or extreme temperatures			0.8%	0.9%		
	Fall	62.2%	59.5%	59.6%	61.5%		
	Struck against	0.9%	1.0%	0.8%			
	Explosion, fire	1.8%	0.5%	1.0%			
Compensation	Under NT$2,000,000	28.0%	33.8%	32.6%	30.0%	23.647	0.005**
	NT$2,000,000–4,000,000	40.0%	18.0%	33.0%	15.7%		
	NT$4,000,000–6,000,000	28.0%	28.8%	22.3%	30.0%		
	Over NT$6,000,000	4.0%	19.4%	12.0%	24.3%		

Table 9 (*continued*)

Factor	Item	NSE (%)				Chi-square test	*P* value
		Self-employment	Subcontracted labor	Temporary employment	Temporary agency work		
Management factors							
Whether the original enterprise established a consultative organization	N	54.8%	32.3%	44.2%	34.4%	11.990	0.007**
	Y	45.2%	67.7%	55.8%	65.6%		
Whether the original enterprise notified the contractor of potential hazards	N	56.5%	33.3%	43.3%	32.8%	12.611	0.006**
	Y	43.5%	66.7%	56.7%	67.2%		
Whether the original enterprise established labor safety and health staff	N	55.9%	24.0%	43.8%	22.9%	44.809	0.000**
	Y	44.1%	76.0%	56.2%	77.1%		
Whether the original enterprise provided labors with safety and health education and training	N	66.7%	43.2%	64.6%	47.7%	31.299	0.000**
	Y	33.3%	56.8%	35.4%	52.3%		
Whether the original enterprise implemented self-inspection	N	73.1%	41.7%	61.7%	44.0%	38.359	0.000**
	Y	26.9%	58.3%	38.3%	56.0%		
Whether the accident contractor provided labors with safety and health education and training	N	94.3%	79.2%	84.5%	68.8%	24.576	0.000**
	Y	5.7%	20.8%	15.5%	31.2%		
Whether the accident contractor prepared safety and health work rules	N	95.5%	90.1%	88.7%	66.1%	49.066	0.000**
	Y	4.5%	9.9%	11.3%	33.9%		
Whether the accident contractor established labor safety and health staff	N	90.9%	69.3%	73.9%	45.9%	51.952	0.000**
	Y	9.1%	30.7%	26.1%	54.1%		
Whether the accident contractor implemented self-inspection	N	95.5%	89.6%	86.8%	72.5%	25.959	0.000**
	Y	4.5%	10.4%	13.2%	27.5%		

Note: 1 USD ≈28.14 NTD (10/25/2021).
2. * Indicates a significance level of 0.05.
**Indicates a significance level of 0.01.

The aforementioned two results show that the majority of workers in the self-employment and temporary employment worker category had not received safety and health training and that the largest number of occupational injuries occurred on small projects. This may be because these two types of NSE workers are mostly engaged in small projects and have a high level of mobility. They are less likely to receive safety training yet engage in high-risk technical work, leading to occupational injuries.

Aside from the fact that occupational injuries involving self-employed workers occur most commonly on repair projects, occupational injuries involving other types of NSE workers most commonly occur on building projects. Repair projects are generally ad hoc and short-term in nature and not under the control of government agencies, making it more difficult to prevent work hazards.

The highest *location of the project site* was in New Taipei City. It is worth noting that in Taipei City, the ratio of occupational injuries involving self-employed workers was found to be much higher than other types of NSE workers. In Hsinchu County, the ratio of occupational injuries involving temporary agency workers is much higher than other types of NSE workers. Occupational injuries involving self-employed and temporary agency workers are clearly regionally concentrated. Table 4 shows that Taipei City only accounts for 4.5% of building permits issued in the Northern Region. Combining the *project type* and *location of the project site* results, occupational injuries involving self-employed workers mostly occurred on repair projects, possibly because repair projects accounted for the majority of projects in Taipei City.

Aside from movement accounting for the most *worker motion* in occupational injuries involving temporary agency workers, applying a force hard without moving accounted for the most *worker motion* in occupational injuries involving other types of NSE workers. This may be because temporary agency workers are unfamiliar with the site or are more involved in non-technical work.

The *environment* was the factor that accounted for the largest number of sources of injury. It is worth noting that, for workers in temporary employment and temporary agency work, more than half of all occupational injuries are caused by the *environment*. This may be because they are temporary workers who do not work at a fixed worksite. Their unfamiliarity with the worksite may cause greater risk.

There was a clear difference in the distribution of *compensation* for occupational injuries among the four types of NSE. It is worth noting that the proportion of self-employed workers receiving NT $6 million or more in compensation was much lower than the other types of NSE (only 4%). In addition, for NSE workers involved in occupational injuries, the largest proportion received less than NT$2 million. It is clear that employers do not take occupational injury coverage for NSE workers seriously. In particular, self-employed workers without an employer received the lowest compensation.

The original enterprise did not establish a consultative organization was found in most cases of occupational injuries involving self-employed workers, while in most cases involving other types of NSE *the original enterprise had established a consultative organization. The original enterprise did not notify the contractor of potential hazards* was found in most cases of occupational injuries involving self-employed workers, while in most cases involving other types of NSE *the original enterprise had notified the contractor of potential hazards. The original enterprise did not establish labor safety and* *health staff* was found in most cases of occupational injuries involving self-employed workers, while in most cases involving other types of NSE *the original enterprise had established labor safety and health staff*. *The original enterprise did not provide laborers with safety and health education and training* was found in most cases of occupational injuries involving self-employed and temporary employment workers. *The original enterprise did not implement self-inspection* was found in most cases of occupational injuries involving self-employed and temporary employment workers.

In terms of safety management of the original enterprise, for occupational injuries involving self-employed workers, in most cases the original enterprise had not *established a consultative organization*, *notified the contractor of potential hazards*, *established labor safety and health staff*, *provided laborers with safety and health education and training*, or *implemented self-inspection*. For occupational injuries involving temporary employment workers, in most case the original enterprise *had not provided laborers with safety and health education and training* or *implemented self-inspection*. For original enterprises that hire self-employed workers, the majority used subcontracting and did not perform on-site management. Moreover, most of the original enterprises did not pay attention to regulations and site safety, which also increases the risk of hazards. For original enterprises that hire temporary employment workers, possibly because the turnover of temporary employment workers is very high and work periods are short, safety and health education and training and self-inspection were not implemented.

In most cases of occupational injuries involving self-employed and temporary employment workers, *the accident contractor did not provide laborers with safety and health education and training*. Regardless of the type of NSE, in most cases of occupational injuries, *the accident contractors did not prepare safety and health work rules*. Aside from *the accident contractor established labor safety and health staff*, in the majority of cases involving temporary agency workers, in most cases of occupational injuries involving other types of NSE, *the accident contractor did not establish labor safety and health staff*. In most cases of occupational injuries, *the accident contractor did not implement self-inspection*.

5. Conclusions

In this study, 1,612 cases of occupational fatalities are analyzed to investigate the resulting characteristics of occupational injuries for NSE workers in the Taiwan construction industry. In order to explore policy impact between NSE and SE, this study focuses not only on general factors, but also on some specific factors related to safety management regulations on construction sites. In terms of the individual factors of the victims, all five attributes were significantly associated with *whether the worker is non-standard worker*. Except for *whether the worker is in the possession of labor insurance status*, the remaining four attributes were significantly associated with the NSE types. In terms of the project factors, except for *location of the project site*, the remaining three project factors were significantly associated with *whether the worker is non-standard worker*. Except for *whether the project is public*, the remaining three project factors were significantly associated with NSE types. In terms of accident factors, all four attributes were significantly associated with *whether the worker is non-standard worker*. Except for accident type, the remaining three accident factors were significantly associated with the NSE types. In terms of management factors, except for *whether the original enterprise established a consultative organization* and *whether the original enterprise notified the contractor of potential hazards*, the remaining seven attributes were significantly associated with *whether the worker is non-standard worker*. All nine management factors are significantly associated with the NSE type.

The results of the analysis show that the hazard characteristics of NSE workers are clearly different from SE workers. Considering the individual factors, the NSE occupational injury rate for older workers (over 60 years old) is higher in the construction industry, especially for self-employed workers taking on technical work. Considering the project factors, NSE workers are more expected to suffer occupational fatalities in the small-scale, non-public, and repair projects. Occupational injuries involving self-employed and temporary agency workers are clearly regionally concentrated. Considering the accident factors, temporary agency workers involved in occupational injuries are most engaged in non-technical work and movement for worker motion with their unfamiliarity with the worksite. Considering the management factors, most enterprises did not perform safety management on construction sites for occupational injuries involving NSE workers, especially for self-employed workers. Obviously, NSE workers face inferior job security and protection. Compared to other NSE workers, self-employed workers receive less security and protection. The results can be used to establish effective occupational safety management policies and programs more efficiently. However, the readers must note that, the limitation of databases may diminish the applicability of the presented results.

Conflicts of interest

The authors declare no potential conflicts of interest with respect to the research, authorship and publication of this article.

Acknowledgements

This work was funded by a grant from the Ministry of Science and Technology in Taiwan (MOST 108-2635-E-163-001 –).

References

Alfers, L., & Rogan, M. (2015). Health risks and informal employment in South Africa: Does formality protect health? *International Journal of Occupational and Environmental Health, 21*(3), 207–215.
Aleksynska, M. (2018). Temporary employment, work quality, and job satisfaction. *Journal of Comparative Economics, 46*(3), 722–735.
Arestis, P., Ferreiro, J., & Gómez, C. (2020). Quality of employment and employment protection. Effects of employment protection on temporary and permanent employment. *Structural Change and Economic Dynamics, 53*, 180–188.
Aronsson, G. (1999). Contingent workers and health and safety. *Work, Employment & Society, 13*(3), 439–459.
Basu, N., Ayelo, P. A., Djogbénou, L. S., Kedoté, M., et al. (2016). Occupational and environmental health risks associated with informal sector activities-selected case studies from West Africa. *New Solutions: Journal of Environmental and Occupational Health Policy, 26*(2), 253–270.
Benach, J., Vives, A., Amable, M., Vanroelen, C., Tarafa, G., & Muntaner, C. (2014). Precarious employment: Understanding an emerging social determinant of health. *Annual Review of Public Health, 35*, 229–35253.
Boyd, D. D., Scharf, M., & Cross, D. (2021). A comparison of general aviation accidents involving airline pilots and instrument-rated private pilots. *Journal of Safety Research, 76*, 127–134.
Bunner, J., Prem, R., & Korunka, C. (2021). Perceived organizational support and perceived safety climate from the perspective of safety professionals: Testing reciprocal causality using a cross-lagged panel design. *Journal of Safety Research, 78*, 1–8.
Benavides, F. G., Benach, J., Muntaner, C., Delclos, G. L., Catot, N., & Amable, M. (2006). Associations between temporary employment and occupational injury: What are the mechanisms? *Occupational and Environmental Medicine, 63*(6), 416–421.
Cummings, K. J., & Kreiss, K. (2008). Contingent workers and contingent health: Risks of a modern economy. *JAMA, 299*(4), 448–450.
Econie, A., & Dougherty, M. L. (2019). Contingent work in the US recycling industry: Permatemps and precarious green jobs. *Geoforum, 99*, 132–141.
Eijkelenboom, A. M., Kim, D. H., & Bluyssen, P. M. (2020). First results of self-reported health and comfort of staff in outpatient areas of hospitals in the Netherlands. *Building and Environment, 177* 106871.
Grunberg, L., Moore, S., & Greenberg, E. (1996). The relationship of employee ownership and participation to workplace safety. *Economic and Industrial Democracy, 17*(2), 221–241.
Kawachi, I. (2008). Globalization and workers' health. *Industrial Health, 46*(5), 421–423.

Kwon, M., Remøy, H., Dobbelsteen, A., & Knaack, U. (2019). Personal control and environmental user satisfaction in office buildings: Results of case studies in the Netherlands. *Building and Environment, 149*, 428–435.
Lilla, M., & Staffolani, S. (2012). Young entrants, temporary jobs and career opportunities: Short-term perspectives of young Italian workers. *Rivista di statistica Ufficiale, 14*(1), 50–60.
Mai, Q. D., Jacobs, A. W., & Schieman, S. (2019). Precarious sleep? Nonstandard work, gender, and sleep disturbance in 31 European countries. *Social Science & Medicine, 237* 112424.
Martinez, G., Cuyper, N. D., & Witte, H. D. (2010). Review of temporary employment literature: Perspectives for research and development in Latin America. *Psykhe, 19*, 61–73.
Pierce, O., Larson, J., & Grabell, M., 2013. How We Calculated Injury Rates for Temp and Non-temp Workers. Retrieved from <https://www.propublica.org/nerds/item/how-we-calculated-injury-rates-for-temp-and-non-temp-workers>.
Probst, T. M. (2002). Layoffs and tradeoffs: Production, quality, and safety demands under the threat of job loss. *Journal of Occupational Health Psychology, 7*(3), 211–220.
Størseth, F. (2006). Changes at work and employee reactions: Organizational elements, job insecurity, and short-term stress as predictors for employee health and safety. *Scandinavian Journal of Psychology, 47*(6), 541–550.
Probst, T. M., & Brubaker, T. L. (2001). The effects of job insecurity on employee safety outcomes: Cross-sectional and longitudinal explorations. *Journal of Occupational Health Psychology, 6*(2), 139–159.
Probst, T. M., Petitta, L., Barbaranelli, C., & Lavaysse, L. M. (2018). Moderating effects of contingent work on the relationship between job insecurity and employee safety. *Safety Science, 106*, 285–293.
Rousseau, D. M., & Libuser, C. (1997). Contingent workers in high risk environments. *California Management Review, 39*(2), 103–123.
Ruiz, M. E., Vives, A., Martínez-Solanas, È., Julià, M., & Benach, J. (2017). How does informal employment impact population health? Lessons from the Chilean employment conditions survey. *Safety Science, 100*, 57–65.
Sakurai, K., Nakata, A., Ikeda, T., Otsuka, Y., & Kawahito, J. (2013). How do employment types and job stressors relate to occupational injury? A cross-sectional Investigation of employees in Japan. *Public Health, 127*, 1012–1020.
Schaufeli, W. B. (2016). Job insecurity research is still alive and kicking twenty years later: A commentary. *Australian Psychologist, 51*(1), 32–35.
Thébaud-Mony, A. (1999). Contracting and subcontracting by the French nuclear power industry. *International Journal of Occupational and Environmental Health, 5*(4), 296–299.

Liao received her B.S. degree in Civil Engineering from National Central University, Jhongli, Taiwan, the M.S. degree in Construction Engineering from National Taiwan University of Science and Technology, Taipei, Taiwan, and the Ph.D. degree in Architecture from National Taiwan University of Science and Technology, Taipei, Taiwan. She is an associate professor at China University of Technology in Taiwan.

Chiang received his B.S. degree in Civil Engineering from National Central University, Jhongli, Taiwan, the M.S. degree in Civil Engineering from National Taiwan University, Taipei, Taiwan.