BMJ Health & Care Informatics

Improving the evidence base for digital health interventions to increase contraception use

Chris Smith [1,2]

To cite: Smith C. Improving the evidence base for digital health interventions to increase contraception use. *BMJ Health Care Inform* 2021;**28**:e100346. doi:10.1136/bmjhci-2021-100346

Accepted 29 March 2021

[1]School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan
[2]Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, UK

Correspondence to
Dr Chris Smith;
christopher.smith@lshtm.ac.uk

Contraception is a lifesaving and essential component of healthcare. Since COVID-19 disrupted routine health service delivery, increased use of digital health (or mHealth) has been required to reduce risks to patients and healthcare workers[1 2] A recent systematic review was, however, unable to draw concrete conclusions on the overall effectiveness of mHealth interventions to increase contraception use in low-income and middle-income countries.[3] A meta-analysis was not possible due to differences between study populations, interventions and outcomes. In the meantime, another trial reported no measurable effect of an mHealth intervention for female sex workers on unintended pregnancy in Kenya, adding to the mixed evidence.[4] A better evidence base for digital health interventions to increase contraception use is required, but how can this be achieved?

First, a common set of study outcome measures for interventions for contraception is required. To date, a variety of approaches to measuring contraception use have been used, to assess current use or adherence over time. While objective measures such as using biological markers or electronic medication monitors are considered less prone to bias, their use is challenging.[5] Thus subjective measures enquiring about self-reported use remain de rigueur for most trials, accepting their known biases. An expert working group could be convened with the aim of harmonising contraceptive trial outcomes in order to reduce heterogeneity between studies and allow future meta-analyses.

Second, a deeper examination of the intervention mode of delivery and resource requirements is required. Distinctions between unidirectional and interactive interventions are important but the intensity and mode of interaction needs further evaluation. The ideal digital health intervention would be fully automated, scalable, safe and effective. It remains to be seen whether this is possible for contraception where there are a wide range of methods, cultural beliefs and nuanced side effects. Further understanding of the degree of personal interaction required and overall resource implications are important considerations for service providers when considering replicability and scale.

Third, now is the time to conduct larger studies that have shown to be effective, or promising but underpowered, and prioritised over trials of new interventions. Trials could be undertaken in different settings, but use the same outcome measures, maintaining the core components of the original intervention, subject to cultural or linguistic adaptions that might be required. This should result in an increased evidence base for service providers wanting to adopt best practices in digital health in their contraception programmes.

Funding The author has 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 iD
Chris Smith http://orcid.org/0000-0001-9238-3202

REFERENCES

1 Nanda K, Lebetkin E, Steiner MJ, *et al*. Contraception in the era of COVID-19. *Glob Heal Sci Pract* 2020;8:166–8.
2 Peek N, Sujan M, Scott P. Digital health and care in pandemic times: impact of COVID-19. *BMJ Health Care Inform* 2020;27:e100166.

3 Aung B, Mitchell JW, Braun KL. Effectiveness of mHealth interventions for improving contraceptive use in low- and middle-income countries: a systematic review. *Glob Heal Sci Pract* 2020;8:813–26.
4 Ampt FH, Lim MSC, Agius PA, *et al*. Effect of a mobile phone intervention for female sex workers on unintended pregnancy in Kenya (WHISPER or SHOUT): a cluster-randomised controlled trial. *The Lancet Glob Heal* 2020;8:e1534–45.
5 Smith C, Edwards P, Free C. Assessing the validity and reliability of self- report data on contraception use in the mobile technology for improved family planning (motif) randomised controlled trial. *Trials* 2018:1–5.

BMJ Health & Care Informatics

Impact of DSMES app interventions on medication adherence in type 2 diabetes mellitus: systematic review and meta-analysis

Dumisani Enricho Nkhoma,[1,2] Charles Jenya Soko,[2] Kondwani Joseph Banda,[3] David Greenfield,[4] Yu-Chuan (Jack) Li ,[5,6,7,8,9] Usman Iqbal [2,6,10]

To cite: Enricho Nkhoma D, Jenya Soko C, Joseph Banda K, *et al*. Impact of DSMES app interventions on medication adherence in type 2 diabetes mellitus: systematic review and meta-analysis. *BMJ Health Care Inform* 2021;**28**:e100291. doi:10.1136/bmjhci-2020-100291

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2020-100291).

Received 30 November 2020
Revised 18 February 2021
Accepted 15 March 2021

For numbered affiliations see end of article.

Correspondence to
Dr Usman Iqbal;
usmaniqbal@tmu.edu.tw

ABSTRACT

Objectives To conduct systematic review and meta-analysis of interventional studies to investigate the impact of diabetes self-management education and support (DSMES) apps on adherence in patients with type 2 diabetes mellitus (T2D).

Methods PubMed, Embase, CENTRAL, Web of Science, Scopus and ProQuest were searched, in addition to references of identified articles and similar reviews. Experimental studies, reported in English, assessing DSMES app intervention's impact on adherence and clinical outcomes of patients with T2D compared with usual care were included. Study bias was assessed using Cochrane Risk of Bias V.2.0 tool. Analysis plan involved narrative synthesis, moderator and meta-analysis.

Results Six randomised controlled trials were included, involving 696 participants (average age 57.6 years, SD 10.59). Mobile apps were mostly used for imputing clinical data, dietary intake or physical activity, and transmitting information to the provider. At 3 months, DSMES apps proved effective in improving medication adherence (standardized mean difference (SMD)=0.393, 95% CI 0.17 to 0.61), glycated haemoglobin (HbA1c) (mean difference (MD)=−0.314, 95% CI −0.477 to −0.151) and Body Mass Index (BMI) (MD=−0.28, 95% CI −0.545 to −0.015). All pooled estimates had low heterogeneity (I^2 0%). Four studies had moderate risk of bias while one each was judged to be low and high risks, respectively.

Conclusion DSMES apps had significant small to moderate effects on medication adherence, HbA1c and BMI of patients with T2D compared with usual care. Apps were described as reliable, easy to use and convenient, though participants were required to be phone literate. Evidence comes from feasibility trials with generally moderate risk of bias. Larger trials with longer follow-up periods using theory-based interventions are required to improve current evidence.

BACKGROUND

The burden of diabetes mellitus (DM) has grown enormously over the years with its prevalence is estimated to be more than 500 million globally in 2018, being one of the leading causes of death worldwide with an estimated economic burden of around US$1.3 trillion as of 2015.[1–3] DM is characterised by two major types: type 1 DM and type 2 diabetes mellitus (T2D), the latter constituting the majority of all known DM cases. Major risk factors for the disease include obesity and lack of exercise with long-term complications leading to stroke, leg amputation, and kidney and heart or eye problems.[4] T2D has no cure and, as such, patients are bound to lifelong treatment of which, if committed to, is associated with prevention of complications, lower medical costs and consequently better quality of life.[5 6]

A commitment to T2D treatment means being fully adherent to medical prescriptions, diet and exercise plans. Treatment adherence is the extent to which a person's behaviour, medication use, and diet or lifestyle changes correspond to agreed recommendations from a healthcare provider.[7] However, treatment adherence in patients with chronic diseases, including DM, has been reported to be suboptimal.[8 9]

Several techniques, such as single-dose regimens, reminders and easy packaging, have been used to improve treatment adherence.[10] One of these techniques, diabetes self-management education and support (DSMES), has been used to cover lifestyle, medication, blood glucose monitoring and other psychosocial aspects of treatment—all of which have been associated with improved health outcomes and reduced medical costs.[11] DSMES is a patient-centred and holistic approach that makes it one of the most ideal techniques to improve treatment adherence of patients with T2D.

Furthermore, delivery of DSMES through technologies such as mobile health (m-health) may improve coverage and convenience for patients. m-health helps to improve adherence including by setting automated reminders

and messages, and simplifying tracking of medication and prescriptions.[10 12] It is the the most common technology adopted by WHO member states.[13] In addition, the American Diabetes Association includes m-health into 'standard of medical care' for DM.[14] Unfortunately despite this endorsement, m-health is often underused in managing DM.[15] There is even lower patronage for apps that would help improve medication adherence.[16]

There is even fewer literature covering the effect of mobile app-delivered interventions that aim at improving adherence in the T2D population. Majority of prior reviews have concentrated on assessing effectiveness of SMS interventions on medication adherence.[10 17] One review did report the effect of app-delivered interventions; however, the study population was heterogeneous with little representation of patients with T2D.[18] Other reviews have aimed at assessing m-health intervention effects on glycaemic control.[19 20]

However, to our knowledge, no review was found to evaluate impact of mobile app-delivered DSMES on adherence specifically in the T2D population. Studies have concentrated on medication adherence rather than lifestyle modification adherence. Furthermore, prior reviews have reported inconsistent findings, lack of clarity on definition and measurement of medication adherence, and inadequate use of theoretical frameworks in the study interventions.[10 21]

Results of the current review would help add clarity to existing literature and offer quantitative evidence on the impact of app-delivered diabetes education and self-management support on treatment adherence of patients with T2D. The review's findings would also offer crucial applicability details to information technology and health professionals involved in efforts to reduce global DM burden.

Thus, we aimed to assess the effectiveness of mobile app-delivered DSMES to improve medication adherence and clinical outcomes in patients with T2D. We conducted a systematic review and meta-analysis of experimental studies to address this goal.

METHODS

Design and data sources

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).[22] The study protocol was registered with Open Science Framework (link: https://osf.io/z6sgk). Research team members (DEN and CJS) searched five academic databases from 1 to 14 August 2020, in two phases, for articles published from inception until 31 July 2020: PubMed, Embase, CENTRAL, Web of Science, ProQuest and Scopus. Medical Subject Headings (MeSH) and free-text terms were both used to combine keywords, as illustrated in the online supplemental file tables S1).

ELIGIBILITY CRITERIA

Studies were included if they had (1) adults, 18 years or older, with T2D; (2) independent analysis of participants with T2D in multidisease studies; (3) reported use of mobile apps giving DSMES as a principal intervention; (4) medication or lifestyle therapy adherence as an outcome; (5) experimental design; and (6) been reported in English. Studies were excluded if control groups used any type of digital technology or did not include routine diabetes care.

Study collection, selection and data extraction

The studies were independently screened (DEN and CJS) with disputes being resolved through consensus with a separate expert reviewer (UI). Variable and characteristics were independently extracted team members (DEN, CJS and KJB) using a standard sheet adopted from a Cochrane Public Health Group template.[23] Missing information was sought after by contacting authors through email. If the data could still not be retrieved, the study was excluded from the quantitative analysis.

QUALITY ASSESSMENT

Study bias was assessed using Cochrane's Risk of Bias V.2.0 tool independently by team members (DEN and KJB).[24] Studies were graded according to their randomisation process, deviation from intended interventions, handling of missing data, measurement of outcomes and reporting of results. Each domain was judged as high, low or moderate risk (some concerns). Overall risk of bias depended on the assessment of the individual domains. Assessment was to be conducted for each outcome. However, outcomes were grouped into primary (adherence) and clinical outcomes (glycated haemoglobin (HbA1c), Body Mass Index (BMI), blood pressure (BP) and total cholesterol). Across-study bias was assessed using a funnel plot and Egger's regression test.

Data analysis

Narrative synthesis

Qualitative analysis involved description of population and intervention characteristics. Items included acceptability, challenges and use of cointerventions.

Quantitative analysis

Fixed-effects meta-analysis was performed using pairwise comparison at 3 months after baseline measurements. Medication and patient adherence were presented using standardized mean difference (SMD) with 95% CIs. Change in HbA1c, BMI, total cholesterol and BP would be expressed using difference in means (MD) with 95% CI. Quantitative analysis was conducted using Comprehensive Meta-Analysis V.3 (Biostat, Englewood, USA). During analysis, an outcome, patient adherence, was created as an aggregate of medication and lifestyle adherence estimates. Preintervention and postintervention score correlation was set at 0.5. For studies presenting results as median and IQR or range, their data were converted into mean and SD.[25 26] SDs were also calculated from 95% CIs.[27]

Heterogeneity

Heterogeneity was assessed using the I^2 statistic. It was expressed as low (less than 25%), moderate (26%–74%) or high (more than 75%),[28] and importantly, interpreted according to its general context. Further explanations of high heterogeneity would be done in moderator analysis.

Moderator analysis

Subgroup analysis aimed to assess measures of effect according to intervention characteristics such as presence of cointervention, theoretical frameworks, content (diabetes self-management education/diabetes self-management support/DSMES), risk of bias assessment presence of comorbidities or complications and frequency (daily/weekly/monthly). We expected to do

metaregression using the following variables: age, gender ratio, sample size, duration of disease and mean HbA1c, BP and BMI levels.

Sensitivity analysis

We aimed to assess the robustness of review estimates. This was performed by changing of preintervention and postintervention value correlation from 0.5 to 0.2 and 0.8, respectively.

RESULTS

A total of 3460 articles were identified, from which 405 were duplicates or published in languages other than English. A total of 3080 articles were screened for eligibility. After screening using title and abstracts, 27 articles underwent full-text review from which 21 studies were excluded due to ineligible population, intervention, study design or outcome. Finally, six studies were included into the review, one of which was excluded from quantitative analysis due to insufficient data.[29] All studies were conducted in Asia and Europe. Further details are illustrated in figure 1.

Study characteristics were assessed for diabetes duration, participant age and gender ratio, comorbidities and medications (table 1). The review population consisted of 696 participants with an average age of 57.6 years (SD 10.59), and 320 (45.98%) were female. Duration of disease since diagnosis ranged from 6 months to 18 years, with all studies reporting presence of concurrent hypertension and dyslipidaemia in varying percentages. One study reported presence of complications such as neuropathy and kidney disease.[30]

Studies were randomised controlled trials (RCTs) with either as a single-blind or open-label trial, and all, except one,[31] were exploratory. One study was conducted in a tertiary clinic,[30] while the rest were conducted in primary healthcare (PHC) settings.

Interventions were used daily with the implementation period ranging from 6 weeks to 6 months. Two studies offered both self-management support and education.[29 32] The rest offered only self-management support. Use of theoretical framework was only reported in one study.[33] Compliance to mobile app interventions ranged from 69.6% to 88.0%.

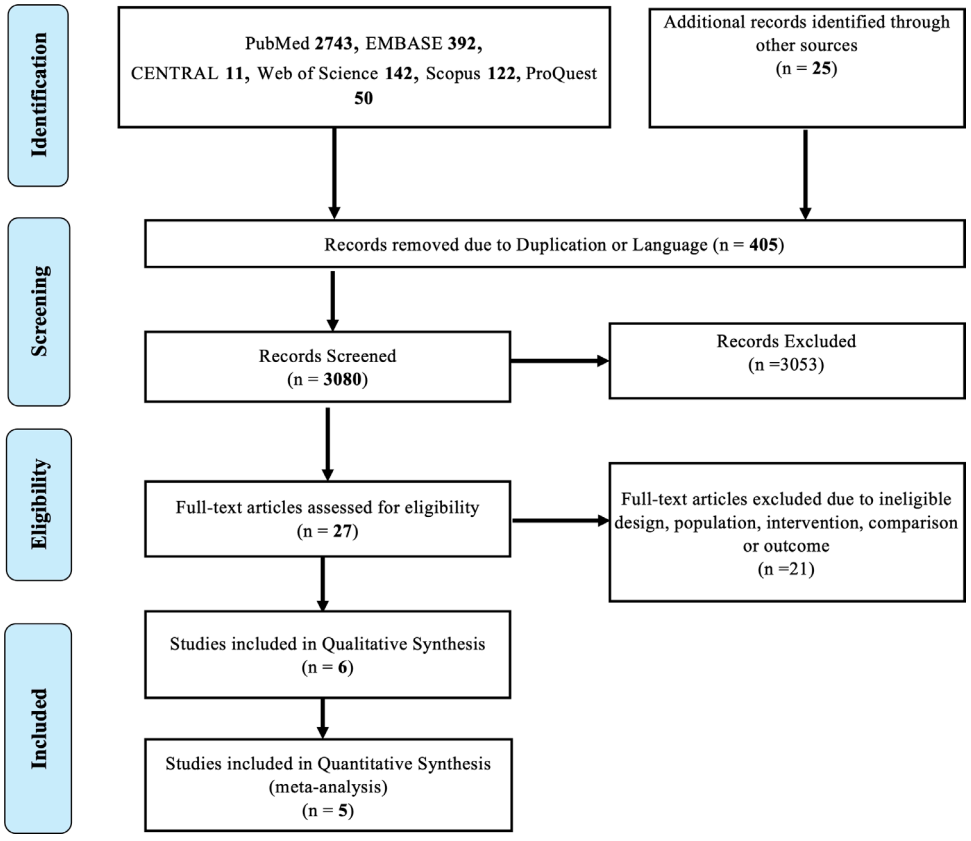

Figure 1 Research framework according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.

Table 1 Characteristics of included studies and interventions

ID	Population	Design	Intervention
ID, country	1. Diabetes duration. 2. Age. 3. Female ratio. 4. Comorbidities. 5. Medication.	1. Setting. 2. Sample size. 3. Intervention duration. 4. Blinding. 5. Design. 6. Control. 7. ROB.	1. Functions. 2. Cointerventions. 3. Theory. 4. DSME/DSMES.
Brath *et al*,[34] Austria	1. NR. 2. 69.4 (SD 4.8). 3. 30 (54.5%). 4. HTN and HCHL. 5. Oral AHs.	1. PHC. 2. I=53, C=53. 3. 6 months. 4. Single. 5. Crossover. 6. Standard blister. 7. High.	1.Received data from e-blister and transmitted to providers' server. 2. e-blister, provider portal and reminder calls. 3. NR. 4. DSMS.
Alonso-Domínguez *et al*,[32] Spain	1. 6.55 (SD 4.64). 2. 69.6 (SD 8.1). 3. 93 (45.6%). 4. HTN and HCHL. 5. Oral AHs.	1. PHC. 2. I=102, C=102. 3. 3 months. 4. Open-label. 5. Parallel. 6. Usual care. 7. Some concerns.	1. Entering food intake and daily exercise data, provide detailed information on nutritional deviations in terms of diet composition and number of calories. 2. Walks, workshops. 3. NR. 4. DSMES.
Huang *et al*,[30] Singapore	1. 14.44 (SD 8.4). 2. 49.08 (SD 11.4). 3. 21 (51.2%). 4. HTN and HCHL. 5. Oral AHs and insulin.	1. TC. 2. I=22, C=19. 3. 3 months. 4. Single. 5. Parallel. 6. Usual care. 7. Some concerns.	1. Medication scheduling, reminder, tracking, data sharing and medication adherence assessments. 2. NR. 3. DSMS.
Kardas *et al*,[29] Poland	1. NR. 2. 59.5 (SD 6.8). 3. 24 (40%). 4. HTN, HCHL and CAD. 5. Oral AHs.	1. PHC. 2. I=30, C=30. 3. 6 weeks. 4. Open-label. 5. Parallel. 6. Usual care. 7. Some concerns.	1. Clinical monitoring, transmission (ECG, glucose, BP and respiration). 2. MEMS. 3. NR. 4. DSMES.
Kleinman *et al*,[33] India	1. 9.67 (SD 8.29). 2. 48.4 (SD 9.2). 3. 30 (30%). 4. HTN and HCHL. 5. Oral AHs and insulin.	1. PHC. 2. I=44, C=46. 3. 6 months. 4. Single. 5. Parallel. 6. Usual care. 7. Low.	1. Reminders, data visualisation, and ongoing support to increase self-care behaviours and to facilitate collaborative care decisions and interactions with providers. 2. Provider portal and app. 3. Health Belief Model (HBM) theory of planned behaviour and Bandura's theory of self-efficacy. 4. DSMS.
Yang *et al*,[31] South Korea	1. NR. 2. 57.6 (SD 10.59). 3. 122 (49.4%). 4. HTN and HCHL. 5. Oral AHs and insulin.	1. PHC. 2. I=150, C=97. 3. 3 months. 4. Open-label. 5. Parallel cluster. 6. Usual care. 7. Some concerns.	1. Clinical monitoring and transmission to provider portals. 2. Short Message Service (SMS),[12] provider portal. 3. NR. 4. DSMS.

Comorbidities: HTN, HCHL and CAD.
Settings: PHC and TC.
Sample size: I and C.
AH, antihyperglycaemic; BP, blood pressure; C, control group; CAD, coronary artery disease; DSMES, diabetes self-management education and support; DSMS, diabetes self-management support; HBM, Health Belief Model; HCHL, hypercholesterolaemia; HTN, hypertension; I, intervention group; MEMS, Medication Event Monitoring System; NR, not reported; PHC, primary healthcare; ROB, risk of bias; SMS, Short Message Service; TC, tertiary care.

Mobile app functions included recording food intake, physical activity and clinical information such as BP and glucose levels.[29–34] Other functions included alerting users if their uploaded data deviated from prespecified standards.[30 32] Intervention apps also helped participants track and schedule medication, visualise and transmit data to providers and act as automated reminders.[29–34] For measurement of clinical data, participants were equipped with instruments such as glucose monitors and strips.

The most common cointervention was provider portals which helped providers review transmitted patient data which they used to offer feedback to participants.[31 33 34] Additional reminders were provided in two studies.[31 34] Brath *et al*[34] offered e-blisters, which helped participants adhere to their medication.[34] Most interventions were guided by physicians. Control groups were offered usual clinic consultations, except for one study in which participants also received standard blisters and medication diaries.[34]

Acceptability was high among participants. Participants commended apps for their speed, ease of use and convenience.[29 30 34] Other participants reported that they liked that providers were familiar with their regimens.[34] Participants indicated that interventions would improve their self-confidence in managing diseases and that they would use app afterwards.[29 30]

In terms of operation ability, participants were expected to have a working knowledge of phones and apps. In addition, researchers offered training to all participants to ensure proper use of interventions. Some of the challenges reported included technical complexity,[32 34] frequent need to recharge gadgets and lack of glucose strips.[29]

Quality of studies

In terms of risk of bias, two studies were deemed as high[34] and low risks.[33] The rest were judged as having some concerns.[29–32] Major concerns were inadequate reporting of randomisation process,[29–31 34] possibility of reporting bias,[30–32] lack of blinding[30–32 34] and lack of optimal handling of missing data.[30 31 34] Risk of bias assessment is illustrated in figure 2. Attrition rates ranged from 0% to 31%. Publication bias was not detected (p=0.398).

Moderator analysis and sensitivity analysis

Due to the small number of studies, subgroup analyses were not performed. Changing precorrelation/postcorrelation from 0.5 to 0.2 and 0.8 did not significantly alter review effect estimates (see online supplemental figures S2–7).

Figure 2 Risk of bias assessment chart. Note: clinical outcomes: blood pressure, glycated haemoglobin, Body Mass Index and total cholesterol.

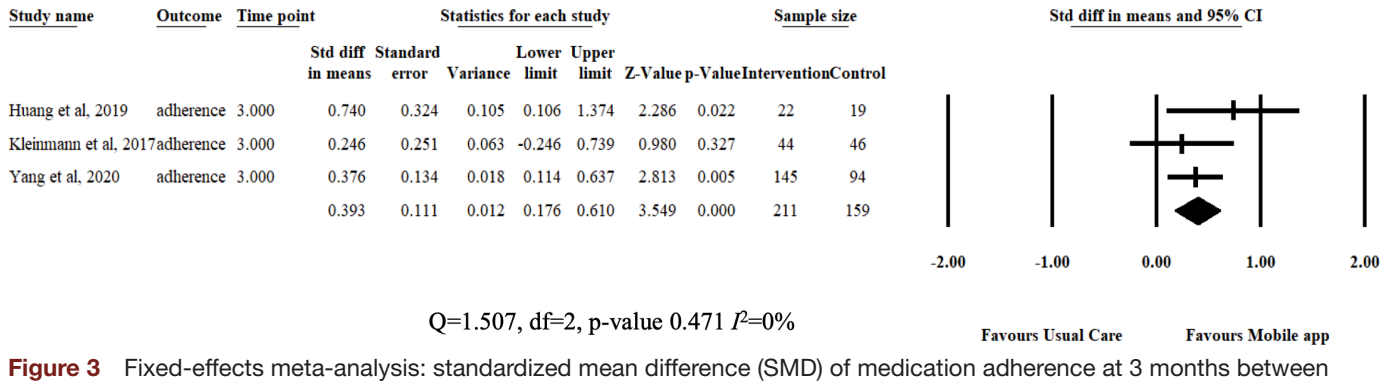

Figure 3 Fixed-effects meta-analysis: standardized mean difference (SMD) of medication adherence at 3 months between diabetes self-management education and support app and usual care groups.

Intervention effect

Review estimates were aggregated from study results at 3 months from baseline measurements. Pooled analysis showed that DSMES app improved medication adherence (SMD=0.393, 95% CI 0.17 to 0.61; I^2=0%) and patient adherence (SMD=0.632, 95% CI 0.17 to 1.094; I^2=83%) compared with usual care. Interventions also reduced Hba1c (MD=−0.314, 95% CI −0.477 to −0.151; I^2=0%) and BMI (MD=−0.28, 95% CI −0.545 to −0.015; I^2=0%). The authors did not pool results for BP (few studies) and total cholesterol (high heterogeneity). Results are illustrated in figures 3–5.

DISCUSSION

After 3 months of DSMES app intervention, participants had a small but significant improvement in medication adherence compared with those receiving usual care. The intervention also had a small and moderate effect on HbA1c and BMI, respectively.

Medication adherence improved significantly after receiving 3 months of intervention. Intervention effect was also highly consistent among students (I^2=0%). Improved medication adherence is a good development as it is associated with positive health outcomes and delay in development of DM complications.[35] However, achieving optimal adherence is an ongoing challenge as patients are required to strictly follow dosages and recommendations from initiation to long-term continuation of disease management throughout their lives.[36] The current review shows that DSMES apps could help by acting as reminders and communication channels, scheduling and calculating doses as well. In a way, they do unburden patients and allow them to sync DM management effortlessly into their lives. A review by Peng *et al* found similar findings in participants with various chronic diseases.[18] The use of simple messages has been widely studied with contrasting results,[10 17] including the use of mobile apps to improve medication adherence.

Lifestyle adherence was measured in one study.[32] Thus, a postanalysis outcome, patient adherence, was made by aggregating medication and lifestyle adherence outcomes. The pooled estimate showed a moderate and significant effect compared with usual care, though there was high heterogeneity observed. Adherence to lifestyle recommendations is important as they play a crucial role in controlling T2D. Patients who manage the diet and increase physical activity are more likely to fare well in terms of health outcomes.[37] Lifestyle therapy adherence is also crucial to controlling other comorbidities such as hypertension, which were present in all studies.[38 39]

Studies all had cointerventions; most commonly, provider portals and independent reminders were used. These components illustrate the importance of patient–provider interaction, which have proven useful in improving other health outcome.[40] The review does not have evidence that this feature, or any other intervention component or population characteristic, was vital in improving adherence partly due to failure to conduct moderator analysis.

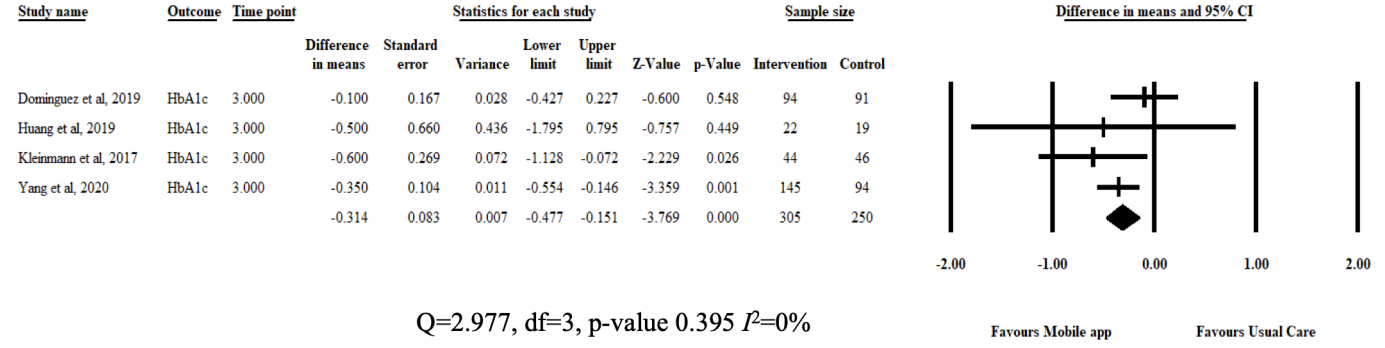

Figure 4 Fixed-effects meta-analysis: mean difference (MD) of glycated haemoglobin at 3 months between diabetes self-management education and support app and usual care groups.

Study name	Outcome	Time point	Statistics for each study							Sample size	
			Difference in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Intervention	Control
Dominguez et al, 2019	BMI	3.000	-0.300	0.152	0.023	-0.598	-0.002	-1.972	0.049	94	91
Huang et al, 2019	BMI	3.000	-3.700	2.779	7.723	-9.147	1.747	-1.331	0.183	22	18
Kleinmann et al, 2017	BMI	3.000	-1.100	1.080	1.166	-3.217	1.017	-1.019	0.308	44	46
Yang et al, 2020	BMI	3.000	-0.090	0.307	0.095	-0.693	0.513	-0.293	0.770	145	94
			-0.280	0.135	0.018	-0.545	-0.015	-2.072	0.038	305	249

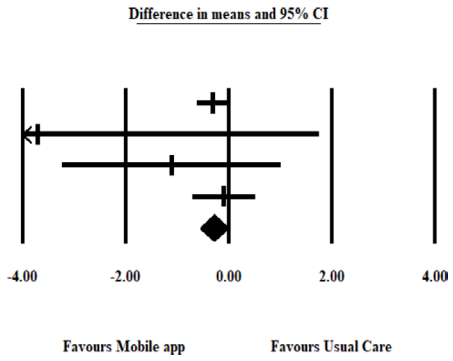

Figure 5 Fixed-effects meta-analysis: mean difference (MD) of Body Mass Index at 3 months between diabetes self-management education and support app and usual care groups.

In terms of DSMES, with advances in technology, healthcare providers are also transitioning from or at least complimenting traditional delivery media with digital tools.[14] As earlier stated, DSMES covers a lot of crucial topics including patient adherence to treatment.[11] However, from the current review, studies show that mobile apps are mostly being used to offer self-management support or in combination with diabetes education.[30–34] It is likely that pure diabetes education services are offered using either online or messaging services. Regardless, DSMES apps have the potential to offer both diabetes education and self-management support.[32] Thus, implementers of DSMES interventions would ideally design a separate component of education using tools like websites and Short Message Service (SMS) or incorporate the education component into a self-management mobile app.

Our pooled effect estimate on HbA1c showed a small but significant reduction compared with usual care. The findings are in line previous literature.[41 42] HbA1c is an important indicator of disease control though its efficacy as a screening tool is still debated.[43] DSMES app functions, such as glucose monitoring and calculation of doses, would have played a role in improving HbA1c in review participants. In addition, medication adherence may have also influenced glycaemic control.[44] Further research is required to shed more light on the association between HbA1c levels and medication adherence.

BMI is a direct indicator of patient adherence to lifestyle recommendations. It is noticeable that both patient adherence and BMI significantly improved, begging the question: how, if at all, are the two outcomes correlated? However, as earlier stated, the review was underpowered to conduct such analysis. BMI is crucial for patients with T2D, as increased BMI is an important risk factor in developing adverse cardiometabolic events and other noncommunicable diseases.[37 45] Managing BMI through lifestyle modifications should be given as much emphasis as medical treatment. Most treatment protocols, of course, include lifestyle management.[46] However, few interventions targeted lifestyle therapies.[32 33] On the other hand, studies that target lifestyle interventions seldom included patients with T2D. Consequently, most of these studies were excluded from the review.

Applicability

Intervention group participants were generally accepting of DSMES apps. Major advantages reported were convenience, ease of use, and being motivational and knowledgeable. Interventions acted as a medium of communication, increasing frequency of interactions from usual month or quarterly consultation to at time weekly. This advantage could possibly help providers spot anomalies faster, address emerging problems earlier and eventually reduce medical costs.

In terms of feasibility, first, for the system to work, there needs technical expertise, and second, patients are required to be digitally literate. The review's population was mostly old adults who might not be very technologically adept.[18] However, studies reported assistance to participants whenever it was required.

Strengths

This review was tailored to the use of DSMES apps by patients with T2D and their impact on adherence. Results of all prespecified outcomes have shown a high consistency among included studies. Studies are all RCTs and were mostly conducted in PHC settings, which is generalisable to most patients with T2D.

Limitations

Several factors need to be taken into consideration. First, studies are mostly feasible trials. Second, few studies meant reviewers could not achieve moderator analysis. The first two limitations could indicate that relatively little research has been conducted on this topic. Third, it is difficult to ascertain which intervention component was vital in improving outcomes. Adherence was self-reported in some studies, raising the possibility of reporter bias. Fourth, use of theoretical frameworks was reported in only one study,[33] although they are known to be important in development of interventions.

Further research

First, DSMES app intervention needs to be tested in larger trials. In addition, the review reports results measured at 3 months. There is a need to study the long-term impact of DSMES apps on both medication and lifestyle therapy

adherence. Furthermore, long-term follow-up would allow the measurement of outcomes, such as persistence. Lastly, use of theoretical frameworks during intervention implementation is highly recommended in future studies.[10]

CONCLUSION

Pooled analysis showed that DSME apps had significant small to moderate effects on medication adherence, HbA1c and BMI of patients with T2D in comparison to usual care. All results had low heterogeneity. Participant feedback showed apps were reliable, easy to use and convenient, though most required to be phone literate. Evidence comes from feasibility trials with generally moderate risk of bias. Larger trials with longer follow-up periods using theory-based interventions are required to improved current evidence.

Author affiliations
[1]Nkhata Bay District Health Office, Nkhata Bay District Hospital, Nkhata Bay, Malawi
[2]Global Health and Development Department, College of Public Health, Taipei Medical University, Taipei, Taiwan
[3]PhD Program in Nursing, College of Nursing, Taipei Medical University, Taipei, Taiwan
[4]Australian Institute of Health Service Management, Tasmanian School of Business and Economics, University of Tasmania, Sydney, New South Wales, Australia
[5]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan
[6]International Center for Health Information Technology (ICHIT), Taipei Medical University, Taipei, Taiwan
[7]Department of Dermatology, Taipei Municipal Wan Fang Hospital, Taipei, Taiwan
[8]International Medical Informatics Association (IMIA), Geneva, Switzerland
[9]Research Center for Artificial Intelligence in Medicine, Taipei Medical University, Taipei, Taiwan
[10]PhD Program of Global Health and Health Security, College of Public Health, Taipei Medical University, Taipei, Taiwan

Twitter Dumisani Enricho Nkhoma @dnkhoma

Contributors DEN and UI conceived the idea of the paper. DEN and CJS searched, screened articles and collected data. DEN and KJB did the quality assessment; DEN and CJS conducted the analysis, supervised by UI; and KJB, DEN and CJS did the paper writing, supervised and edited by KJB, DG, Y-CL and UI.

Funding This work was supported in part by Ministry of Science and Technology (MOST), Taiwan (ROC), under grant MOST109-2221-E-038-015.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The review collected secondary information reported in published studies in which consent was already sort from study participants. Therefore, no ethical approval was required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No available data to share. The current systematic review and meta-analysis study used previous published research papers which have been cited in the manuscript.

ORCID iDs
Yu-Chuan (Jack) Li http://orcid.org/0000-0001-6497-4232
Usman Iqbal http://orcid.org/0000-0002-0614-123X

REFERENCES

1 Kaiser AMYB, Zhang N, Der Pluijm WVAN. Global prevalence of type 2 diabetes over the next ten years (2018-2028). *Diabetes* 2018;67:202.
2 World Health Organization. *Fact sheet: the top 10 causes of death*. Geneva, Switzerland: World Health Organization, 2017.
3 Bommer C, Sagalova V, Heesemann E, *et al*. Global economic burden of diabetes in adults: projections from 2015 to 2030. *Diabetes Care* 2018;41:963–70.
4 World Health Organization. *Global report on diabetes*. Geneva: World Health Organization, 2016.
5 Asche C, LaFleur J, Conner C. A review of diabetes treatment adherence and the association with clinical and economic outcomes. *Clin Ther* 2011;33:74–109.
6 Lerman I. Adherence to treatment: the key for avoiding long-term complications of diabetes. *Arch Med Res* 2005;36:300–6.
7 Organization WH. Adherence to long-term therapies: evidence for action: World Health organization 2003.
8 Cramer JA. A systematic review of adherence with medications for diabetes. *Diabetes Care* 2004;27:1218–24.
9 Lam WY, Fresco P. Medication adherence measures: an overview. *Biomed Res Int* 2015;2015:1–12.
10 Farmer AJ, McSharry J, Rowbotham S, *et al*. Effects of interventions promoting monitoring of medication use and brief messaging on medication adherence for people with type 2 diabetes: a systematic review of randomized trials. *Diabet Med* 2016;33:565–79.
11 Beck J, Greenwood DA, Blanton L, *et al*. 2017 national standards for diabetes self-management education and support. *Diabetes Educ* 2018;44:35–50.
12 Dayer L, Heldenbrand S, Anderson P, *et al*. Smartphone medication adherence apps: potential benefits to patients and providers. *J Am Pharm Assoc* 2013;53:172–81.
13 World Health Organization. eHealth Geneva: World Health organization, 2019. Available: https://www.who.int/ehealth/en/ [Accessed 1 Apr 2019].
14 American Diabetes Association. 7. Diabetes Technology: Standards of Medical Care in Diabetes-2019. *Diabetes Care* 2019;42:S71–80.
15 Jahns R. Currently, only 1.2% of diabetics that have a smartphone use a diabetes app, 2018. Available: http://research2guidance com/currently-only-12-of-diabetics-that-have-a-smartphone-use-a-diabetes-app-until-2018-the-share-will-rise-to-78-globally/ [Accessed 16 Aug 2015].
16 Adams A, Shankar M, Tecco H. 50 things we now know about digital health consumers. *Rock Health* 2016.
17 Thakkar J, Kurup R, Laba T-L, *et al*. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. *JAMA Intern Med* 2016;176:340–9.
18 Peng Y, Wang H, Fang Q, *et al*. Effectiveness of mobile applications on medication adherence in adults with chronic diseases: a systematic review and meta-analysis. *J Manag Care Spec Pharm* 2020;26:550–61.
19 Lunde P, Nilsson BB, Bergland A, *et al*. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. *J Med Internet Res* 2018;20:e162.
20 Wu X, Guo X, Zhang Z. The efficacy of mobile phone Apps for lifestyle modification in diabetes: systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2019;7:e12297.
21 Vermeire E, Wens J, Van Royen P, *et al*. Interventions for improving adherence to treatment recommendations in people with type 2 diabetes mellitus. Cochrane database of systematic reviews 2005.
22 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
23 Cochrane Public Health. Review authors: cochrane collaboration, 2019. Available: https://ph.cochrane.org/review-authors

24 Sterne JAC, Savović J, Page MJ, *et al*. RoB 2: a revised tool for assessing risk of bias in randomised trials 2019;366:l4898.
25 Wan X, Wang W, Liu J, *et al*. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol* 2014;14:135.
26 Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 2005;5:13.
27 Higgins JP, Green S. *Cochrane Handbook for systematic reviews of interventions*. John Wiley & Sons, 2011.
28 Higgins JPT, Thompson SG, Deeks JJ, *et al*. Measuring inconsistency in meta-analyses 2003;327:557–60.
29 Kardas P, Lewandowski K, Bromuri S. Type 2 diabetes patients benefit from the COMODITY12 mHealth system: results of a randomised trial. *J Med Syst* 2016;40:259.
30 Huang Z, Tan E, Lum E, *et al*. A smartphone APP to improve medication adherence in patients with type 2 diabetes in Asia: feasibility randomized controlled trial. *JMIR Mhealth Uhealth* 2019;7:e14914.
31 Yang Y, Lee EY, Kim H-S, *et al*. Effect of a mobile Phone-Based Glucose-Monitoring and feedback system for type 2 diabetes management in multiple primary care clinic settings: cluster randomized controlled trial. *JMIR Mhealth Uhealth* 2020;8:e16266.
32 Alonso-Domínguez R, García-Ortiz L, Patino-Alonso MC, *et al*. Effectiveness of a multifactorial intervention in increasing adherence to the Mediterranean diet among patients with diabetes mellitus type 2: a controlled and randomized study (EMID study). *Nutrients* 201910.3390/nu11010162. [Epub ahead of print: 14 Jan 2019].
33 Kleinman NJ, Shah A, Shah S, *et al*. Improved medication adherence and frequency of blood glucose Self-Testing using an m-Health platform versus usual care in a multisite randomized clinical trial among people with type 2 diabetes in India. *Telemed J E Health* 2017;23:733–40.
34 Brath H, Morak J, Kästenbauer T, *et al*. Mobile health (mHealth) based medication adherence measurement - a pilot trial using electronic blisters in diabetes patients. *Br J Clin Pharmacol* 2013;76:47–55.
35 Krass I, Schieback P, Dhippayom T. Adherence to diabetes medication: a systematic review. *Diabet Med* 2015;32:725–37.
36 Vrijens B, De Geest S, Hughes DA, *et al*. A new taxonomy for describing and defining adherence to medications. *Br J Clin Pharmacol* 2012;73:691–705.
37 Klein S, Sheard NF, Pi-Sunyer X, *et al*. Weight management through lifestyle modification for the prevention and management of type 2 diabetes: rationale and strategies. A statement of the American diabetes association, the North American association for the study of obesity, and the American Society for clinical nutrition. *Am J Clin Nutr* 2004;80:257–63.
38 Castaneda C, Layne JE, Munoz-Orians L, *et al*. A randomized controlled trial of resistance exercise training to improve glycemic control in older adults with type 2 diabetes. *Diabetes Care* 2002;25:2335–41.
39 Esposito K, Maiorino MI, Bellastella G, *et al*. A journey into a Mediterranean diet and type 2 diabetes: a systematic review with meta-analyses. *BMJ Open* 2015;5:e008222.
40 Chatterjee S, Davies MJ, Heller S, *et al*. Diabetes structured self-management education programmes: a narrative review and current innovations. *Lancet Diabetes Endocrinol* 2018;6:130–42.
41 Liu K, Xie Z, Or CK. Effectiveness of mobile App-Assisted self-care interventions for improving patient outcomes in type 2 diabetes and/or hypertension: systematic review and meta-analysis of randomized controlled trials. *JMIR Mhealth Uhealth* 2020;8:e15779.
42 Hou C, Carter B, Hewitt J, *et al*. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and grade of 14 randomized trials. *Diabetes Care* 2016;39:2089–95.
43 Larese J. When is hemoglobin A1c inaccurate in assessing glycemic control. NYU Langone internal medicine Blog, faculty peer reviewed 2012.
44 Wabe NT, Angamo MT, Hussein S. Medication adherence in diabetes mellitus and self management practices among type-2 diabetics in Ethiopia. *N Am J Med Sci* 2011;3:418–23.
45 Look AHEAD Research Group, Wing RR. Long-Term effects of a lifestyle intervention on weight and cardiovascular risk factors in individuals with type 2 diabetes mellitus: four-year results of the look ahead trial. *Arch Intern Med* 2010;170:1566–75.
46 American Diabetes Association. 4. Lifestyle Management: Standards of Medical Care in Diabetes-2018. *Diabetes Care* 2018;41:S38–50.

BMJ Health & Care Informatics

Equity in essence: a call for operationalising fairness in machine learning for healthcare

Judy Wawira Gichoya,[1,2] Liam G McCoy,[3] Leo Anthony Celi ,[4,5,6] Marzyeh Ghassemi[7,8,9]

To cite: Wawira Gichoya J, McCoy LG, Celi LA, *et al*. Equity in essence: a call for operationalising fairness in machine learning for healthcare. *BMJ Health Care Inform* 2021;**28**:e100289. doi:10.1136/bmjhci-2020-100289

Received 22 November 2020
Revised 07 February 2021
Accepted 09 February 2021

For numbered affiliations see end of article.

Correspondence to
Liam G McCoy;
liam.mccoy@mail.utoronto.ca

INTRODUCTION

Machine learning for healthcare (MLHC) is at the juncture of leaping from the pages of journals and conference proceedings to clinical implementation at the bedside. Succeeding in this endeavour requires the synthesis of insights from both the machine learning and healthcare domains, in order to ensure that the unique characteristics of MLHC are leveraged to maximise benefits and minimise risks. An important part of this effort is establishing and formalising processes and procedures for characterising these tools and assessing their performance. Meaningful progress in this direction can be found in recently developed guidelines for the development of MLHC models,[1] guidelines for the design and reporting of MLHC clinical trials,[2 3] and protocols for the regulatory assessment of MLHC tools.[4 5]

But while such guidelines and protocols engage extensively with relevant technical considerations, engagement with issues of fairness, bias and unintended disparate impact is lacking. Such issues have taken on a place of prominence in the broader ML community,[6–9] with recent work highlighting issues such as racial disparities in the accuracy of facial recognition and gender classification software,[6 10] gender bias in the output of natural language processing models[11 12] and racial bias in algorithms for bail and criminal sentencing.[13] MLHC is not immune to these concerns, as seen in disparate outcomes from algorithms for allocating healthcare resources,[14 15] bias in language models developed on clinical notes[16] and melanoma detection models developed primarily on images of light-coloured skin.[17] Within this paper, we will examine the inclusion of fairness in recent guidelines for MLHC model reporting, clinical trials and regulatory approval. We highlight opportunities to ensure that fairness is made fundamental to MLHC, and examine ways how this can be operationalised for the MLHC context.

FAIRNESS AS AN AFTERTHOUGHT?

Model development and trial reporting guidelines

Several recent documents have attempted, with varying degrees of practical implication, to enumerate guiding principles for MLHC. Broadly, these documents do an excellent job of highlighting artificial intelligence (AI)-specific technical and operational concerns, such as how to handle human-AI interaction, or how to account for model performance errors. Yet as outlined in table 1, references to fairness are either conspicuously absent, made merely in passing, or relegated to supplemental discussion.

Notable examples are the recent the Standard Protocol Items: Recommendations for Interventional Trials-AI (SPIRIT-AI)[2] and Consolidated Standards of Reporting Trials-AI (CONSORT-AI)[3] extensions, which expand prominent guidelines for the design and reporting of AI clinical trials to include concerns relevant to AI. While the latter states in the discussion that 'investigators should also be encouraged to explore differences in performance and error rates across population subgroups',[3] there is no more formal inclusion of the concept into the guideline itself. Similarly, the announcement papers for the upcoming Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis-ML (TRIPOD-ML)[18] andStandards for Reporting of Diagnostic Accuracy Studies AI Extension (STARD-AI)[19] guidelines for model reporting do not allude to these issues (though we wait in anticipation for their potential inclusion in the final versions of these guidelines). While recently published guidelines from the editors of

Table 1 Fairness in recently released and upcoming guidelines

Guideline	How is fairness included?
Reporting guidelines	
Development and Reporting of Prediction Models: Guidance for Authors From Editors of Respiratory, Sleep, and Critical Care Journals[1]	Discussion of the risk of unfairness is included in https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7161722/bin/ccm-48-0623-s001.docx but not the main document.
TRIPOD-ML (Announcement Statement Only)[18]	No explicit mention.
STARD-AI (Announcement Statement Only)[19]	No explicit mention.
Checklist for Artificial Intelligence in Medical Imaging[31]	Bias discussed, but not clearly in the context of fairness with respect to differential performance or impact between patient groups.
Clinical Trial Guidelines	
CONSORT-AI Extension[3]	Fairness is brought up in the discussion section but not included explicitly in any of the guideline checklist points.
SPIRIT-AI Extension[2]	No explicit mention.

CONSORT-AI, Consolidated Standards of Reporting Trials–Artificial Intelligence; SPIRIT-AI, Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence; STARD-AI, Standards for Reporting of Diagnostic Accuracy Studies–Artificial Intelligence ; TRIPOD-ML, Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis–Machine Learning.

respiratory, sleep and critical care medicine journals engage with the concept in an exemplary fashion, the depth of their discussion is relegated to a supplementary segment of the paper.[1]

Regulatory guidance

Broadly, the engagement of prominent regulatory bodies with MLHC remains at a preliminary stage, and engagement with fairness tends to be either minimal or vague. The Food and Drug Administration in the USA has made significant strides towards modernisation of its frameworks for the approval and regulation of software-based medical interventions, including MLHC tools.[5] Their documents engage broadly with technical concerns, and criteria for effective clinical evaluation, but entirely lack discussion of fairness or the relationship between these tools and the broader health equity context.[20] The Canadian Agency for Drugs and Technologies in Health has explicitly highlighted the need for fairness and bias to be considered, but further elaboration is lacking.[21]

The work of the European Union on this topic remains at a broad stage.[4] While their documents do make reference to principles of 'diversity, non-discrimination and fairness', they do so in a very broad manner without any clearly operationalised specifics.[22 23] The engagement of the UK with MLHC is relatively advanced, with several prominent reports engaging with the topic,[24–26] and an explicit 'Code of Conduct for Data-Driven Healthcare Technology'[27] from the Department of Health and Social Care that highlights the need for fairness. However, the specifics of this regulatory approach are still being decided, and no clear guidance has yet been put forth to clarify these principles in practice.[28] MLHC as a whole would benefit from increased clarity and force in regulatory guidance from these major agencies.[29]

OPERATIONALISING FAIRNESS IN MLHC PRACTICE

If fairness is an afterthought in the design and reporting of MLHC papers and trials, as well as regulatory processes, it is likely to remain an afterthought in the development and implementation of MLHC tools. If MLHC is going to prove effective for— and be trusted by—a diverse range of patients, fairness cannot be a post-hoc and after-the-fact consideration. Nor is it sufficient for fairness to be a vague abstraction of academic importance but ineffectual consequence. The present moment affords a tremendous opportunity to define MLHC such that fairness is integral, and to ensure that this commitment is reflected in model reporting guidelines, clinical trial guidelines and regulatory approaches.

However, moving from vague commitments of fairness to practical and effective guidance is far from a trivial task. As work in the machine learning community has demonstrated, fairness has multiple definitions which can occasionally be incompatible,[7] and bias can arise from a complex range of sources.[30] Operationalisation of fairness must be context-specific, and embeds the relevant values in a field. We call for concerted effort from the MLHC community, and in particular the groups responsible for the development and propagation of guidelines, to affirm a commitment to fairness in an explicit and operationalised fashion. Similarly, we call on the various regulatory agencies to establish clear minimum standards for AI fairness. In box 1, we highlight a non-exhaustive series of recommendations that are likely to be beneficial as the MLHC community engages in this endeavour.

Box 1 Recommendations for operationalising fairness

Recommendations

- Engage members of the public and in particular members of marginalised communities in the process of determining acceptable fairness standards.
- Collect necessary data on vulnerable protected groups in order to perform audits of model function (eg, on race, gender).
- Analyse and report model performance for different intersectional subpopulations at risk of unfair outcomes.
- Establish target thresholds and maximum disparities for model function between groups.
- Be transparent regarding the specific definitions of fairness that are used in the evaluation of a machine learning for healthcare (MLHC) model.
- Explicitly evaluate for disparate treatment and disparate impact in MLHC clinical trials.
- Commit to postmarketing surveillance to assess the ongoing real-world impact of MLHC models.

CONCLUSION

Values are embedded throughout the MLHC pipeline, from the design of models, to the execution and reporting of trials, to the regulatory approval process. Guidelines hold significant power in defining what is worthy of emphasis. While fairness is essential to the impact and consequences of MLHC tools, the concept is often conspicuously absent or ineffectually vague in emerging guidelines. The field of machine MLHC has the opportunity at this juncture to render fairness integral to the identity field. We call on the MLHC community to commit to the project of operationalising fairness, and to emphasise fairness as a requirement in practice.

Author affiliations
[1]Department of Radiology & Imaging Sciences, Emory University, Atlanta, Georgia, USA
[2]Fogarty International Center, National Institutes of Health (NIH), Bethesda, Maryland, USA
[3]Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada
[4]Laboratory for Computational Physiology, Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts, USA
[5]Division of Pulmonary Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[6]Department of Biostatistics, Harvrd T.H. Chan School of Public Health, Boston, Massachusetts, USA
[7]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada
[8]Department of Medicine, University of Toronto, Toronto, Ontario, Canada
[9]Vector Institute for Artificial Intelligence, Toronto, Ontario, Canada

Twitter Judy Wawira Gichoya @judywawira, Liam G McCoy @liamgmccoy and Leo Anthony Celi @MITCriticalData

Contributors Initial conceptions and design: JWG, LGM, MG and LAC. Drafting of the paper: LGM, JWG, MG and LAC. Critical revision of the paper for important intellectual content: JWG, LGM, MG and LAC.

Funding Division of Electrical, Communications and Cyber Systems (1928481), National Institute of Biomedical Imaging and Bioengineering (EB017205).

Competing interests MG acts as an advisor to Radical Ventures in Toronto.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD
Leo Anthony Celi http://orcid.org/0000-0001-6712-6626

REFERENCES

1 Leisman DE, Harhay MO, Lederer DJ, *et al*. Development and reporting of prediction models: guidance for authors from editors of respiratory, sleep, and critical care journals. *Crit Care Med* 2020;48:623–33.
2 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.
3 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.
4 Cohen IG, Evgeniou T, Gerke S, *et al*. The European artificial intelligence strategy: implications and challenges for digital health. *Lancet Digit Health* 2020;2:e376–9.
5 FDA. Artificial intelligence and machine learning in software as a medical device, 2020. Available: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device [Accessed 11 Oct 2020].
6 Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. *Conference on Fairness, Accountability and Transparency*, 2018:77–91.
7 Gajane P, Pechenizkiy M. On formalizing fairness in prediction with machine learning. arXiv:171003184 [cs, stat], 2018. Available: http://arxiv.org/abs/1710.03184 [Accessed 20 Sept 2020].
8 et alMehrabi N, Morstatter F, Saxena N. A survey on bias and fairness in machine learning. arXiv:190809635 [cs], 2019. Available: http://arxiv.org/abs/1908.09635 [Accessed 11 Oct 2020].
9 De-Arteaga M, Romanov A, Wallach H. Bias in bios: a case study of semantic representation bias in a High-Stakes setting. *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19. Published online*, 2019:120–8.
10 Klare BF, Burge MJ, Klontz JC, *et al*. Face recognition performance: role of demographic information. *IEEE Transactions on Information Forensics and Security* 2012;7:1789–801.
11 Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. *Science* 2017;356:183–6.
12 Bordia S, Bowman SR. Identifying and reducing gender bias in word-level language models. arXiv:190403035 [cs], 2019. Available: http://arxiv.org/abs/1904.03035 [Accessed 30 Jan 2021].
13 Huq AZ. Racial equity in algorithmic criminal justice. *Duke LJ* 2018;68:1043.
14 Obermeyer Z, Powers B, Vogeli C, *et al*. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447–53.
15 Benjamin R. Assessing risk, automating racism. *Science* 2019;366:421–2.
16 Zhang H, AX L, Abdalla M. Hurtful words: quantifying biases in clinical contextual word embeddings. *Proceedings of the ACM Conference on Health, Inference, and Learning. CHIL '20. Association for Computing Machinery*, 2020:110–20.
17 Adamson AS, Smith A. Machine learning and health care disparities in dermatology. *JAMA Dermatol* 2018;154:1247–8.
18 Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet* 2019;393:1577–9.
19 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.
20 Ferryman K. Addressing health disparities in the food and drug administration's artificial intelligence and machine learning regulatory framework. *J Am Med Inform Assoc* 2020;27:2016–9.
21 Mason A, Morrison A, Visintini S. *An overview of clinical applications of artificial intelligence*. Ottawa: CADTH, 2018.
22 Commission E. *COM(2019) 168 final: building trust in human-centric artificial intelligence*, 2019.
23 Commission E. *White paper on artificial intelligence–a European approach to excellence and trust*, 2020.
24 Tankelevitch L, Ahn A, Paterson R. *Advancing AI in the NHS*, 2018.
25 Fenech M, Strukelj N, Buston O. *Ethical, social, and political challenges of artificial intelligence in health*. London: Wellcome Trust Future Advocacy, 2018.
26 Topol E. *The Topol review: preparing the healthcare workforce to deliver the digital future*. Health Education England, 2019.
27 Department of Health and Social Care. Code of conduct for data-driven health and care technology, 2019. Available: https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology [Accessed 1 Aug 2020].
28 NHS. Regulating AI in health and care - Technology in the NHS. Available: https://healthtech.blog.gov.uk/2020/02/12/regulating-ai-in-health-and-care/ [Accessed 12 Oct 2020].
29 Parikh RB, Obermeyer Z, Navathe AS. Regulation of predictive analytics in medicine. *Science* 2019;363:810–2.
30 Suresh H, Guttag JV. A framework for understanding unintended consequences of machine learning. arXiv:190110002 [cs, stat], 2020. Available: http://arxiv.org/abs/1901.10002 [Accessed 20 Sept 2020].
31 Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (claim): a guide for authors and reviewers. *Radiology* 2020;2:e200029.

BMJ Health & Care Informatics

Electronic display of a patient treatment over time: a perspective on clinicians' burn-out

Valentina Lichtner ,[1,2] Melissa Baysari[3]

To cite: Lichtner V, Baysari M. Electronic display of a patient treatment over time: a perspective on clinicians' burn-out. *BMJ Health Care Inform* 2021;**28**:e100281. doi:10.1136/bmjhci-2020-100281

Received 02 December 2020
Accepted 03 February 2021

[1]UCL School of Pharmacy, University College London, London, UK
[2]Leeds University Business School, University of Leeds, Leeds, UK
[3]Faculty of Medicine and Health, Charles Perkins Centre, The University of Sydney, Sydney, New South Wales, Australia

Correspondence to
Dr Valentina Lichtner;
v.lichtner@leeds.ac.uk

INTRODUCTION

Burn-out, a state of mental exhaustion caused by one's profession, is particularly acute among clinicians, especially in the USA.[1–3] Oncology clinicians seem to be particularly affected.[4–8] It is now acknowledged that use of electronic health record systems (EHRs), contributes to clinicians' dissatisfaction and burn-out[9–13] predominantly via the excessive requirements posed on clinicians for data entry.[14] A recent special issue in the *Journal of the American Medical Informatics Association* included a number of papers that focused on quantitative measures of time spent on data entry and numbers of entries,[15–18] all demonstrating that quantity of work is a key contributor to burn-out. In this perspective piece, we highlight and explain how there are additional mechanisms for EHRs to induce burn-out, such as the organisation of information on screen, and poor support for cognitive tasks.[19 20] Specifically, EHR interface designs tend to fragment information,[21] making it difficult to 'get the full picture' of a patient case, thus increasing the cognitive burden for tasks associated with clinical decision making.[22] The lack of an effective display of a patient's treatment as a whole may thus be one of the contributing factors of clinician burn-out.[4] In our view, it is not only the quantity of EHR related work that leads to burn-out, but the quality of the work—the cognitive burden.

An exemplar case of the phenomenon of fragmentation of information about a patient treatment and the consequences for clinicians' cognitive burden is provided by the design of EHR systems for cancer care, and more specifically the design of electronic oncology treatment regimens. These are linked, multidisciplinary, longitudinal records, not too dissimilar to electronic clinical pathways, linking preset tasks or orders over time.[23–25] Systems of this kind are also used, for example, in stroke care.[26] Electronic pathway functionalities are implemented across healthcare settings to standardise care and thus improve quality and safety, and are intended to support not only single clinical tasks, but a series of linked tasks by teams of clinicians managing a patient illness.[23] Major integrated EHR systems such as EPIC[27] or Cerner Millennium[28] include functionalities of this kind, giving hospitals the possibility of linking orders along clinical guidelines. Given the longitudinal, interdependent and multidisciplinary aspect of the information involved, electronic pathways and electronic regimens pose challenges to the organisation of information and the design of information displays.

CLINICAL PROTOCOLS AND DECISION MAKING IN CANCER CARE

Treatment regimens for most patients with cancer, especially in paediatrics, are complex. Medications are administered to patients over cycles, over months or years. At each cycle or dose, clinicians (re)assess the patient response to treatment and make decisions on whether and how to continue with the regimen. Medications are potentially toxic and clinical decisions involve high risk for the patient. The toxic effects of some medications manifest over a patient life-time, and result from the cumulation of multiple doses over time.

In travelling from the time of diagnosis to the end of treatment, clinical protocols provide a path, or a thread, for clinicians to follow (or deviate from). At each step, an assessment is made of both 'where the patient is at'[29] in the treatment journey—for example, how the patient is responding to treatment, how far along in the treatment they are—and how close or distant the patient treatment and trajectory are to the initial protocol path. As an oncologist in a children hospital explained to us, '…to synthesise a journey…

Surgery															
	Drug	X Y	X	X Y		X Y	X	X Y		X Y	X	X Y		X Y	X	X Y
	Weeks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	Cycle no.	1		2		3		4		5		6		7		8

X = drug 1; Y = drug 2

Figure 1 Example of a chemotherapy protocol roadmap, simplified.

at least two years, sometimes ten. … we need to maintain a thread of continuity, … not continuous together, but joined, so you can follow that particular thread… I want to see … what is the treatment the patient is getting today … in terms of where does it fall within the narrative of this patient's treatment'.[29]

Traditionally, paper-based oncology protocols provide 'roadmaps' that summarise graphically the treatment over a period of time—tests, drugs and doses, distributed over a number of cycles (figure 1).

These paper-based versions of the roadmap can be printed and easily annotated with a patient's actual treatment. Annotations show both what the protocol requires and any variations needed to address a patient response. For example, dose modifications, delayed doses or cycles.

FROM PAPER TO ELECTRONIC, FROM PATTERN RECOGNITION TO COMPUTATION

EHRs have been increasingly implemented and used in cancer care in the USA and other parts of the world. Their functionalities for linking orders, allow reproducing clinical protocols in electronic order sets. As human factors experts, we studied a recent implementation of an EHR for cancer care in Australia. We evaluated the implementation of one of these systems (a Cerner system) in a paediatric hospital,[29 30] and carried out a (confidential) expert human factors review of a number of chemotherapy management systems for a commissioning body. In our research, we noticed how clinicians experienced difficulties with the electronic regimens' information display, and in making sense of the patient treatment in relation to the relevant protocol. An oncologist told us, for example, how with the electronic record, '… it's very hard to [see] that they haven't got an extra dose at day 15 or they've missed a dose at day seven'.[29] A junior physician made explicit the burn-out effect of the effort to recombine fragmented information of a patient's history of treatment: '… to try and find a patient's cumulative anthracyclines dose [in the electronic record]… [was] more exhausting than night shifts exhausting'.[29]

Electronic versions of the protocols 'translate' a treatment plan into a series of orders (events or tasks), arranged as lines of text, nested (partly visible) into cycles (aggregating lines of text). This creates a fragmentation of the regimen as a whole. Intended variations to the protocol are to be annotated in digital notes, only available when clicking on each cycle, and superseding other notes which the EHR automatically archives in the patient record. As we learnt in our research, the problem with this design is that since the notes 'supersede each other', any changes made to the treatment are difficult to identify and explain: '…someone would make a change 6 months ago to a particular dose of something, and you'd have no idea why and no idea when it happened' (oncologist, paediatrics).[29]

To make sense of the information retrieved from each record, the clinician has to keep each data piece in memory and assemble a mental map of how they relate as a whole. This is a cognitive task known as 'computation'.[31]

Paper-based roadmaps in oncology make immediately visible to clinicians 'where the patient is at' in the protocol as a whole and any deviations, easing navigation of a patient treatment. In this respect, the use of this tool is not unlike navigational tasks performed by operators in non health-related disciplines. It can be said of oncology roadmaps what has been said of US Navy ship charts[32] that in roadmaps, every element of treatment (eg, a chemotherapy dose), has a specifiable 'address', and the relationships of all elements of treatment to the others 'are implicitly represented'; they 'introduce a perspective' on the whole and on the position and 'motion' of the patient across the whole; 'standing over a chart, one has a 'bird's eye' view …'. Thus, 'Having the chart […] makes this [navigation] task much easier. For example, […], displacement […] [off] track can be measured directly. The information regarding the next course is ready at hand and need only be read off the chart after the position has been plotted. […] the number of yards to the next turn [or the next dose] need not be measured; it is available by simple inspection'.[32]

Paper-based roadmaps support decision making through pattern recognition, while electronic protocols transform the task into computation. That is, with an EHR system, clinicians are required to compute the patient's position on a treatment journey based on information visible across rows and screens, a cognitively demanding task, in a high-risk environment. This is the opposite of what human factors traditionally recommend for safe engineering of sociotechnical systems. For example, it is the opposite of what was reported in aeroplane cockpit design in the 1990s[31] —instruments for landing were designed to transform pilots' mental computation into pattern recognition, easing the landing task and reducing potential for errors.

While some have attempted the implementation of computerised roadmaps reproducing and enhancing the traditional paper versions,[33] in our experience, EHR systems rarely provide a diagrammatic representation of encoded protocols akin to a roadmap. They provide detailed information about each cycle and dose, but this information is often fragmented; they document what has been ordered and administered to the patient, but not how close or distant from the protocol this was.

Here, we argue that the lack of effective displays of a complex treatment as a whole, in support of high risk decisions, may be one of the contributing factors of clinician burn-out in oncology.[4] It is interesting to note that

interventions targeting burn-out, even those targeted at the organisational rather than individual level, typically aim to enhance clinicians' resilience—their capacity to sustain stress—through training, wellness monitoring or a reduction in working hours.[2 3 6] Interventions rarely focus on improving EHR design. When EHR redesign is advocated for, it is usually in relation to alleviating the burden of clerical tasks,[19 34] not improving system usability, nor converting mental computation into pattern recognition. Greater usability and less cognitive burden may be generated by incorporating functions into EHR which work on pattern recognition, rather than computation, resulting in reduced complexity, faster and less error-prone tasks, ultimately benefiting patient safety and reducing clinician burn-out. One recent example of this was an attempt to design a medication timeline for chronic disease patients, which resulted in improved physician performance on medication-related tasks when piloted.[22] We recommend future work targeting clinician burn-out in oncology follow this path and focus on redesigning EHRs, using a human factors approach,[35] to support complex navigation work. There is a need for research on how to design, and automatically generate, digital oncology roadmaps of a patient treatment in EHR. These roadmaps must be easy to navigate and must support pattern recognition. Research can also show whether and how lessons learnt from oncology are applicable to EHRs in other clinical contexts, where an overview of a patient's trajectory, pathway or treatment journey is required.

Twitter Valentina Lichtner @VLichtner

Acknowledgements We thank Bethany Van Dort (University of Sydney) for her assistance with figure 1.

Contributors Both authors have contributed equally to this work.

Funding VL received funding from the European Union's Horizon 2020 research and innovation programme under the Marie-Skłodowska Curie grant agreement number 740 131.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; internally peer reviewed.

ORCID iD

Valentina Lichtner http://orcid.org/0000-0003-3956-3743

REFERENCES

1 National Academies of Sciences Engineering and Medicine. *Taking action against clinician burnout: a systems approach to professional well-being*. Washington, DC: The National Academies Press, 2019.
2 West CP, Dyrbye LN, Erwin PJ, *et al*. Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. *Lancet* 2016;388:2272–81.
3 Lacy BE, Chan JL. Physician burnout: the hidden health care crisis. *Clin Gastroenterol Hepatol* 2018;16:311–7.
4 Hlubocky FJ, Back AL, Shanafelt TD. Addressing burnout in oncology: why cancer care clinicians are at risk, what individuals can do, and how organizations can respond. *Am Soc Clin Oncol Educ Book* 2016;35:271–9.
5 Murali K, Banerjee S. Let's address burnout in oncologists and reimagine the way we work. *Nat Rev Clin Oncol* 2019;16:1–2.
6 Murali K, Banerjee S. Burnout in oncologists is a serious issue: what can we do about it? *Cancer Treat Rev* 2018;68:55–61.
7 Gajra A, Bapat B, Jeune-Smith Y, *et al*. Frequency and causes of burnout in US community oncologists in the era of electronic health records. *JCO Oncol Pract* 2020;16:e357–65.
8 Burki TK. Oncologists burnout in the spotlight. *Lancet Oncol* 2018;19:e238.
9 Melnick ER, Dyrbye LN, Sinsky CA, *et al*. The association between perceived electronic health record usability and professional burnout among US physicians. *Mayo Clinic Proceedings* 2020;95:476–87.
10 Gardner RL, Cooper E, Haskell J, *et al*. Physician stress and burnout: the impact of health information technology. *J Am Med Inform Assoc* 2019;26:106–14.
11 Khairat S, Coleman C, Ottmar P, *et al*. Association of electronic health record use with physician fatigue and efficiency. *JAMA Netw Open* 2020;3:e207385–e85.
12 Bakken S. Building the evidence base on health information technology-related clinician burnout: a response to impact of health information technology on burnout remains unknown-for now. *J Am Med Inform Assoc* 2019;26:1158–58.
13 Bakken S. Can informatics innovation help mitigate clinician burnout? *J Am Med Inform Assoc* 2019;26:93–4.
14 Kroth PJ, Morioka-Douglas N, Veres S, *et al*. Association of electronic health record design and use factors with clinician stress and burnout. *JAMA Netw Open* 2019;2:e199609.
15 Lourie EM, Utidjian LH, Ricci MF, *et al*. Reducing electronic health record-related burnout in providers through a personalized efficiency improvement program. *J Am Med Inform Assoc* 2020. doi:10.1093/jamia/ocaa248. [Epub ahead of print: 09 Nov 2020].
16 Hron JD, Lourie E. Have you got the time? challenges using vendor electronic health record metrics of provider efficiency. *J Am Med Inform Assoc* 2020;27:644–6.
17 Adler-Milstein J, Zhao W, Willard-Grace R, *et al*. Electronic health records and burnout: time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. *J Am Med Inform Assoc* 2020;27:531–8.
18 Hilliard RW, Haskell J, Gardner RL. Are specific elements of electronic health record use associated with clinician burnout more than others? *J Am Med Inform Assoc* 2020;27:1401–10.
19 Sutton JM, Ash SR, Al Makki A, *et al*. A daily Hospital progress note that increases physician usability of the electronic health record by facilitating a problem-oriented approach to the patient and reducing physician clerical burden. *Perm J* 2019;23. doi:10.7812/TPP/18-221. [Epub ahead of print: 14 Jun 2019].
20 Johnson C, Johnston D, Crowle P. *EHR usability toolkit: A background report on usability and electronic health records [AHRQ Publication No. 11-0084-EF]*. Rockville, MD: Agency for Healthcare Research Quality, 2011.
21 Coiera E, Ash J, Berg M. The unintended consequences of health information technology revisited. *Yearb Med Inform* 2016;1:163–9.
22 Belden JL, Wegier P, Patel J, *et al*. Designing a medication timeline for patients and physicians. *J Am Med Inform Assoc* 2019;26:95–105.
23 Gooch P, Roudsari A. Computerization of workflows, guidelines, and care pathways: a review of implementation challenges for process-oriented health information systems. *J Am Med Inform Assoc* 2011;18:738–48.
24 Peleg M. Computer-interpretable clinical guidelines: a methodological review. *J Biomed Inform* 2013;46:744–63.
25 Dong W, Huang Z. A method to evaluate critical factors for successful implementation of clinical pathways. *Appl Clin Inform* 2015;6:650–68.
26 Ballard DW, Kim AS, Huang J, *et al*. Implementation of computerized physician order entry is associated with increased thrombolytic administration for emergency department patients with acute ischemic stroke. *Ann Emerg Med* 2015;66:601–10.
27 Epic Systems Corporation. Available: https://www.epic.com/
28 Cerner Corporation. Available: https://www.cerner.com/
29 Lichtner V, Franklin BD, Dalla-Pozza L, *et al*. Electronic ordering and the management of treatment interdependencies: a qualitative study of paediatric chemotherapy. *BMC Med Inform Decis Mak* 2020;20:193.
30 Baysari MT, van Dort BA, Prgomet M, *et al*. The efficiency-thoroughness trade-off after implementation of electronic medication management: a qualitative study in paediatric oncology. *Int J Qual Health Care* 2020;32:511–6.

31 Hutchins E. How a Cockpit Remembers its speeds. *Cogn Sci* 1995;19:265–88.
32 Hutchins E. *Cognition in the wild*. Cambridge, Mass.: MIT Press, 1995.
33 Allen SW, Hayashi RJ, Jones SJ, *et al*. Development of electronic chemotherapy Roadmaps for pediatric oncology patients. *J Pediatr Oncol Nurs* 2018;35:314–9.
34 Shanafelt TD, Dyrbye LN, Sinsky C, *et al*. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. *Mayo Clin Proc* 2016;91:836–48.
35 Carayon P, Hoonakker P. Human factors and usability for health information technology: old and new challenges. *Yearb Med Inform* 2019;28:71–7.

BMJ Health & Care Informatics

Comparison of identifiable and non-identifiable data linkage: health technology assessment of MitraClip using registry, administrative and mortality datasets

Kim Keltie [1,2] Paola Cognigni [1] Sam Gross,[3] Samuel Urwin [2] Julie Burn [1] Helen Cole [4] Lee Berry,[5] Hannah Patrick,[5] Andrew Sims [1,2]

To cite: Keltie K, Cognigni P, Gross S, *et al*. Comparison of identifiable and non-identifiable data linkage: health technology assessment of MitraClip using registry, administrative and mortality datasets. *BMJ Health Care Inform* 2021;**28**:e100223. doi:10.1136/bmjhci-2020-100223

Received 11 August 2020
Revised 21 January 2021
Accepted 08 March 2021

[1]Northern Medical Physics and Clinical Engineering, Newcastle Upon Tyne Hospitals NHS Trust, Newcastle Upon Tyne, UK
[2]Translational and Clinical Research Institute, University of Newcastle upon Tyne, Newcastle upon Tyne, UK
[3]Data Management Services, NHS Digital, Leeds, UK
[4]The Northern Health Science Alliance, Manchester, UK
[5]Observational Data Unit, National Institute for Health and Care Excellence, London, UK

Correspondence to
Dr Andrew Sims;
andrew.sims5@nhs.net

ABSTRACT

Objectives The UK MitraClip registry was commissioned by National Health Service (NHS) England to assess real-world outcomes from percutaneous mitral valve repair for mitral regurgitation using a new technology, MitraClip. This study aimed to determine longitudinal patient outcomes by linking to routine datasets: Hospital Episode Statistics (HES) Admitted Patient Care (APC) and Office of National Statistics.

Methods Two methods of linkage were compared, using identifiable (NHS number, date of birth, postcode, gender) and non-identifiable data (hospital trust, age in years, admission, discharge and operation dates, operation and diagnosis codes). Outcome measures included: matching success, patient demographics, all-cause mortality and subsequent cardiac intervention.

Results A total of 197 registry patients were eligible for matching with routine administrative data. Using identifiable linkage, a total of 187 patients (94.9%) were matched with the HES APC dataset. However, 21 matched individuals (11.2%) had inconsistencies across the datasets (eg, different gender) and were subsequently removed, leaving 166 (84.3%) for analysis. Using non-identifiable data linkage, a total of 170 patients (86.3%) were uniquely matched with the HES APC dataset. Baseline patient characteristics were not significantly different between the two methods of data linkage. The total number of deaths (all causes) identified from identifiable and non-identifiable linkage methods was 37 and 40, respectively, and the difference in subsequent cardiac interventions identified between the two methods was negligible.

Conclusions Patients from a bespoke clinical procedural registry were matched to routine administrative data using identifiable and non-identifiable methods with equivalent matching success rates, similar baseline characteristics and similar 2-year outcomes.

Summary

What is already known?

- Few, if any, published studies report on the strengths and limitations of different methods of linking registries to routine administration datasets, to inform health technology assessments.
- Both identifiable and non-identifiable linkage of registries with routine data sets are possible.

What does this paper add?

- Two independent methods of linking a clinical registry with Hospital Episode Statistics data enabled reliable longer term outcomes to be obtained.
- Identifiable and non-identifiable linkage are complementary.
- Transparent description of the two methods facilitates use of the techniques by other researchers.

INTRODUCTION

Evidence-based clinical guidelines, such as those published by the UK National Institute for Health and Care Excellence (NICE), prioritise randomised controlled trials (RCTs) over other study types. However, because RCTs are costly, lengthy and may fail to identify rare adverse events, alternative study designs using real-world data have a role to play in technology assessment.[1 2]

With its Commissioning through Evaluation (CtE) scheme, National Health Service (NHS) England enabled a limited number of patients to access promising interventions that were not currently funded by the NHS, while collecting clinical and patient experience data within a formal evaluation programme.[3] Analysis of this real-world evidence, combined with evidence from clinical trials, informed NHS England's commissioning decision. Included in the CtE programme was percutaneous mitral valve repair for mitral regurgitation using MitraClip (Abbott, Illinois, USA). Data collection was mandated in a bespoke procedural registry (UK MitraClip registry), from which in-hospital and short-term outcomes

for the procedure were reported. To obtain longer term outcome data (readmission rates, subsequent interventions, adverse events and mortality outcomes), linkage of identifiable registry data to routinely collected Hospital Episode Statistics (HES) and Office of National Statistics (ONS) mortality data was also conducted by NHS Digital.[4]

However, transcription errors in patient identifiers and missing data in any data source are known to influence data linkage and may bias results.[5] Therefore, in order to assess the quality of the data linkage, and potential impact of any linkage errors, this study aimed to develop an alternative method, linking non-identifiable data from the MitraClip registry to pseudonymised extracts from HES Admitted Patient Care (APC) and ONS mortality datasets. This allowed a comparison of matching success rates between the two linkage methods (using identifiable and non-identifiable data) and additionally, a comparison of longitudinal outcomes such as subsequent cardiac intervention and all-cause mortality.

METHODS

Registry data collection

The MitraClip registry was commissioned by NHS England and opened on 1 October 2014. Patients were eligible for the MitraClip CtE scheme if they had stage 2, 3 or 4 mitral regurgitation of functional/ischaemic or degenerative aetiology (ie, excluding rheumatic heart disease), and were deemed high risk or were turned down for conventional mitral valve surgery. MitraClip implantation was either standalone or alongside percutaneous coronary intervention (staged or concurrent).

Identifiable data were collected in the MitraClip registry without explicit patient consent, via section 251 of the National Health Service Act 2006 (17/CAG/0153 (Previously CAG 10-07(b)/2014)).

Linkage using identifiable data

Patient identifiers were extracted from the MitraClip registry by the data controller (the National Institute for Cardiovascular Outcomes Research, NICOR) on 5 April 2018 and sent to NHS Digital (data supplied under DARS-NIC-151212-B5Z3R agreement). Records were linked by NHS Digital to the HES APC dataset and the ONS mortality dataset, using an eight-step deterministic matching proprietary algorithm based on NHS number, date of birth, gender and postcode (figure 1A).[6] Data from HES included all admissions from matched patients with hospital discharge dates between 1 April 2008 and 1 March 2018. Data from ONS included all reported deaths from matched patients until 4 April 2018. Records from patients having registered type 2

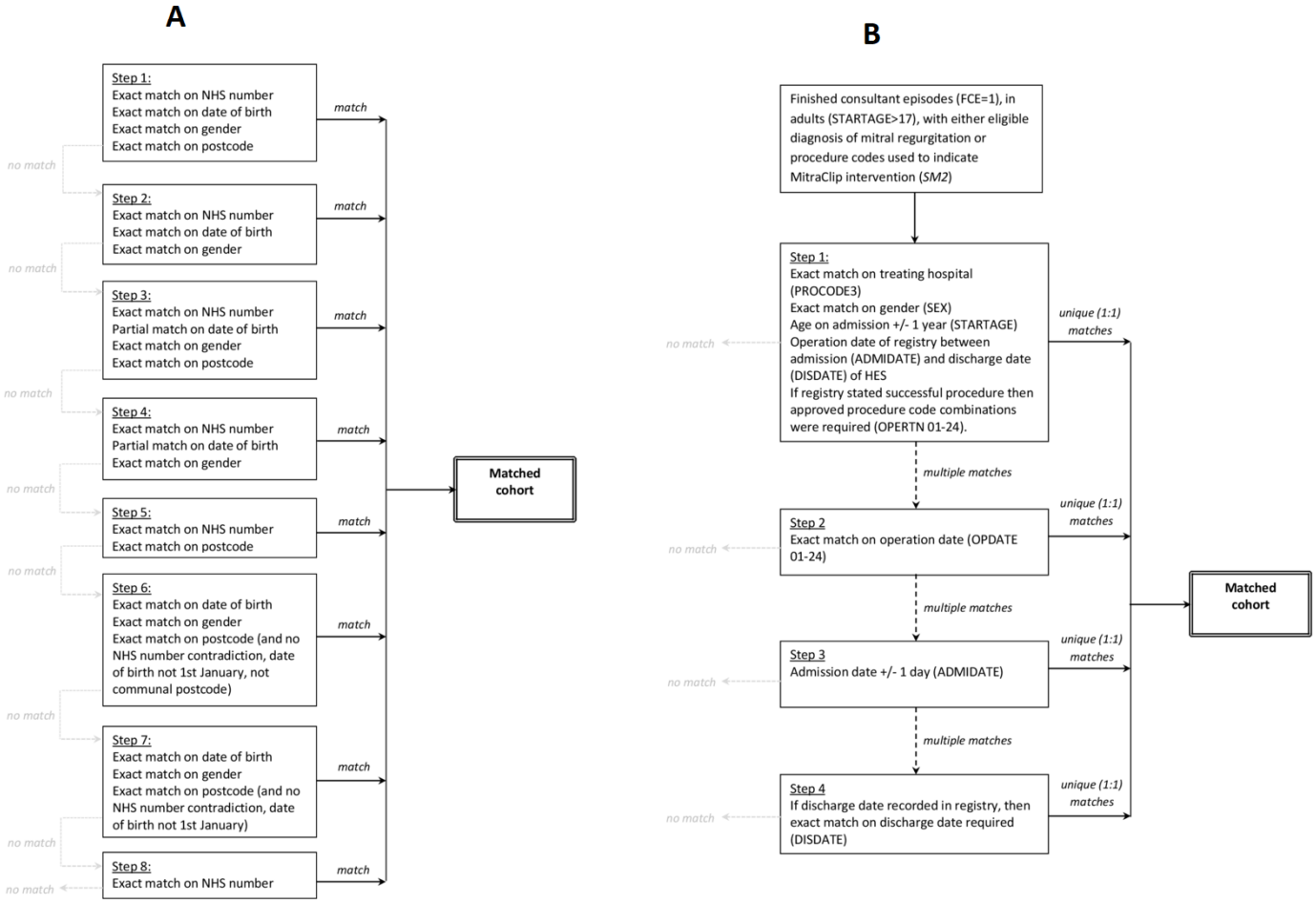

Figure 1 Matching steps used during (A) identifiable and (B) non-identifiable linkage. NHS: National Health Service.

opt-outs (ie, those not wishing their confidential patient information to be used for purposes other than their individual care) were removed from both extracts by NHS Digital before releasing the linked data back to the data controller. Patient identifiers (in registry, HES and ONS data) were then replaced with a 'Study ID' by NICOR (as data controller) and submitted to the study team (as the named co-data processor), figure 1A.

Matched records from the registry, HES and ONS were reviewed by the study team, and those with conflicting demographic and administrative details were flagged to indicate potential errors in matching (ie, matching to an incorrect patient). The demographics of those with and without inconsistencies were then compared to confirm that exclusion of potentially mismatched records would not introduce bias into the linked data results. All patient-level information provided by NHS Digital resulting from the identifiable linkage was deleted by the study team on expiry of the data sharing agreement prior to the commencement of linkage using non-identifiable data, enabling an independent assessment of methods.

Linkage using non-identifiable data

Separately, the study team had access to pseudonymised HES and ONS mortality datasets for all patients admitted to hospital in England, under Data Access Request Service (DARS) agreement DAR-NIC-170211-Z1B4J. These data were supplied via NHS Digital's managed extract service and were saved on a secure SQL server within Newcastle upon Tyne Hospitals.

An anonymised data extract was taken by the data controller from the UK MitraClip registry on 5 April 2018 and sent to the study team. For patients with more than one MitraClip procedure entered into the registry, the most recent MitraClip procedure was used for matching.

From the HES dataset, individual episodes of care satisfying the following criteria were deemed eligible for matching to the registry: (1) finished consultant episodes with a discharge date between 1 April 2014 and 1 March 2018 (to match the time frame of identifiable linkage); (2) a diagnosis of mitral insufficiency or a procedure code indicating mitral valve repair (see online supplemental file 1); (3) age over 17 years; (4) treatment at specific NHS Trust who submitted data to the UK MitraClip registry (see online supplemental file 2).

Individual episodes of care were aggregated into admissions (also known as spells) using the spell identification number (SUSSPELLID). Data cleaning was carried out to remove admissions if the spell number was missing or invalid. Non-unique admissions (where the same spell number was assigned to different patients) were also removed. Age was determined using the age on admission. Spells with missing discharge date were assigned the end date of the last episode in the admission (see online supplemental file 3).

Individual patient matching between the anonymised registry extract and the eligible admissions from HES was performed by a four-step algorithm using the following variables: treating hospital, gender, age, admission date, procedure codes, procedure dates, discharge date (figure 1B). At each step, patients with no matches were excluded while those with multiple matches proceeded to the next step. Unique (1:1) matches from all steps were combined to give the final matching HES cohort.

The cohort was followed longitudinally by extracting all subsequent episodes of care from HES APC (discharged on or before 1 March 2018), and ONS mortality records (dated on or before 4 April 2018), using the unique patient identifier (ENCRYPTED_HESID) determined from the non-identifiable matching process. This ensured that the study period was the same for the identifiable and the non-identifiable linkage methods.

Extended follow-up of the cohort was also determined using the latest pseudonymised HES APC and ONS data available (discharge date or date of death up to and including 31 March 2020) to Newcastle upon Tyne Hospitals under the managed extract service.

Statistical analysis

Storage and querying of HES data were conducted using SQL (MariaDB). All scripts for case ascertainment, cleaning, processing and statistical analyses were written in the statistical programming language R.[7]

Descriptive statistics were used to describe the baseline characteristics of patients matched using the identifiable and non-identifiable data linkage methods, including: gender, age, diabetes status, critical pre-op status, mitral regurgitation aetiology, admission method.

Long-term outcomes were determined for all matched patients. Each patient was followed in the HES APC and ONS datasets from the date of the MitraClip procedure until the date of discharge (or for patients where MitraClip device was not successfully implanted, until the end of the procedural admission), the date of death or the latest date included in the linked HES data. Kaplan-Meier analysis was conducted for total all-cause mortality. Specific cardiac interventions following mitral valve implantation (eg, mitral valve intervention, cardiac pacemaker insertion and implantation of a cardioverter defibrillator) were identified from readmissions by searching for the Office of Population Censuses and Surveys (OPCS) procedure codes described in online supplemental file 4.

Role of the funding source

NHS England commissioned and funded a fixed number of MitraClip procedures within the CtE scheme. NHS England also commissioned NICE to facilitate the evaluation through Newcastle External Assessment Centre. Staff at NICE contributed to the design and conduct of the study, interpretation of the results, review and approval of the manuscript. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

RESULTS

Records of 278 MitraClip procedures were extracted from the registry by the data controller; 79 did not meet the study inclusion criteria (multiple reasons for exclusion may apply: 57 with non-eligible indications, 56 with concomitant treatment, 7 with unknown/none/moderate mitral regurgitation, 4 due to rheumatic aetiology, 4 missing/inconsistent procedural dates) leaving 199 procedures from 197 patients eligible for matching. The earliest procedure was conducted in January 2015 and the latest in January 2018.

From identifiable linkage, a total of 187 patients (187/197, 94.9%) were matched with the HES APC dataset, however 21 of the matched patients (11.2%) had inconsistencies across the datasets (eg, different gender) (online supplemental file 5). There were no significant differences in baseline characteristics between patients who passed the internal quality checks (n=166) and those who did not (n=21) in terms of age, sex, body surface area, diabetes, critical pre-op status, Canadian Cardiovascular Society angina status, New York Heart Association classification dyspnoea status, Killip class of heart failure, Canadian Study of Health and Aging frailty score, aetiology of mitral regurgitation and Katz index of independence. Therefore, following identifiable linkage, a total of 166 patients (84.2% matching) remained for subsequent analysis.

Using non-identifiable data, a total of 169 patients (169/197, 85.8%) were uniquely matched with the HES APC dataset using a four-step matching algorithm (online supplemental file 5). Any patients with conflicting information across data sources would not have passed the matching algorithm, and therefore all patients matched via this method remained for subsequent analysis.

Table 1 Baseline characteristics for matched cohorts

	Identifiable linkage (n=166)	Non-identifiable linkage (n=169)
Female gender	52 (31.3%)	50 (26.6%)
Median age, years (Q1:Q3)	78.5 (69:85)	78 (68:85)
Diabetes	30 (18.1%)	34 (20.1%)
Critical pre-op status	3 (1.8%)	4 (2.4%)
Functional/ischaemic aetiology	100 (61.7%)	103 (60.9%)
Urgent/emergency procedure	20 (12.0%)	19 (11.2%)

*Note that diabetes status which included any of the following was converted to presence of diabetes: diabetes (dietary control), diabetes (oral medicine), diabetes (insulin), newly diagnosed diabetes.

Table 2 Outcomes for matched cohorts

	Identifiable linkage (n=166)	Non-identifiable linkage (n=169)
Total patient days followed	99 241	89 736
Follow-up per patient, days		
Mean (SD)	597.8 (283.1)	531.0 (307.1)
Median (Q1:Q3)	610 (409:807)	560 (355:764)
Range	2–1128	0–1128
All-cause deaths	37	42
1-year survival	0.873 (0.823 to 0.925) (n=133)	0.833 (0.778 to 0.893) (n=122)
2-year survival	0.773 (0.706 to 0.847) (n=58)	0.725 (0.652 to 0.807) (n=50)
Subsequent intervention, patients		
Mitral valve replacement	6	6
Mitral valve repair	2	2
Other mitral intervention	1	0
Cardiac pacemaker insertion	10	12
Cardioverter defibrillator	2	3

Baseline patient characteristics for matched patients using the two methods of data linkage is described in table 1.

Median length of follow-up was 610 and 560 days following identifiable (n=166) and non-identifiable (n=169) linkage methods, respectively. The total number of deaths (all causes) was 37 and 42 across the two linked datasets, respectively, and the difference in subsequent cardiac interventions identified between the two methods was negligible, table 2.

Extended follow-up using non-identifiable linkage and latest available data enabled a median (Q1:Q3) length of follow-up of 1161 (597:1458) days, where a total of 78 deaths were recorded. Survival at 1-year (n=136) 0.841 (0.786 to 0.899), 2-year (n=119) 0.736 (0.671 to 0.807), 3-year (n=96) 0.617 (0.546 to 0.697) and 4-year (n=42) 0.499 (0.422 to 0.590) follow-up are shown in figure 2.

DISCUSSION

Key points

Our study transparently describes the identifiable and non-identifiable linkage of a bespoke clinical procedural registry to routine healthcare data sets. Equivalent matching rates were achieved using the two methods, and similar baseline characteristics and 2-year outcomes were demonstrated for patients matched by either technique.

Strengths and limitations of linkage using identifiable data

To collect identifiable patient data, explicit approvals are required by an independent body: either by an

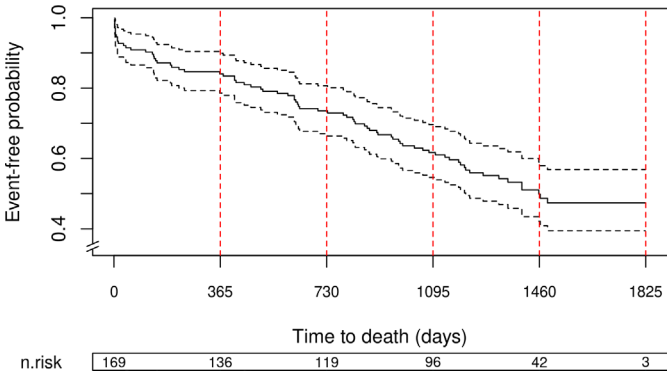

Figure 2 Kaplan-Meier survival curve derived from extended follow-up using non-identifiable linkage.

ethics board alongside a Caldicott Guardian with patient consent, or as in the case of the MitraClip registry, section 251 of the National Health Service Act 2006 approval such that the common law duty of confidentiality is temporarily lifted. All of these processes ensure that minimal patient identifiable information is collected for a specified purpose and safeguarding of those data. Although there are additional governance steps to undertake and the application process for obtaining matched data can be time consuming and costly, a key benefit of using patient identifiers (like NHS number) for subsequent matching to other data sources is that it does not require any clinical or clinical coding knowledge on the part of the researcher, with the process being fully automated and conducted independently by a trusted third party.

However, matching rates using identifiable sources may be reduced by certain factors; type 2 patient opt-outs are applied to identifiable HES data, with a rate of 2.7% reported nationally in March 2019.[8] Furthermore, matching multiple data sources which rely solely on patient identifiers is sensitive to manual transcription errors (eg, mistyping of NHS number, date of birth and/or postcode) or missing data.[9] Therefore, as demonstrated in this study, following the results of automated matching, an additional sense check/validation stage is still required to ensure matching to the correct person.

Strengths and limitations of linkage using non-identifiable data

To conduct linkage using non-identifiable data, variables including demographic details (gender, age) and administrative data (treating hospital dates, type of procedure) are used, thus combining matching and validation all within a single process. There is no requirement for identifiable information, which is beneficial from a data security and information governance perspective. If no identifiable information is being collected, then there is also no requirement to seek section 251 approval where explicit consent has not been obtained.

However, additional skills are required to conduct linkage using non-identifiable information. Access to clinical and clinical coding expertise is necessary to provide insight into the clinical pathway, to determine relevant procedure and diagnostic codes for analyses and to identify the relevant subgroup for matching. Knowledge of HES data quality and cleaning processes is also required. For example, an individual patient is not always assigned the same identifier and, rarely, different patients may be assigned the same identifier. Hagger-Johnson *et al* previously demonstrated that using the patient identifier 'HESID' generated by NHS Digital resulted in a false match rate of 0.2% and missed match rate of 4.1% in paediatric intensive care records in England, leading to a under-estimate in readmission rates.[10] Additionally, analysis of spells (collections of care episodes within a single admission) reveals inconsistencies that point to underlying data quality problems, such as duplications of records, missing information or inconsistencies, or activity recorded after death.[11] This necessitates additional cleaning of HES data to identify duplicated, inconsistent or missing spell information, and overlapping spells all of which allow removal of ineligible patients prior to matching.

We have demonstrated that non-identifiable data linkage works well for procedures with well-defined OPCS/International Classification of Diseases (ICD) coding combinations such as the MitraClip procedure, but this may not be the case for all clinical interventions; for example, those which have less specific clinical coding, but also those conducted out with the inpatient/day-case setting (where quality and completeness of clinical coding may differ), and high volume procedures where the likelihood of having multiple patients treated with same age and gender on same day are high. Further research is required to confirm that the matching success reported for the NHS England MitraClip registry is achievable for a range of other interventional procedures. Access to national pseudonymised administrative data was required to conduct the non-identifiable data linkage. With this comes responsibility for protecting the confidentiality of information for all potentially matching patients as well as for the cohort of interest. Each study applying data linkage of multiple data sources using non-identifiable information must recognise the potential for re-identification.[12] Our linkage algorithm shows that unique matching is possible with non-identifiable fields, suggesting that pseudonymised extracts carry confidentiality implications comparable to identifiable datasets. For this reason, all uses of HES data, including anonymous and pseudonymous matching proposals are reviewed by an independent panel (Independent Group Advising on the Release of Data, IGARD) to ensure safeguarding of patient data and subsequent data handling and processing by the approved institution are subject to audit by NHS Digital. Additionally, the guidelines for publishing results of such studies should be followed.[13] Other initiatives for safe data linkage of identifiable data do exist, for example, the 'Separation of functions' offered by the Scottish Informatics and Linkage Collaboration.[14]

Overall strengths and limitations of linking to administrative datasets

The overall strengths of data linkage between clinical registries and routine data have been well documented, such as richer clinical information and the estimation of reporting bias.[15–17] The main benefit of data linkage in this study, however, was the ability to conduct long-term comprehensive follow-up across all NHS Trusts in England. It has been reported that MitraClip appears to confer immediate improvements in cardiac indices of patients in line with published trial data,[4] but long-term outcomes have not been published. Our techniques have delivered longer term outcome data and offer the capability for analyses to be repeated at 5 and 10 years as required, demonstrating a way of conducting active surveillance of medical procedures using routine data sources.[18] This has the potential to improve understanding of the safety and efficacy of an intervention (particularly where long-term complications are likely to be detected outside the centre responsible for an intervention) and thereby to inform refinement of procedural selection to maximise long-term effectiveness.

Our study validated outcomes following MitraClip intervention by two separate techniques of data linkage between a clinical registry and routine healthcare data. However, the quality of collected data is crucial to the success of any data linkage. Clinical registries and routine administrative datasets both require high quality data submission as a single mismatch between the two can force the exclusion of a patient from the study. This work has highlighted the many benefits of data linkage to routine databases and thus strongly advocates the adoption of high quality data entry protocols and data validation in registries intended for health technology assessment.

This study has highlighted several lessons to be learnt for future linkage of clinical data to routine administrative data, whichever linkage method is used. As far as registry design is concerned, improvements could be achieved by incorporating input data validation. Also, to ensure easier identification during matching, mandatory coding of procedures using pre-specified OPCS codes should be used by treating hospitals contributing data to clinical registries. Comorbidities and adverse events could also be captured in registries using ICD codes, which would also be beneficial when conducting matching. Interim data linkage to identify potential data entry errors before final linkage is advisable.

Conclusion

This study has demonstrated the linkage of patient data from a bespoke clinical registry with routine healthcare databases via two equivalent methods to gain comprehensive and reliable follow-up from a cardiac intervention conducted in the 'real world' NHS in England. Linking to administrative data, by either method, would limit the administrative burden of future observational research. Here we have described a novel method that uses non-identifiable data, and in cases where robust clinical coding can be specified, it could be applied to other hospital interventions. Studies using the technique must recognise the potential for re-identification. In this study, both data sources were pseudonymised but in cases where the study team may have access to identifiable patient information consent must be obtained to undertake linkage.

Furthermore, to ensure robust and generalisable results from clinical registries and routine databases, data validation at the point of data entry and data cleaning following linkage are essential steps in the analysis methods. Interim data linkage to identify and correct potential data quality issues during the course of data collection are also strongly recommended.

Acknowledgements HES and ONS data held by NHS Digital (formerly the UK NHS Health and Social Care Information Centre, HSCIC) have been used to help complete the analysis © 2019. Reused with the permission of the NHS Digital/HSCIC. All rights reserved. Clinical coding advice was provided by the clinical coding department within Newcastle upon Tyne Hospitals NHS Foundation Trust. NHS England's specialised services Clinical Panel for granting permission to use the Commissioning through Evaluation data.

Contributors All authors conceived and designed the study. KK, PC, SU, JB, HC and AS analysed the data and were responsible for the statistical analysis. All authors interpreted the data. All authors contributed to the manuscript.

Funding The UK MitraClip registry and the identifiable data linkage with NHS Digital under DARS-NIC-151212-B5Z3R agreement was funded by NHS England. NICE were contracted by NHS England to oversee the Commissioning through Evaluation scheme. The Newcastle upon Tyne Hospitals NHS Foundation Trust hosts an External Assessment Centre funded by NICE. The Newcastle uponon Tyne Hospitals NHS Foundation Trust funded the national anonymised extracts provided under DAR-NIC-170211-Z1B4J agreement.

Competing interests The Newcastle upon Tyne Hospitals NHS Foundation Trust, the employing institution of KK, PC, SU, JB, HC and AS, is contracted as External Assessment Centre to the NICE Medical Technologies Evaluation Programme (MTEP) and is contracted by Academic Health Science Network North East and North Cumbria to develop methodologies and case studies relating to'evaluation in practice' in the context of using routine healthcare datasets and, where appropriate, clinical registries, to assess outcomes and adoption of novel interventions. AS reports grants from NIHR and Wellcome Trust and outside the submitted work. KK reports grants from NIHR outside the submitted work. HP and LB are employed by NICE and were contracted by NHS England to oversee the Commissioning through Evaluation scheme. SG is employed by NHS Digital. HP and LB are employed by NICE and were contracted by NHS England to oversee the Commissioning through Evaluation scheme. No other financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Hospital Episodes Statistics data to reproduce results are available from NHS Digital via a formal application process. The MitraClip CtE registry data is owned by NHS England; applications to access the data can be made to NHS England's specialised services Clinical Panel.

ORCID iDs
Kim Keltie http://orcid.org/0000-0001-5108-6279
Paola Cognigni http://orcid.org/0000-0001-5418-3103
Samuel Urwin http://orcid.org/0000-0003-4304-1358
Julie Burn http://orcid.org/0000-0002-2300-2725
Helen Cole http://orcid.org/0000-0003-3036-3374
Andrew Sims http://orcid.org/0000-0002-9553-7278

REFERENCES

1 Charlton V. NICE and Fair? Health Technology Assessment policy under the UK's National Institute for Health and Care Excellence, 1999-2018. *Health Care Analysis* 2019:1–35.
2 Bowman L, Baras A, Bombien R, *et al*. Understanding the use of observational and randomized data in cardiovascular medicine. *Eur Heart J* 2020;41:2571–8 https://doi.org/10.1093/eurheartj/ehaa020
3 NHS England. Commissioning through evaluation. Available: https://www.england.nhs.uk/commissioning/spec-services/npc-crg/comm-eval/ [Accessed 12th Aug 2019].
4 Willits I, Keltie K, Urwin S. Nice observational data unit: commissioning through evaluation (CTE) percutaneous mitral valve leaflet repair for mitral regurgitation (MitraClip), 2018. Available: https://www.england.nhs.uk/commissioning/wp-content/uploads/sites/12/2019/04/Percutaneous-mitral-valve-leaflet-repair-for-mitral-regurgitation-Commissioning-through-Evaluation-Report.pdf [Accessed 12th Aug 2019].
5 Harron K, Dibben C, Boyd J, *et al*. Challenges in administrative data linkage for research. *Big Data Soc* 2017;4:205395171774567–12.
6 Padmanabhan S, Carty L, Cameron E, *et al*. Approach to record linkage of primary care data from clinical practice research Datalink to other health-related patient data: overview and implications. *Eur J Epidemiol* 2019;34:91–9.
7 R Core Team. R: A language and environment for statistical computing [Internet]. R Foundation for Statistical Computing, Vienna, Austria. Available: http://www.R-project.org/ [Accessed 13th Aug 2019].
8 NHS Digital. National data Opt-out statistics, 2019. Available: https://digital.nhs.uk/data-and-information/publications/statistical/national-data-opt-out/march-2019/ndop-mar19 [Accessed 13th Sep 2019].
9 Hagger-Johnson G, Harron K, Goldstein H, *et al*. Probabilistic linking to enhance deterministic algorithms and reduce linkage errors in hospital administrative data. *BMJ Health & Care Informatics* 2017;24:891.
10 Hagger-Johnson G, Harron K, Fleming T, *et al*. Data linkage errors in hospital administrative data when applying a pseudonymisation algorithm to paediatric intensive care records. *BMJ Open* 2015;5:e008118.
11 Keltie K, Elneil S, Monga A, *et al*. Complications following vaginal mesh procedures for stress urinary incontinence: an 8 year study of 92,246 women. *Sci Rep* 2017;7:12015.
12 Problems with care.data and other stories…. *British Medical Journal* 2015;351:h4613.
13 NHS Digital. Hospital episode statistics (Hes) analysis guide, December, 2019. Available: https://digital.nhs.uk/binaries/content/assets/website-assets/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hes_analysis_guide_december_2019-v1.0.pdf [Accessed Last accessed 15th January 2021].
14 Scottish Informatics and Linkage Collaboration. Data linkage safeguards. available from. Available: http://www.datalinkagescotland.co.uk/data-linkage-safeguards [Accessed Last accessed 15th January 2021].
15 Boyd JH, Randall SM, Ferrante AM, *et al*. Accuracy and completeness of patient pathways--the benefits of national data linkage in Australia. *BMC Health Serv Res* 2015;15:312.
16 Burn J, Sims AJ, Keltie K, *et al*. Procedural and short-term safety of bronchial thermoplasty in clinical practice: evidence from a national registry and hospital episode statistics. *Journal of Asthma* 2017;54:872–9.
17 Field K, Kosmider S, Johns J, *et al*. Linking data from hospital and cancer registry databases: should this be standard practice? *Intern Med J* 2010;40:566–73.
18 Garies S, Youngson E, Soos B, *et al*. Primary care EMR and administrative data linkage in Alberta, Canada: describing the suitability for hypertension surveillance. *BMJ Health Care Inform* 2020;27:e100161.

BMJ Health & Care Informatics

Building resiliency in emergency room physicians: anticipating the next catastrophe

Vimla L Patel [1], Rahul Shidhaye,[2] Parvati Dev,[3] Edward H Shortliffe [4]

To cite: Patel VL, Shidhaye R, Dev P, *et al*. Building resiliency in emergency room physicians: anticipating the next catastrophe. *BMJ Health Care Inform* 2021;**28**:e100343. doi:10.1136/bmjhci-2021-100343

Received 09 March 2021
Revised 24 March 2021
Accepted 29 March 2021

[1]Center for Cognitive Studies in Medicine and Public Health, New York Academy of Medicine, New York, New York, USA
[2]Pravara Institute of Medical Sciences, Loni, Maharashtra, India
[3]SimTabs, Los Altos Hills, California, USA
[4]Vagelos College of Physicians and Surgeons, Columbia University, New York, New York, USA

Correspondence to
Dr Edward H Shortliffe; ted@shortliffe.net

Although media coverage of the COVID-19 crisis may have gradually desensitised us to the daily updates on numbers of cases and deaths, broadcast interviews with distraught front-line health workers continue to evoke an empathic response as we hear about the traumatic conditions of their experiences. It is natural to ask how greater foresight and preparation might have mitigated their emotional distress. In the major pandemic epicentres, even the most experienced emergency physicians confide that they have never before tried to deal with the stress of caring for the pressing needs of overwhelming volumes of patients with gut-wrenching illness and rapid progression of disease, amplified by inadequate staffing and a shortage of essential supplies. These conditions create internal struggles and unusual mental health challenges for the affected physicians.[1]

The director of the COVID-19 response at the University of California-San Francisco Medical Centre emergency department recently explained how the lockdown and burnout are still wearing on her and her colleagues: "Nine Months Into It, the Adrenaline Is Gone and It's Just Exhausting".[2] It is accordingly appropriate to ask how our emergency physicians and other front-line health professionals might have been better prepared for the ongoing unique, high intensity and unexpected circumstances.[3] We propose here an approach to learning from the current situation so that, with the development of new immersive training technologies, our emergency health workers might better anticipate and be more resilient when dealing with intense and prolonged experiences that may entail watching their coworkers become ill and even die (similar to war-related post-traumatic stress disorder (PTSD) and its associated sequelae).[4]

Specialised knowledge and skills of emergency physicians distinguish them as experts in dealing with complex situations under uncertainty and time pressure. When such physicians are confronted with unexpected situations, their organised knowledge allows them to respond adaptively and flexibly to the constraints of a rapidly changing environment, demonstrating instinctive responses to high-intensity circumstances. The novelty of the required decisions or interventions, the simultaneity of multiple such challenges and a barrage of expectations from others, including fellow workers, patients and their families, exacerbate the situation. Such an environment, with constant trade-offs, can contribute to significant emotional distress and burnout.[4 5]

Others have drawn parallels between preparing emergency physicians for extreme circumstances and training airline pilots to handle unique and unforeseen emergencies that can occur in the air.[6] Commercial pilots undergo innovative virtual training and evaluation to develop and maintain their skills, and the industry has developed and leveraged advanced simulation methods to mimic the experience of real-world flight operations. Such simulators are designed not only to train pilots regarding new aircraft or instruments but also to give them experiences designed to prepare them for unexpected situations, and the cognitive pressures, that can arise mid-flight, on landing or during take-off. While recognising that there are differences between aviation and healthcare, we believe that our recent COVID-19 experience suggests that emergency physicians may benefit from similar training and evaluation on simulators designed to capture the emergency room experience and its most stressful aspects, thereby building resiliency.

Medicine has recognised that the useful role of simulations in clinical education, including the creation of simulated patients for learning how to interview and examine

patients and how to reach accurate diagnoses.[7] In particular, providing a foundational framework, Gaba pioneered the use of aviation simulation concepts in training and evaluating anaesthesiologists for the rare but extreme situations that can arise in the operating room.[8] His work used instrumented mannequins in realistic operative suites with observational tools that captured the trainee's dialogue and actions while they were handling routine simulated cases that were interrupted by unexpected and diagnostically challenging crisis situations. As is expected, such simulators can document both the errors that occur during a routine case and the recognition and correction processes that typically follow among the more expert clinicians.[9]

Current safety strategy trains physicians to anticipate potential threats rather than preparing them to deal with the unpredictability of real-life emergency situations, where unexpected and surprising conditions can make it difficult to maintain cognitive control. A different training approach is needed to prepare individuals to be unprepared!

The design of modern simulators requires a deep understanding of both the technical skills that are being explored and the nature of potential stresses and how they can affect accuracy and error management. Emergency medicine can be described as a *complex adaptive system*[10] and any attempt to develop simulators for training and evaluation purposes would need to be preceded by a careful assessment of how such clinicians behave emotionally and psychologically, especially as they make patient-care decisions at times of high-intensity crises. The lessons we can learn by studying current health workers, challenged by the COVID-19 crisis, are akin to the data analyses that aviation performs using contents of 'black boxes' retrieved after accidents. Such studies are difficult to perform during actual crises, however, when clinicians are unable to cooperate and observational assessments are intrusive and may be constrained by privacy concerns. Yet retrospective data capture, based on interviews with physicians' weeks or months after the trauma, is unlikely to capture accurately the pressures of actual events, which get blunted and reconstructed in memory as a function of time.

Real-world scenario-based immersive simulations for physician trainees can be designed to elicit an emotional response similar to their response in real emergency medicine situations. In carefully constructed scenarios, these psychological stresses will induce real physiological manifestations, thereby allowing actual measurement of emotional responses similar to those in emergency situations. Training on such simulators over time, with appropriate supportive behavioural feedback, is likely to help physicians to develop resiliency to cope with such situations when they are encountered in real emergencies. Repeated exposure to such psychologically stressful situations, even in training simulations, does raise some concerns about the ethical and psychological implications of the long-term impact of such practice on ER physicians.[11] Therefore, training for unexpected, high acuity crisis events should take into account both the training to expect the unexpected and the training for stress management skills leading to resilience.

Virtual simulation (VS) methods are considered a viable platform for high-stakes training and assessment.[12] As such, immersive virtual reality (VR) can be used for training in high-acuity, low-frequency events, including disaster and mass casualty events such as the response to the influx of patients with COVID-19.[13] It is a significant additional benefit that they provide convenience as well as the ability to scale and distribute simulations widely with lower costs.

The medical use of VS is not new, particularly in managing the mental health of returning military personnel. However, with the introduction of cheaper, more accessible and sophisticated simulation technology, there is a new emphasis on the clinical use of VR and augmented reality (AR). It has been shown that people can be exposed to anxiety, phobias and characteristics of PTSD using immersive technology to train them to manage trauma before they encounter these events in the real world.[14 15] For example, veterans from Afghanistan have often had heart-thumping real-life experiences that can be recreated by an immersion programme called *Bravemind*, which leverages a VR system created almost 16 years ago to confront veterans with simulated experiences and to build resilience to such events if they are encountered in the future.[16] Subsequent VR programmes have also been used to incorporate traumatic cues from actual combat events realistically to build AR simulations for treating combat-related PTSD.[17] This trauma management programme's efficacy has been assessed through a randomised controlled trial of Iraq and Afghanistan veterans and active-duty military personnel with combat-related PTSD. During these virtual training experiences, users were monitored physiologically before and after the training.

The current heart-rending interviews of front-line physicians and other health workers demonstrate the value of capturing the nature of the emotional and cognitive challenges close to the time when they have been experienced. We accordingly suggest that preparedness for the next significant crises, whether a pandemic like COVID-19 or something totally different and unexpected, requires us to begin with the capture of detailed information regarding the current psychological and emotional challenges and what has driven them. This preparedness will also allow for the development of crisis management simulators inspired by those used in the training and preparation of pilots or members of the military.

The problems we have described are likely to be even more severe in low resource settings with fragile health systems. Although this would generally include low and middle-income countries, it also refers to rural, remote settings in high-income countries. In such environments, availability of trained health workers and fiscal resources for innovative simulation-based training are constrained

and there is often poor governance, leadership and delineation of roles and responsibilities. Preparation may also be more dependent on counselling and frugal educational efforts that must be undertaken with an understanding of how demographic as well as resource characteristics may aggravate some of the mental health challenges associated with the kinds of stress discussed here.[18] As in high resource settings, it is cognitive and emotional preparation, rather than straightforward knowledge of emergency medicine practices, that is needed to help emergency health workers as they build resilience for potential disasters that may lie ahead.

Twitter Vimla L Patel @VimlaLPatel and Edward H Shortliffe @tshortliffe

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.

ORCID iDs
Vimla L Patel http://orcid.org/0000-0003-1656-6642
Edward H Shortliffe http://orcid.org/0000-0001-5201-6176

REFERENCES

1 Lieberman JA. My colleague's death, and our collective psychological pain. New York Daily News. Available: https://www.nydailynews.com/opinion/ny-oped-our-colleagues-death-our-collective-pain-20200429-sghgwq2t3ncf3poulmnpyr3kw4-story.html [Accessed 29 Apr 2020].
2 Barry-Jester AM. Kaiser Health News (KHN), December 21, 2020. Available: https://khn.org/news/article/nine-months-into-it-the-adrenaline-is-gone-and-its-just-exhausting/
3 National Academy of Medicine's Action Collaborative on Clinician Well-Being and Resilience. Available: https://nam.edu/publications-from-the-action-collaborative-on-clinician-well-being-and-resilience/
4 Wears R, Perry S, Anders S. Resilience in the emergency department (Chapter 15). In: Hollnagel E, Nemeth SP, Dekker S, eds. *Remaining sensitive to the possibility of failure*. Aldershot, Hampshire, UK: Ashgate Publishing, 2008: 193–210.
5 Patel VL, Kaufman D, Cohen T, eds. *Cognitive informatics in health and biomedicine: case studies on critical care, complexity and errors*. London: Springer, 2014.
6 Amalberti R, Auroy Y, Berwick D, *et al*. Five system barriers to achieving ultrasafe health care. *Ann Intern Med* 2005;142:756–64.
7 Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. *Med Educ* 2009;43:303–11.
8 Gaba DM, DeAnda A. A comprehensive anesthesia simulation environment: re-creating the operating room for research and training. *Anesthesiology* 1988;69:387–94.
9 Patel VL, Kannampallil TG, Shortliffe EH. Role of cognition in generating and mitigating clinical errors. *BMJ Qual Saf* 2015;24:468–74.
10 Plsek PE, Greenhalgh T. The challenge of complexity in health care. *BMJ* 2001;323:625–8.
11 Gaba DM. Simulations that are challenging to the psyche of participants: how much should we worry and about what? *Simul Healthc* 2013;8:4–7.
12 Youngblood P, Harter PM, Srivastava S, *et al*. Design, development, and evaluation of an online virtual emergency department for training trauma teams. *Simul Healthc* 2008;3:146–53.
13 Andreatta PB, Maslowski E, Petty S, *et al*. Virtual reality triage training provides a viable solution for disaster-preparedness. *Acad Emerg Med* 2010;17:870–6.
14 Waldrop MM. News feature: virtual reality therapy set for a real renaissance. *Proc Natl Acad Sci U S A* 2017;114:10295–9.
15 Wiederhold BK, Miller I, Wiederhold MD. Augmenting behavioral healthcare: mobilizing services with virtual reality and augmented reality. In: Rivas H, Wac K, eds. *Digital Health. Health Informatics Series*. London: Springer, 2018.
16 Rizzo A, Buckwalter JG, John B, *et al*. Strive: stress resilience in virtual environments: a pre-deployment VR system for training emotional coping skills and assessing chronic and acute stress responses. *Stud Health Technol Inform* 2012;173:379–85.
17 Beidel DC, Frueh BC, Neer SM, *et al*. Trauma management therapy with virtual-reality augmented exposure therapy for combat-related PTSD: a randomized controlled trial. *J Anxiety Disord* 2019;61:64–74.
18 Moitra M, Rahman M, Collins PY, *et al*. Mental health consequences for healthcare workers during the COVID-19 pandemic: a scoping review to draw lessons for LMICs. *Front Psychiatry* 2021;12:602614.