BMJ Health & Care Informatics

Use of informatics to characterise the exposome of COVID-19

Fernando Martin-Sanchez,[1] Miguel Atienza-Maderuelo,[2] Guillermo Lopez-Campos,[3] Paloma Collado[4]

To cite: Martin-Sanchez F, Atienza-Maderuelo M, Lopez-Campos G, *et al*. Use of informatics to characterise the exposome of COVID-19. *BMJ Health Care Inform* 2021;**28**:e100371. doi:10.1136/bmjhci-2021-100371

FM-S and MA-M are joint first authors.

Received 08 June 2021
Accepted 26 July 2021

[1]National School of Public Health, Carlos III Health Institute, Madrid, Spain
[2]IMIENS, UNED, Madrid, Spain
[3]Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, UK
[4]Psychobiology, UNED, Madrid, Spain

Correspondence to
Dr Fernando Martin-Sanchez; fmartin@isciii.es

THE IMPORTANCE OF EXPOSOME IN PRECISION MEDICINE

Exposome refers to the set of external exposures that affect an individual's health from conception to death.[1] This term includes all non-genetic risk factors which, interacting with the genetic endowment (genome), produce a state of health or disease (phenome). Considering these non-genetic data in precision medicine contributes to improving clinical decision making and biomedical research. It is still too frequent to find in the literature the description of research projects on certain diseases that only refer to the study of the genotype and the phenotype (https://www.icpermed.eu/en/icpermed-medicine.php). The same can be observed in database resources (https://www.ncbi.nlm.nih.gov/gap/). This indicates that either the exposome is neglected or the phenome is mistakenly considered to include the exposome (eg, smoking).

INTRODUCING THE CONCEPT OF EXPOTYPE

There is a scarcity of accepted standards to represent human exposure data. Therefore, representing individuals' exposome data constitutes a challenge for biomedical informatics, considering that it represents a source of big data that changes throughout time and space.[2] To overcome this, the concept of expotype emerges as a fundamental term to facilitate a holistic approach to health.

Introducing the concept of *expotype*[3] is a non-trivial issue, and so it is presented here by analogy with the terms *genotype* and *phenotype*. As the genome represents the complete DNA of a human organism, including all its genes, the term genotype refers to the genetic information of one specific individual. Likewise, the concept of phenome represents all those features presented by a human organism, and the phenotype is defined as the structural and/or functional characteristics that can be observed about an individual, produced by the interaction between its genotype and its context. Finally, by analogy, the *expotype would represent a specific set of exposures accumulated by an individual during a certain time/space window.* In this sense, to understand how diseases are developed, it is mandatory to assess how the environmental risk factors (expotypes) interact with the genomic information of an individual (genotype), generating specific phenotypes.[4] Collecting standardised data about individual genotypes, expotypes and phenotypes in research repositories offers the possibility of conducting integrative studies about the pathophysiology of complex diseases.

DESCRIPTION OF A COVID-19 EXPOTYPE

Multiple knowledge domains need to be integrated to efficiently address current health challenges, such as the COVID-19 pandemic. SARS-CoV-2 has put great pressure on health systems, and it is an appropriate example of how external exposure plays a key role in affecting human health.

Building a COVID-19 expotype requires, besides the virus biology, considering all those factors related to its transmission and development, gathering individual data for health and research purposes. In general, environmental factors such as temperature and humidity play a key role in viral transmission. Recent studies show that ecological factors (policy, health behaviours, physical environment and clinical care) are associated with COVID-19 case fatality rate.[5] Furthermore, exposure to ultraviolet radiation favours the synthesis of vitamin D in the body, which plays an important role in the immune defence against viral and bacterial infections.[6] Additionally, a relevant correlation has been found between exposure to particulate matter (pollution) and COVID-19 death rates.[7] On the one hand, pollution actively contributes to comorbidity, worsening the health of those exposed and therefore their prognosis

fighting the virus. On the other hand, particulate matter may also be a possible virus carrier. Other environmental factors such as noise pollution and tobacco use have also been related to COVID-19.

SARS-CoV-2 effects are also related to social, cultural and behavioural factors. Measuring time spent by individuals outdoors and indoors is essential, given that most of the infections are related to extended human contact periods, mostly occurring in locations such as restaurants or gyms. Safety measures such as using masks or maintaining social distance are also behavioural risk factors to be analysed in conjunction with the expotype data.

Individuals with obesity show greater likelihood of requiring special care and increased risk of death.[8] Moreover, research has shown how food consumption influences brain neurotransmitter systems, affecting mood. Bidirectionally, mood can also influence food choices, leading to repetitive cycles of overeating, weight gain and depression. Hence, data related to eating habits also form part of COVID-19's expotype structure,[9] whereas psychological status would be part of the phenotype.

REPRESENTING EXPOTYPES WITH THE ISO-EN13606 MODEL (ARCHETYPES)

Today, digital health methods allow us to track individual exposure data. However, a standard model to process these data is still lacking. We present here a mechanism for data representation based on the concept of expotype, meaning specific exposures of an individual during a certain time/space window, and its representation using *archetypes*, based on the CEN/ISO-EN13606 data model international standard.[10] Such standard data models will allow formalising and managing knowledge as it evolves, easily adapting it to new research and development needs.

Twitter Fernando Martin-Sanchez @fermarsan

Contributors FM-S and GL-C conceived of the presented idea. FM-S and MA-M developed the data models. PC provided the biomedical framework. MA-M wrote the manuscript with support from FM-S, PC and GL-C. All authors provided critical feedback and contributed to the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; externally peer reviewed.

REFERENCES

1 Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2005;14:1847–50.
2 Cui Y, Balshaw DM, Kwok RK, *et al*. The Exposome: embracing the complexity for discovery in environmental health. *Environ Health Perspect* 2016;124:A137–40.
3 Martin-Sanchez FJ, Lopez-Campos GH. The new role of biomedical informatics in the age of digital medicine. *Methods Inf Med* 2016;55:392–402.
4 Martin-Sanchez F, Bellazzi R, Casella V, *et al*. Progress in characterizing the human Exposome: a key step for precision medicine. *Yearb Med Inform* 2020;29:115–20.
5 Kathe NJ, Wani RJ. Determinants of COVID-19 case fatality rate in the United States: spatial analysis over one year of the pandemic. *J Health Econ Outcomes Res* 2021;8:51–62.
6 Borella E, Nesher G, Israeli E, *et al*. Vitamin D: a new anti-infective agent? *Ann N Y Acad Sci* 2014;1317:76–83.
7 Sciomer S, Moscucci F, Magrì D, *et al*. SARS-CoV-2 spread in northern Italy: what about the pollution role? *Environ Monit Assess* 2020;192:325.
8 Dietz W, Santos-Burgoa C. Obesity and its implications for COVID-19 mortality. *Obesity* 2020;28:1005.
9 Colomina M-T, Sánchez-Santed F, Conejo NM, *et al*. The Psychoexposome: a holistic perspective beyond health and disease. *Psicothema* 2018;30:5–7.
10 Tapuria A, Kalra D, Curcin V. Feasibility of using en 13606 clinical archetypes for defining Computable phenotypes. *Stud Health Technol Inform* 2020;270:228–32.

BMJ Health & Care Informatics

Turbulence health systems: engineering a rapidly adaptive health system for times of crisis

Enrico Coiera , Jeffrey Braithwaite

To cite: Coiera E, Braithwaite J. Turbulence health systems: engineering a rapidly adaptive health system for times of crisis. *BMJ Health Care Inform* 2021;**28**:e100363. doi:10.1136/bmjhci-2021-100363

Received 08 June 2021
Accepted 03 August 2021

Australian Institute of Health Innovation, Macquarie University, Sydney, New South Wales, Australia

Correspondence to
Professor Enrico Coiera;
enrico.coiera@mq.edu.au

The coronavirus epidemic has provided us a teachable moment in how to run a healthcare system. Dealing with this pandemic has required near-wartime levels of health system reconfiguration and reinvention—often with heroic effort and unprecedented commitment of resources. By the time this pandemic is over, we will have learnt much about how to manage the next one.[1 2] Innovations in contact tracing, vaccine engineering, accelerated research translation through open science and in public health interventions will all make managing a future pandemic easier.

Our problem is that the next crisis will not be like the last one. Indeed, it is not hard to imagine a future where the health system suffers shock after shock from climate change-triggered mass events like floods and fires, to new pandemics, and social and geopolitical unrest. If it took a heroic response to get us through COVID-19, how could a health system survive many such challenges and in short order? Even resilient systems eventually fracture under repeated stress.

Somehow, we will need to engineer a health system that can bend and stretch its resources and capabilities, quickly and repeatedly. Like a chameleon, it will need to reconfigure to different contexts, never having a 'true' shape. It will be liminal, defined by the potential behaviours it could exhibit. We might call this a 'turbulence system'—one purpose-designed to operate across a wide variety of conditions and to efficiently and effectively reconfigure itself to meet different demands.[3] Turbulence systems would recognise threatening system shocks as early as possible to provide time to reconfigure and will deploy a range of strategies to first prepare for, and then ride out, a system-wide shock.

How will a turbulence system in healthcare function? If it were an immune system, it would rely on antibodies to respond to those pathogens it has seen before. When faced with a novel threat, non-specific immune responses hold the fort until such antibodies are developed. It is the same for a turbulence system. A new pandemic will largely be dealt with using the infrastructure and know-how built for the previous one. Other shocks, however, may require new ideas and new tools.

This dynamic balancing between precommitted resources using pre-emptively designed strategies with just-in-time strategies to handle novel situations is a familiar one. In health informatics, the tools we use are drawn from a continuum of general-purpose technologies which focus on communication, through to highly designed digital systems crafted to support well-understood tasks.[4] When at its best, our day-to-day work blends the use of tools like electronic records and decision support systems to deal with the expected, and a range of generalist tools like voice and text communication to deal with novelty, complexity and inefficiency.

So, what generalist tools and strategies might a turbulence-ready health system adopt? The first tranche are tools that aggregate and analyse data at every stage of a system shock, from early warning to finale. If we are to make decisions under uncertainty, then access to high-quality data that can convert the unknown to the known will be critical.[5] Data sharing is essential—a feature of the early stages of the COVID-19 pandemic was a deep asymmetry of distribution and quality of pandemic data when some nations had few cases and others struggled to cope. The many technical, privacy, ethical and commercial barriers to capturing and sharing clinical data sets remain formidable, whether at the level of health service, region, nation or globally. If we are to shock-proof our national health systems, we must accelerate our data capture and sharing efforts.

Managing data synthesis and interpretation is also challenging. The COVID-19 Living Evidence Project had indexed about 160 000

peer-reviewed publications by the end of May 2021.[6] Manual analysis of such a volume of science, of varying provenance and quality, with much duplication and reuse of the same patient data, has stretched our sense-making capacity. Automated tools will eventually do the bulk of this evidence synthesis using methods from artificial intelligence. Systematic reviews will be automated[7 8]; clinical trial and observational data will be analysed on the fly,[9] answering both patient-specific questions as well as issues of policy, public health and system configuration, often in real time. We will synthesise quantitative and qualitative data, and this data fusion will have to accommodate a variety of data types including measurements, text, speech and images.[10]

While we cannot predict all the questions that will be asked of us, we must be prepared to quickly find the answers. This task cannot be solved by automation alone, especially when system shocks present novel challenges. Coalitions of humans and machines will come together physically or virtually at interagency 'fusion centres' to overcome compartmentalisation.[11] This will permit us to synthesise different streams of emerging evidence, ask critical questions, simulate possible scenarios, formulate strategies, and deploy people and resources. Achieving this will require an integrated and sophisticated communication infrastructure. The command structure that governs this will have aspects of top-down and bottom-up control, but will most likely need a flexible middle-out approach.[12]

Highly specific strategies will complement these general ones. Rapid system reconfiguration and surge capacity will require a rapid response infrastructure to be in place. The emergence of mRNA technologies and the capacity to rapidly design and produce novel vaccines or other molecules may be among the greatest rapid-response gifts of the COVID-19 pandemic.[13] Machine learning and other artificial intelligence methods will help optimise resource allocation across health services, predicting patient demand and identifying spare capacity.[14] Traditionally backroom systems for supply-chain management become front-line weapons in a crisis. The COVID-19 pandemic is as much a story about early failures to manufacture, secure and deploy appropriate personal protective equipment and ventilators, as it is of panic buying of foodstuffs and consumer goods. We have much to learn from companies like Amazon on how to manage supply lines and shift supplies to where they are most needed as early as possible.[15] We also need information systems to quickly identify functionally equivalent products or services that are substitutable or reconfigurable in a time of crisis, from finding spare hospital beds that can be used for intensive care patients to computer chips to build urgently needed devices. We also need to know how to make technologies interoperate 'just enough' to work together, even when they are built to different standards.

This may all seem too hard and not worth the price. Engineering a turbulence-ready health system will push our understanding of engineering complex adaptive systems and require new investment into an already resource-stretched sector. As memories of COVID-19 begin to fade, the desire to return to 'normal' will be strong. However, we do not have that luxury. Climate change looms over us all. New pandemics are certain. Global financial and geopolitical uncertainty is increasing. Old political allegiances are fracturing. If we wish to flourish or even just make do in this emerging reality, we will need to do more than create a learning health system that learns from the past. We must build a health system prepared to face that which cannot be foreseen.

Contributors EC conceived of and wrote the first draft, and JB provided critical feedback and revised the draft.

Funding This study was funded by National Health and Medical Research Council (Centre for Research Excellence in Digital Health).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

ORCID iD

Enrico Coiera http://orcid.org/0000-0002-6444-6584

REFERENCES

1 Braithwaite J, Tran Y, Ellis LA, *et al*. The 40 health systems, COVID-19 (40HS, C-19) study. *Int J Qual Health Care* 2021;33:mzaa113.
2 Durski KN, Osterholm M, Majumdar SS, *et al*. Shifting the paradigm: using disease outbreaks to build resilient health systems. *BMJ Glob Health* 2020;5:e002499.
3 Coiera E, Magrabi F, Sintchenko V. The centre for health informatics at the University of new South Wales-a clinical informatics research centre. *Yearbook of Medical Informatics* 2007;16:141–8.
4 Coiera E. When conversation is better than computation. *J Am Med Inform Assoc* 2000;7:277–86.
5 Basit MA, Lehmann CU, Medford RJ. Managing pandemics with health informatics: successes and challenges. *Yearb Med Inform* 2021. doi:10.1055/s-0041-1726478. [Epub ahead of print: 21 Apr 2021].
6 Project C-OA. Living evidence on COVID-19 2020. Available: https://ispmbern.github.io/COVID-19/living-review/
7 Tsafnat G, Dunn A, Glasziou P, *et al*. The automation of systematic reviews. *BMJ* 2013;346:f139.
8 van Dinter R, Tekinerdogan B, Catal C. Automation of systematic literature reviews: a systematic literature review. *Information and Software Technology* 2021;136:106589.
9 Gallego B, Walter SR, Day RO, *et al*. Bringing cohort studies to the bedside: framework for a 'green button' to support clinical decision-making. *J Comp Eff Res* 2015;4:191–7.
10 Rapport F, Braithwaite J. Are we on the cusp of a fourth research paradigm? predicting the future for a new approach to methods-use in medical and health services research. *BMC Med Res Methodol* 2018;18:131.
11 Catano V, Gauger J. *Information fusion: intelligence centers and intelligence analysis. Information sharing in military operations*. Springer, 2017: 17–34.
12 Coiera E. Building a national health it system from the middle out. *J Am Med Inform Assoc* 2009;16:271–3.
13 Meng C, Chen Z, Li G, *et al*. Nanoplatforms for mRNA therapeutics. *Advanced Therapeutics* 2021;4:2000099.
14 Strömblad CT, Baxter-King RG, Meisami A, *et al*. Effect of a predictive model on planned surgical duration accuracy, patient wait time, and use of presurgical resources: a randomized clinical trial. *JAMA Surg* 2021;156:315–21.
15 National Academies of Sciences, Engineering, and Medicine. *Crisis standards of care: ten years of successes and challenges: proceedings of a workshop*. Washington, DC: The National Academies Press, 2021. https://doi.org/10.17226/25767

BMJ Health & Care Informatics

A step-by-step guide to peer review: a template for patients and novice reviewers

Liz Salmi [1], Charlotte Blease [2,3]

To cite: Salmi L, Blease C. A step-by-step guide to peer review: a template for patients and novice reviewers. *BMJ Health Care Inform* 2021;**28**:e100392. doi:10.1136/bmjhci-2021-100392

Received 23 April 2021
Accepted 29 July 2021

[1]General Medicine and Primary Care, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[2]Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[3]Harvard Medical School

Correspondence to
Ms Liz Salmi;
lsalmi@bidmc.harvard.edu

While relatively novel, patient peer review has the potential to change the healthcare publishing paradigm. It can do this by helping researchers enlarge the pool of people who are welcome to read, understand and participate in healthcare research. Academic journals who are early adopters of patient peer review have already committed to placing a priority on using person-centred language in publicly available abstracts and focusing on translational and practical research.

A wide body of literature has shown that including people with lived experiences in a truly meaningful way can improve the quality and efficiency of health research. Traditionally considered only as 'subjects' of research, over the last 10–15 years, patients and care partners have increasingly been invited to contribute to the design and conduct of studies. Established institutions are increasingly recognising the distinctive expertise patients possess—many patients have acquired deep insights about their conditions, symptoms, medical treatments and quality of healthcare delivery. Among some funders, including the views of patients is now a requirement to ensure research proposals are meaningful to persons with the lived experience of illness. Further illustrating these developments, patients are now involved in reviewing and making recommendations as part of funding institutions, setting research agendas and priorities, being funded for and leading their own research and leading or coauthoring scholarly publications, and are now participating in the peer review process for academic journals.[1–5] Patients offer an outsider's perspective within mainstream healthcare: they have fewer institutional, professional or social allegiances and conflicts of interest—factors recognised as compromising the quality of research. Patient involvement is essential to move away from rhetorical commitments to embrace a truly patient-centred healthcare ecosystem where everyone has a place at the table.

As people with lived health experiences climb a ladder of engagement in patient–researcher partnerships, they may be asked to act as peer reviewers of academic manuscripts. However, many of these individuals do not hold professional training in medicine, healthcare or science and have never encountered the peer review process. Little guidance exists for patients and care partners tasked with reviewing and providing input on manuscripts in search of publication.

In conversation, however, even experienced researchers confess that learning how to peer review is part of a *hidden curriculum* in academia—a skill outlined by no formal means but rather learnt by mimicry.[6] As such, as they learn the process, novices may pick up bad habits. In the case of peer review, learning is the result of reading large numbers of academic papers, occasional conversations with mentors or commonly "trial by fire" experienced via reviewer comments to their own submissions. Patient reviewers are rarely exposed to these experiences and can be at a loss for where to begin. As a result, some may forgo opportunities to provide valuable and highly insightful feedback on research publications. Although some journals are highly specific about how reviewers should structure their feedback, many publications—including top-tier medical journals—assume that all reviewers will know how to construct responses. Only a few forward-thinking journals actively seeking peer review from people with lived health experiences currently point to review tips designed for experienced professionals.[7]

As people with lived health experiences are increasingly invited to participate in peer review, it is essential that they be supported in this process. The peer review template for patients and novice reviewers (table 1) is a

Table 1 Peer review template for patients and other novice reviewers

Name of journal	Insert the name of the journal here
The journal's area of focus	*Type the area of focus here (eg, oncology and health literacy)*
Title of manuscript	*Insert the title of the manuscript you are reviewing*
Link to review website	*Paste here a link to the journal's online form where you need to submit your review*
GENERAL TIPS FOR A GOOD REVIEW	
► Be constructive. Think about advice or recommendations you can make to improve the paper ► Keep the review short; 2–3 paragraphs in total are enough ► *Add tips you learn here*	
WRITING PROMPTS	
1. Summarise what the paper is about in two to three sentences	
Example: "This is an interview study of 53 people living with metastatic cancer about their perspective on physicians' use of the computer during follow-up visits. The findings are similar to other studies the authors cite (basically, most patients don't seem to mind when doctors are using the computer). The study question was developed in partnership with the hospital's patient–family advisory council."	
Write your summary here	
2. Summarise your opinion of the manuscript and what the authors may need to address	
Example: "What makes this paper interesting is that it was conducted at a community hospital and not at a major cancer centre. Assuming the oncology clinic also serves people with many different types of cancer, my main suggestion is to pare down the paper and make THAT the thrust of the findings: for example, 53 patients' attitudes towards computers in the examination room at community hospitals are similar to those of patients who receive care at major cancer centres. Beyond consulting the hospital PFAC at the outset, the authors did not mention working with patients on any other aspects of the study—please elaborate more on how else patient advisors may have been involved."	
Write your summary here	
3. Major comments: provide feedback on major aspects of the paper	
Comments here will depend on the paper, and patient reviewers should feel comfortable knowing their most important insights might be reflective of their lived experiences—you are not expected to comment on methods or statistics. Things to think about here may include the following: Did the authors give enough background to justify why the research question was important? Were the authors clear about their objectives? Did you notice any problems with the results? Did the authors detail the strengths and limitations of the study? Were the conclusions supported by the research? Was anything missing from the paper? Were the figures and/or tables clearly laid out? Do you have any suggestions on how to make the paper more useful for patient readers?	
Write your comments here	
4. Provide feedback on the quality of the writing	
Think about the following: Was the writing clear? Was the writing grammatically correct? Was the referencing complete? Detail any minor comments such as stylistic issues, missing references, typos or queries you think the reviewers need to address *Example:* "The tone and writing style of this manuscript are chaotic; I suggest one of the authors review and edit it one more time so it reads like it is coming from one voice."	
Give your writing feedback here	
5. Make a specific recommendation to the journal's editor	
Options may include the following: ► Accept for publication with minor revisions. ► Accept for publication with major revisions. ► Reject for publication. Be clear whether you recommend 'reject' or 'no revisions'. *Example:* "To editor: The purpose and implementation of the study are incomprehensible. It's not just the writing there is no discernible study design."	
Write your recommendation and justification for that recommendation here	
6. Share a statement of limitations with the editor and/or authors (*optional*)	
If there is a technical aspect of the manuscript in which you felt unprepared/unqualified to comment on, it is OK to be candid with the journal editor and/or authors. Adding a statement like this is uncommon, but such feedback is important for fair and honest review *Example:* "To editor: Aspects of this manuscript I am unable to comment on include statistical analyses and medical ethics."	
Comment on your own review limitations here	

series of steps designed to create a workflow for the main components of peer review. A structured workflow can help a reviewer organise their thoughts and create space to engage in critical thinking. The template is a starting point for anyone new to peer review, and it should be modified, adapted and built on for individual preferences and unique journal requirements. Peer reviews are commonly submitted via website portals, which vary widely in design and functionality; as such, reviewers are encouraged to decide how to best use the template on a case-by-case basis. Journals may require reviewers to copy and paste responses from the template into a journal website or upload a clean copy of the template as an attachment. *Note: If uploading the review as an attachment, remember to remove the template examples and writing prompts.*

It is important to point out that patient reviewers are not alone in facing challenges and a steep learning curve in performing peer review. Many health research agendas and, as a result, publications straddle disciplines, requiring peer reviewers with complementary expertise and training. Some experts may be highly equipped to critique particular aspects of research papers while unsuited to comment on other parts. Curiously, however, it

is seldom a requirement that invited peer reviewers admit their own limitations to comment on different dimensions of papers. Relatedly, while we do not suggest that all patient peer reviewers will be equipped to critique every aspect of submitted manuscripts—though some may be fully competent to do so—we suggest that candour about limitations of expertise would also benefit the broader research community.

As novice reviewers gain experience, they may find themselves solicited for a growing number of reviews, much like their more experienced counterparts or mentors.[8] Serving as a patient or care partner reviewer can be a rewarding form of advocacy and will be crucial to harnessing the feedback and expertise of persons with lived health experiences. As we move into a future where online searches for information are a ubiquitous first step in searching for answers to health-related questions, patient and novice reviewers may become the much-needed link between academia and the lay public.

Twitter Liz Salmi @TheLizArmy and Charlotte Blease @crblease

Acknowledgements LS thanks the experienced and novice reviewers who encouraged her to publish this template.

Contributors Both authors contributed substantially to the manuscript. LS conceived the idea and design and drafted the text. CB refined the idea and critically revised the text.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests The authors have read and understood the BMJ policy on declaration of interests and declare the following interests: LS is a member of the BMJ Patient Advisory Panel, serves as a BMJ patient reviewer and is an ad hoc patient reviewer for the Patient-Centered Outcomes Research Institute; CB is a Keane OpenNotes scholar; both LS and CB work on OpenNotes, a philanthropically funded research initiative focused on improving transparency in healthcare.

Patient consent for publication Not required.

Provenance and peer review Commissioned; externally peer reviewed.

ORCID iDs
Liz Salmi http://orcid.org/0000-0003-3798-7438
Charlotte Blease http://orcid.org/0000-0002-0205-1165

REFERENCES

1 Engage with Us. Patient-Centered outcomes research Institute (PCORI), 2019. Available: https://www.pcori.org/engagement/engage-us [Accessed 16 Apr 2021].
2 Coordinating Center for Clinical Trials - Patient Advocate Steering Committee [Internet]. National Cancer Institute. U.S. Department of Health and Human Services. National Institute of Health. Available: https://www.cancer.gov/about-nci/organization/ccct/steering-committees/patient-advocate [Accessed 16 Apr 2021].
3 Transforming Healthcare through Innovative and Impactful Research [Internet]. Peer Reviewed Cancer Research Program, Congressionally Directed Medical Research Programs. U.S. Department of Defense, 2021. Available: https://cdmrp.army.mil/prcrp/default [Accessed 16 Apr 2021].
4 Riggare S. Patient researchers - the missing link? *Nat Med* 2020;26:1507.
5 Richards T, Snow R, Schroter S. Logging The BMJ's "patient journey" [Internet]. The BMJ. British Medical Journal Publishing Group, 2015. Available: https://www.bmj.com/content/351/bmj.h4396 [Accessed 16 Apr 2021].
6 Mahood SC. Medical education: beware the hidden curriculum. *Can Fam Physician* 2011;57:983–5.
7 Guidance for BMJ Patient and Public Reviewers [Internet]. The BMJ. Available: https://www.bmj.com/about-bmj/resources-reviewers/guidance-patient-reviewers [Accessed 16 Apr 2021].
8 Rubenstein J, Salmi L. How to Cultivate a "Friendtorship" [Internet]. CLOSLER. Miller Coulson Academy of Clinical Excellence, Johns Hopkins Medicine, 2021. Available: https://closler.org/lifelong-learning-in-clinical-excellence/how-to-cultivate-a-friendtorship [Accessed 16 Apr 2021].

BMJ Health & Care Informatics

Review of study reporting guidelines for clinical studies using artificial intelligence in healthcare

Susan Cheng Shelmerdine ,[1] Owen J Arthurs,[1] Alastair Denniston,[2] Neil J Sebire [3]

To cite: Shelmerdine SC, Arthurs OJ, Denniston A, *et al*. Review of study reporting guidelines for clinical studies using artificial intelligence in healthcare. *BMJ Health Care Inform* 2021;**28**:e100385. doi:10.1136/bmjhci-2021-100385

Received 22 April 2021
Accepted 09 August 2021

[1]Radiology, Great Ormond Street Hospital NHS Foundation Trust, London, UK
[2]Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK
[3]Digital Research, Informatics and Virtual Environments Unit (DRIVE), London, UK

Correspondence to
Dr Susan Cheng Shelmerdine; susie.shelmerdine@gmail.com

ABSTRACT

High-quality research is essential in guiding evidence-based care, and should be reported in a way that is reproducible, transparent and where appropriate, provide sufficient detail for inclusion in future meta-analyses. Reporting guidelines for various study designs have been widely used for clinical (and preclinical) studies, consisting of checklists with a minimum set of points for inclusion. With the recent rise in volume of research using artificial intelligence (AI), additional factors need to be evaluated, which do not neatly conform to traditional reporting guidelines (eg, details relating to technical algorithm development). In this review, reporting guidelines are highlighted to promote awareness of essential content required for studies evaluating AI interventions in healthcare. These include published and in progress extensions to well-known reporting guidelines such as Standard Protocol Items: Recommendations for Interventional Trials-AI (study protocols), Consolidated Standards of Reporting Trials-AI (randomised controlled trials), Standards for Reporting of Diagnostic Accuracy Studies-AI (diagnostic accuracy studies) and Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis-AI (prediction model studies). Additionally there are a number of guidelines that consider AI for health interventions more generally (eg, Checklist for Artificial Intelligence in Medical Imaging (CLAIM), minimum information (MI)-CLAIM, MI for Medical AI Reporting) or address a specific element such as the 'learning curve' (Developmental and Exploratory Clinical Investigation of Decision-AI) . Economic evaluation of AI health interventions is not currently addressed, and may benefit from extension to an existing guideline. In the face of a rapid influx of studies of AI health interventions, reporting guidelines help ensure that investigators and those appraising studies consider both the well-recognised elements of good study design and reporting, while also adequately addressing new challenges posed by AI-specific elements.

INTRODUCTION

Recent, rapid developments in computational technologies and increased volumes of digital data for analysis have resulted in an unprecedented growth in research activities relating to artificial intelligence (AI), particularly within healthcare. This volume of work has even led to several high impact journals launching their own subjournals within the 'AI healthcare' field (eg, *Nature Machine Intelligence*,[1] *Lancet Digital Health*,[2] *Radiology: Artificial Intelligence*).[3] High-quality research should be accompanied by transparency, reproducibility and validity of techniques for adequate evaluation and translation into clinical practice. Standardised reporting guidelines help researchers define key components of their study, ensuring that relevant information is provided in the final publication.[4] Studies pertaining to algorithm development and clinical application of AI however, have brought unique challenges and added complexities in how such studies are reported, assessed and compared in relation to elements that are not conventionally prespecified in traditional reporting guidelines. This could lead to missing information and high risk of hidden bias. If these actual or potential limitations are not identified, then it may lead to tacit approval through publication which in turn may support premature adoption of new technologies.[5 6] Conversely well-designed, well-delivered studies that are poorly reported may be judged unfavourably due to being adjudged to have a high risk of bias, simply due to a lack of information.

Inadequacies of reporting of AI clinical studies are increasingly well-recognised. In 2019, a systematic review by Liu *et al*[7] reviewed over 20 500 articles, but found that fewer than 1% of these were sufficiently robust in their design and reporting allowing independent reviewers to have confidence in their claims. Similarly Nagendran *et al*[8] identified high levels of bias in the field. In another study,[9] it was reported that only 6% of over 500 eligible radiological-AI research publications performed any external validation of their models, and none used multicentre or prospective data collection. Similarly most studies using machine learning (ML) models

for medical diagnosis[10] did not have adequate detail on how these were evaluated nor sufficient detail for these to be reproduced. Inconsistencies in how ML models from electronic health records have also been reported, with details regarding race and ethnicity of participants omitted in 64% of studies, and only 12% of models being externally validated.[11]

In order to address these concerns, adapted research reporting guidelines based on the well-established EQUATOR Network (Enhancing the QUAlity and Transparency Of health Research)[12 13] and de novo recommendations by individual societies have been published, with a greater relevance for AI research. In this review, we highlight those that will cover the majority of healthcare focused AI-related studies, and explain how they differ to the well-known guidance for non-AI related clinical work. Our intention is to raise awareness of how such studies should be structured, thereby improving the quality of future submissions and providing a helpful aid for researchers, peer reviewers and editors.

In compiling a detailed, yet relevant list of study guidelines, we reviewed the EQUATOR network[13] website for those containing the terms AI, ML or deep learning. A separate search was also conducted using Medline, Scopus and Google Scholar databases for publications using the same search terms with the addition of 'reporting guideline', 'checklist' or 'template'. Opinion pieces were excluded. Articles were included where the description of the recommendations were provided, and published at time of the search (March 2021).

TYPES OF RESEARCH REPORTING GUIDELINES

An ideal reporting guideline should be a clear, structured tool with a minimum list of key information to include within a published scientific manuscript. The EQUATOR Network[13] is the international 'standard bearer' for reporting guidelines, committed to improving 'the reliability and value of published health research literature by promoting transparent and accurate reporting and wider use of robust reporting guidelines'. Since the landmark publication of Consolidated Standards of Reporting Trials (CONSORT),[14] the network has overseen the development and publication of a number of guidelines that address other types of study design (eg, diagnostic accuracy studies). The EQUATOR guidelines are centrally registered (available via a core library) which ensures adherence to robust methodology of development and avoids redundancy of parallel initiatives to address the same issue. Importantly these guidelines are not medical specialty specific but are focused on the type of study, which helps ensure that there is a consistent approach and quality for addressing the same study design. It is recognised that certain specific scenarios may require specific extensions to these guidelines. For example, the increasing recognition of the importance of patient-reported outcomes (PROs) has led to the development of Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT-PRO)[15] and CONSORT-PRO.[16] In a similar way, the specific attributes of AI as an intervention, has led to a number of AI extensions, both published and in process, which build on the robust methodology of the original EQUATOR guidelines, while ensuring AI-specific elements are also addressed.

In parallel to the work of the EQUATOR network, a number of experts and institutions have developed their own recommendations for good practice and reporting. In contrast, these start with the intervention (ie, AI) rather than the study type (ie, randomised controlled trial (RCT)), and therefore, cover essentially the same territory. They vary in depth, and there can be differences in nuance depending on their primary purpose. For example some have originated from the need to support reviewers and editorial staff ('is this complete and is it good enough?'), whereas others are addressing at building a shared understanding of appropriate design and delivery ('this is what good looks like').

Given the number of different reporting guidelines in this area, there is value in setting them in context to help support users in understanding which is most appropriate for a particular setting (table 1). Ultimately the most important elements of a high-quality study are contained within the methodology of the study design itself and not within the intervention. It is these elements that help minimise the major biases that all studies must address. In line with leading journals, we would, therefore, recommend starting with the guideline that addresses that particular study design (eg, CONSORT[14] for an RCT). If an AI extension is already in existence for that study type then these are clearly appropriate for that study (eg, CONSORT-AI).[17–19] If no such -AI extension exists then we recommend using the appropriate EQUATOR guideline (eg, Standards for Reporting of Diagnostic Accuracy Studies (STARD)[20] for diagnostic accuracy studies), but supplementing with AI-specific elements recommended in other guidelines (eg, SPIRIT-AI,[21–23] CONSORT-AI[17–19] or the non-EQUATOR guidelines described below). Indeed all the guidelines considered here contain valuable insights into the specific challenges of AI studies, and are recommended reading into good practice for design and reporting.

EQUATOR NETWORK GUIDELINES

Clinical trials protocols

The quality of a study and the trustworthiness of its findings, starts at the design phase. The study protocol should contain all elements of the study design, sufficient for independent groups to carry out the study and expect replicability. Prepublication of the study protocol, helps avoid biases such as post-hoc assignment of the primary outcome in which the triallist can 'cherry pick' one of a number of outcomes that point in the desired direction.

Guidance for recommended items to include in a trial protocol are provided by the SPIRIT Statement (latest version published in 2013),[24] which has been recently

Table 1 Summary of reporting guidelines for common study types used in radiological research, and their corresponding guideline extensions where these involve artificial intelligence

Study design	Reporting guideline	Latest version	AI-related extension	Date of AI-extension published
Clinical Trial Protocol	SPIRIT	2013	SPIRIT-AI	September 2020
Diagnostic Accuracy Studies	STARD	2015	STARD-AI	Expected 2021
			CLAIM	March 2020
			MINIMAR	June 2020
Prediction models for diagnostic or prognostication purposes	TRIPOD	2015	TRIPOD –AI/ML	Expected 2021
	PROBAST	2019	PROBAST-ML	Expected 2021
Randomised Controlled Trials (Interventional Study Design)	CONSORT	2010	CONSORT-AI	September 2020
Systematic reviews and meta-analyses	PRISMA PRISMA-DTA	2009 2018	None planned or announced	
Critical appraisal and data extraction of publications relating to prediction models	CHARMS	2014	Applicable to machine learning	
Evaluation of human factors in early algorithm deployment	Not applicable		DECIDE-AI	Expected 2021/2022

AI, artificial intelligence; CHARMS, Checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies; CLAIM, Checklist for Artificial Intelligence in Medical Imaging; CONSORT, Consolidated Standards of Reporting Trials; DECIDE-AI, Developmental and Exploratory Clinical Investigation of Decision-support systems driven by Artificial Intelligence; DTA, Diagnostic Trials of Accuracy; MINIMAR, Minimum Information for Medical AI Reporting; ML, machine learning; PRISMA, Preferred Reporting Items for Systematic Review and Meta-analysis; PROBAST, Prediction model Risk Of Bias Assessment Tool; SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials; STARD, Standards for Reporting of Diagnostic Accuracy Studies; TRIPOD, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.

adapted for trials with an AI-related focus, termed the 'SPIRIT-AI' guideline.[21–23] This adaptation includes an additional 15 items (12 extensions, 3 elaborations) to the existing 33-item SPIRIT 2013 guideline. The key differences are outlined in table 2, mostly focused on the methodology of the trial, (accounting for eight extensions, one elaboration) with emphasis on inclusion/exclusion of data and participants, dealing with poor quality data and how the AI intervention will be applied to and benefit clinical practice.

Clinical trials reports

While most AI studies are currently at early-phase validation stages, those evaluating the use of 'AI-interventions' in real world setting are fast emerging, and will become of increasing importance, since these are required for real-world clinical benefit demonstration. RCTs are the exemplar study design in providing a robust evidence basis for efficacy and safety of a given intervention, with the CONSORT statement, 2010 version[14] providing a 25-item checklist for the minimum reporting content in such studies. An adapted version, entitled the 'CONSORT-AI' extension[17–19] was published in September 2020 for 'AI intervention' studies. This includes an additional 14 items (11 extensions, 3 elaborations) to the existing CONSORT 2010 statement, the majority of which (8 extensions, 1 elaboration) relate to the study participants and details of the 'AI intervention' being evaluated, which are similar to those additions already described in the SPIRIT-AI extension. Specific key differences in the new guideline are outlined in table 3. Although not specific for AI interventions, some aspects of the checklist Template for Intervention Description and Replication, 2014[25] may be a helpful addition when reporting details of the interventional elements of a study (ie, as an extension of item 5 of the CONSORT 2010 statement or as item 11 of the SPIRIT 2013 statement). These include details regarding any modifications of the intervention during a study, including how and why certain aspects were personalised or adapted. There are currently no publicly proposed plans to publish an 'AI' extension to this guideline to the best of our knowledge.

Diagnostic accuracy studies

The STARD statement, 2015 version[20] is the most widely accepted reporting standard for diagnostic accuracy studies. A steering group has been established to devise an AI-specific extension to the latest version of the 30-item STARD statement (called the STARD-AI extension.[26] At the time of writing this is undergoing an international consensus survey among leaders in the AI field for suggested adaptations and pending publication.

Prediction models

Extensions to reporting guidelines describing prediction models that use ML have been announced, and are anticipated for publication soon. These include adapted versions of the 'Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis' (TRIPOD), 2015 version,[27] which will be entitled

Table 2 Additional items proposed for studies relating to AI intervention clinical protocols within the SPIRIT-AI statement (in addition to the SPIRIT 2013 statement)

Section	Item no	SPIRIT 2013 item	Amendment	SPIRIT-AI item
Administrative information				
Title	1	Descriptive title identifying the study design, population, interventions and if applicable, trial acronym	Elaboration	Indicate that the intervention involves artificial intelligence/machine learning and specify the type of model.
			Elaboration	Specify the intended use of the AI intervention.
Introduction				
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	Extension	Explain the intended use of the AI intervention in the context of the clinical pathway, including its purpose and its intended users (eg, healthcare professionals, patients, public).
			Extension	Describe any pre-existing evidence for the AI intervention.
Methods: Participants, interventions and outcomes				
Study Setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained	Extension	Describe the onsite and offsite requirements needed to integrate the AI intervention into the trial setting.
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)	Elaboration	State the inclusion and exclusion criteria at the level of participants.
			Extension	State the inclusion and exclusion criteria at the level of the input data.
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered	Extension	State which version of the AI algorithm will be used.
			Extension	Specify the procedure for acquiring and selecting the input data for the AI intervention.
			Extension	Specify the procedure for assessing and handling poor quality or unavailable input data.
			Extension	Specify whether there is human-AI interaction in the handling of the input data, and what level of expertise is required for users.
			Extension	Specify the output of the AI intervention.
			Extension	Explain the procedure for how the AI intervention's output will contribute to decision making or other elements of clinical practice.
Methods: Monitoring				
Harms	22	Plans for collecting, assessing, reporting and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	Extension	Specify any plans to identify and analyse performance errors. If there are no plans for this, justify why not.
Access to data	29	Statement of who will have access to the final trial dataset and disclosure of contractual agreements that limit such access for investigators	Extension	State whether and how the AI intervention and/or its code can be accessed, including any restrictions to access or reuse.

Table adapted from Cruz Rivera *et al*.[21–23] Items within the SPIRIT 2013 statement that have not changed for the SPIRIT-AI statement have been omitted.
AI, artificial intelligence; SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials.

Table 3 Additional criteria to be included for studies relating to AI interventions within the CONSORT-AI statement (in addition to the CONSORT 2010 statement)

Section	Item no	CONSORT 2010 item	Amendment	CONSORT-AI item
Title and abstract				
Title and abstract	1a	Identification as a randomised trial in the title	Elaboration	Indicate that the intervention involves artificial intelligence/machine learning in the title and/or abstract and specify the type of model.
	1b	Structured summary of trial design, methods, results and conclusions	Elaboration	State the intended use of the AI intervention within the trial in the title and/or abstract.
Introduction				
Background and objectives	2a	Scientific background and explanation of rationale	Extension	Explain the intended use of the AI intervention in the context of the clinical pathway, including its purpose and its intended users (eg, healthcare professionals, patients, public).
Methods				
Participants	4a	Eligibility criteria for participants	Elaboration	State the inclusion and exclusion criteria at the level of participants.
			Extension	State the inclusion and exclusion criteria at the level of the input data.
	4b	Settings and locations where the data were collected	Extension	Describe how the AI intervention was integrated into the trial setting, including any onsite or offsite requirements.
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	Extension	State which version of the AI algorithm was used.
			Extension	Describe how the input data were acquired and selected for the AI intervention.
			Extension	Describe how poor quality or unavailable input data were assessed and handled
			Extension	Specify whether there was human–AI interaction in the handling of the input data, and what level of expertise was required of users.
			Extension	Specify the output of the AI intervention.
			Extension	Explain how the AI intervention's outputs contributed to decision making or other elements of clinical practice.
Results				
Harms	19		Extension	Describe results of any analysis of performance errors and how errors were identified, where applicable. If no such analysis was planned or done, justify why not.
Discussion				
Funding	25		Extension	State whether and how the AI intervention and/or its code can be accessed, including any restrictions to access or re-use.

Table adapted from Liu *et al*.[17–19] Items within the CONSORT 2010 statement that have not been changed for the CONSORT-AI statement have been omitted.
AI, artificial intelligence; CONSORT, Consolidated Standards of Reporting Trials.

'TRIPOD-AI',[28 29] and supported by the 'Prediction model Risk Of Bias Assessment Tool' (PROBAST, 2019 version)[30] which is proposed to be entitled PROBAST-ML.[28 29]

Human factors

Another upcoming guideline, focused on the evaluation of the 'human factors' in algorithm implementation, has been announced: the checklist (Developmental and Exploratory Clinical Investigation of Decision-support systems driven by AI).[31] This checklist is intended for use in early small-scale clinical trials that evaluate and provide information on how algorithms may be used in practice, bridging the gap between the algorithm development/validation stage (which would follow TRIPOD-AI, STARD-AI or Checklist for Artificial Intelligence in Medical Imaging (CLAIM)), but before large-scale clinical trials of AI interventions (where the CONSORT-AI would be used). Publication is anticipated to be late 2021 or early 2022.

Systematic reviews

Given the increasing volume of radiological AI-related research for a growing variety of conditions and clinical settings, it is also likely that we will encounter more systematic reviews and meta-analyses that aim to aggregate the evidence from studies in this field (eg, recent

publications have already emerged that summarise research regarding the role of AI in COVID-19.[32–34] At present, the 'Preferred Reporting Items for Systematic Reviews and Meta-analyses' (PRISMA), 2009[35] guidelines are the most established for systematic reviews and meta-analyses, with a modified version specifically tailored for meta-analyses relating to diagnostic test accuracies (ie, the PRISMA-Diagnostic Trials of Accuracy (DTA), 2018).[36] Currently, there have not been any announcements for an update to these guidelines for AI-related systematic reviews or meta-analyses, and therefore, it is suggested that the PRSIMA 2009[35] or PRISMA-DTA 2018[36] guidance should be followed.

In the planning stages for conducting systematic reviews of prediction models, the 'Checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies' (CHARMS, 2014[37] was developed by the Cochrane Prognosis Methods Group. This was not intentionally created for publications relating to AI per se, but applicable to a wide range of studies, which also happen to include the evaluation of ML models. The developers provide the checklist to help authors frame their review question, design and extract relevant items from published reports of prediction models and guide assessment of risk of bias (rather than in the analysis of these). This checklist will, therefore, be useful to those who wish to plan a review of AI tools that provide a 'risk score' or 'probability of diagnosis'. A tutorial on how to carry out a 'CHARMS analysis' for prognostic multivariate models with real-life worked examples has been published[38] and may be a helpful resource for readers wishing to carry out similar work. It is worth noting that the authors of CHARMS still recommend reference to the PRISMA 2009[35] and PRISMA-DTA 2018[36] statements for the reporting and analysis of trial results, in conjunction with their own checklist for planning of the review design.

OTHER (NON-EQUATOR NETWORK) GUIDELINES

Alternative guidelines have been published by expert interest groups and endorsed by different specialty societies. A few are described here to supplement further reading and interest.

The Radiological Society of North America recently published the 'CLAIM'[39] in 2020, containing elements of the STARD 2015 guideline and applicable for trials addressing a wide spectrum of AI applications using medical images (eg, classification, reconstruction, text analysis, work flow optimisation). This checklist comprises of 42 items, of which 6 are new (pertaining to model design and training), 8 are extensions of pre-existing STARD 2015 items, 14 items are elaborations (mostly relating to methods and results) and 14 items remain the same. Particular emphasis is given to data, the reference standard of 'ground truth' and the precise development and methodology of the AI algorithm being tested. These are listed in further detail in table 4, where differences to the STARD 2015 are highlighted. Care should be taken to avoid any confusion with another similarly named checklist entitled 'minimum information about clinical AI modelling' (MI-CLAIM),[40] which is less of a reporting guideline but a document outlining required shared understanding in the development and evaluation of AI models aimed to serve clinical and data scientists), repository managers and model users.

It is also worth noting that the American Medical Informatics Association produced a set of guidelines in 2020 termed the 'MI for Medical AI Reporting' (MINIMAR),[41] specific to studies reporting the use of AI solutions in healthcare. Rather than a list of items for manuscript writing, this guidance provides suggestions for details pertaining to data sources used in algorithm development and their intended usage, spread across four key subject areas (ie, study population and setting, patient demographics, model architecture and model evaluation). There are many similarities with the aforementioned CLAIM checklist, although the key differences include the granularity by which the MINIMAR suggests researchers should explicitly state participant demographics (eg, ethnicity and socioeconomic status, rather than just age and sex) and how code and data can be shared with the wider community.

FURTHER READING

There is an increasing need to build a cadre of researchers and reviewers with sufficient domain knowledge of technical aspects (including limitations and risk) and of the principles of good trial methodology (including areas of potential bias, analysis issues, etc). There is also a need for ML experts and clinical trial communities to increasingly learn each other's language, to ensure accurate and precise communication of concepts, and enable comparison between studies. A number of reviews are highlighted here for further reading[42–46] along with work[47] explaining different evaluation metrics used in AI and ML studies. It is also worth bearing in mind the wider clinical and ethical context of how any AI tool would fit into our existing clinical pathways and healthcare systems.[48]

CONCLUSION

In conclusion, this article has provided readers an overview of changes to standard clinical reporting guidelines specific for AI-related studies. The fundamental basics of describing the trial setup, inclusion and exclusion criteria, detailing the study methodology and standards used, together with details on algorithm development, should create transparency and address reproducibility. Those which are most relevant for a particular healthcare specialty will depend on the type of research being conducted in that particular field (eg, guidelines for AI-related diagnostic accuracy trials may be more relevant for radiological or pathological specialties, whereas those addressing patient outcomes with the aid of an

Table 4 Criteria for the CLAIM checklist for diagnostic accuracy studies using AI

Section	Item no	STARD 2015 item	Amendment	CLAIM item
Title and abstract				
Title	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values or AUC).	Elaboration	Identification as a study of AI methodology, specifying the category of technology used (eg, deep learning).
Abstract	2	Structured summary of study design, methods, results and conclusions.	Same	
Introduction				
Background	3	Scientific and clinical background, including the intended use and clinical role of the index test.	Elaboration	Scientific and clinical background, including the intended use and clinical role of the AI approach.
Objectives	4	Study objectives and hypotheses.	Same	
Methods				
Study design	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study).	Same	
			Extension	Study goal, such as model creation, exploratory study, feasibility study, non-inferiority trial.
Participants	6	Eligibility criteria (inclusion/exclusion).	Extension	State data sources.
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry).	Same	
	8	Where and when potentially eligible participants were identified (setting, location and dates).		
	9	Whether participants formed a consecutive, random or convenience series.	Extension	Data preprocessing steps.
			Extension	Selection of data subsets, if applicable.
			Extension	Definitions of data elements, with references to common data elements.
			Extension	Deidentification methods.
Test methods	10b	Reference standard, in sufficient detail to allow replication.	Elaboration	Definition of 'ground truth' (ie, reference standard), in sufficient detail to allow replication.
			Elaboration	Source of ground truth annotations; qualifications and preparation of annotators.
			Elaboration	Annotation tools.
	11	Rationale for choosing the reference standard (if alternatives exist).	Same	
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing prespecified from exploratory.	Elaboration	Measurement of inter-rater and intrarater variability; methods to mitigate variability and/or resolve discrepancies for ground truth.
Model			New	Detailed description of model, including inputs, outputs, all intermediate layers and connections.
			New	Software libraries, frameworks, and packages.
			New	Initialisation of model parameters (eg, randomisation, transfer learning).

Continued

Table 4 Continued

Section	Item no	STARD 2015 item	Amendment	CLAIM item
Training			New	Details of training approach, including data augmentation, hyperparameters, number of models trained.
			New	Method of selecting the final model.
			New	Ensembling techniques, if applicable
Analysis	14	Methods for estimating or comparing measures of diagnostic accuracy.	Elaboration	Metrics of model performance.
	16	How missing data on the index test and reference standard were handled.	Same	
	17	Any analyses of variability in diagnostic accuracy, distinguishing prespecified from exploratory.	Elaboration	Statistical measures of significance and uncertainty (eg, CIs).
			Elaboration	Robustness or sensitivity analysis.
			Elaboration	Methods for explainability or interpretability (eg, saliency maps) and how they were validated.
			Elaboration	Validation or testing on external data.
	18	Intended sample size and how it was determined.	Same	
			Extension	How data were assigned to partitions; specify proportions.
			Extension	Level at which partitions are disjoint (eg, image, study, patient, institution).
Results				
Participants	19	Flow of participants, using a diagram.	Same	
	20	Baseline demographic and clinical characteristics of participants.	Elaboration	Demographic and clinical characteristics of cases in each partition.
Test results	23	Cross tabulation of the index test results (or their distribution) by the results of the reference standard.	Elaboration	Performance metrics for optimal model(s) on all data partitions.
	24	Estimates of diagnostic accuracy and their precision (such as 95% CIs).	Same	
	25	Any adverse events from performing the index test or the reference standard.	Elaboration	Failure analysis of incorrectly classified cases.
Discussion				
Limitations	26	Study limitations, including sources of potential bias, statistical uncertainty and generalisability.	Same	
Implications	27	Implications for practice, including the intended use and clinical role of the index test.	Same	
Other Information				
Registration	28	Registration no and name of registry.	Same	
Protocol	29	Where the full study protocol can be accessed.	Same	
Funding	30	Sources of funding and other support; role of funders.	Same	

This is based on the STARD 2015 guidelines,[20] demonstrating which aspects are new, the same or elaborated on. Items not included in the CLAIM checklist (which were previously present in the STARD guideline) have been removed. Table adapted from Bossuyt *et al*[20] and Mongan *et al*.[39]

AI, artificial intelligence; CLAIM, Checklist for Artificial Intelligence in Medical Imaging; STARD, Standards for Reporting of Diagnostic Accuracy Studies.

AI algorithm may be more relevant for oncological or surgical specialties).

Although the reporting guidelines outlined may seem comprehensive, there remain areas that will need to be addressed, such as for economic health evaluation of AI-tools and algorithms (many are currently developed for 'pharmacoeconomic evaluations'.[49] It is likely that future guidelines may take the form of an extension to the widely used CHEERS guidance (Consolidated Health Economic Evaluation Reporting Standards[50 51] available via the EQUATOR network.[13] Nevertheless, a wide variation in opinion regarding the most appropriate economic evaluation guideline already exists for non-AI related tools, and this may be reflected in future iterations of such guidelines depending on how the algorithms are funded in different healthcare systems.[52]

The current guidelines outlined here will likely continue to be updated in the light of new understanding of the specific challenges of AI as an intervention and, how traditional study designs and reports need to be adapted.

Funding OJA is funded by a National Institute for Health Research (NIHR) Career Development Fellowship (NIHR-CDF-2017-10-037). SS, OJA and NJS receive funding from the Great Ormond Street Children's Charity and the Great Ormond Street Hospital NIHR Biomedical Research Centre. AD receives funding from Health Data Research UK. An initiative funded by UK Research and Innovation, Department of Health and Social Care (England) and the devolved administrations and leading medical research charities.

Disclaimer The funding source(s) did not have any direct involvement in the methodology, design or write-up of this review article.

Competing interests None declared.

Patient and public involvement statement Not required

Patient consent for publication Not required.

Ethics approval Not required

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study.

ORCID iDs
Susan Cheng Shelmerdine http://orcid.org/0000-0001-6642-9967
Neil J Sebire http://orcid.org/0000-0001-5348-9063

REFERENCES

1 More than machines. *Nat Mach Intell* 2019;1.
2 The Lancet Digital Health. A digital (r)evolution: introducing The Lancet Digital Health. *Lancet Digit Health* 2019;1:e1.
3 Kahn CE, Charles E, Kahn J. Artificial intelligence, real radiology. *Radiol Artif Intell* 2019;1:e184001.
4 Moher D. Reporting guidelines: doing better for readers. *BMC Med* 2018;16:233.
5 Bluemke DA, Moy L, Bredella MA, *et al*. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers-from the radiology editorial board. *Radiology* 2020;294:487–9.
6 CONSORT-AI and SPIRIT-AI Steering Group. Reporting guidelines for clinical trials evaluating artificial intelligence interventions are needed. *Nat Med* 2019;25:1467–8.
7 Liu X, Faes L, Kale AU, *et al*. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health* 2019;1:e271–97.
8 Nagendran M, Chen Y, Lovejoy CA, *et al*. Artificial intelligence versus clinicians: systematic review of design, reporting Standards, and claims of deep learning studies. *BMJ* 2020;368:m689.
9 Kim DW, Jang HY, Kim KW, *et al*. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. *Korean J Radiol* 2019;20:405–10.
10 Yusuf M, Atal I, Li J, *et al*. Reporting quality of studies using machine learning models for medical diagnosis: a systematic review. *BMJ Open* 2020;10:e034568.
11 Bozkurt S, Cahan EM, Seneviratne MG, *et al*. Reporting of demographic data and representativeness in machine learning models using electronic health records. *J Am Med Inform Assoc* 2020;27:1878–84.
12 Altman DG, Simera I, Hoey J, *et al*. EQUATOR: reporting guidelines for health research. *Lancet* 2008;371:1149–50.
13 The EQUATOR Network. Enhancing the quality and transparency of health research. Available: https://www.equator-network.org [Accessed 22 Mar 2021].
14 Moher D, Hopewell S, Schulz KF, *et al*. Consort 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
15 Calvert M, Kyte D, Mercieca-Bebber R, *et al*. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. *JAMA* 2018;319:483–94.
16 Calvert M, Blazeby J, Altman DG, *et al*. Reporting of patient-reported outcomes in randomized trials: the CONSORT pro extension. *JAMA* 2013;309:814–22.
17 Liu X, Cruz Rivera S, Moher D, *et al*. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health* 2020;2:e537–48.
18 Liu X, Rivera SC, Moher D. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *BMJ* 2020;370:m3164.
19 Liu X, Cruz Rivera S, Moher D, *et al*. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020;26:1364–74.
20 Bossuyt PM, Reitsma JB, Bruns DE, *et al*. Stard 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology* 2015;277:826–32.
21 Cruz Rivera S, Liu X, Chan A-W, *et al*. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Lancet Digit Health* 2020;2:e549–60.
22 Cruz Rivera S, Liu X, Chan A-W, *et al*. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med* 2020;26:1351–63.
23 Rivera SC, Liu X, Chan A-W, *et al*. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *BMJ* 2020;370:m3210.
24 Chan A-W, Tetzlaff JM, Gøtzsche PC, *et al*. Spirit 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ* 2013;346:e7586.
25 Hoffmann TC, Glasziou PP, Boutron I, *et al*. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;348:g1687.
26 Sounderajah V, Ashrafian H, Aggarwal R, *et al*. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: the STARD-AI steering group. *Nat Med* 2020;26:807–8.
27 Collins GS, Reitsma JB, Altman DG, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.
28 Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet* 2019;393:1577–9.
29 Andaur Navarro CL, Damen JAAG, Takada T, *et al*. Protocol for a systematic review on the methodological and reporting quality of prediction model studies using machine learning techniques. *BMJ Open* 2020;10:e038832.
30 Moons KGM, Wolff RF, Riley RD, *et al*. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med* 2019;170:W1.
31 DECIDE-AI Steering Group. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. *Nat Med* 2021;27:186–7.

32 Albahri OS, Zaidan AA, Albahri AS, *et al*. Systematic review of artificial intelligence techniques in the detection and classification of COVID-19 medical images in terms of evaluation and benchmarking: taxonomy analysis, challenges, future solutions and methodological aspects. *J Infect Public Health* 2020;13:1381–96.
33 Li WT, Ma J, Shende N, *et al*. Using machine learning of clinical data to diagnose COVID-19: a systematic review and meta-analysis. *BMC Med Inform Decis Mak* 2020;20:247.
34 Syeda HB, Syed M, Sexton KW, *et al*. Role of machine learning techniques to tackle the COVID-19 crisis: systematic review. *JMIR Med Inform* 2021;9:e23811.
35 Liberati A, Altman DG, Tetzlaff J, *et al*. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
36 McInnes MDF, Moher D, Thombs BD, *et al*. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018;319:388–96.
37 Moons KGM, de Groot JAH, Bouwmeester W, *et al*. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the charms checklist. *PLoS Med* 2014;11:e1001744.
38 Palazón-Bru A, Martín-Pérez F, Mares-García E, *et al*. A general presentation on how to carry out a CHARMS analysis for prognostic multivariate models. *Stat Med* 2020;39:3207–25.
39 Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (claim): a guide for authors and reviewers. *Radiol Artif Intell* 2020;2:e200029.
40 Norgeot B, Quer G, Beaulieu-Jones BK, *et al*. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat Med* 2020;26:1320–4.
41 Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, *et al*. MINIMAR (minimum information for medical AI reporting): developing reporting standards for artificial intelligence in health care. *J Am Med Inform Assoc* 2020;27:2011–5.
42 Luo W, Phung D, Tran T, *et al*. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res* 2016;18:e323.
43 Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *NPJ Digit Med* 2020;3:126.
44 Kocak B, Kus EA, Kilickesmez O. How to read and review papers on machine learning and artificial intelligence in radiology: a survival guide to key methodological concepts. *Eur Radiol* 2021;31:1819–30.
45 Faes L, Liu X, Wagner SK, *et al*. A clinician's guide to artificial intelligence: how to critically appraise machine learning studies. *Transl Vis Sci Technol* 2020;9:7.
46 Do S, Song KD, Chung JW. Basics of deep learning: a radiologist's guide to understanding published radiology articles on deep learning. *Korean J Radiol* 2020;21:33–41.
47 Handelman GS, Kok HK, Chandra RV, *et al*. Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. *AJR Am J Roentgenol* 2019;212:38–43.
48 McCradden MD, Joshi S, Mazwi M, *et al*. Ethical limitations of algorithmic fairness solutions in health care machine learning. *Lancet Digit Health* 2020;2:e221–3.
49 Sullivan SM, Wells G, Coyle D. What guidance are economists given on how to present economic evaluations for policymakers? A systematic review. *Value Health* 2015;18:915–24.
50 Husereau D, Drummond M, Petrou S, *et al*. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health* 2013;16:231–50.
51 Husereau D, Drummond M, Petrou S, *et al*. Consolidated health economic evaluation reporting standards (cheers) statement. *Value Health* 2013;16:e1–5.
52 Sharma D, Aggarwal AK, Downey LE, *et al*. National healthcare economic evaluation guidelines: a Cross-Country comparison. *Pharmacoecon Open* 2021;5:349–64.

BMJ Health & Care Informatics

Predictive modeling of COVID-19 case growth highlights evolving racial and ethnic risk factors in Tennessee and Georgia

Jamieson D Gray,[1] Coleman R Harris,[1,2] Lukasz S Wylezinski,[1,3] Charles F Spurlock, III [1,3]

To cite: Gray JD, Harris CR, Wylezinski LS, *et al*. Predictive modeling of COVID-19 case growth highlights evolving racial and ethnic risk factors in Tennessee and Georgia. *BMJ Health Care Inform* 2021;**28**:e100349. doi:10.1136/bmjhci-2021-100349

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2021-100349).

Received 25 February 2021
Accepted 19 July 2021

[1]Decode Health, Inc. and IQuity Labs, Inc, Nashville, Tennessee, USA
[2]Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA
[3]Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

Correspondence to
Dr Charles F Spurlock, III; chase.spurlock@vanderbilt.edu

ABSTRACT

Introduction The SARS-CoV-2 (COVID-19) pandemic has exposed the need to understand the risk drivers that contribute to uneven morbidity and mortality in US communities. Addressing the community-specific social determinants of health (SDOH) that correlate with spread of SARS-CoV-2 provides an opportunity for targeted public health intervention to promote greater resilience to viral respiratory infections.

Methods Our work combined publicly available COVID-19 statistics with county-level SDOH information. Machine learning models were trained to predict COVID-19 case growth and understand the social, physical and environmental risk factors associated with higher rates of SARS-CoV-2 infection in Tennessee and Georgia counties. Model accuracy was assessed comparing predicted case counts to actual positive case counts in each county.

Results The predictive models achieved a mean R^2 of 0.998 in both states with accuracy above 90% for all time points examined. Using these models, we tracked the importance of SDOH data features over time to uncover the specific racial demographic characteristics strongly associated with COVID-19 incidence in Tennessee and Georgia counties. Our results point to dynamic racial trends in both states over time and varying, localized patterns of risk among counties within the same state. For example, we find that African American and Asian racial demographics present comparable, and contrasting, patterns of risk depending on locality.

Conclusion The dichotomy of demographic trends presented here emphasizes the importance of understanding the unique factors that influence COVID-19 incidence. Identifying these specific risk factors tied to COVID-19 case growth can help stakeholders target regional interventions to mitigate the burden of future outbreaks.

INTRODUCTION

In January 2021, Tennessee and Georgia reported over 1,637,000 cases and 23,848 deaths due to COVID-19. Hispanic individuals comprise 14% of the states' population but represent 25% of confirmed cases, suggesting race and ethnicity are associated with case growth.[1] To explore this association, we sought to combine publicly available COVID-19 data and proprietary social determinants of health (SDOH), which measure certain physical, social, economic and demographic characteristics, to build and tune machine learning models predicting COVID-19 incidence in Tennessee and Georgia. Our objective was to accurately predict COVID-19 case growth, while investigating model-based relevance of racial demographic features influencing these predictions over time. We hypothesised that SDOH features significantly influence COVID-19 incidence and that underlying risk patterns associated with county-level race and ethnicity data features influence prediction accuracy.

METHODS

Our approach combined publicly available COVID-19 case, hospitalization and death metrics with county-specific SDOH data.[2 3] Information sources for the study included the State of Tennessee Department of Health, State of Georgia Department of Health, the Johns Hopkins Coronavirus Research Center and the US Census database. Feature engineering and feature selection, including interaction terms, were employed to define the data inputs that best represent changes in COVID-19 incidence over time. We developed novel features that included offset and normalised case growth as well as dynamic time window features derived from state health department data from July 2020 to January 2021. Time independent enrichment data, including both quantitative and qualitative SDOH with demographic information, were joined into this feature set to generate county-specific data.[4] Data were aggregated

from multiple sources to minimize the impact of any implicit bias and missing values removed or ignored depending on model type. The outcome for predictive modelling was defined as the future relative case growth normalised to the population in Tennessee and Georgia counties. We performed a grid search of generalised linear and tree-based machine learning models, training and testing each model with 4–6 weeks of historical COVID-19 case data to generate predictions using the most recent data available. From the ~50 regression models that we built for each time point, models were chosen using cross-validation model metrics (eg, R^2, Tweedie deviance) and prediction accuracy for COVID-19 case growth.[5] We identified the top third of Tennessee and Georgia counties

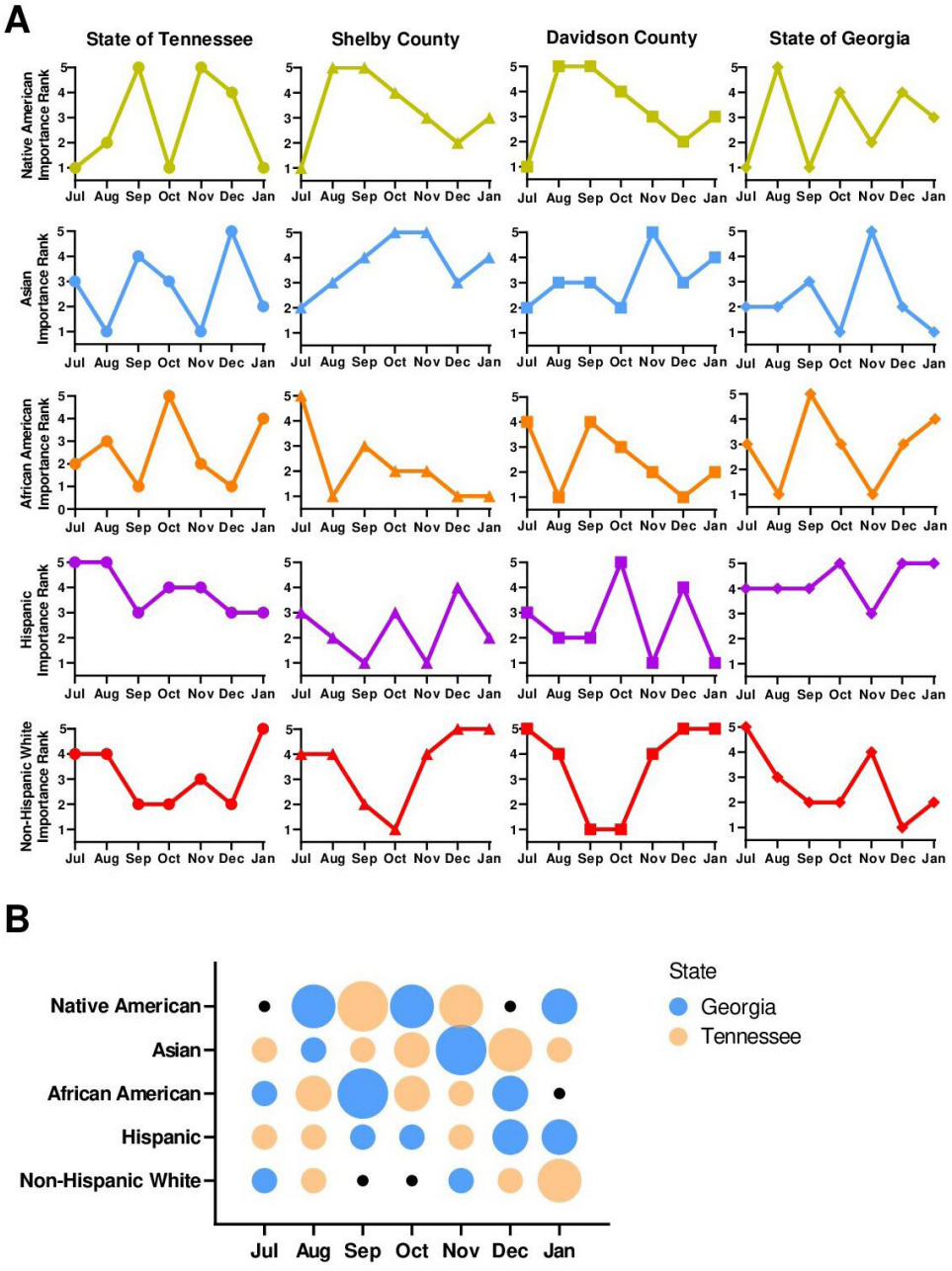

Figure 1 Influence of racial demographic features linked to COVID-19 case growth exhibit dynamic shifts over time in Tennessee and Georgia. (A) Relative rank of demographic feature importance across top predictive models is reported for the entire state of Tennessee (●) and the two most populous counties in Tennessee, Shelby County (▲) and Davidson County (■) as well as the state of Georgia (◆). A score of 5 on the importance rank indicates the most important demographic feature relative to the other four demographic features. Groups include Native American (●), Asian (●), African American (●), Hispanic (●) and non-Hispanic white (●). (B) Differences in the rank of demographic feature importance in Tennessee and Georgia over time. The colour of the bubble (Tennessee (●); Georgia (●)) indicates the state that exhibited a higher importance rank of the specific demographic feature for predicting COVID-19 case growth. Black dots (●) designate months where the two states displayed the same importance rank for an individual demographic feature. The size of the bubbles shows the difference in importance of each demographic feature between the two states. Larger bubbles connote greater difference in importance.

at highest risk for case growth and assessed our prediction accuracy versus actual case growth over time. Finally, we analysed each data feature's impact at the state and county level to track the rank order of demographic data features that drove COVID-19 case growth at each time point.

RESULTS

Candidate models for Tennessee and Georgia achieved excellent metrics across all time points including a mean R^2 value of 0.998 (Tennessee and Georgia), mean Tweedie deviance of 0.003 (Tennessee) and 0.002 (Georgia) as well as a mean absolute error of 0.357 (Tennessee) and 0.337 (Georgia) (online supplemental figure 1A). Prediction accuracy was >90% in all models across both states when compared with actual case growth (online supplemental figure 1B).

Racial demographics produced variable trends at both the state and county levels. The two most populous counties in Tennessee, Shelby and Davidson, revealed an identical pattern of importance for Native American racial demographics in determining future case growth while exhibiting differences among Asians. Shelby County displayed a gradual increase in importance in the Asian demographic, while Davidson County saw a more pronounced spike between October and November. Comparing racial demographic importance at the Tennessee state level versus individual counties yields similar patterns (non-Hispanic white) as well as contrasting trends (African American). Further, the Hispanic ethnicity risk factor trends across Tennessee differed from the individual Tennessee counties' more acute fluctuation of Hispanic risk factor importance (figure 1A).

Additionally, similarities and differences in racial demographic trends extend across state borders. While Hispanic ethnicity displayed the most meaningful importance in Tennessee during July and August, Georgia saw a similar increase in importance starting in September. Comparison of the two states' top racial demographic drivers showed a potential macropattern in which the most important driver for one state often preceded its rise to top importance in the other (figure 1A,B).

DISCUSSION

Our study developed highly accurate modeling to predict COVID-19 case growth and discover associations between state and county SDOH characteristics connected to risk for future spread of infection. Analysis of the most influential racial and ethnic demographic data at each time point discovered localized, evolving patterns of risk that correlate with state-level and county-level SARS-CoV-2 case growth. These patterns can shift dramatically month to month, increasing or decreasing over time and vary by geography, even among similarly sized counties within a state or between two neighbouring states. The state-specific and county-specific modelling results we describe for Tennessee and Georgia may bias or limit the validity of extrapolating our specific modelling results to other localities. However, the approach is extensible to all US states and counties.

Early identification of the specific SDOH risk drivers tied to disease outcomes in a pandemic could help decision-makers promote health equity and deliver targeted interventions to mitigate disease risk in vulnerable populations. Closing the loop to address certain SDOH risk factors also enhances community resilience to future viral respiratory infections.[6]

Applications of this approach extend beyond acute respiratory infection to chronic disease outcomes. A growing percentage (approximately 10%) of patients infected with SARS-CoV-2 develop long COVID.[7] These patients experience prolonged, debilitating symptoms months after infection and emergence or exacerbation of chronic illness. Targeted approaches to mitigate spread of disease can lessen future acute and chronic disease burden.

Twitter Jamieson D Gray @jamiesongray, Coleman R Harris @colemanrharris and Charles F Spurlock, III @cfspurlock

Acknowledgements The authors would like to acknowledge Cody Heiser for providing a constructive review of the manuscript.

Contributors CFS had full access to all of the data in the study, devised the concept and study design and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors took part in acquisition, analysis and interpretation of the data along with drafting and revising the manuscript.

Funding This work was supported by Decode Health, Inc., IQuity Labs, Inc., and grants from the National Institutes of Health (AI124766, AI129147 and AI145505).

Competing interests JDG, LSW and CFS are shareholders in IQuity Labs, Inc. (Nashville, Tennessee) and Decode Health, Inc. (Nashville, Tennessee). IQuity Labs, Inc. develops blood-based RNA tools to aid in the diagnosis and treatment of human disease. Decode Health, Inc. develops artificial intelligence approaches to predict chronic and infectious disease risk in patient populations.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Charles F Spurlock, III http://orcid.org/0000-0001-9015-6321

REFERENCES

1 The COVID tracking project. racial data Dashboard. Available: https://covidtracking.com/race/dashboard [Accessed 25 Nov 2020].
2 Johns Hopkins University Coronavirus Resource Center. COVID-19 United States cases by County. Available: https://coronavirus.jhu.edu/us-map [Accessed 25 Nov 2020].

3 Vest JR, Ben-Assuli O. Prediction of emergency department revisits using area-level social determinants of health measures and health information exchange information. *Int J Med Inform* 2019;129:205–10.
4 Kolak M, Bhatt J, Park YH, *et al*. Quantification of neighborhood-level social determinants of health in the continental United States. *JAMA Netw Open* 2020;3:e1919928.
5 Muhlestein WE, Akagi DS, Chotai S, *et al*. The impact of presurgical comorbidities on discharge disposition and length of hospitalization following craniotomy for brain tumor. *Surg Neurol Int* 2017;8:220.
6 Graves E, Weiss A, Rickles M, *et al*. Community resiliency to COVID-19 in a subset of US communities. Available: https://about.sharecare.com/wp-content/uploads/2020/08/National_COVID-whitepaper_PROOF_08.28.20.pdf [Accessed 25 Nov 2020].
7 Greenhalgh T, Knight M, A'Court C, *et al*. Management of post-acute covid-19 in primary care. *BMJ* 2020;370:m3026.

BMJ Health & Care Informatics

Is home pulse oximeter monitoring for COVID-19 feasible in low-income and low-middle-income countries?

Sabina Rijal [1], Bipin Poudel [2]

To cite: Rijal S, Poudel B. Is home pulse oximeter monitoring for COVID-19 feasible in low-income and low-middle-income countries? *BMJ Health Care Inform* 2021;**28**:e100465. doi:10.1136/bmjhci-2021-100465

Accepted 17 August 2021

[1]Shree Military Hospital Nepalgunj, Banke, Nepal
[2]Patan Academy of Health Sciences, Kathmandu, Nepal

Correspondence to
Bipin Poudel;
bipinpoudel@pahs.edu.np

We thank Gootenberg *et al*[1] for sharing the article on 'Developing a pulse oximetry home monitoring protocol for patients suspected with COVID-19 after emergency department discharge' and raise concerns over feasibility of such monitoring in low-income countries (LIC) and lower-middle-income countries (LMIC) in terms of availability of devices and ability to use them correctly.

Irrespective of economic status of country, hospitals have been overwhelmed with COVID-19 cases. The use of pulse oximeter at home may be feasible in countries where devices are readily available and affordable, but this seems like a far-fetched idea in resource-limited countries. With only $6 and $26 to spend per person for COVID-19-related social protection in LIC and LMIC, respectively, many people from those areas are struggling for basic needs like food.[2] The idea of distributing pulse oximeter (each of which costs $20–$48) with added cost of telephone calls, from emergency social fund to each suspected person discharged form emergency department in such countries, is challenging without international support.[1 3] Shortage of standard devices is a rising problem in such countries. Even the ones that are available in the market are either not quality-certified or expensive.[4] This adds to the financial burden without any potential benefit while increasing false reassurance.

The literacy rate of 61% and 76% in LIC and LMIC, respectively, which is below the global literacy rate of 86%, limits its applicability in such areas.[5] Remaining population would not be able to comprehend the readings. Even those who can read the numbers are not aware of the correct method to use it. Factors like poor perfusion, dye, pigmentation, movement of hands, etc lead to false reading.[3] Failure to identify potential sources of error adds to the problem.

The challenges faced by LIC and LMIC as mentioned above need to be addressed first to ensure sustainability of programme. Financial aid from international fraternity and provision of quality as well as cost control of the devices from government level may be helpful in solving economic hardship. Governments can mobilise community health volunteers for creating awareness about the correct use of oximeters. They can monitor the oxygen saturation of suspected or confirmed COVID-19 cases by visiting individual houses while taking necessary precautions. Overall, the protocol given by Gootenberg *et al* provides a safe and effective framework for use of pulse oximeter to identify silent hypoxia.[1]

Contributors SR and BP: concept. SR: literature review, manuscript writing, editing and review. BP: manuscript editing and review. Both authors have approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; internally peer reviewed.

ORCID iDs
Sabina Rijal http://orcid.org/0000-0003-3867-2761
Bipin Poudel http://orcid.org/0000-0001-8194-6081

REFERENCES

1 Gootenberg DB, Kurtzman N, O'Mara T, *et al*. Developing a pulse oximetry home monitoring protocol for patients suspected with COVID-19 after emergency department discharge. *BMJ Health Care Inform* 2021;28:e100330.

2 In developing countries, the COVID-19 crisis has not affected everyone equally [Internet]. World Bank Blogs, 2021. Available: https://blogs.worldbank.org/voices/developing-countries-covid-19-crisis-has-not-affected-everyone-equally [Accessed 30 Jul 2021].
3 Greenhalgh T, Knight M, Inda-Kim M, *et al*. Remote management of covid-19 using home pulse oximetry and virtual ward support. *BMJ* 2021;372:n677.
4 MK M. Why many of the oximeters you're buying in India may not be reliable [Internet]. The News Minute, 2021. Available: https://www.thenewsminute.com/article/why-many-oximeters-youre-buying-india-may-not-be-reliable-135313 [Accessed 30 Jul 2021].
5 World Bank Group. Literacy rate, adult total (% of people ages 15 and above) | Data [Internet]. The World Bank, 2021. Available: https://data.worldbank.org/indicator/SE.ADT.LITR.ZS [Accessed 30 Jul 2021].