BMJ Health & Care Informatics

Adoption of electronic health record systems to enhance the quality of healthcare in low-income countries: a systematic review

Misganaw Tadesse Woldemariam ,[1,2] Worku Jimma[1]

To cite: Woldemariam MT, Jimma W. Adoption of electronic health record systems to enhance the quality of healthcare in low-income countries: a systematic review. *BMJ Health Care Inform* 2023;**30**:e100704. doi:10.1136/bmjhci-2022-100704

Received 19 November 2022
Accepted 21 May 2023

[1]Department of Information Science, Jimma University Institute of Technology, Jimma, Ethiopia
[2]Department of Information and Communication Technology, Bonga College of Education, Bonga, Ethiopia

Correspondence to
Misganaw Tadesse Woldemariam;
misganawt2012@gmail.com

ABSTRACT

Background Electronic health record (EHR) systems are mentioned in several studies as tools for improving healthcare quality in developed and developing nations. However, there is a research gap in presenting the status of EHR adoption in low-income countries (LICs). Therefore, this study systematically reviews articles that discuss the adoption of EHR systems status, opportunities and challenges for improving healthcare quality in LICs.

Methods We used Preferred Reporting Items for Systematic Reviews and Meta-Analyses in articles selected from PubMed, Science Direct, IEEE Xplore, citations and manual searches. We focused on peer-reviewed articles published from January 2017 to 30 September 2022, and those focusing on the status, challenges or opportunities of EHR adoption in LICs. However, we excluded articles that did not consider EHR in LICs, reviews or secondary representations of existing knowledge. Joanna Briggs Institute checklists were used to appraise the articles to minimise the risk of bias.

Results We identified 12 studies for the review. The finding indicated EHR systems are not well implemented and are at a pilot stage in various LICs. The barriers to EHR adoption were poor infrastructure, lack of management commitment, standards, interoperability, support, experience and poor EHR systems. However, healthcare providers' perception, their goodwill to use EMR and the immaturity of health information exchange infrastructure are key facilitators for EHR adoption in LICs.

Conclusion Most LICs are adopting EHR systems, although it is at an early stage of implementation. EHR systems adoption is facilitated or influenced by people, environment, tools, tasks and the interaction among these factors.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Research findings show that electronic health record (EHR)/EMR is being implemented in low-income countries (LICs) despite various challenges influencing its success. However, no empirical evidence is built on systematically collected and analysed studies across LICs that could be used to develop a better implementation strategy.

WHAT THIS STUDY ADDS

⇒ The study identified that LICs are struggling to adopt EHR systems, but they are failing at the initial stages due to people-related barriers, environment-related barriers, infrastructure-related barriers and poor integration of the system with people.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study could help LICs to properly adopt and use EHR systems considering the barriers identified.

INTRODUCTION

According to the WHO definition, quality of healthcare is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes.[1] Currently, with advancements in digital technology, most of the work in the healthcare sector is becoming digitised and efficient.[2] This could significantly improve the quality of healthcare[3 4] compared with the traditional approach. The electronic health record (EHR) system is at the forefront of implementation in healthcare institutions to enhance the quality of healthcare.[5]

The EHR system is a digital way of capturing, storing, and using patient information by authorised healthcare providers to deliver healthcare services effectively.[6] EHR systems enable data-driven clinical decision-making to improve healthcare quality. Gatiti *et al*[7] noted that the proper adoption of EHR systems could boost the quality of healthcare by enhancing patient safety and ensuring effective, efficient, timely, equitable and patient-centred care.

Despite the benefits of EHR systems, problems or unintended consequences are hampering the successful adoption and use of EHR systems in healthcare settings. The most common are physician burn-out,[8–10] failure of expectations,[8] EHR market saturation,[8] innovation vacuum,[8] data obfuscation,[8] interoperability,[11] privacy in data sharing,[12]

protracted to complete tasks,[13] interruption of tasks and workarounds at point of care[13] and misalignment of technology and clinical context.[11] In addition to these, DeWane *et al*[14] and Gagnon *et al*[15] noted data duplication errors during decision-making, intermittent system delays and workflow interruptions as unintended consequences of EHR systems. Generally, unintended consequences could have a severe impact on the diagnostic and therapeutic processes undertaken by healthcare professionals at points of care, eventually jeopardising patients' safety and well-being.[13]

EHR system has been used in developed countries since its inception in the USA in the 1960s.[16] Since then, its impact in enhancing the quality of healthcare has been clear both in the developed and developing world. In developed countries, where EHR systems have undergone an established implementation strategy, there is increased success and health worker satisfaction and decreased delays and chances of usability being compromised.[17] However, despite increased use in developed countries, multiple studies conducted in developing countries indicated the adoption of an EHR system is still lagging[18]; hence, multiple factors play a role in technology adoption and use. A study conducted in Kenya, Ghana, Nigeria, South Africa and Saudi Arabia indicated that EHR adoption is challenged by inadequate training,[19–23] poor infrastructure,[19 21–23] lack of technical support,[19 21–23] poor communication between users[21] and absence of regulations and implementation framework.[22] Furthermore, the findings from Jung *et al*[24] showed that EHR implementation is not an easy task even for countries advancing from developing to developed, let alone developing countries.

EHR implementation or adoption in most low-income countries (LICs) is lagging and affected by multifaceted challenges. Some of these barriers are economy,[25 26] infrastructure[25] and policy.[26] In addition to these, healthcare professionals' readiness,[27] poor collaboration among stakeholders,[28] and relying on software provided by non-governmental organizations (NGOs)[28] are affecting EHR adoption in LICs. However, due to the development of open-source systems, support from international donors and homegrown software development campaign[29]; EHR adoption in LICs is becoming feasible and a future direction.

On top of this, there is a research gap in identifying the existing situation of EHR adoption in LICs despite some efforts made in low-middle-income and middle-income countries. Therefore, this review aimed to examine the status, challenges and opportunities of adopting EHR systems to enhance the quality of healthcare delivery in LICs. It is hoped that the review will provide effective support for the local developers, healthcare providers, different stakeholders and funders in the course of developing or adopting EHR systems. We conducted the review based on the following questions:

RQ1: What is the status of HER systems adoption in LICs?

RQ2: What are the challenges influencing the adoption of EHR systems in LICs?

RQ3: What opportunities are facilitating the adoption of EHR systems in LICs?

METHODS

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 checklist was used to conduct this review.[30]

Eligibility criteria

We used the inclusion and exclusion criteria presented in table 1 to identify articles that meet the study objectives.

Information sources and search strategy

PubMed, Science Direct and IEEE Xplore were the electronic databases used for the literature search. We conducted the search using keywords based on four concepts, namely "electronic health record," "adoption," "quality of healthcare," and "developing countries." Medical subject heading (MeSH) terms were also used to supplement the keyword search in the PubMed database, hence it is a controlled vocabulary thesaurus used for indexing articles. We conducted forward and backward citation searches on significant search results and manual searches on health informatics journals found in developing countries. We presented the search strategies in table 2.

Table 1 Inclusion and exclusion criteria

	Criteria
Inclusion	1. Articles that present the status, challenges and opportunities of EHR adoption in LICs
	2. Articles published in English starting from January 2017 to 30 September 2022
	3. Peer-reviewed journal articles
Exclusion	1. Articles do not explicitly discuss EHR adoption, its challenges and opportunities in LICs
	2. Articles on EHR adoption in countries other than LICs
	3. Books, book chapters, conference papers, symposiums, review articles and non-English scripts

EHR, electronic health record; LICs, low-income countries.

Table 2 Information sources and search strategy

Date of the search	Database	Search query	Filters	Search result
28 September 2022	PubMed	(("Electronic Health Records"[MeSH Terms] OR "electronic health record*"[Title/Abstract] OR "electronic medical record*"[Title/Abstract] OR "computerized medical record*"[Title/Abstract] OR "EHR"[Title/Abstract] OR "EMR"[Title/Abstract]) AND ("Adoption"[Title/Abstract] OR "application"[Title/Abstract] OR "utilization"[Title/Abstract] OR "acceptance"[Title/Abstract] OR "implementation"[Title/Abstract]) AND ("Quality of Health Care"[MeSH Terms] OR "Health Care Quality"[Title/Abstract] OR "Quality of Healthcare"[Title/Abstract] OR "Healthcare Quality"[Title/Abstract] OR "Quality of Care"[Title/Abstract] OR "Care Quality"[Title/Abstract] OR "pharmacy audit*"[Title/Abstract] OR "Audit Pharmacy"[Title/Abstract]) AND ("Developing Countries"[MeSH Terms] OR "developing countr*"[Title/Abstract] OR "developing nation*"[Title/Abstract] OR "economically developing nation*"[Title/Abstract] OR "economically developing countr*"[Title/Abstract] OR "emergent nation*"[Title/Abstract] OR "least developed countr*"[Title/Abstract] OR "low income countr*"[Title/Abstract] OR "underdeveloped nation*"[Title/Abstract])) AND (2017:2022[pdat])	Year of publication between 2017 and 2022	39
28 September 2022	Science Direct	Year: 2017–2022 Title, abstract, keywords: ("electronic health records" OR EMR OR EHR) AND (adoption OR implementation) AND "Quality of healthcare" AND ("developing countries" OR "low-income countries" OR "developing nations") Article type: Research articles	Year of publication between 2017 and 2022, Research articles	44
28 September 2022	IEEE Xplore	("All Metadata":"electronic health record*" OR "All Metadata":"electronic medical record*" OR "All Metadata":"computerized medical record*" OR "All Metadata":EHR OR "All Metadata":EMR) AND ("All Metadata":"developing countr*" OR "All Metadata":"low income countr*") You Refined By: Content-Type: Journals, Early Access Articles Year: 2017-2022	Journals, early access articles, Year of publication between 2017 and 2022	12
29 September 2022	Citation search+other journals	"electronic health records" AND "name of LIC" OR "electronic medical records" AND "name of LIC"	Year of publication between 2017 and 2022, empirical research articles	14

Selection process

We imported the search results from all databases and citation searches into EndNote to begin the selection process. First, we removed duplicate records. After doing so, we screened the remaining records to detect subject relevance with the research objectives considering their title and, or abstract. Next, full-text articles were identified for retrieval. Finally, articles that fulfilled the inclusion criteria were selected for qualitative analysis and synthesis. The author (WJ) validated the entire selection process to ensure its accuracy.

Data collection process, data items, analysis and synthesis

The data collection process started by identifying the main concepts in the three research questions that appear as results or findings in each of the reviewed articles. This approach formed a conceptual basis for data extraction under the corresponding heading in a Microsoft Word document. The headings include the authors' name, publication year, research design, data collection methods, data analysis techniques, study population, sample size and sampling techniques. Moreover, the findings from each of the studies included were extracted as EHR functions, challenges, opportunities and healthcare quality indicators addressed. Content analysis was used to organise related concepts under the categories EHR in LICs, challenges of EHR adoption in LICs, opportunities of EHR in LICs and EHR and healthcare quality in LICs. Finally, narrative synthesis and ordering of the evidence were conducted in each of the four categories by comparing and contrasting with previous studies conducted on the topics.

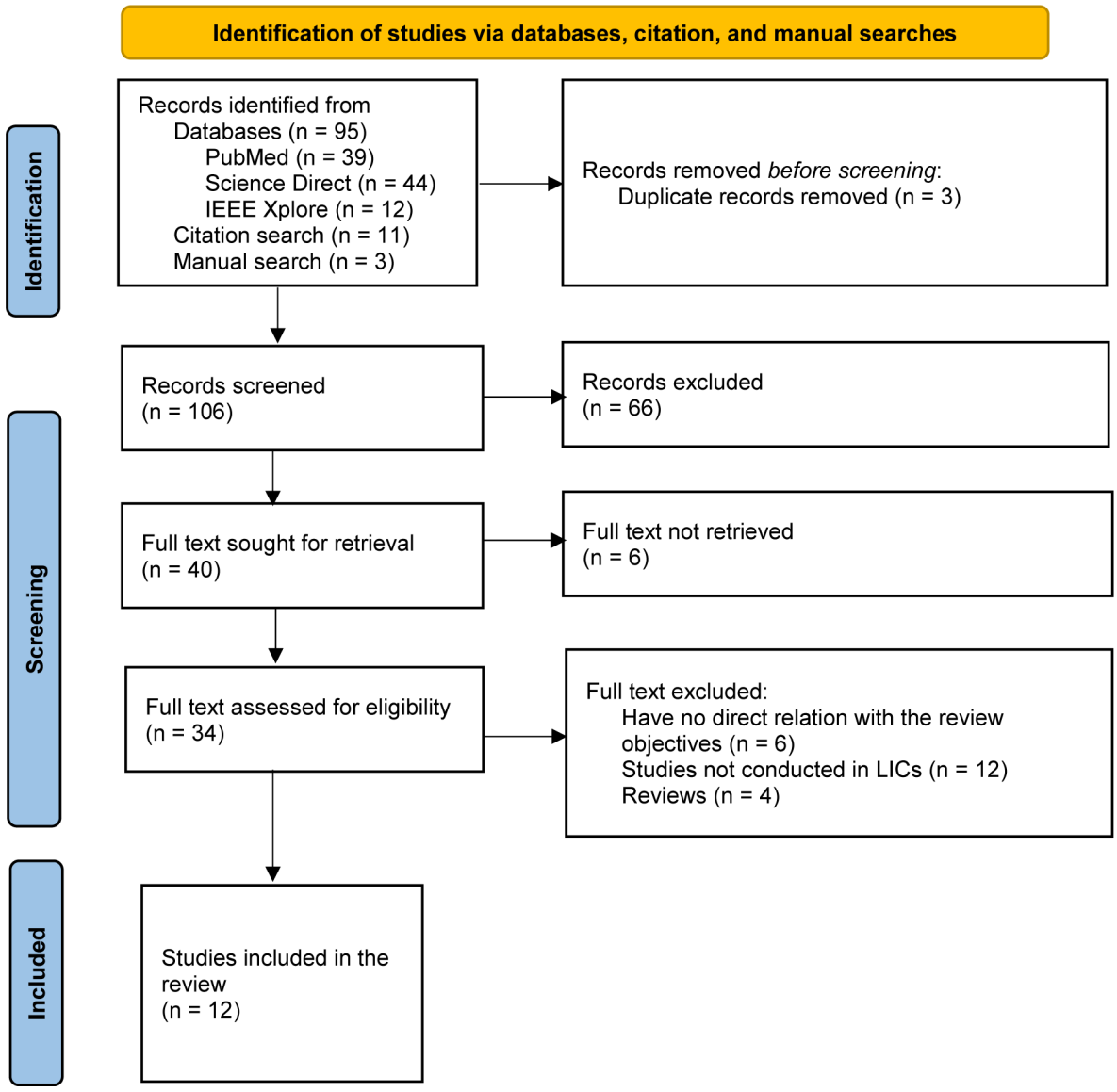

Figure 1 PRISMA 2020 flow diagram illustrating the overall selection process to show studies included and excluded (modified from Page *et al*[30]). LICs, low-income countries; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Risk of bias

Each study included in the review was subject to an appraisal using the Joanna Briggs Institute checklists.[31 32] Accordingly, we selected and included studies with an optimum score based on the requirements in the checklist. Further, to avoid selection bias, we strictly followed the protocol. In doing so, to some extent, we managed the risk of bias in selection, analysis and reporting.

RESULTS

Study selection

As presented in figure 1, we retrieved 109 records following the search strategy defined. We removed three records that were duplicates. Further, we excluded 66 records after reviewing titles and, or abstracts. Out of 40 studies sought for retrieval, we discarded six as a result of not finding their full text. Out of 34 studies accessed for eligibility, we included 12, which qualified for the inclusion criteria.

Study characteristics

In this review, we used the world bank classification of 2023 to identify LICs.[33] The studies selected systematically from this group were four from Ethiopia, two from Uganda and one from each remaining country: Gabon, Rwanda, Malawi, Sierra Leone, Angola and LICs altogether (Kenya, Rwanda, Uganda and Mozambique). Based on the type of study; five were quantitative, two were qualitative, three were mixed-type, one was agile software development and one was situational analysis. The details of each study are presented in table 3.

EHR in low-income countries

In this review, 9 of the 12 studies showed some of the major functions of EHR/EMR in LICs. The first one is

Table 3 Study characteristics

Author(s)	Country	Article title	Study design/method of data collection	Data analysis technique
Bagayoko *et al*[42]	Gabon	Implementation of a national electronic health information system in Gabon: a survey of healthcare providers' perceptions	Cross-sectional survey/ questionnaire	Logistic regression
Bisrat *et al*[43]	Ethiopia	Implementation challenges and perception of care providers on Electronic Medical Records at St. Paul's and Ayder Hospitals, Ethiopia	Cross-sectional survey/ questionnaire and Interview	Descriptive analysis and thematic analysis
Fraser *et al*[40]	Rwanda	User Perceptions and Use of an Enhanced Electronic Health Record in Rwanda With and Without Clinical Alerts: Cross-sectional Survey	Cross-sectional survey/ interviews, observation and free text	Thematic analysis and descriptive analysis
Liang *et al*[39]	Uganda	A Locally Developed Electronic Health Platform in Uganda: Development and Implementation of Stre@mline	Cross-sectional survey/ questionnaire	Descriptive analysis
Mkalira Msiska *et al*[44]	Malawi	Factors affecting the utilisation of electronic medical records system in Malawian central hospitals	Cross-sectional survey/ questionnaire and Interview	Descriptive analysis and χ^2 test
Oumer *et al*[38]	Ethiopia	Utilisation, Determinants and Prospects of Electronic Medical Records in Ethiopia	Cross-sectional survey/ questionnaire	Descriptive analysis, bivariate and multivariate logistic regression
Oza *et al*[34]	Sierra Leone	Development and Deployment of the OpenMRS-Ebola Electronic Health Record System for an Ebola Treatment Centre in Sierra Leone	Agile software development/ questionnaire	Not mentioned
Robbiati *et al*[35]	Angola	Improving TB Surveillance and Patients' Quality of Care Through Improved Data Collection in Angola: Development of an Electronic Medical Record System in Two Health Facilities of Luanda	Not mentioned/meetings, interviews, site visits and observation	Situational analysis
Were *et al*[41]	Kenya, Rwanda, Uganda and Mozambique	mUzima Mobile Electronic Health Record (EHR) System: Development and Implementation at Scale	Not mentioned/mHealth evidence reporting assessment checklist	Not mentioned
Ahmed *et al*[37]	Ethiopia	Intention to use electronic medical record and its predictors among healthcare providers at referral hospitals, north-West Ethiopia, 2019: using unified theory of acceptance and use technology 2 (UTAUT2) model	Cross-sectional explanatory/ questionnaire and in-depth interview	Structural Equation Model, χ^2 test and thematic analysis
Kabukye *et al*[54]	Uganda	User Requirements for an Electronic Medical Records System for Oncology in Developing Countries: A Case Study of Uganda	Qualitative study/FGD and interview	Content and thematic analysis
Ngusie *et al*[36]	Ethiopia	Healthcare providers' readiness for electronic health record adoption: a cross-sectional study during pre-implementation phase	Cross-sectional/questionnaire	Multivariate logistic regression

FGD, focus group discussion.

the OpenMRS-Ebola, which was implemented in Sierra Leone. The system can track patients' vital signs, medication, intravenous fluid ordering and monitoring, laboratory results, and clinician notes, and export data for clinical decision-making.[34] EMR systems are being used to enhance tuberculosis surveillance and control in Angola.[35]

In Ethiopia, studies were conducted to assess the healthcare providers' technological and organisational readiness and the level of EHR adoption. The findings indicated that the overall readiness of healthcare providers was inadequate.[36] Ahmed *et al*[37] noted that 39.8% of healthcare providers surveyed showed a score above the mean intention to use EMR in northwest Ethiopia. Whereas, a study by Oumer *et al*[38] in eastern Ethiopia showed optimal EMR usage level. These findings portray that EHR systems are not adopted as expected to address quality healthcare in the country.

In Uganda, a locally developed EHR platform (Stre@mline) is highly accepted and used despite implementation challenges.[39] The system can monitor patients, control stock levels, provide early warning and capture prescription errors. Similarly, Fraser *et al*[40] noted that OpenMRS in Rwanda supports healthcare delivery by

Table 4 Barriers and facilitators to EHR adoption in LICs

	EHR system adoption in LICs to enhance the quality of healthcare	
Work system factors	Barriers	Facilitators
People	Awareness, experience, resistance, lack of training	Providing alerts Perception to use EHR
Environmental	Interoperability with other systems, finance, absence of explicit policy, lack of standards, lack of management commitment, quality of a system	Immaturity of health information exchange infrastructure in LICs
Tools	Poor infrastructure	
Tasks		The partnership among stakeholders to design and adopt EHR systems.
Interaction between people, environments, tools and tasks	Poor integration of the EHR system with people, infrastructure, functions and other existing systems.	

EHR, electronic health record; LICs, low-income countries.

managing patient records, making informed decisions, and providing useful alerts and reminders. Finally, mUzima is a mobile-based EMR system that is providing quality healthcare in countries like Kenya, Rwanda, Uganda and Mozambique.[41]

Barriers and facilitators to EHR adoption in LICs

Five studies identified barriers to EHR adoption in LICs, as presented in table 4. Dominantly, lack of training,[40 42–44] poor infrastructure,[40 43 44] lack of management commitment,[40 43] lack of standards[40 43 44] and absence of interoperability[43] are the barriers observed. Bagayoko *et al*[42] identified the quality of a system, support, information, actual use, satisfaction and impact as potential barriers. Oza *et al*[34] showed that inconsistency in EHR systems creates an enormous challenge. In addition to this argument, experience is another barrier to adopting EHR as a health professional over 5 years of experience had two times higher odds of using EMR than early career workers.[38] Overall, in most LICs, EHR adoption exists in the preimplementation phase.[36]

Four studies identified facilitators to HER adoption in LICs, as presented in table 4. Bisrat *et al*[43] found 70%–95% of healthcare providers have a favourable perception of using EMR systems. Similarly, Oumer *et al*[38] identified that about 85% of healthcare professionals demonstrated goodwill in using EMR systems. Fraser *et al*[40] noted the role of EHR systems in supporting patient care by providing alerts ahead of complications. The immaturity of health information exchange infrastructure in many LICs provides an opportunity to enhance EHR systems by incorporating mobile-based systems.[41]

Table 4 illustrates people and environment-related factors are both facilitating and impeding EHR adoption in LICs. While tool-related factors influence, task-related factors are facilitating EHR adoption. Overall, poor integration of EHR among the work systems factors affects EHR adoption. The absence of facilitators under tools and interaction among the four work systems indicated an insufficiency of technology and lack of management support to facilitate EHR adoption, respectively.

EHR and healthcare quality in LICs

EHR systems are improving healthcare delivery in both developing and developed countries. An empirical work reported from Rwanda,[40] Uganda[39] and Malawi[44] showed that EHR systems improve healthcare by managing patient information, supporting informed decisions and providing useful alerts. In developed nations, EHR-based clinical trials are providing evidence about treatment strategies, patient safety, care and health policy decisions.[45] Based on the WHO definition, this review considered seven quality indicators of healthcare: Safety, effectiveness, people-centredness, timeliness, efficiency, equity and integrated service.[1]

Safety

In terms of safety, the finding presented by Fraser *et al*[40] signified the role of openMRS system in supporting patient care by providing alerts. Additionally, Liang *et al*[39] mentioned the significance of EHR systems in maintaining patient safety features, which in turn has improved care for more than 60 000 patients in Uganda. This indicates implementation of an EHR system is highly important to ensure patient safety.

Effectiveness

In this review, Mkalira Msiska *et al*[44] noted EMR systems help generate more accurate information that can reduce medical errors. This could improve the decision-making capability of healthcare workers for effective patient management. Liang *et al*[39] mentioned that the locally developed EHR platform is capable of managing patient information and related healthcare services. Further, Fraser *et al*[40] indicated the effectiveness of openMRS despite the infrastructure limitation in Rwanda. These all assertions prove the significance of adopting EHR systems in delivering effective healthcare services.

People-centredness

In this review, the findings of Liang *et al*[39] reported that the partnership between healthcare providers and developers is significant to the design and adoption of user-centred technologies. The mUzima (mobile health) application is an example of how technologies can be used to promote healthcare for people at large.[41] The finding also indicated the adoption of mUzima across multiple LICs and for numerous core healthcare domains. These findings depict, EHR systems that are well communicated with the users during the design and adoption phases would yield a better outcome.

Timeliness

The review identified the benefits of EHR systems in facilitating contacting patients to ensure good ongoing care in place.[39] Mkalira Msiska *et al*[44] finding affirmed the introduction of EMR systems in Malawi healthcare helped to assess patients within a short period. Similarly, a survey by Oumer *et al*[38] found that 75% of health professionals agreed EMRs can improve timely patient care. These findings affirm the significance of EHR systems in providing timely care for patients in LICs.

Efficiency

The findings from Liang *et al*[39] indicated that EHR platforms play a crucial role in improving clinical efficiency. This could help healthcare professionals to carry out their duty on time and help patients not to wait too long to get treatments. Further, Mkalira Msiska *et al*[44] noted that the EMR system is more efficient in assessing more patients in a short period than traditional systems. Thus, adopting EHR systems can help improve healthcare quality by providing efficient services.

Equity

In this review, Were *et al*[41] stressed the use of the EHR system in delivering healthcare services by avoiding geographical barriers. The study identified that such systems could extend the reach of EHR systems within resource-limited settings as opposed to siloed mhealth applications. Further, Mkalira Msiska *et al*[44] underlined the significance of EHR systems in reaching every patient awaiting healthcare services with no bias. EHR systems provide healthcare services without geographical, economical and social limitations.

Integrated

In this review, the findings of Oza *et al*[34] showed that OpenMRS is the most comprehensive, adaptable clinical EHR built for a low-resource setting. The system is interoperable with other EHR systems to provide integrated healthcare services. Liang *et al*[39] noted that EHR platforms are being used to support patient care, live control of stock medicines, forward warnings to pharmacists and recognise prescription errors before causing harm. These findings elucidate the role of EHR systems in providing integrated quality healthcare services for patients.

DISCUSSION

Status of EHR adoption, challenges and opportunities in LICs

This review aims to examine the status, challenges and opportunities of adopting EHR systems to enhance the quality of healthcare delivery in LICs. In most LICs, donors provide support to establish EHR systems, which usually fail for many reasons. For example, in Ethiopia, the Smart Care system, which is supported by donors, is not functioning at full scale as expected due to low economic readiness.[46] It is failing at a pilot stage in many of the hospitals where the system is implemented.[43] Further, Ngusie *et al*[36] noted that, in most LICs, EHR implementation exists at the preimplementation stage. This affirms that countries should first identify organisational, technological, social and economic readiness before adopting EHR systems.[46]

However, in countries such as Uganda, locally developed EHR platforms are being used to enhance patient care.[39] The openMRS system in Rwanda is also making a notable influence in supporting healthcare delivery by providing informed decisions, alerts and reminders.[40] Further, studies conducted in Sierra Leone and Angola indicated that open-source EMR systems are enhancing clinical care and clinical decision-making.[34 35] These findings show that EHR systems are currently being practised in LICs despite the challenges reported. It is also in line with the findings reported in low-income and-middle-income countries.[47] Therefore, LICs should work hard towards adopting open-source EMR systems which fit the shortcomings of the economy and user-friendliness.

Most of the challenges for the failure of EHR adoption in LICs were lack of training, infrastructure, management commitment, standards, consistency, interoperability, quality of systems, support, use, information, satisfaction and impact of the system.[34 40 42–44] Further, Oumer *et al*[38] added the impact of healthcare providers' experience on affecting EHR adoption as experienced have twice higher odds of using EMR than early career workers. Most of these challenges are similar to those reported in studies conducted in middle-income countries.[19–23] Furthermore, a scoping review of studies published between 2005 and 2020 on PubMed, Web of Science, IEEE Xplore and ACM Digital Library reported similar challenges as the current study.[48] Therefore, every LICs needs to develop strategies, legislations, regulations and a framework of implementation that can address the mentioned challenges before adopting or implementing EHR systems.

Moreover, EHR adoption might pose unanticipated challenges to existing healthcare systems if not managed appropriately. Windle *et al*[49] in their findings indicated the perception of clinicians on the impact of EHR in impeding the workflow and communication, and prolonging their workday. EHR implementation causes physician burn-out due to contributing factors like increased documentation, which are significantly underestimated.[50] These challenges need critical attention and should be addressed during the preimplementation phase.

Despite the various factors influencing the success of EHR adoption, there are opportunities that can

maximise its potential. The most important scenario is a good perception of healthcare providers in using EMR systems.[43] Also, most healthcare professionals are open-minded about using such systems whenever deployed or adopted.[38] Moreover, the health information exchange infrastructure in LICs is immature or absent. These findings are in line with those mentioned in the studies conducted by Amend *et al*[51] which considers stakeholder readiness, change management, accessibility and ownership, EHR structure and external factors as key facilitators for EHR adoption.

Multiple studies indicated the impact of EHR systems in capitalising on quality in healthcare delivery.[34 38–41 44] Studies conducted in countries other than LICs indicated the significance of EHR systems in enhancing the quality of healthcare in terms of safety, effectiveness, people-centredness, timeliness, efficiency, equity and provision of integrated services.[52 53] This study portrayed a clear image of EHR systems adoption status, challenges and opportunities in LICs to enhance the quality of healthcare delivery.

Conclusion

EHR adoption is at early stage in most LICs, with different types of EHRs being used. It is facilitated or influenced by people, environment, tools, tasks and the interaction among these four factors. Unanticipated challenges such as physician burn-out are creating a challenge in slowing down EHR adoption.

Strengths

The review followed a protocol to select and synthesise relevant studies on the topic. Further, it identified research gaps to be addressed by future researchers. Overall, because of absence of previous systematic reviews in LICs, the findings could help develop implementation strategies and policies.

Limitations

The search result was vulnerable to various problems, such as reporting bias or lack of enough research outputs from LICs, as only studies from eight countries out of 28 were included. Additionally, a literature search was conducted only on PubMed, Science Direct, IEEE Xplore and journals of health informatics in developing countries. However, the quality of the studies was not compromised by following the review protocol.

Implications for practice, policy and future research

The review findings suggest all actors involved in EHR systems should collaborate effectively to yield a better outcome in healthcare delivery. This can be supported through EHR adoption policies, which are currently missing in many countries. Future research should focus on comparative studies on the practice of EHR systems in developing and developed countries.

Twitter Misganaw Tadesse Woldemariam @Misganaw2012

Acknowledgements We would like to thank Jimma University, Institute of Technology for arranging the schedule and essential resources altogether to complete this work.

Contributors In this systematic review, MTW contributed to the conception, design, database searching, full-text screening and write-up of the manuscript. WJ contributed by evaluating and assuring the quality of the review's study selection, analysis, synthesis and edition of the manuscript. The final draft of the manuscript was read, edited, and approved for publication by MTW and WJ.

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 applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

ORCID iD

Misganaw Tadesse Woldemariam http://orcid.org/0000-0001-9906-3009

REFERENCES

1 WHO. Fact sheet: quality health services. World Health Organization; 2020. Available: https://www.who.int/news-room/fact-sheets/detail/quality-health-services [Accessed 6 Oct 2022].
2 Chakraborty S, Bhatt V, Chakravorty T. Impact of Digital technology adoption on care service orchestration, agility, and responsiveness. *Int J Sci Technol Res (New Delhi)* 2020;9:4581–6.
3 Bauer AM, Thielke SM, Katon W, *et al*. Aligning health information Technologies with effective service delivery models to improve chronic disease care. *Prev Med* 2014;66:167–72.
4 Kutney-Lee A, Sloane DM, Bowles KH, *et al*. Electronic health record adoption and nurse reports of usability and quality of care: the role of work environment. *Appl Clin Inform* 2019;10:129–39.
5 Aldosari B. Patients' safety in the era of EMR/EHR automation. *Informatics in Medicine Unlocked* 2017;9:230–3.
6 Muhaise H, Kareyo DM, Muwanga-Zake P. Factors influencing the adoption of electronic health record systems in developing countries: A case of Uganda. *ASRJETS* 2019;61:160–6.
7 Gatiti P, Ndirangu E, Mwangi J, *et al*. Enhancing Healthcare quality in hospitals through electronic health records: A systematic review. *J Health Inform Dev Ctries* 2021;15.
8 Colicchio TK, Cimino JJ, Del Fiol G. Unintended consequences of nationwide electronic health record adoption: challenges and opportunities in the post-meaningful use era. *J Med Internet Res* 2019;21:e13313.
9 Downing NL, Bates DW, Longhurst CA. Physician burnout in the electronic health record era: are we ignoring the real cause. *Ann Intern Med* 2018;169:50–1.
10 Tajirian T, Stergiopoulos V, Strudwick G, *et al*. The influence of electronic health record use on physician burnout: cross-sectional survey. *J Med Internet Res* 2020;22:e19274.
11 Alami H, Lehoux P, Gagnon M-P, *et al*. Rethinking the electronic health record through the quadruple aim: time to align its value with the health system. *BMC Med Inform Decis Mak* 2020;20:32.
12 Shah SM, Khan RA. Secondary use of electronic health record: opportunities and challenges. *IEEE Access* 2020;8:136947–65.
13 Ogundaini O, de la Harpe R, McLean N. Unintended consequences of technology-enabled work activities experienced by Healthcare professionals in tertiary hospitals of sub-Saharan Africa. AJSTID 2022;14:876–85.
14 DeWane M, Waldman R, Waldman S. Cell phone etiquette in the clinical arena: A professionalism imperative for Healthcare. *Curr Probl Pediatr Adolesc Health Care* 2019;49:79–83.
15 Gagnon M-P, Ngangue P, Payne-Gagnon J, *et al*. M-health adoption by Healthcare professionals: a systematic review. *J Am Med Inform Assoc* 2016;23:212–20.
16 Net Health. The history of electronic health records (Ehrs) - updated. 2022. Available: https://www.nethealth.com/the-history-of-electronic-health-records-ehrs/

17 Aguirre RR, Suarez O, Fuentes M, *et al*. Electronic health record implementation: A review of resources and tools. *Cureus* 2019;11:e5649.
18 Parks R, Wigand RT, Othmani MB, *et al*. Electronic health records implementation in Morocco: challenges of silo efforts and recommendations for improvements. *Int J Med Inform* 2019;129:430–7.
19 Adedeji P, Irinoye O, Ikono R, *et al*. Factors influencing the use of electronic health records among nurses in a teaching hospital in Nigeria. *J Health Inform Dev Ctries* 2018;12.
20 Alboliteeh M. A cross-sectional study of nurses' perception toward utilization and barriers of electronic health record. *MJHR* 2022;26.
21 Essuman LR, Apaak D, Ansah EW, *et al*. Factors associated with the utilization of electronic medical records in the Eastern region of Ghana. *Health Policy and Technology* 2020;9:362–7.
22 Katurura MC, Cilliers L. Electronic health record system in the public health care sector of South Africa: A systematic literature review. *Afr J Prim Health Care Fam Med* 2018;10:e1–8.
23 Muinga N, Magare S, Monda J, *et al*. Digital health systems in Kenyan public hospitals: a mixed-methods survey. *BMC Med Inform Decis Mak* 2020;20:2.
24 Jung SY, Lee K, Lee H-Y, *et al*. Barriers and Facilitators to implementation of nationwide electronic health records in the Russian far East: A qualitative analysis. *Int J Med Inform* 2020;143:104244.
25 Irshad R, Ghafoor N. Infrastructure and economic growth: evidence from lower middle-income countries. *J Knowl Econ* 2023;14:161–79.
26 Steinbach R. *growth in low-income countries: evolution, prospects, and Policies*. July 2019.
27 Awol SM, Birhanu AY, Mekonnen ZA, *et al*. Health professionals' readiness and its associated factors to implement electronic medical record system in four selected primary hospitals in Ethiopia. *Adv Med Educ Pract* 2020;11:147–54.
28 Tolera A, Oljira L, Dingeta T, *et al*. Electronic medical record use and associated factors among Healthcare professionals at public health facilities in dire Dawa, Eastern Ethiopia: A mixed-method study. *Front Digit Health* 2022;4:935945.
29 Colicchio TK, Cimino JJ. Twilighted Homegrown systems: the experience of six traditional electronic health record developers in the post–meaningful use era. *Appl Clin Inform* 2020;11:356–65.
30 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
31 Lockwood C, Munn Z, Porritt K. Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. *Int J Evid Based Healthc* 2015;13:179–87.
32 Moola S, Munn Z, Tufanaru C, *et al*. Chapter 7: systematic reviews of etiology and risk in. In: Aromataris E, Munn Z, eds. *JBI Manual for Evidence Synthesis: JBI*. 2020.
33 World Bank. Countries classified as low-income countries (2023 World Bank). 2023. Available: https://worldpopulationreview.com/country-rankings/low-income-countries
34 Oza S, Jazayeri D, Teich JM, *et al*. Development and deployment of the Openmrs-Ebola electronic health record system for an Ebola treatment center in Sierra Leone. *J Med Internet Res* 2017;19:e294.
35 Robbiati C, Tosti ME, Mezzabotta G, *et al*. Improving TB surveillance and patients' quality of care through improved data collection in Angola: development of an electronic medical record system in two health facilities of Luanda. *Front Public Health* 2022;10:745928.
36 Ngusie HS, Kassie SY, Chereka AA, *et al*. Healthcare providers' readiness for electronic health record adoption: a cross-sectional study during pre-implementation phase. *BMC Health Serv Res* 2022;22:282.
37 Ahmed MH, Bogale AD, Tilahun B, *et al*. Intention to use electronic medical record and its predictors among health care providers at referral hospitals, North-West Ethiopia, 2019: using unified theory of acceptance and use technology 2(Utaut2) model. *BMC Med Inform Decis Mak* 2020;20:207.
38 Oumer A, Muhye A, Dagne I, *et al*. Utilization, determinants, and prospects of electronic medical records in Ethiopia. *Biomed Res Int* 2021;2021:2230618.
39 Liang L, Wiens MO, Lubega P, *et al*. A locally developed electronic health platform in Uganda: development and implementation of Stre@Mline. *JMIR Form Res* 2018;2:e20.
40 Fraser HSF, Mugisha M, Remera E, *et al*. User perceptions and use of an enhanced electronic health record in Rwanda with and without clinical alerts: cross-sectional survey. *JMIR Med Inform* 2022;10:e32305.
41 Were MC, Savai S, Mokaya B, *et al*. mUzima mobile electronic health record (EHR) system: development and implementation at scale. *J Med Internet Res* 2021;23:e26381.
42 Bagayoko CO, Tchuente J, Traoré D, *et al*. Implementation of a national electronic health information system in Gabon: a survey of Healthcare providers' perceptions. *BMC Med Inform Decis Mak* 2020;20:202.
43 Bisrat A, Minda D, Assamnew B, *et al*. Implementation challenges and perception of care providers on electronic medical records at St. Paul's and Ayder hospitals, Ethiopia. *BMC Med Inform Decis Mak* 2021;21:306.
44 Mkalira Msiska KE, Kumitawa A, Kumwenda B. Factors affecting the utilisation of electronic medical records system in Malawian central hospitals. *Malawi Med J* 2017;29:247–53.
45 Abdel-Kader K, Jhamb M. EHR-based clinical trials: the next generation of evidence EHR-based clinical trials. *Clin J Am Soc Nephrol* 2020;15:1050–2.
46 Fanta GB, Pretorius L, Erasmus L. Hospitals' readiness to implement sustainable Smartcare systems in Addis Ababa, Ethiopia. 2019 Portland International Conference on Management of Engineering and Technology (PICMET); Portland, OR, USA.
47 Kabukye JK, de Keizer N, Cornet R. Assessment of organizational readiness to implement an electronic health record system in a low-resource settings cancer hospital: A cross-sectional survey. *PLOS ONE* 2020;15:e0234711.
48 Tsai CH, Eghdam A, Davoody N, *et al*. Effects of electronic health record implementation and barriers to adoption and use: A. *Life* 2020;10:327.
49 Windle JR, Windle TA, Shamavu KY, *et al*. Roadmap to a more useful and usable electronic health record. *Cardiovascular Digital Health Journal* 2021;2:301–11.
50 Starren JB, Tierney WM, Williams MS, *et al*. A retrospective look at the predictions and recommendations from the 2009 AMIA policy meeting: did we see EHR-related clinician burnout coming *J Am Med Inform Assoc* 2021;28:948–54.
51 Amend J, Eymann T, Kauffmann AL, *et al*. Deriving Facilitators for electronic health record implementation: A systematic literature review of opportunities and challenges. ECIS 2022 Research Papers2022;
52 Acholonu RG, Raphael JL. The influence of the electronic health record on achieving equity and eliminating health disparities for children. *Pediatr Ann* 2022;51:e112–7.
53 Parzen-Johnson S, Kronforst KD, Shah RM, *et al*. Use of the electronic health record to optimize antimicrobial Prescribing. *Clin Ther* 2021;43:1681–8.
54 Kabukye JK, Koch S, Cornet R, *et al*. User requirements for an electronic medical records system for oncology in developing countries: A case study of Uganda. *AMIA Annu Symp Proc* 2017;2017:1004–13.

BMJ Health & Care Informatics

Designing and implementing mHealth technology: the challenge of meeting the needs of diverse communities

Vimla L. Patel [1,2] Edward H. Shortliffe [2]

To cite: Patel VL, Shortliffe EH. Designing and implementing mHealth technology: the challenge of meeting the needs of diverse communities. *BMJ Health Care Inform* 2023;**30**:e100813. doi:10.1136/bmjhci-2023-100813

VLP and EHS contributed equally.

Received 17 June 2023
Accepted 20 June 2023

[1]Center for Cognitive Studies in Medicine and Public Health, New York Academy of Medicine, New York, New York, USA
[2]Department of Biomedical Informatics, Vagelos College of Physicians and Surgeons, Columbia University, New York, New York, USA

Correspondence to
Dr Vimla L. Patel;
vpatel@nyam.org

With the widespread adoption of mobile technologies, including in the developing world, there has been enthusiastic exploration of ways that such devices can support care delivery and management in a wide variety of settings. mHealth was accordingly introduced as a general term for the use of such devices, and especially mobile phones, to support the practice of medicine and promotion of public health. The most common application of mHealth has involved the use of mobile devices to communicate with patients or healthy individuals. The goal has been to educate them about health promotion and disease prevention, or to assist with remote patient monitoring and care delivery, either through direct interaction with patients or with health workers. Mobile technology has recently appeared rapidly in low-income and middle-income nations (as defined by the World Bank economic criteria). Middle-income and (especially) low-income countries face various constraints in their healthcare systems, such as a severe lack of human, physical and fiscal resources, as well as highly significant burdens of disease and extreme poverty. Additionally, healthcare access to many parts of society is generally low in these countries.

The potential to lower informational and transactional health-related costs improves when the populace has greater access to mobile phones— typically available in urban settings but also important in rural areas where the communications infrastructure may be suboptimal or absent. These factors have motivated discussions regarding how greater access to mobile phone technology can be leveraged to mitigate the numerous pressures faced by healthcare systems in developing countries. There has been a substantial involvement of informatics professionals in discussions, both as researchers and as system builders. Their work has greatly enhanced our understanding of the optimal strategies and designs for building technical solutions that can be successfully introduced to, and adopted by, some of the most challenging user communities and healthcare settings on the globe.

There are several challenges in mHealth, which have included limited access to mobile devices and constrained cellular or internet connectivity. Even when these problems are addressed, data privacy and security, variations in literacy, cultural factors, and attitudes towards technology can profoundly constrain the use of mobile devices for health delivery and information management. Additionally, even when literacy concerns have been addressed (for both patients and health workers), the lack of standardisation in mHealth interventions can obstruct the best of intentions by those who seek to apply the technology. Literacy issues often mandate that the task undertaken on mobile technology be kept as simple as possible, enhancing efficiency to reach a larger portion of the needy population and reducing opportunities for error. In addition, new research often enlightens our understanding of the considerations that should guide ongoing work in the area.

Burka *et al*[1] present their user-centred design (UCD) approach to designing a digital information system to support chronic disease management (hypertension) in four low-income and middle-income countries in South Asia and Africa. UCD is an iterative design process in which developers focus on the users and their needs in each phase to create usable and accessible products. Particular attention is paid to usability goals (crucial for acceptance), user characteristics, environment, tasks and the workflow surrounding use of the anticipated product. In the study, the authors applied this design approach to create a simple, offline-first,

mobile application for providers to use when recording data during patients' clinical visits, linked to a web-based dashboard that can be used to monitor programme performance. This offline application, focusing on data acquisition and simple guidance, ensures the continuous functionality of the application, even if there is a temporary loss of network connectivity. The article highlights the creation of a fast and easy-to-use hypertension management system, aimed at managing the providers' time constraints by minimising data entry and focusing on key performance indicators. Their goal has been to reach scale successfully in low-resource settings. The application did scale rapidly over 4 years to reach more than 11 400 primary care facilities in the four countries with over three million patients enrolled. This is an impressive result since such scaling usually takes much longer. The authors summarise four key design principles that they believe account for this success: speed/ease of use, minimal data entry, ability to do basic work offline and inclusion of minimalistic requirements designed to address key indicators of quality improvement.

In a second paper, Schretzlmaier *et al*[2] conduct a cross-sectional validation study to evaluate (in two German-speaking countries, Germany and Austria) the extended Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) model[3] for predicting mHealth acceptance, using mobile diabetes applications as an example. The authors found that even though UTAUT2 has been well established in the information technology sector to predict a system's acceptance by its intended users, the original UTAUT2 should be extended by two additional constructs: 'perceived disease threat' and 'trust'. These allow the model to predict mHealth acceptance more effectively. The perceived disease threat is an individual's awareness of the risk and limitations of the disease for their health, especially with chronic diseases like diabetes. They offer a detailed analysis, based on an extensive survey of patients with diabetes who were users of mobile applications, to show that awareness of risk is a significant driver for achieving consistent acceptance and use of mobile health applications. Trust in the technology (ie, that it would not fail if used in their care) was also shown to be a key factor in acceptance. However, the augmented model, with the two additional factors included, while improved over the base model, still could not consistently predict mHealth acceptance.

The two studies analyse user acceptance in very different ways that demonstrate the complexity of the task when one endeavours to introduce mHealth technology for routine use by either patients or providers. In one study, the emphasis is on providers in low-income to middle-income countries, demonstrating that successful implementation is possible if the four key factors are addressed. It uses minimum, immediately relevant patient information for data entry, with limited access to other health information. Sometimes simple technology suffices when attempting to reach a large population quickly. In the other study, the emphasis is on patients in more advanced countries, where they use mHealth to participate in managing their own care. Here, the emphasis of the analysis, and new insights, involve the importance of their own perceptions of the threats of the disease to their own health and their trust in the technology itself. The study offers methodological innovations that can be used to refine the current model for technology evaluation and acceptance.

The two studies are not contradictory, but they demonstrate the complexity of issues that need to be addressed when assessing and designing for user acceptance of mHealth technology. There will clearly be no single solution for all countries, cultures, levels of literacy, disease settings and fiscal environments. As a result, there remain myriad opportunities for study, assessment and development of targeted guidelines that will assist those who seek to engage a wide variety of healthcare communities with mHealth solutions.

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 applicable.

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 Burka D, Gupta R, Moran AE, *et al*. Keep it simple: designing a User- centred Digital information system to support chronic disease management in low/middle-income countries. *BMJ Health Care Inform* 2023;30:e100641.
2 Schretzlmaier P, Hecker A, Ammenwerth E. Extension of the unified theory of acceptance and use of technology 2 model for predicting mHealth acceptance using diabetes as an example: a Crosssectional validation study. *BMJ Health Care Inform* 2022;29:e100640.
3 Ammenwerth E. Technology acceptance models in health Informatics: TAM and UTAUT. *Stud Health Technol Inform* 2019;263:64–71.

BMJ Health & Care Informatics

Digital health in Tasmania – improving patient access and outcomes

Usman Iqbal ,[1,2,3] Warren Prentice,[1] Anthony Lawler[1,4]

To cite: Iqbal U, Prentice W, Lawler A. Digital health in Tasmania – improving patient access and outcomes. *BMJ Health Care Inform* 2023;**30**:e100802. doi:10.1136/bmjhci-2023-100802

Received 10 May 2023
Accepted 03 June 2023

[1]Department of Health, Hobart, Tasmania, Australia
[2]School of Population Health, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia
[3]Department of Global Health and Health Security, College of Public Health, Taipei Medical University, Taipei, Taiwan
[4]School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, Tasmania, Australia

Correspondence to
Adj. Associate Professor Usman Iqbal;
usman.iqbal@unsw.edu.au

With digital health's potential to transform healthcare delivery, Australia is investing significantly to improve healthcare quality and efficiency.[1] Investment is guided by national and global strategies, developed by the Australian Digital Health Agency[1] and the WHO,[2] respectively. The WHO strategy aims to promote the use of digital technologies to improve health outcomes and reduce inequalities, supporting countries in developing their own digital health strategies. The Australian Government is driving several national initiatives, leveraging technology and data to improve patient outcomes and quality of care.[3] While Australian states and territories are at varying stages of their digital health journeys, each has invested in a strategy or framework to guide their digital technology adoption.

Tasmania's digital health strategy harmonises with broader national and international efforts, harnessing the power of technology to improve healthcare. Australia's smallest state, Tasmania has a population of approximately 528 000, with the proportion of people aged 65+ projected to increase from 15% to 24% by 2058.[4] This demographic shift has important healthcare implications, particularly in aged care service demand. In 2022, Tasmania released its digital health strategy,[5] outlining priorities and an investment of US$476 million over the next 10 years for digital health advancements delivered through a state-wide integrated care platform. The strategy fosters multisector, interprofessional collaboration spanning regions, settings and disciplines to provide a longitudinal patient centred view. This includes real-time, secure communication and information exchange between primary, community, acute, subacute and aged care settings.

The strategy centres on state-wide fully integrated electronic medical records (EMRs) within the Tasmanian health system, facilitating real-time provider access to comprehensive patient information. EMRs can significantly improve the care, accuracy and efficiency, reduce risk of errors and adverse medication events, and drive improved clinical outcomes.[6] In addition to EMRs and interoperable digital health infrastructure, delivery will include an integrated prehospital electronic patient care record and the expansion of telehealth services. Telehealth and Virtual Care services particularly benefit patients in rural or remote areas, where in-person care is difficult to access. Telehealth services prevent unnecessary hospital visits, reducing cost and the burden on emergency departments. Additionally, the use of wearable technology and mobile health apps empowers patients to manage their own health, make informed decisions about their care and communicate with their healthcare providers from the comfort of their own homes.[7]

With an emerging data-driven healthcare landscape, leveraging big data and artificial intelligence (AI) through robust, validated and unbiased models enables more personalised, predictive and preventive care models.[8] AI can revolutionise healthcare by improving diagnosis and treatment speed, accuracy and cost-effectiveness. With abundant EMRs data, AI algorithms can learn to identify patterns and make predictions beyond human ability, such as analysing medical images to identify signs of disease earlier.[9] Additionally, routine task automation will reduce administrative burdens on healthcare professionals, such as scheduling appointments and managing electronic health records.[10 11]

While adopting digital solutions is essential to improve care, digital health transformation is not a panacea. Evolving challenges such as a lack of digital health literacy, steep clinician and consumer learning curves and the digital divide may limit realisable benefits. The incorporation of AI is likely to engender data privacy, security and possibly ethical concerns.[12] Tasmania's Health Data Strategy addresses governance challenges by

implementing strict protocols and processes to protect patient interests, ensure data confidentiality, promote evidence-based care and facilitate informed clinician–patient decision-making.[13] Moreover, the strategy incorporates lessons learnt, advice and guidance from other jurisdictions, informing the initial investment. The involvement of all stakeholders, including consumers, families, carers, the clinical and non-clinical workforce, community sector organisations, primary health, private providers and the Tasmanian and Australian governments, is instrumental in shaping this strategy. Supporting this, digital health investment will bolster and propel Tasmania's *One Health Culture Program*[14] by harnessing technology to enhance collaboration, problem-solving, risk-sharing, empowerment and mutual respect among participants. This will enable the programme to more effectively leverage diverse backgrounds, experiences, knowledge and skills, advancing the One Health aspiration. Further, digital health initiatives can drive organisational change through streamlined processes, improved data management and innovation in healthcare delivery.

The future impact of digital health initiatives on healthcare in Tasmania, and Australia more generally, will be significant. With digital health transformation, Tasmanian health will deliver more efficient and effective healthcare services, well positioned for ongoing innovation. Patient-centred digitalisation of healthcare services has the potential to bridge divides, shorten distances and improve outcomes. By investing in and implementing such initiatives, the country and the state can ensure that all Australians have access to high-quality, efficient and effective healthcare.

Contributors Initial conception design and drafting of the manuscript: UI, WP and AL. Drafting the manuscript: UI, WP and AL. Critical revision of the paper: UI, WP and AL.

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 applicable.

Ethics approval Not applicable.

Provenance and peer review Commissioned; externally peer reviewed.

ORCID iD

Usman Iqbal http://orcid.org/0000-0002-0614-123X

REFERENCES

1 Australian Digital Health Agency. National digital health strategy and framework for action. 2021. Available: https://www.digitalhealth.gov.au/about-us/strategies-and-plans/national-digital-health-strategy-and-framework-for-action
2 Global strategy on digital health 2020-2025, licence: CC BY-NC-SA 3.0 IGO. Geneva World Health Organization; 2021.
3 Australian Commission on Safety and Quality in Health Care. E-health safety. 2023. Available: https://www.safetyandquality.gov.au/our-work/e-health-safety [Accessed 15 Apr 2023].
4 Tasmanian Government Department of Health, Our Healthcare Future. Drivers of Tasmania's future population health needs; 2022. Available: https://www.health.tas.gov.au/sites/default/files/2022-06/Drivers%20of%20Tasmania%27s%20Future%20Population%20Health%20Needs_0.pdf
5 Digital health transformation strategy - improving patient outcomes (2022 - 2032). 2022. Available: https://www.health.tas.gov.au/sites/default/files/2022-06/Digital%20Health%20Transformation%20Improving%20Patient%20Outcomes_0.pdf [Accessed 15 Apr 2023].
6 Bates DW, Cheng H-Y, Cheung NT, *et al*. 'Improving smart medication management': an Online expert discussion. *BMJ Health Care Inform* 2022;29:e100540.
7 Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. *N Engl J Med* 2020;382:e82.
8 Iqbal U, Celi LA, Li Y-C. Correction: how can artificial intelligence make medicine more preemptive? *J Med Internet Res* 2020;22:e23645.
9 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
10 Vanguri RS, Luo J, Aukerman AT, *et al*. Multimodal integration of Radiology, pathology and Genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. *Nat Cancer* 2022;3:1151–64.
11 Liu N, Xie F, Siddiqui FJ, *et al*. Leveraging large-scale electronic health records and interpretable machine learning for clinical decision making at the emergency department: protocol for system development and validation. *JMIR Res Protoc* 2022;11:e34201.
12 Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics* 2021;22:122.
13 Seastedt KP, Schwab P, O'Brien Z, *et al*. Global healthcare fairness: We should be sharing more, not less, data. *PLOS Digit Health* 2022;1:e0000102.
14 Department of Health Tasmania. Strategic priorities 2021-23. 2021. Available: https://www.health.tas.gov.au/sites/default/files/2021-12/Strategic_Priorities_2021-23_DoHTasmania.pdf [Accessed 15 Apr 2023].

BMJ Health & Care Informatics

ChatGPT in glioma adjuvant therapy decision making: ready to assume the role of a doctor in the tumour board?

Julien Haemmerli ,[1] Lukas Sveikata,[2,3,4] Aria Nouri,[1] Adrien May,[1] Kristof Egervari,[5] Christian Freyschlag,[6] Johannes A Lobrinus,[5] Denis Migliorini,[7] Shahan Momjian,[1] Nicolae Sanda,[2] Karl Schaller,[1] Sebastien Tran,[8] Jacky Yeung,[9] Philippe Bijlenga[1]

To cite: Haemmerli J, Sveikata L, Nouri A, *et al*. ChatGPT in glioma adjuvant therapy decision making: ready to assume the role of a doctor in the tumour board? *BMJ Health Care Inform* 2023;**30**:e100775. doi:10.1136/bmjhci-2023-100775

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2023-100775).

JH and LS contributed equally.

Received 03 April 2023
Accepted 21 June 2023

For numbered affiliations see end of article.

Correspondence to
Dr Julien Haemmerli;
julien.haemmerli@hcuge.ch

ABSTRACT

Objective To evaluate ChatGPT's performance in brain glioma adjuvant therapy decision-making.

Methods We randomly selected 10 patients with brain gliomas discussed at our institution's central nervous system tumour board (CNS TB). Patients' clinical status, surgical outcome, textual imaging information and immuno-pathology results were provided to ChatGPT V.3.5 and seven CNS tumour experts. The chatbot was asked to give the adjuvant treatment choice, and the regimen while considering the patient's functional status. The experts rated the artificial intelligence-based recommendations from 0 (complete disagreement) to 10 (complete agreement). An intraclass correlation coefficient agreement (ICC) was used to measure the inter-rater agreement.

Results Eight patients (80%) met the criteria for glioblastoma and two (20%) were low-grade gliomas. The experts rated the quality of ChatGPT recommendations as poor for diagnosis (median 3, IQR 1–7.8, ICC 0.9, 95% CI 0.7 to 1.0), good for treatment recommendation (7, IQR 6–8, ICC 0.8, 95% CI 0.4 to 0.9), good for therapy regimen (7, IQR 4–8, ICC 0.8, 95% CI 0.5 to 0.9), moderate for functional status consideration (6, IQR 1–7, ICC 0.7, 95% CI 0.3 to 0.9) and moderate for overall agreement with the recommendations (5, IQR 3–7, ICC 0.7, 95% CI 0.3 to 0.9). No differences were observed between the glioblastomas and low-grade glioma ratings.

Conclusions ChatGPT performed poorly in classifying glioma types but was good for adjuvant treatment recommendations as evaluated by CNS TB experts. Even though the ChatGPT lacks the precision to replace expert opinion, it may serve as a promising supplemental tool within a human-in-the-loop approach.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Advanced artificial intelligence (AI) language models, such as ChatGPT, are quickly evolving and have the potential to incorporate multi-modal medical information and assist with complicated medical decision-making.

WHAT THIS STUDY ADDS

⇒ The use of AI in making therapeutic decisions for central nervous system tumours has not been fully explored. This study aims to assess the effectiveness of AI compared with expert recommendations in aiding complex brain tumour decision-making, providing valuable insights into the potential and limitations of AI in this field.

⇒ This study shows that an AI language model was successful in suggesting adjuvant treatment plans for glioma patients. However, the model had difficulty accurately identifying glioma subtypes and only achieved moderate success in taking patients' functional status into account when making recommendations.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ While AI language models like ChatGPT cannot currently replace the opinions of medical experts, they may serve as a useful supplementary tool in aiding complex brain tumour decisions when used as part of a human-in-the-loop approach.

INTRODUCTION

Artificial intelligence (AI) is attracting a lot of interest in the present era of personalised medicine.[1–3] Since novel drug discovery, surgical robotics or complex interdisciplinary oncological therapy decisions are time-consuming and resource-demanding, innovative AI-based language models may enhance the performance of healthcare ecosystems.[4–6] Recently, a novel general-purpose AI chatbot, called ChatGPT-3.5 (Generative Pretrained Transformer 3.5), was launched, spurring mixed reactions of curiosity and scepticism from the scientific community.[7–11]

ChatGPT is an AI-powered chat interface which results in a language model that uses unsupervised learning and generates human-like text. It allows humans to satisfy their curiosity by engaging in a dialogue using various questions and prompts.[12] Although

the chatbot was not designed to deliver medical knowledge, it allows one to chat on specific medical topics and provides answers with a tone of authority as one would interact with an expert. Nevertheless, chatGPT has some limitations such as the availability of online data until September 2021 and that it sometimes provides incorrect although plausible-sounding answers[13] possibly limiting its use in medical settings.

Neuro-oncolgy has significantly evolved in parallel with new research advances.[14] For instance, the treatment of high-grade gliomas has been extensively studied for the last 20 years to offer a longer survival rate for affected individuals.[15 16] Furthermore, the consideration of the patient's clinical status, age, and comorbidities have been included in novel trials to optimise treatment protocols.[17] Low-grade gliomas which account for approximately 20% of all gliomas are even more heterogenous and adjuvant treatment is based on their complex molecular profile.[18–20] In order to deliver the best treatment strategies for glioma patients, central nervous system (CNS) tumour boards (TB) arose implicating a multidisciplinary team composed of neurosurgeons, oncologists, neurologists, pathologists, radiation oncologists and neuroradiologists.[21] TBs are, however, mobilising an extensive amount of resources, which might be challenging to apply in every scenario. In this regard, AI-assisted decision-making could prove helpful in delivering personalised treatment strategies.[22]

Given the promise of AI in using vast amounts of knowledge to synthesise information and provide recommendations, we investigated whether ChatGPT had a role to play in CNS TB regarding glioma patient adjuvant therapy decision-making. We hypothesised that ChatGPT would perform as well as CNS TB experts in providing glioma subtype diagnosis and adjuvant treatment strategy in line with the current guidelines.[23]

METHODS

Patients' selection

We randomly selected 10 glioma cases from our institutional CNS TB registry from 2014 to 2022. During this period a total of 215 brain glioma cases were evaluated. Inclusion criteria were: (1) new onset or recurrent supratentorial glioma, (2) surgical treatment was performed (removal or biopsy), (3) CNS TB recommendation and (4) informed consent was available. Exclusion criteria were: (1) a presence of brain metastasis, (2) extra-axial tumours and (3) glioma involving the brainstem or the spinal cord.

Dialogue with ChatGPT

Electronic patients' records were retrospectively reviewed. From 1 February to 14 February 2023, 10 case summaries were presented to ChatGPT (V.3.5, February 2023). A separate chat session was used for each case and was presented concisely with information on age, sex, medical history, symptoms, textual imaging results, surgical outcome, tumour resection extent, histopathological and molecular examination results. No diagnosis nor patient identification information was provided to ChatGPT. The questionnaire was modelled after a real-life TB panel discussion format. Two questions were asked to ChatGPT: (1) 'what is the best adjuvant treatment?', (2) 'what would be the regimen of radiotherapy and chemotherapy for this patient?'. ChatGPT's answers were collected. The same case information and a complete chat transcript were provided to the experts (online supplemental material 1). As a quality control measure, we asked the chatbot to provide the presumed diagnosis, which was consistent with its initial spontaneousresponse for each case.

CNS TB and experts' selection

Our institutional CNS TB is composed of neuro-oncologists, radio-oncologists, radiologists, neurosurgeons, neuropathologists and neurologists. We considered our institutional CNS TB as a reference, as its decisions are evidence-based and are supported by a multidisciplinary consensus. Every patient with CNS oncological disease admitted to our institution is presented at this multidisciplinary meeting. For the purpose of this study, five experts from our CNS TB (two neuropathologists, one neurosurgeon, one radio-oncologist and one oncologist) and two external independent experts (two neurosurgeons from Europe and North America) evaluated ChatGPT's output with regard to the formal decision of the CNS TB.

Studied parameters

The experts were asked to rank ChatGPT's answers for each of the 10 cases. The CNS TB decisions were used as the gold standard. The experts were asked to evaluate the ChatGPT's output on a scale between 0 and 10, where '0' indicated complete disagreement, '10' indicated complete agreement and '5' a neutral answer ('neither agreement nor disagreement'). The experts had to evaluate ChatGPT's answers regarding the diagnosis, the proposed treatment, the consideration of the patient's functional status to support adjuvant therapy, the proposed regimen of adjuvant therapy and the overall accuracy of ChatGPT with respect to its answers. Finally, the experts were asked to provide their opinion on the possible place of AI in interdisciplinary CNS tumour decision-making. The experts were provided with a questionnaire to rate ChatGPT's performance in providing the diagnosis of specific glioma types, adjuvant treatment recommendations, adjuvant therapy regimen, how well the chatbot integrated the overall functional status of the patient into the decision-making and the overall quality of the recommendations provided. Figure 1 summarises the study workflow. Online supplemental material 2 presents the questions asked to the experts. Finally, the agreement between experts was evaluated.

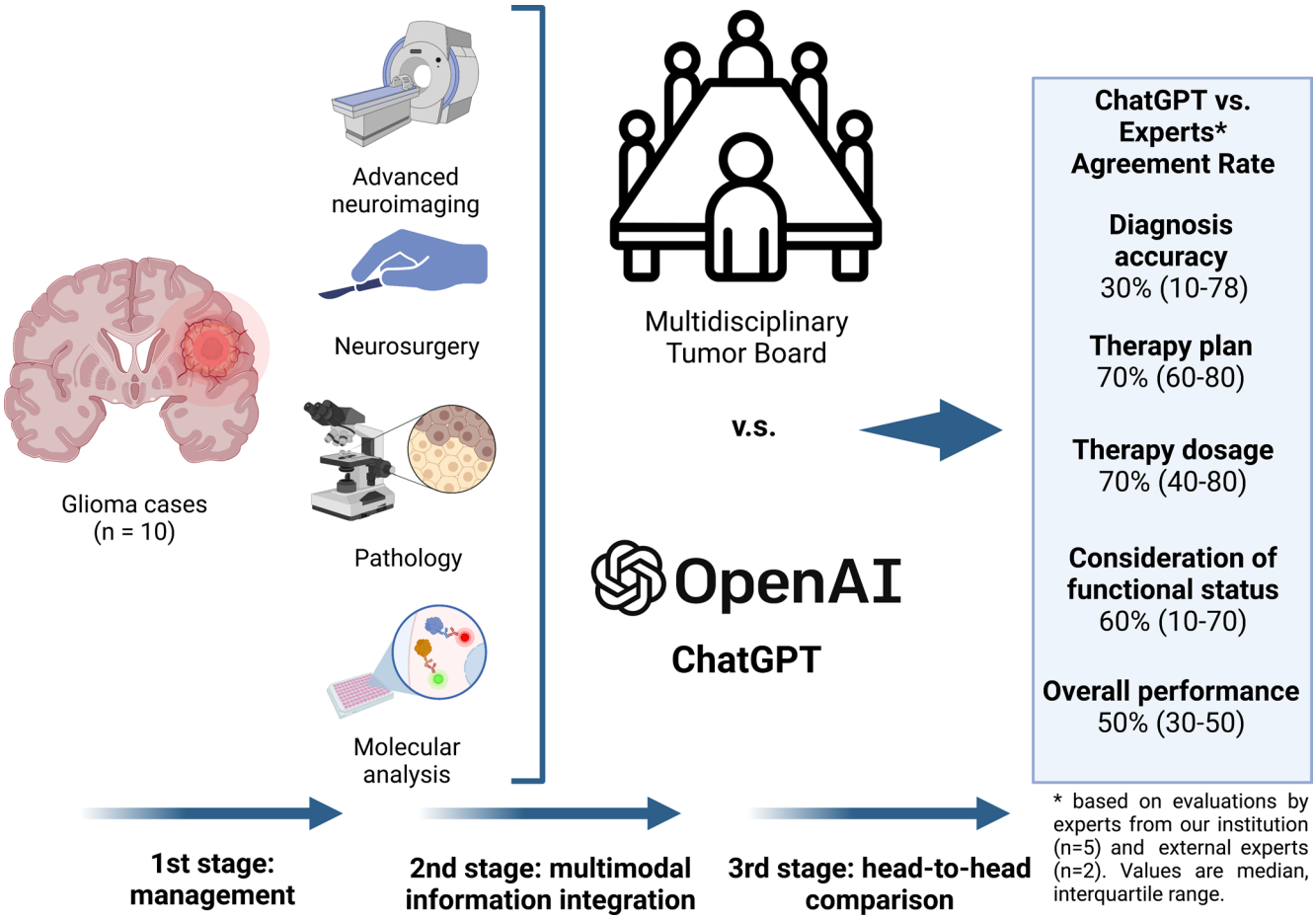

Figure 1 Summary of study workflow. Ten patients were randomly selected from our institutional central nervous system (CNS) tumour board (TB) registry. All cases received state-of-the-art preoperative and postoperative glioma workups. Third, a summary of the anonymised case, including clinical, textual imaging information and immunohistological findings were presented to the ChatGPT, as it would be done at the CNS TB. Seven experts compared ChatGPT's output and the TB recommendations. The results represent the median experts' rating with the IQR. The figure was created with BioRender.com.

Statistics

We used R V.3.6.1 for the statistical analysis. The randomisation process was performed using function floor(runif). Ordinal variables were presented as median with IQR and were compared using a Mann-Whitney U test when appropriate. Experts' rating score between 0 and 3 was considered poor, 4 and 6 as moderate, 7 and 8 as good, and 9 and 10 as excellent. The intraclass correlation coefficient (ICC) was used to evaluate the agreement between the experts (two-way random effects, absolute agreement, multiple raters average, ICC $(2,k)$).[24] An ICC <0.5 was considered as poor, ≥0.5 and <0.75 as moderate, ≥0.75 and <0.9 as good and ≥0.9 as excellent agreement.[24] Hypothesis testing was considered significant at p value <0.05 (two-sided).

RESULTS

ChatGPT's output

ChatGPT provided the diagnosis for suspected glioma type, recommendations for adjuvant treatment plan, regimen for radiotherapy and chemotherapy, and consideration of functional status for all 10 cases. Regarding the first question 'what is the best adjuvant treatment', ChatGPT started the dialogue by giving its appreciation of the diagnosis. Based on the patient summary, it correctly recognised and classified the tumours as glioma in all cases and suggested the tumour type (eg, low-grade glioma, grade II or III astrocytoma, glioblastoma). Of note, no alternative diagnosis such as brain metastasis or extra-axial brain tumour was proposed. ChatGPT then recommended 'the best adjuvant treatment […]' or 'the standard of care for glioblastoma […]'. Concerning the second question 'what would be the regimen of radiotherapy and chemotherapy for this patient', ChatGPT provided a recommendation for all cases. However, a complete regimen of radiotherapy (greys in fractions over weeks) was provided in 70% of the cases, and a complete regimen of chemotherapy (medication and doses) in 50% of cases.

For both questions, ChatGPT nuanced its answers for all cases by mentioning the need to adjust the treatment according to the patient's individual preferences and functional status, although never specifying alternatives. Finally, ChatGPT mentioned the need to confirm

its treatment suggestion with a multidisciplinary team in 80% of the cases.

Experts' opinion and agreement

Seven experts rated ChatGPT's output regarding the diagnosis, recommendations for therapy and regimen, the consideration of the patient's functional status and ChatGPT's overall performance. Rater 6 only rated the diagnosis accuracy and treatment recommendations for case 2 and did not rate the output regarding the consideration of the functional status nor the regimen of adjuvant therapy (the expert preferred to remain in their scope of practice).

Figure 2 demonstrates the inter-rater agreement for each evaluated outcome. Concerning the diagnosis, ChatGPT's output was evaluated as poor with a median score of 3 (IQR 1–7.8) with excellent agreement between the experts (ICC 0.9, 95% CI 0.7 to 1.0). For the adjuvant therapy, the ChatGPT recommendations were evaluated as good with a median score of 7 (IQR 6–8) and a good agreement (ICC 0.8, 95% CI 0.4 to 0.9). The adjuvant

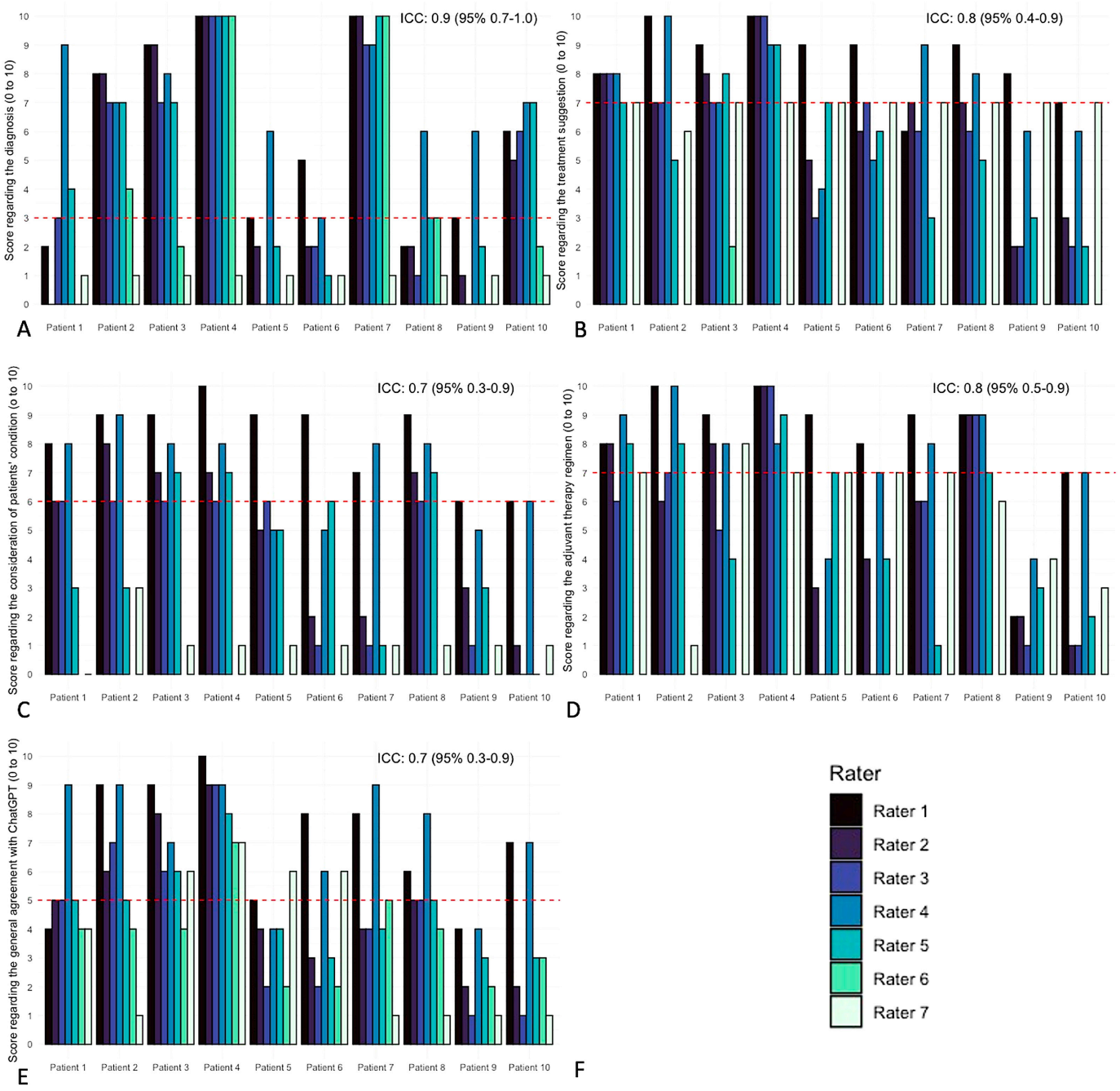

Figure 2 Barplots representing the ratings per patient and per expert, regarding (A) the diagnosis, (B) the adjuvant treatment recommendation, (C) the consideration of the patient's functional status, (D) the regimen of the adjuvant therapy, (E) ChatGPT's overall performance, (F) the legend. ICC, intraclass correlation coefficient (from 0 to 10, 95% CI). The dashed red line represents the median value of the experts' rating.

therapy regimen was evaluated as good with a median score of 7 (IQR 4–8) and good expert agreement (ICC 0.8, 95% CI 0.5 to 0.9). Regarding ChatGPT's output on the consideration of the patient's functional status, the experts rated the recommendations as moderate with a median score of 6 (IQR 1–7) and a moderate agreement (ICC 0.7, 95% CI 0.3 to 0.9). Finally, the global evaluation of ChatGPT's output accuracy was moderate and scored 5 (IQR 3–5) with a moderate expert agreement (ICC 0.7, 95% CI 0.3 to 0.9). Six experts (86%) evaluated ChatGPT's role in a CNS TB as useful if the AI-based system can evolve and learn. One rater (14%) evaluated ChatGPT's role in a CNS TB as useful, but only in specific circumstances.

There was no significant difference between experts' ratings in glioblastoma (8/10) and two low-grade glioma cases.

DISCUSSION

In this study, we assessed the performance of ChatGPT, an AI-based language model, in providing treatment recommendations for glioma patients. To the best of our knowledge, this is the first study aiming to evaluate this novel chatbot within the framework of CNS tumour multidisciplinary decision-making. While ChatGPT demonstrated proficiency in accurately identifying cases as gliomas, it displayed limited precision in identifying specific tumour subtypes. Furthermore, the tool's recommendations regarding treatment strategy and regimen were rated as good, while the ability to incorporate functional status in its decision-making process as moderate.

Rationale for CNS TB

Oncological patients discussed in the multidisciplinary CNS TB are more likely to benefit from a preoperative and postoperative staging and are more likely to receive the optimal adjuvant treatment.[25 26] Barbaro *et al* presented the foundations of neuro-oncology and the need for multidisciplinary expertise in order to embrace the multiple disease aspects in CNS tumour-affected patients.[14] The authors highlighted the prerogatives and missions of a CNS TB: (1) neuro-oncology, neurosurgery, radiation oncology, neuropathology, neurology and radiology are specialties necessary to compose the CNS TB; (2) the expert consortium's main goal is to propose a collaborative treatment plan; (3) the development of novel clinical trials. Furthermore, a single-centre prospective evaluation of a CNS TB showed that the experts' consortium influences the clinical management of patients suffering from a brain tumour through high-impact decisions.[27] However, the organisation of CNS TB is limited by economic costs, time expenditure, resource availability and the limited presence of TB across the geographic and socioeconomic strata.[26] New AI-based tools with underlying deep learning, such as ChatGPT, might represent a valuable complement or at least offer some help to centres lacking expertise or resources.

ChatGPT ready to assume the role of the doctor?

Two questions were asked ChatGPT that corresponded to the main aim of a CNS TB discussion: 'what is the best adjuvant treatment?', and 'what would be the regimen of radiotherapy and chemotherapy for this patient?'. ChatGPT scored well on both parameters, but its responses were less accurate on other parameters such as incorporating the functional status of the patient, and glioma subtype diagnostic accuracy. Regarding the latter, the output provided by the chatbot was often incorrect (ie, pleiomorphic astrocytoma instead of glioblastoma in one case), or not detailed enough (ie, no distinction between grade II or III astrocytoma). On the other hand, the adjuvant treatment suggestion and its regimen were rated as good. In future studies, it may be worth exploring alternative questioning methods that align better with how chatbots process information. This approach could potentially lead to more accurate results.

In this cohort, 80% of the included patients were diagnosed with glioblastoma (WHO grade IV). In the literature, the treatment of glioblastoma WHO IV has been extensively studied.[15–17 19 23 28] AI models used by ChatGPT are trained on a large dataset of information found online including websites, journals and digitalised books. It is thus comprehensible that ChatGPT's output regarding the adjuvant treatment and its regimen related to glioblastoma is of better quality because the underlying knowledge base is well-documented. To this extent, ChatGPT's performance is mediocre regarding recommendations that are based on less extensive knowledge base. The consideration of patient functional status was rated as moderate, even though the clinical preoperative and postoperative state of the included cases was presented to ChatGPT. This consideration is much less documented in the literature as only a few clinical trials studied adjuvant therapy for glioblastoma in patients with impaired functional status or in older adults.[17]

Strengths and limitations

Our results provide valuable information on the potential of human-AI interfaces in medical decision-making. To test the chatbot's performance, we have used glioma cases which represent a homogenous sample of tumour cases which allowed us to test the performance in this setting but limited the generalisability of our findings to other tumour types. Of note, ChatGPT's recommendations were conscientiously mitigated with disclosure statements that it was not designed to provide medical advice, which presents another limitation in a medical setting. Notwithstanding, it might be seen as an opportunity if similar algorithms would be designed specifically for this purpose. Given this, at the moment we cannot appreciate the full potential of ChatGPT in CNS TB. Notwithstanding this limitation, one could imagine that AI chatbots, with pursued development in the medical field, could hold great promise to complement the classic CNS TB workflow. Another limitation lies in the fact that the chatbot's knowledge relies on content from

the internet limited to 2021. Although information on more novel research developments in neuro-oncology were not accessible for the chatbot, this should not have impacted its recommendations for standard clinical care. If the chatbot had access to information on new clinical trials, it could greatly aid the therapeutic discussion and potentially lead to new development directions . Finally, ChatGPT recommendations cannot be taken at face value without specialist verification since it is not uncommon for the chatbot to provide erroneous information.[13] In language models such as ChatGPT, a phenomenon known as 'hallucinations' frequently occurs and can span from rather benign, for example, providing plausible but non-existent scientific references, to very dangerous medical scenarios, such as recommending an ineffective or harmful treatment.[2] Therefore, whether used to inform medical or other high-stake decisions, at this stage it is indispensable that the output is verified by a human professional. Finally, our study relied on textual neuroimaging information and did not involve a quantitative AI imaging analysis which could be a potential area of development.

Further developments

Six of the seven experts evaluated ChatGPT as useful if the system could learn and improve. This notion is supported by the medical community as AI is growing and holds immense promise in medicine.[2 6 29–31] However, since its launch in November 2022, ChatGPT has raised scepticism in the scientific community regarding threats to the originality of scientific work.[10 11 32–35] Another consideration is the risk that AI chatbots may be prone to bias or commit omissions and errors in the interpretation of medical information. Due to these shortcomings, AI-based systems in medicine should be used with a human-in-the-loop approach.

Even if our results suggest a reserved rating for ChatGPT's performance on glioma subtype diagnosis and multi-modal information integration, AI-based chatbots may be a promising supplement in TB decision-making. Future studies could explore ways to refine ChatGPT's functionality, such as incorporating more patient-specific data and refining its ability to provide nuanced recommendations based on the clinical context. Furthermore, future developments in the ChatGPT interface could introduce the ability to read medical imaging, such as preoperative and postoperative brain MRI, which could enormously improve its diagnostic ability and treatment recommendations.

Nonetheless, our results highlight the potential utility of ChatGPT in facilitating clinical decision-making. Chatbots could be used to quickly provide information related to a patient's medical history, differential diagnosis, relevant diagnostic tests, experimental treatment options and potential side effects. Furthermore, we intentionally provided the chatbot with only one conversation log. Thus, it is possible that further interaction and additional discussion with the chatbot may have yielded increased performance.

However, ChatGPT's ability to provide medical information was restricted as it did not have access to the latest clinical trial findings. This was because it lacks live internet access and access to research databases.[28] Overcoming these barriers and facilitating AI access to the newest scientific information, could be a potential direction of future development as the novel clinical trials are a crucial part of a CNS TB discussion.[14] AI-based chatbots could have the potential to integrate the newest trial and bench science information into multidisciplinary decision-making and help TB direct patients to potential applicable treatments.

AI language models are evolving at a tremendous speed, and by the time of the publication of this manuscript, a newer ChatGPT V.4.0 was introduced, offering a more versatile conversational tool. It is possible that future updates may include a neuro-imaging analysis tool, which would greatly enhance the complexity of AI tools available for the medical field.

CONCLUSION

We have evaluated the performance of the novel AI-based language generator ChatGPT in glioma-related treatment recommendations. ChatGPT correctly identified the cases as CNS tumours but lacked precision on tumour subtype. The treatment strategy and regimen recommendations were rated as good; however, it lacked the ability to nuance its recommendations when taking into consideration the functional status. Overall, our findings suggest that ChatGPT has potential as an adjunct to the multidisciplinary TB decision workflow within a human-in-the-loop approach, provided that further algorithmic advancements are made in the medical domain.

Author affiliations
[1]Department of Clinical Neurosciences, Division of Neurosurgery, Geneva University Hospitals, Geneva, Switzerland
[2]Department of Clinical Neurosciences, Division of Neurology, Geneva University Hospitals, Geneva, Switzerland
[3]Institute of Cardiology, Lithuanian University of Health Sciences, Kaunas, Lithuania
[4]Faculty of Medicine, University of Geneva, Geneva, Switzerland
[5]Department of Pathology and Immunology, Geneva University Hospitals, Geneva, Switzerland
[6]Department of Neurosurgery, Medical University of Innsbruck, Innsbruck, Austria
[7]Department of Oncology, Geneva University Hospitals, Geneva, Switzerland
[8]Department of Radiation Oncology, Geneva University Hospitals, Geneva, Switzerland
[9]Department of Neurosurgery, Yale University School of Medicine, New Haven, Connecticut, USA

Twitter Lukas Sveikata @lsveikata

Contributors JH and LS contributed equally to this work. JH acts as a guarantor of the published work.

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 applicable.

Ethics approval All included patients signed the local general consent for reuse of clinical data for glioma patients (Geneva University Hospitals).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iD

Julien Haemmerli http://orcid.org/0000-0001-5998-8541

REFERENCES

1 Bhinder B, Gilvary C, Madhukar NS, *et al*. Artificial intelligence in cancer research and precision medicine. *Cancer Discov* 2021;11:900–15.
2 Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism* 2017;69S:S36–40.
3 Kulkarni S, Seneviratne N, Baig MS, *et al*. Artificial intelligence in medicine: where are we now? *Acad Radiol* 2020;27:62–70.
4 Lee J, Yoon W, Kim S, *et al*. Biobert: a pre-trained BIOMEDICAL language representation model for BIOMEDICAL text mining. *Bioinformatics* 2020;36:1234–40.
5 Liu Z, Roberts RA, Lal-Nag M, *et al*. AI-based language models powering drug discovery and development. *Drug Discov Today* 2021;26:2593–607.
6 Shimizu H, Nakayama KI. Artificial intelligence in oncology. *Cancer Sci* 2020;111:1452–60.
7 Biswas S. Chatgpt and the future of medical writing. *Radiology* 2023;307:e223312.
8 Else H. Abstracts written by Chatgpt fool scientists. *Nature* 2023;613:423.
9 Huh S. Are Chatgpt's knowledge and interpretation ability comparable to those of medical students in Korea for taking a Parasitology examination?: a descriptive study. *J Educ Eval Health Prof* 2023;20:1.
10 The lancet Digital health null. Chatgpt: friend or foe? *Lancet Digit Health* 2023;5:e102.
11 van Dis EAM, Bollen J, Zuidema W, *et al*. Chatgpt: five priorities for research. *Nature* 2023;614:224–6.
12 ChatGPT. Available: https://chat.openai.com
13 Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI Chatbot for medicine. reply. *N Engl J Med* 2023;388:10.1056/NEJMc2305286#sa3.
14 Barbaro M, Fine HA, Magge RS. Foundations of neuro-oncology: A Multidisciplinary approach. *World Neurosurg* 2021;151:392–401.
15 Stupp R, Mason WP, van den Bent MJ, *et al*. Radiotherapy plus concomitant and adjuvant Temozolomide for glioblastoma. *N Engl J Med* 2005;352:987–96.
16 Stupp R, Taillibert S, Kanner A, *et al*. Effect of tumor-treating fields plus maintenance Temozolomide vs maintenance Temozolomide alone on survival in patients with glioblastoma: A randomized clinical trial. *JAMA* 2017;318:2306–16.
17 Malmström A, Grønberg BH, Marosi C, *et al*. Temozolomide versus standard 6-week radiotherapy versus Hypofractionated radiotherapy in patients older than 60 years with glioblastoma: the Nordic randomised, phase 3 trial. *Lancet Oncol* 2012;13:916–26.
18 Wang TJC, Mehta MP. Low-grade glioma radiotherapy treatment and trials. *Neurosurg Clin N Am* 2019;30:111–8.
19 Hervey-Jumper SL, Berger MS. Maximizing safe resection of Low- and high-grade glioma. *J Neurooncol* 2016;130:269–82.
20 Ryken TC, Parney I, Buatti J, *et al*. The role of radiotherapy in the management of patients with diffuse low grade glioma: A systematic review and evidence-based clinical practice guideline. *J Neurooncol* 2015;125:551–83.
21 Snyder J, Schultz L, Walbert T. The role of tumor board conferences in neuro-oncology: a nationwide provider survey. *J Neurooncol* 2017;133:1–7.
22 Maron JL. Personalizing therapies and targeting treatment strategies through Pharmacogenomics and artificial intelligence. *Clin Ther* 2021;43:793–4.
23 Louis DN, Perry A, Reifenberger G, *et al*. The 2016 world health organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol* 2016;131:803–20.
24 Koo TK, Li MY. A guideline of selecting and reporting Intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155–63.
25 Pillay B, Wootten AC, Crowe H, *et al*. The impact of Multidisciplinary team meetings on patient assessment, management and outcomes in oncology settings: A systematic review of the literature. *Cancer Treat Rev* 2016;42:56–72.
26 Berardi R, Morgese F, Rinaldi S, *et al*. Benefits and limitations of a Multidisciplinary approach in cancer patient management. *Cancer Manag Res* 2020;12:9363–74.
27 Ameratunga M, Miller D, Ng W, *et al*. A single-institution prospective evaluation of a neuro-oncology Multidisciplinary team meeting. *J Clin Neurosci* 2018;56:127–30.
28 Bagley SJ, Kothari S, Rahman R, *et al*. Glioblastoma clinical trials: Current landscape and opportunities for improvement. *Clin Cancer Res* 2022;28:594–602.
29 Deng Y, Qin H-Y, Zhou Y-Y, *et al*. Artificial intelligence applications in pathological diagnosis of gastric cancer. *Heliyon* 2022;8:e12431.
30 Connor CW. Artificial intelligence and machine learning in Anesthesiology. *Anesthesiology* 2019;131:1346–59.
31 Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. *J Intern Med* 2020;288:62–81.
32 Thorp HH. Chatgpt is fun, but not an author. *Science* 2023;379:313.
33 Kitamura FC. Chatgpt is shaping the future of medical writing but still requires human judgment. *Radiology* 2023;307:e230171.
34 Tools such as Chatgpt threaten transparent science; here are our ground rules for their use. *Nature* 2023;613:612.
35 Doshi RH, Bajaj SS, Krumholz HM. Chatgpt: temptations of progress. *Am J Bioeth* 2023;23:6–8.

BMJ Health & Care Informatics

Delivering on NIH data sharing requirements: avoiding Open Data in Appearance Only

Hope Watson,[1] Jack Gallifant ,[2] Yuan Lai,[3,4] Alexander P Radunsky,[5,6,7] Cleva Villanueva,[8] Nicole Martinez,[9] Judy Gichoya ,[10] Uyen Kim Huynh,[11] Leo Anthony Celi [12,13]

To cite: Watson H, Gallifant J, Lai Y, *et al*. Delivering on NIH data sharing requirements: avoiding Open Data in Appearance Only. *BMJ Health Care Inform* 2023;**30**:e100771. doi:10.1136/bmjhci-2023-100771

Received 23 March 2023
Accepted 29 May 2023

For numbered affiliations see end of article.

Correspondence to
Dr Jack Gallifant;
jack.gallifant@nhs.net

ABSTRACT

Introduction In January, the National Institutes of Health (NIH) implemented a Data Management and Sharing Policy aiming to leverage data collected during NIH-funded research. The COVID-19 pandemic illustrated that this practice is equally vital for augmenting patient research. In addition, data sharing acts as a necessary safeguard against the introduction of analytical biases. While the pandemic provided an opportunity to curtail critical research issues such as reproducibility and validity through data sharing, this did not materialise in practice and became an example of 'Open Data in Appearance Only' (ODIAO). Here, we define ODIAO as the intent of data sharing without the occurrence of actual data sharing (eg, material or digital data transfers).

Objective Propose a framework that states the main risks associated with data sharing, systematically present risk mitigation strategies and provide examples through a healthcare lens.

Methods This framework was informed by critical aspects of both the Open Data Institute and the NIH's 2023 Data Management and Sharing Policy plan guidelines.

Results Through our examination of legal, technical, reputational and commercial categories, we find barriers to data sharing ranging from misinterpretation of General Data Privacy Rule to lack of technical personnel able to execute large data transfers. From this, we deduce that at numerous touchpoints, data sharing is presently too disincentivised to become the norm.

Conclusion In order to move towards Open Data, we propose the creation of mechanisms for incentivisation, beginning with recentring data sharing on patient benefits, additional clauses in grant requirements and committees to encourage adherence to data reporting practices.

INTRODUCTION

Six years on from the development of the FAIR data principles[1] (Findability, Accessibility, Interoperability, and Reusability), the recent deployment of the NIH data sharing mandate is a significant step towards increasing the reproducibility and robustness of research that has long eluded the data science community.[2–4] From January 2023, NIH intramural investigators will be required to prospectively plan for the management and sharing of scientific data, and must submit a data management and sharing (DMS) for each new grant.[2] At a minimum data supporting a publication must be shared at the time of dissemination, and other scientific data released at the end of the research project or protocol. This mandate facilitates an ecosystem-wide shift in mindset surrounding data sharing, creating a culture that places efficient accumulation of knowledge and, ultimately, patients first.[5 6]

Unfortunately, previous initiatives have encountered several obstacles and resistance, as data sharing is not as simple as is often implied.[7] The COVID-19 pandemic highlighted this issue, demanding the public reporting of health data at a scale unlike any other. Continuous monitoring of the quality of care and international comparisons was vital.[8] The clinical and academic communities were also desperate for patient-level data that researchers could evaluate to identify trends and treatments. A significant volume of preprints of questionable reliability transpired, which had no way of validating results.[9] Furthermore, there has been a lack of precise results from trials published with significant duplication resulting, for example, across all registered COVID-19 research studies on CT.gov, only 3% had reported results in July 2022 despite 53% being past completion dates.[10]

It is therefore vital to realise that data sharing is fraught with difficulties spanning technical, legal and organisational risks. Even though Open Access is increasingly supported by many, data sharing is less prevalent.[11] During the pandemic, incentives for sharing were high and the dangers of witholding data were equally significant. Yet, there was limited improvement in the wider system that encourages and facilitates data sharing. Instead, the notion of open data is shrouded in complexity and deemed far to

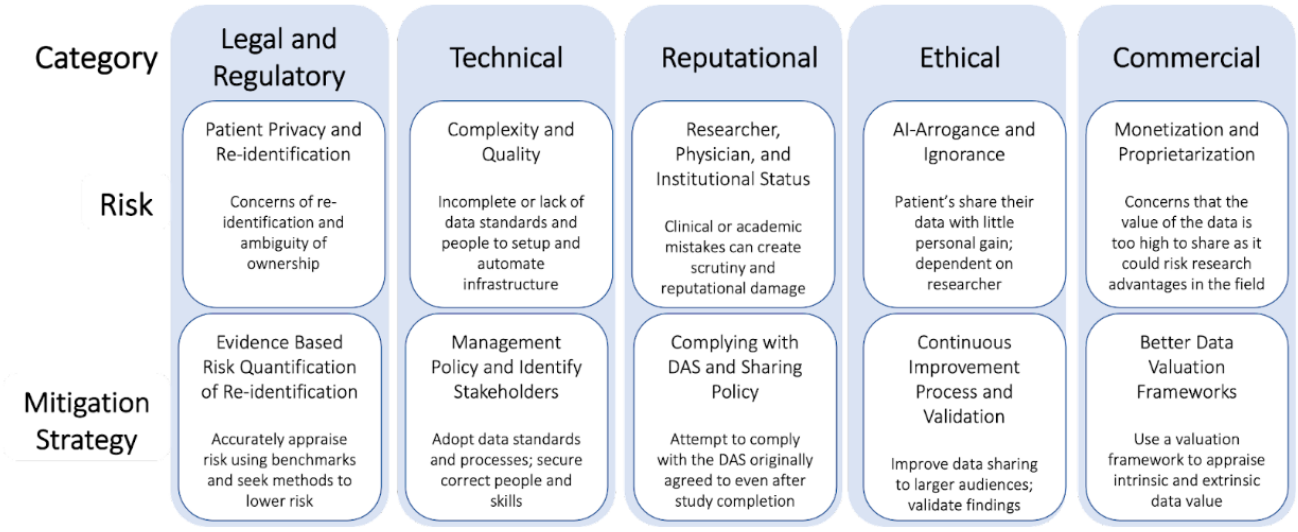

Figure 1 Data sharing risk and mitigation framework. DAS, data availability statement.

risk. Here, we coin the term Open Data in Appearance Only (ODIAO) defined as the intent of data sharing, but without any actual data sharing occurrence (eg, material or digital data transfers).[12 13]

Data sharing has been debated for many years and across industries, where barriers to distribution have been laid out by The Open Data Institute (ODI), particularly in their 5-year strategy (2023–2028).[14] The ODI notes several vital developments that must be overcome to facilitate data sharing and build stakeholder trust. In addition, this mirrors guidelines from the recent NIH DMS plan, which focuses on improving safe data management and its sharing.[15] Both documents aim to accelerate health research, improve transparency and reduce biases transmitted to downstream tasks. In this review, we explore and summarise key lessons from these two critical reports on data sharing risks and barriers. In order to prevent ODIAO, we sought to harmonise and incorporate these key factors into one overarching framework for data sharing that could be used to deliver the recent NIH initiative (figure 1), addressing each factor in turn.

LEGAL AND REGULATORY

Patient privacy and reidentification

The goals of a variety of health and data laws, such as Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health (HITECH) Act, are to protect patient privacy and to create a unified digital infrastructure to improve quality, safety and cost of care.[16] While these initiatives have incentivised practices such as electronic health record (EHR) adoption, there are also penalties and fines for breaches and patient identification[17] that are largely used as reasons not to share data. The risk of reidentification is frequently dependent on knowing pieces of information about a patient outside of bounds of the deidentified data. This includes other publicly available dataset or personal knowledge. A recent study showed that by using publicly available newspaper data to match names to anonymised patient records in statewide hospital data 28% of names in Maine and 34% of names in Vermont were able to be uniquely matched to one hospitalisation. After redacting the same data to HIPAA Safe Harbor standards the linkage rate was reduced to 3.2% and 10.6% reidentification for Maine and Vermont, respectively.[18] The linkage of hospital data poses privacy risks because it allows previously unknown information within the hospitalisation record including other patient diagnoses to be known such as mental health, addiction or disabilities. Another key example may be an uncommon patient diagnosis code currently onward where a person other than any healthcare practitioner overseeing the patient's care could reference the patient by their diagnosis and then correctly identify the patient by only searching for the patient's International Classification of Diseases (ICD) diagnosis code. The latter example is a violation of protected health information (PHI) practices under the HITECH act and represents the most common cause for a HIPAA breach known as 'data snooping'. Rare disease ICD codes may also be considered quasi-identifiers when combining data with patient forums.[19]

In most cases of deidentified medical resources, a potential data consumer must request access to the database and complete ethical research conduct certifications. Although reidentification cannot be completely mitigated, it is worth considering the possibility of identifying a person's health information without deidentified research data at all. For example, if a person is active on a public patient disease forum and states their disease (ie, myasthenia gravis), general field of work (ie, accounting) and geotagged to their city (ie, Boston). Cross-referencing these data with public records and social media may be enough to reasonably infer information on the person

without deidentified research data at all. In instances such as the above example outlined, privacy may be entirely reliant on the deniable plausibility of being any single individual, known as k-anonymity privacy. This example highlights that with or without 'identifiable' health record data, individual's health data can be vulnerable to wide-scale reidentification using data shared directly by individuals 'consensual', shared via data brokers, or found in the 'public domain'.

Evidence-based risk quantification of reidentification

The two main components used to quantify risk are the probability and severity of the event. In order to approximate the quantification of the true risk of reidentification factors outside of the data itself need to be considered. The first is the motivation for reidentifying the research dataset. We argue the incentive is lower to breach data for data that is able to simply be requested. In this way, research data are often different from breaching commercial data for usage, such as financial fraud and identity theft. Before being granted access to a research dataset, the user requesting access typically must accept the institution's data use agreement (DUA). The DUA is linked to information about the user including identifiable information and specifies the intended purpose for the data and how it may not be used. DUAs most common term and condition is to make no attempt to learn the identity of any person or establishment within the data, and sanctions for violating the DUA is considered a felony with charges such as imprisonment or fines (the National Center for Health Statistics is imprisonment up to 5 years and US$250 000 fine).[20]

While reidentification of deidentified does pose a risk to patients, this risk is often systematically overestimated and confused with data exfiltration. In a systematic review of healthcare data reidentification, 14 studies were identified, and 2 studies had been deidentified using standards-based methods.[21] Interestingly, of the 14 reidentification studies, 11 were carried out by researchers, 2 were informed court judgements and 1 by a journalist supporting our claim that reidentifaction and data exfiltration are commonly conflated and confused. Within one of study standards-based methods commissioned by the US Department of Health and Human Services, it found that only 0.013% of the records could be reidentified, while the other study in the UK used survey data that could only be obtained under very strict confidentiality conditions to reidentify information (that would violate a DUA). Another publication that analysed motor vehicle accidents (MVA) specifically due to newspaper coverage found that even when targeting this specific patient population. The data analysed from the Buffalo, NY area found that by cross-referencing seven indirect identifiers 0.88% of the MVA patients were able to be reidentified compared with the 0.0017% of all database patients.[22] While this difference in patient populations represents a huge increase in relative risk of reidentification, it is worth noting that consideration of both having (1) stricter deidentification standards in more easily identifiable subpopulations such as MVA and rare disease (and further verified by statistical expertise where possible) and (2) how publicly available information is reported in outlets such as newspapers. One publication found that by knowing 15 demographic attributes, 99.98% of the population could be reidentified.[23] However, not all attributes were found to have the same level of uniqueness where attributes such as race, gender and citizenship did not give a considerable lift to the reidentification accuracy, additionally, highly unique, and therefore, identifying pieces of information such as the full date of birth and zip code were included in the analysis would not satisfy HIPAA Safe Harbor standards of deidentification. Finally, we acknowledge that reidentification efforts outside of research activity would be less likely to be published in the first place, particularly if the goal is for information gain to be used in an advantageous or illicit way.

Within an organisation, each person who works with data has responsibility to understand the data risks and mitigation through proper HIPAA training and data transfer processes. While the HHS has outlined the methods for deidentification standards as Safe Harbor or expert deidentification,[24] what constitutes 'very small risk' rightfully remains subjective. Such questions an organisation may need to ask are: Who is responsible for this risk assessment and mitigation? How is this risk evaluated? Do these individuals correctly estimate the risk associated with data breach? Do they consider the use of this data to increase the likelihood of data breach? Organisations will view these risks differently; however, by standardising the approach to each of these questions, a systematic approach can be repeatedly performed. Thus, allowing more accurate depiction of risk that can be more readily quantified with the intention of more frequent data sharing in the future. The NIH DMS mandates data sharing being conducted under their funding; the development of an organisational approach to risk monitoring is a necessary accompaniment that would build trust and prevent ODIAO.

TECHNICAL

Complexity and quality

Due to the rapid adoption and large-scale deployment of digital technology in our society, Big Data and related analytics have become ubiquitous for supporting decisions and operations. However, the volume, variety, velocity and veracity of new data bring new complexities and concerns on information quality. A typical example is the ongoing challenge of data sharing in cities, which are frequently used and combined within healthcare research. Thanks to the invention of low-cost sensors, cloud computing and personal digital devices, cities nowadays enjoy rich information resources to assist data-driven decision-making and automated operation. However, the digitisation of urban systems and internet society bring new social and technical complexities. New

information types and data formats create technical and social complexities for sharing data. One practical challenge is a lack of computing expertise for properly creating, managing, processing and exchanging data. A previous study investigating data landscape in US cities reveals a significant variety and disparity of data formats in city open data, particularly the structured, tabular data (eg, less than 40% of city data in Boston are in tabular format).[25]

Successful data diplomacy practice only starts from data sharing but completes with effective information integration and implementation. Even though multiple parties are willing to share data, a lack of standard in definition and classification may still cause data integration failures, preventing greater value creation. One example is the digital building permitting system in US cities. One recent study investigated building permit data in eight major US cities and found various terminology and classifications, although the data are publicly available and report similar information.[26] Such a lack of data standards brings difficulties in quality evaluation and integrated analytics across multiple cities. Beyond the technical barriers, additional non-technical concerns involve unexpected social impact. For example, several cities previously have published aggregated academic performance data by school districts without personal-identifiable information. However, such information-sharing caused concerns on creating discriminations and biases towards specific neighbourhoods, particularly on the local housing value estimation and property market appreciation. Such unwanted consequences to certain communities and population groups create additional complexities in data sharing and information publication.

Data management policy

To cut through the data complexity and quality issues in data sharing development of a data management policy and identifying the correct stakeholders is crucial. The goal of a data management policy is to deliver the right data to the right user at the right time with the lowest possible cost and friction. The data management policy outlines considerations such as what data standards are followed, where the data will be stored, what requirements there are to access the data, how data will be accessed, what the time frame of the data is from, how to join the data and schema information, and include data descriptions and data dictionaries. For example, the Medical Information Mart for Intensive Care (MIMIC-III) database is accessible through both Google Cloud Platform and Amazon Web Services, is accessed through PhysioNet, Collaborative Institutional Training Initiative (CITI) training is required, and date fields and filters are stated within the data itself. An important component of the data management plan to highlight is adoption of data standards. Data standards are documented agreements on representation, format, definition, structuring, tagging, transmission, manipulation, use and management of data.[27] By implementing data standards prior to collecting data where possible, the amount of data governance and structure can be reduced by avoiding remapping and standardising data at a later time. Currently, the data cleaning stage of a project takes the most amount of time, by implementing data standards code, presentations, publications and information quality can be reproduced and validated in less time than without data standards. A notable data standardisation is the Fast Healthcare Interoperability Resources, which is a standard for healthcare data exchange that addresses many aspects of health from diagnostics and medications to claims and genomics.[28]

Creating a village mindset

There are several key stakeholders that must work together with a 'village mindset' in order to make data sharing possible.[29] Here, we outline specific roles, but a single person may represent multiple skill sets and contribute to the data diplomacy ecosystem. Generally for an institution to make a data transfer, there will be approval and strategy, legal, technical and considerations. Our aim is to help organisations accurately identify gaps in people and skills that, if bridged, will facilitate more standardised and swifter data sharing. A data engineer is able to extract the data from the source system, create data quality metrics, filter and aggregate the data, and set up the means in which it will be transferred. For small data sizes, the transfers could be set up through simple cloud storage sharing (ie, box). Larger datasets may require cloud computing (S3, Redshift, Blob, etc) and a Secure File Transfer Protocol or managed roles to control access and port over data. The chief data officer (CDO) is a senior executive responsible for the stewardship, utilisation and governance of an organisation's data. Typically, the CDO's approval is required to sign off on entering into a data sharing agreement. An organisation may also have a chief information officer (CIO) instead as the reviewer. The chief privacy officer (CPO) is responsible for developing, implementing and maintaining policies designed to protect employee and patient data from unauthorised access. Such policies could involve technical access controls to only certain internal personnel or could be non-technical such as running HIPAA and PHI training at regular cadences. General and legal counsel will be involved in the approval and may be responsible for running training to explain the nuances in HIPAA policies such as explaining risk differences between covered entities and business associates. The CDO or CIO work alongside the CPO to create a data sharing agreement outlining the possible risks and synergies from the agreement. Once the format and transfer means are agreed on, the data engineer is able to execute on creating the correct dataset and setting up transfer ports. If the data sharing is maintained through a data sharing platform, there will be additional technical personnel involved, such as cybersecurity and site reliability engineers that are not elaborated on in the scope of this work.

While technical advances continue to be made, the complexity of the data being used and the types of agreements being made continue to grow. Clear standards for data sharing must be provided by governing bodies but must also be set locally as well for internal processes. Organisations must involve a wide range of disciplines and backgrounds in this process to maximise the chance of data usage and prevent data siloing that can lead to ODIAO.

REPUTATIONAL

Researcher, physician and institutional status

A personal barrier to withholding data can be found in the lack of willingness for errors to be found. What is a completely natural response, however, merely delays the time at which the mistake is uncovered. As failure to replicate results sparks investigation. This is both a waste of time and resources as well as potentially putting patient lives at stake. So, although it may appear that refusal to share data avoids the risk of academic or commercial scrutiny. Refusal to share data does not ultimately protect reputation; it masks issues and impedes discovery, innovation and discourse over time. Retraction Watch, part of the Center for Scientific Integrity, has reported a significant rise in the number of retractions each month,[30] particularly since the COVID-19 pandemic.[31] Echoing the number of high-profile cases of fraudulent research.[32] The distribution of data and code would normalise corrections, improve patient safety and reduce duplication of work that attempts to replicate results.

Duplication of research also carries another risk, data breaches. Data breaches are infrequent but can be significant, affecting a large number of patients. Although, in an open data environment, more data will be public, similar volumes of research will still be conducted. By increasing access to standardised and secure data environments, a higher proportion of research would be hypothetically performed in a regulated and secure setting. This relies on data sharing being appropriately regulated to shift the burden of risk from the researchers to the governing organisation.

Complying with data availability statement and regulation

The purpose of data availability statements (DAS) is to provide information regarding where the data supporting the findings in a published article can be found, and if and how they can be accessed. These policies are part of a broader movement to encourage open science and data sharing. Depending on the types of data involved, however, there can be tension between the data sharing promoted through DAS and privacy regulations. Qualitative and mixed-methods research, for example, may contain data that is difficult to sufficiently anonymise in order to prevent deductive disclosure.[33] Recent studies have found, though, that many researchers do not comply with what they set out in their DAS, and even that there was not a difference in compliance rates for articles that have a DAS compared with those that do not.[34 35] Notably, the study found that 80% of corresponding authors did not reply to the contacts for a data request, and of the 20% that did respond, only 50% shared the data. Overall, this yielded a 93% non-response rate or decline to share data.

The General Data Privacy Rule (GDPR) enacted by the European Union (EU) gave stronger privacy protections to individuals by requiring stronger consent and providing new rights to be forgotten and for data portability. While there were initial concerns that the GDPR would impede scientific data sharing, the final version included exemptions that supported data sharing for scientific research.[36] With more complex collaborative arrangements for scientific data sharing, though, there can be a need to establish clearer roles in the data sharing networks under the GDPR.[37] A 2021 report found that the GDPR was having a negative impact on oncology and other types of health research, in part because it hampers the sharing of data outside of the EU, thus making it more difficult to share data as part of international collaborative health research.[38 39] Therefore, both correct interpretation of GDPR and identification of stakeholder responsibilities is necessary.

The new NIH DMS Policy requirement will combine the expectations of proper data management and sharing by formalising the plan as part of its application process. This includes considerations for: describing the data types; related tools, software and/or code; data standards; data preservation, access and associated timelines; and access, distribution or reuse considerations.[15]

COMMERCIAL

Monetisation and proprietarisation

In the last decade, data based startups, academic spin-outs turned companies, and patents on data processing have become more common[40] and data have been referred to as the new oil, by Clive Humby as early as 2006, in the digital and information age. Data sharing can be seen as a risk to both monetisation and proprietarisation if the data asset is core to the research or product. We argue that although data in itself may have some inherent value, it is a building block to higher value insights requiring context to become information, meaning to become knowledge and insight to become wisdom. Each of these stages to transform data into solutions to real-world problems and helping patients requires personnel with specialised technical and subject matter expertise.

For others, the prospect of making institutional data accessible to those outside of the organisation will allow others to benefit from the data, and this may be viewed as the loss of an asset without compensation. This is compared other groups that charge researchers, institutions and industry licensing fees for data access. In the same study analysing DAS, some corresponding authors proposed or expected coauthorship for use of the data, representing an expectation of proprietarisation on

secondary analysis.[35] These types of expectations in the research system make it difficult for analysis to refute original claims or propose divergent hypotheses.

Better data valuation frameworks

Current data valuation approaches for institutions and organisations are ambiguous and vague at best and non-existent at its worst. The idea that the value of data solely resides in what another party would be willing to pay is reductionistic and typically represents only a small fraction of the data's value. Data value would be better valued by its ability for the data to optimise an operation or act in support of a larger product or process.[41] For example, a hospital may want to optimise hospital bed capacity and use parameters such as transfers, unscheduled admissions and unoccupied beds to derive an optimisation model.[42] Making these data used to create the optimisation model available on request through a DAS does not automatically mean that the data will be insightful, generalisable, or actionable to other hospitals for their gain. Finally, by making data available through a DAS, it does not lower significant barriers such as highly specialised personnel, team size, legal assistance and cloud compute costs that usually make data monetisation and proprietarisation possible.

When valuing a data asset, instead of assigning an absolute nebulous worth to the data, it is best to contextualise the data asset in terms of its utility for the problem trying to be solved.[43] Factors to include in data valuation may consist of the data's: strategy, features, size, granularity, quality, standards and processes to create a more meaningful understanding of data utility.

Organisations and researchers must find a middle ground where they are rewarded for efforts in dataset collection, curation and storage yet still maximise access to data that has the potential to improve patient outcomes. The maturation of DAS' and guidelines such as the NIH DMS will help to safeguard the inevitable competition of monetisation through scarcity and beneficial impacts of data sharing.

PSYCHOLOGICAL

AI arrogance and ignorance

The current system means that the risk for sharing one's data is high, with little personal gain. Despite the fact these risks are real to the institution, the failure to disclose data does not eliminate the risk; it merely transfers the risk from the institution to the patients being treated based on the research. Thus, those who we claim to be helping must carry the risk for our own arrogance and ignorance, which may be worse than fatal, where one's data may worsen the outcomes of another human being who 'does not look like you'. This problem can be further exacerbated by reasoning that Artificial Intelligence (AI) methods such as synthetic minority oversampling technique (SMOTE) will simply 'fix' issues such as sex and race data imbalances. AI has introduced new effective methods, such as SMOTE that can forward medical and social issues, but is not a 'cure all' and is instead a specific methodological tool.[44] Current popular interpretation methods such as local interpretable model-agnostic explanations and SHapley Additive exPlanations have respective limitations such as model reduction to an alternative localised linear or probability values for covariates that are in reality collinear.[45] These limitations are not a sole reason to discard them, but be thoughtfully instead of blindly executed. Methods that intersect AI and causal frameworks that perform counterfactual scenarios about outcomes based on attributes should not be implemented indiscriminately on features conditional on each other.[46] For example, if you wanted to understand a survivor expectancy of a male patient if instead they were female, other attributes such as occupation, income level, age and race would need to be considered holistically.

Historically, tools and software used for research are specified in publications, but code sharing is newer and less frequently incorporated as part of the publication or supplement. As AI and coding are linked, so is AI arrogance and lack of code sharing and transparency. Much like the DAS, code is available on request. While the true availability of the data outlined in DAS statements has begun to be researched, code sharing is not specifically well researched and is likely more researched in specifically computational journals.[47] While tools and software may by nature use Graphical User Interfaces (GUIs) that cannot be automatically reproduced by being run, coding scripts are. While code sharing is possible through Git and providers such GitHub and GitLab there are legal, technical and reputational risks associated with sharing source code. These can span from how deidentification is conducted to critiques ways the code is more methodically robust, scalable or elegant (few lines of code). By turning the research focus back to patient centricity, the risks posed by code sharing are smaller compared with the issues of non-reproducibility and model improvements.

Continuous improvement process and validation

A discontinuous and stochastic approach dominates current quality improvement, however, a mindset shift towards a data-centric and systems-based methodology should be leveraged in the future. In order to make data sharing a more frequent reality that acts in service of the patient, incremental change at the organisation, researcher and data set level are required. A continuous improvement process for data sharing means iterating on the parts of the process that cause failure. It is distinct from the data management plan; while a data management plan is created before or during data sharing and primarily completed once the data is shared, a continuous improvement process is cyclical. While a continuous improvement process has technical aspects, it is driven by considerations of an organisation to serve both the patient and research community.[48]

Typically, the data sharing process begins with how a data sharing inquiry is received and to whom, the

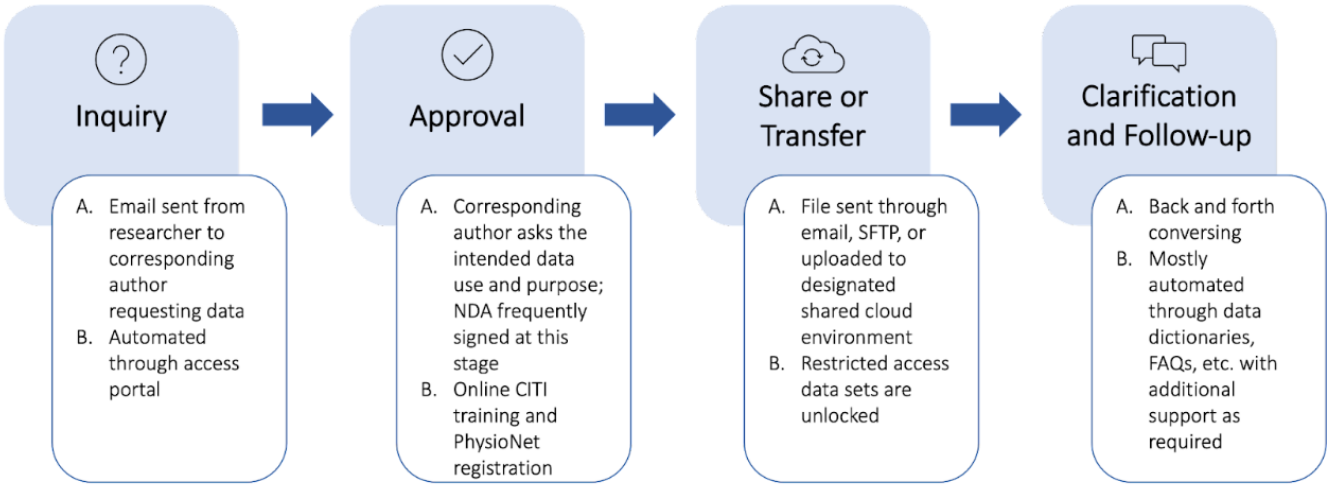

Figure 2 Data sharing process with manual and automated scenarios, A and B, respectively. Non-Disclosure Agreement, NDA; CITI, Collaborative Institutional Training Initiative; SFTP, Secure File Transfer Protocol; FAQs, Frequently Asked Questions.

approval process, the data transfer and/or sharing, and clarification and follow-up support. The goal of a continuous improvement process for data sharing means first designing with data in mind and iterating on the pain points for greater data dissemination.[49] Figure 2 illustrates two possible process flows for a four-step data sharing process, A and B. Scenario A represents what the data sharing process looks like without any online data repository or portal and scenario B represents where a repository or portal solution has been implemented. From the inquiry to clarification and follow-up scenario A has many more communication and back and forth touchpoints between the corresponding author and the researcher making the request. Scenario B outlines the type of data sharing process that is possible when a continuous improvement process is implemented with the patient and research community in mind.

From these two scenarios, we can glean that through a continuous improvement process there are opportunities to potentially automate and reduce the time and effort exerted to share data (figure 3). A continuous improvement process laid out by an institution may consist of multiple aims such as to use a trusted research database portal, begin adopting data standards of the field prior to data collection of an experiment, and incorporate a deidentification requirement for project completion with the intent of data sharing. By placing data sharing as a goal to be met in service of the patient and research community, it is less likely to be considered and after thought or extra work with low incentivisation for the researcher. A continuous improvement process is not seen as complete, as new needs arise whether making the data sizes more accessible or creating documentation for frequently asked questions about the data set, the process is aimed to give the best possible experience in sharing and understanding the data. An exemplar system that lifts the onus of data sharing from the researcher entirely is MIMIC. The MIMIC data are accessible via PhysioNet, where data sets are categorised as open, restricted or credentialed. For restricted data sets, including the latest version of MIMIC, CITI training must be completed, user information and completing the DUA are required. Additionally, data dictionaries, release notes specifying incorrect data and subsequent corrections, and directions for how to join commonly created data views are documented for MIMIC.[50]

Data sharing currently emphasises the ability to garner better scientific reproducibility, but validation is equally if not more important. From a treatment perspective, it is imperative to prove clinical efficacy such as AI enabled treatment recommendations created from longitudinal analysis of demographic, symptom and vital sign data. By putting the patient first, AI then refocuses itself as a tool, where clinical safety and efficacy supersedes importance of AI interpretability and explainability.[51] Where AI is used to lead to treatment enhancement indirectly in medical imaging analysis or organisation of unstructured EHR data, validation of the accuracy of the method, the degree of utility and ability to generalise is where patients can benefit. To mitigate AI research, arrogance and ignorance, goals need to be oriented so there is a direct relationship from the patient providing their data to improvements in health outcomes.

The recent NIH initiative forces the sharing of such data and, thus, we hope, a change in mindset that promotes humility and transparency. The development of continuous and systematic approaches to quality improvement are a beneficiary of such a mindset. Further, it shares the same psychological sentiment that drives data sharing and would discourage ODIAO.

CONCLUSION

There is a growing acknowledgement that data sharing is likely in patients' best interest; however, we identified five key barriers that can oppose data sharing and

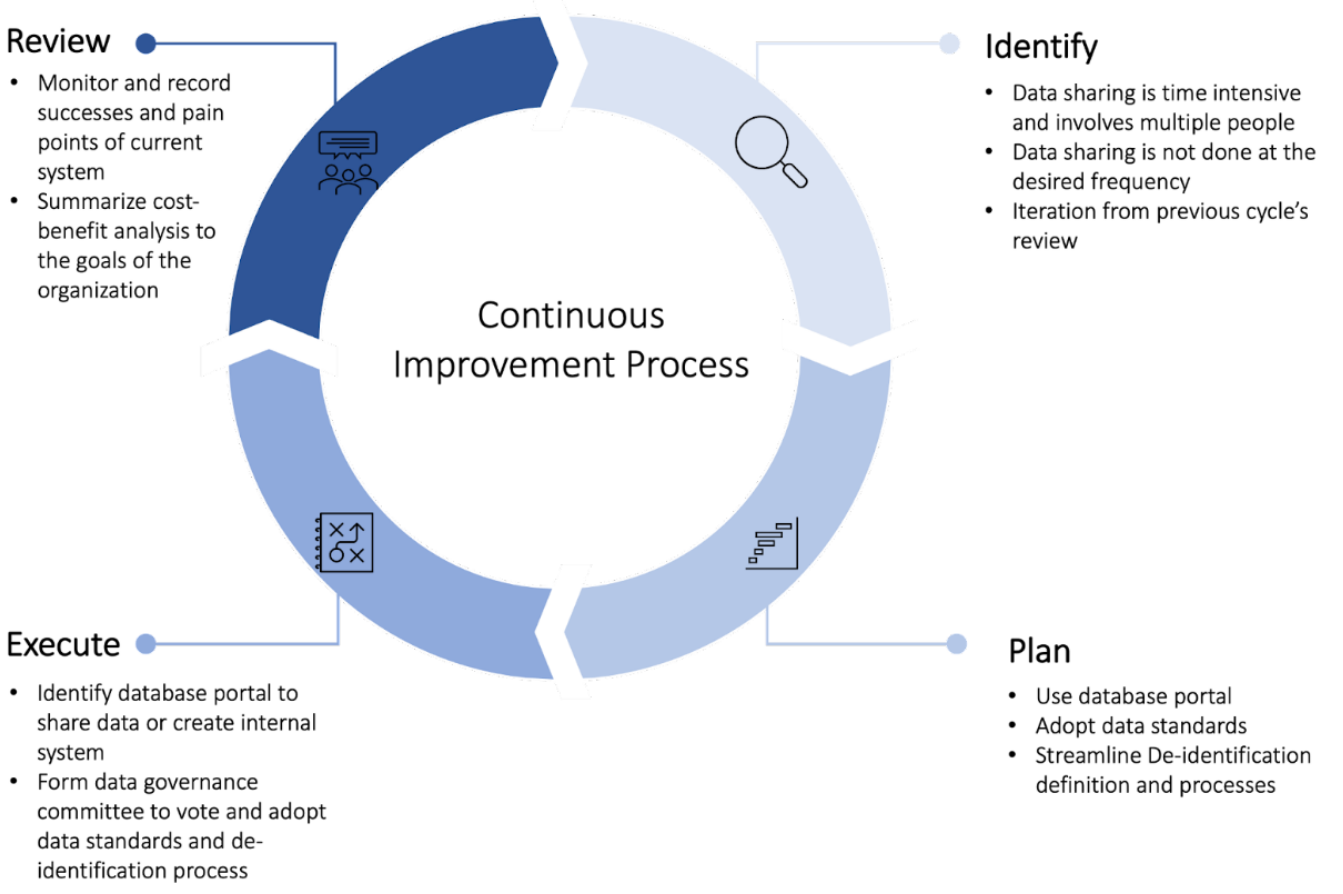

Figure 3 Continuous improvement process for more ubiquitous data sharing.

lead to ODIAO. A mindset shift is required to prioritise patient-centred research in a system where data are a valuable asset and mitigate real patient privacy risks that need to be quantified. In order to realise the benefits of data sharing while navigating such risks, the NIH 2023[3] mandate must be actively supported by a village mindset that cultivates the talents of all stakeholders. The postpandemic world needs data sharing to become a cornerstone of health research, to safeguard against the implementation of harmful treatments and algorithms. Moreover, to encourage public data sharing, there must be incentives driven from the bottom up starting with the patients themselves. The NIH DMS is a valuable start to this and strongly opposes ODIAO.

Author affiliations
[1]DBT Labs, Boston, Massachusetts, USA
[2]Department of Critical Care, Guy's and St Thomas' Hospitals NHS Trust, London, UK
[3]Department of Urban Planning and Design, Tsinghua University, Beijing, China
[4]New York University Marron Institute of Urban Management, New York, New York, USA
[5]University of Texas Southwestern Medical Center, Dallas, Texas, USA
[6]Southern Methodist University, Dallas, Texas, USA
[7]Heidelberg Institute of Global Health, Heidelberg University, Heidelberg, Germany
[8]Department of Medicine, Instituto Politécnico Nacional, Ciudad de Mexico, Mexico
[9]Center for Biomedical Ethics, Stanford University School of Medicine, Stanford, California, USA
[10]Department of Radiology, Emory University, Atlanta, Georgia, USA
[11]UNICEF, New York, New York, USA
[12]Laboratory for Computational Physiology, Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts, USA
[13]Division of Pulmonary Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

Twitter Judy Gichoya @judywawira, Uyen Kim Huynh @ukhuynh and Leo Anthony Celi @MITCriticalData

Contributors HW: conceptualisation, methodology, writing—original draft preparation, writing—reviewing and editing. JG: conceptualisation, methodology, writing—original draft preparation, writing—reviewing and editing. YL: conceptualisation, methodology, writing—original draft preparation, writing—reviewing and editing. APR: conceptualisation, methodology, writing—original draft preparation, writing—reviewing and editing. CV: methodology, writing—reviewing and editing. NM: methodology, writing—reviewing and editing. JG: methodology, writing—reviewing and editing. UKH: methodology, writing—reviewing and editing. LAC: conceptualisation, methodology, writing—reviewing and editing, supervision.

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 applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

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

ORCID iDs
Jack Gallifant http://orcid.org/0000-0003-1306-2334
Judy Gichoya http://orcid.org/0000-0002-1097-316X
Leo Anthony Celi http://orcid.org/0000-0001-6712-6626

REFERENCES

1 Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al*. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 2016;3:160018:160018.:.
2 Liu J, Carlson J, Pasek J, *et al*. Promoting and enabling reproducible data science through a reproducibility challenge. *Harvard Data Science Review* 2022.
3 NOT-OD-21-013: final NIH policy for data management and sharing. Available: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html [Accessed 13 Feb 2023].
4 Parkes DC. Building a more robust data science, toward a more robust world. *Harvard Data Science Review* 2022.
5 Martone M, Nakamura R. Changing the culture on data management and sharing: getting ready for the new NIH data sharing policy. *Harvard Data Science Review* 2022.
6 Tabak L, Jorgenson L, Martone M, *et al*. Conversation with Dr. Lawrence Tabak and Dr. Lyric Jorgenson on the NIH perspective on data sharing and management. *Harvard Data Science Review* 2022.
7 Lowenberg D. Recognizing our collective responsibility in the Prioritization of open data Metrics. *Harvard Data Science Review* 2022.
8 Smith PC, Mossialos E, Papanicolas I, *et al*. *Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects*. Cambridge University Press, 2010.
9 Watson C. Rise of the preprint: how rapid data sharing during COVID-19 has changed science forever. *Nat Med* 2022;28:2–5.
10 Dron L, Kalatharan V, Gupta A, *et al*. Data capture and sharing in the COVID-19 pandemic: a cause for concern. *Lancet Digit Health* 2022;4:e748–56.
11 Bjaalie JG, Goble C, Sansone S-A, *et al*. Perspectives on data sharing and the new NIH policy from the European Union. *Harvard Data Science Review* 2022.
12 Gallifant J, Zhang J, Del Pilar Arias Lopez M, *et al*. Artificial intelligence for mechanical ventilation: systematic review of design, reporting standards, and bias. *Br J Anaesth* 2022;128:343–51.
13 Watson C. Many researchers say they'll share data - but don't. *Nature* 2022;606:853.
14 ODI strategy 2023–2028. Available: https://www.theodi.org/about-the-odi/odi-strategy-2023-2028/ [Accessed 03 May 2023].
15 Writing a data management & sharing plan [Data Sharing]. 2022. Available: https://sharing.nih.gov/data-management-and-sharing-policy/planning-and-budgeting-DMS/writing-a-data-management-and-sharing-plan [Accessed 08 Aug 2022].
16 Rights (OCR) O for C. HITECH act enforcement interim final rule. HHS.gov; 2009. Available: https://www.hhs.gov/hipaa/for-professionals/special-topics/hitech-act-enforcement-interim-final-rule/index.html [Accessed 09 Aug 2022].
17 Office for Civil Rights. HIPAA violation cases - updated [HIPAA J]. 2022. Available: https://www.hipaajournal.com/hipaa-violation-cases/ [Accessed 09 Aug 2022].
18 Yoo JS, Ra Thaler A, Thaler L, *et al*. Risks to patient privacy: a re-identification of patients in Maine and Vermont statewide hospital data [Technol Sci]. 2018. Available: https://techscience.org/a/2018100901/ [Accessed 08 Aug 2022].
19 Gow J, Moffatt C, Blackport J. Participation in patient support forums may put rare disease patient data at risk of re-identification. *Orphanet J Rare Dis* 2020;15:226.
20 Data access - data user agreement. 2020. Available: https://www.cdc.gov/nchs/data_access/restrictions.htm [Accessed 10 Aug 2022].
21 El Emam K, Jonker E, Arbuckle L, *et al*. A systematic review of re-identification attacks on health data. *PLOS ONE* 2011;6:e28071.
22 Janmey V, Elkin PL. Re-identification risk in HIPAA de-identified datasets: the MVA attack. *AMIA Annu Symp Proc* 2018;2018:1329–37.
23 Rocher L, Hendrickx JM, de Montjoye Y-A. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun* 2019;10:3069.
24 Rights (OCR) O for C. Guidance regarding methods for de-identification of protected health information in accordance with the health insurance portability and accountability act (HIPAA) privacy rule [HHS.gov]. 2012. Available: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html [Accessed 11 Aug 2022].
25 Barbosa L, Pham K, Silva C, *et al*. Structured open urban data: understanding the landscape. *Big Data* 2014;2:144–54.
26 Lai Y, Kontokosta CE. Topic modeling to discover the thematic structure and spatial-temporal patterns of building renovation and adaptive reuse in cities. *Computers, Environment and Urban Systems* 2019;78:101383.
27 US EPA O. Learn about data standards. 2015. Available: https://www.epa.gov/data-standards/learn-about-data-standards [Accessed 12 Aug 2022].
28 Index - FHIR V4.3.0. Available: https://www.hl7.org/fhir/index.html [Accessed 11 Aug 2022].
29 Borgman CL, Bourne PE. Why it takes a village to manage and share data. *Harvard Data Science Review* 2022;4.
30 Oransky I. Retractions are increasing, but not enough. *Nature* 2022;608:9.
31 Moradi S, Abdi S. Pandemic publication: correction and Erratum in COVID-19 publications. *Scientometrics* 2021;126:1849–57.
32 Piller C. Blots on a field? [Science]. 2022. Available: https://www.science.org/content/article/potential-fabrication-research-images-threatens-key-theory-alzheimers-disease
33 Tsai AC, Kohrt BA, Matthews LT, *et al*. Promises and pitfalls of data sharing in qualitative research. *Soc Sci Med* 2016;169:191–8.
34 Federer LM, Belter CW, Joubert DJ, *et al*. Data sharing in PLOS ONE: an analysis of data availability statements. *PLOS ONE* 2018;13:e0194768.
35 Gabelica M, Bojčić R, Puljak L. Many researchers were not Compliant with their published data sharing statement: a mixed-methods study. *J Clin Epidemiol* 2022;150:33–41.
36 Meszaros J. The conflict between privacy and scientific research in the GDPR. 2018 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC); 2018:1–6
37 Becker R, Thorogood A, Bovenberg J, *et al*. Applying GDPR roles and responsibilities to scientific data sharing. *International Data Privacy Law* 2022;12:207–19.
38 ALLEA (European Federation of Academies of Sciences and Humanities), FEAM (Federation of European Academies of Medicine), EASAC (European Academies' Science Advisory Council). International sharing of personal health data for research. DE: ALLEA; 2021.
39 Gourd E. GDPR Obstructs cancer research data sharing. *Lancet Oncol* 2021;22:592.
40 Teare G. How seed funding has exploded in the past 10 years [Crunchbase News]. 2021. Available: https://news.crunchbase.com/venture/seed-funding-startups-top-vc-firms-a16z-nea-khosla/ [Accessed 12 Aug 2022].
41 Tufiş M, Boratto L. Toward a complete data valuation process. challenges of personal data. *J Data and Information Quality* 2021;13:1–7.
42 Ravaghi H, Alidoost S, Mannion R, *et al*. Models and methods for determining the optimal number of beds in hospitals and regions: a systematic scoping review. *BMC Health Serv Res* 2020;20:186.
43 Infonomics - Monetize, manage & measure information [Gartner]. Available: https://www.gartner.com/en/publications/infonomics [Accessed 12 Aug 2022].
44 Takaoka K. AI implementation science for social issues: pitfalls and tips. *J Epidemiol* 2022;32:155–62.
45 Molnar C. Chapter 9 local model-agnostic methods [Interpretable Machine Learning]. Available: https://christophm.github.io/interpretable-ml-book/local-methods.html [Accessed 12 Aug 2022].
46 Lundberg I, Johnson R, Stewart BM. What is your Estimand? Defining the target quantity connects statistical evidence to theory. *Am Sociol Rev* 2021;86:532–65.
47 Cadwallader L, Papin JA, Mac Gabhann F, *et al*. Collaborating with our community to increase code sharing. *PLOS Comput Biol* 2021;17:e1008867.
48 Fecher B, Friesike S, Hebing M. What drives academic data sharing *PLOS ONE* 2015;10:e0118053.
49 First, design for data sharing [Nature Biotechnology]. Available: https://www.nature.com/articles/nbt.3516 [Accessed 12 Aug 2022].
50 Johnson A, Pollard T, Mark R. MIMIC-III clinical database. *PhysioNet* 2015.
51 Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health* 2021;3:e745–50.

BMJ Health & Care Informatics

Twenty-year follow-up of promising clinical studies reported in highly circulated newspapers: a meta-epidemiological study

Aran Tajika ,[1] Yasushi Tsujimoto,[1,2] Akira Onishi,[3] Yusuke Tsutsumi ,[4,5] Satoshi Funada ,[1] Yusuke Ogawa,[6] Nozomi Takeshima,[1] Yu Hayasaka,[1] Naotsugu Iwakami,[1] Toshi A Furukawa [1]

To cite: Tajika A, Tsujimoto Y, Onishi A, *et al*. Twenty-year follow-up of promising clinical studies reported in highly circulated newspapers: a meta-epidemiological study. *BMJ Health Care Inform* 2023;**30**:e100768. doi:10.1136/bmjhci-2023-100768

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2023-100768).

Received 23 March 2023
Accepted 29 May 2023

For numbered affiliations see end of article.

Correspondence to
Dr Aran Tajika;
aran.tajika28@gmail.com

ABSTRACT

Objectives Researchers have identified cases in which newspaper stories have exaggerated the results of medical studies reported in original articles. Moreover, the exaggeration sometimes begins with journal articles. We examined what proportion of the studies quoted in newspaper stories were confirmed.

Methods We identified newspaper stories from 2000 that mentioned the effectiveness of certain treatments or preventions based on original studies from 40 main medical journals. We searched for subsequent studies until June 2022 with the same topic and stronger research design than each original study. The results of the original studies were verified by comparison with those of subsequent studies.

Results We identified 164 original articles from 1298 newspaper stories and randomly selected 100 of them. Four studies were not found to be effective in terms of the primary outcome, and 18 had no subsequent studies. Of the remaining studies, the proportion of confirmed studies was 68.6% (95% CI 58.1% to 77.5%). Among the 59 confirmed studies, 13 of 16 studies were considered to have been replicated in terms of effect size. However, the results of the remaining 43 studies were not comparable.

Discussion In the dichotomous judgement of effectiveness, about two-thirds of the results were nominally confirmed by subsequent studies. However, for most confirmed results, it was impossible to determine whether the effect sizes were stable.

Conclusions Newspaper readers should be aware that some claims made by high-quality newspapers based on high-profile journal articles may be overturned by subsequent studies within the next 20 years.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ When newspapers cite the results of clinical research articles, they sometimes misrepresent the results based on exaggerated expectations.
⇒ Studies with higher levels of evidence may overturn the results of clinical research.

WHAT THIS STUDY ADDS

⇒ The results of clinical research articles were relatively stable in papers in which the citation source was properly listed in the newspaper article.
⇒ However, the results of approximately one-third of the papers were overturned in the following two decades.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Journalists should be careful in accurately reporting clinical research articles and stating the sources of their citations.
⇒ Readers should be aware that more than a few claims made in highly circulated newspapers based on high-profile journal articles may still be overturned by subsequent studies.

INTRODUCTION

As people's health awareness has increased, newspapers have covered more stories about health and medicine. These stories feature many diseases, including cancer, stroke, infectious diseases and mental disorders. Some sensationalise the fear and frustration of the disease, while others provide hope for new treatments or preventative measures. These stories are often based on articles published in medical journals. The important points of these articles are summarised and presented clearly in newspaper stories for the general public.

However, the media coverage often exaggerates fear and hopes.[1] For example, a phase I uncontrolled study of a new cancer drug published in the *New England Journal of Medicine* showed some effects in one subgroup. Newspapers reported that this treatment produced highly promising results.[1] However, studies cited in newspaper stories are sometimes overturned. Gonon[2] investigated the 'top 10' most frequently reported studies on attention deficit hyperactivity disorder (by newspapers) and compared these results with

those of subsequent studies. Two studies were confirmed, four attenuated, three refuted and one was neither confirmed nor refuted.

When the strength of the research design is considered, randomised controlled trials (RCTs) and their meta-analyses provide the strongest evidence for treatment decisions. However, newspapers are more likely to report observational studies (OSs) than RCTs.[3] Notably, exaggeration often begins with medical journal articles themselves.[1] One problem with studies with weak evidence is that the reproducibility of the results is low. Ioannidis conducted a simulation study and noted that a meta-analysis of good-quality RCTs and adequately powered RCTs assumed a reproducibility of 85%, but only 23% for underpowered RCTs and approximately 20% for adequately powered OSs.[4] Ioannidis[5] identified studies cited more than 1000 times in high-impact factor (IF) journals in general and internal medicine. When these studies were compared with subsequent studies that theoretically had better-controlled designs, only half of the RCTs and none of the OSs were replicated. Furthermore, when statistically significant and extremely favourable initial reports of intervention effects were examined, it was found that the majority of such large treatment effects emerged from small studies. When additional trials were performed, the effect sizes typically became much smaller.[6] When newspapers report and overestimate the results of these initially promising studies, the information that reaches the public may be doubly overstated.

This study investigated the trustworthiness of medical news. We examined whether newspaper reports were confirmed through subsequent studies that examined the same clinical questions. In other words, we examined how much caution general readers need to exercise when reading newspaper reports on medical research.

METHODS

Selection of newspaper stories and original studies

We selected four quality papers (two from the USA and two from the UK) and four non-quality papers (two from the USA and two from the UK) with the highest circulation according to the Audit Bureau of Circulations[7] and Alliance for Audited Media.[8] We examined these two newspaper types for several reasons. Generally, quality papers are believed to have higher quality reporting than non-quality papers,[9] which tend to focus on readers' emotions rather than on the veracity of the reports.[10] However, when we consider the respective circulations of the two types of papers, non-quality papers have as many readers as quality papers; they sometimes have more power to lead public opinion.[11]

We selected newspaper articles that quoted main medical journals. First, we selected medical journals from the following two fields: 'general and internal medicine' and 'public, environmental and occupational health' according to their journal IF on Journal Citation Reports. In addition, we selected the 20 journals in each field with the highest IFs for 2000. We ultimately selected 40 medical journals as an ad hoc set of representative medical journals that might meet the public interest. Next, we searched the LexisNexis database,[12] which contains stories from prominent newspapers worldwide. We used the names of 40 medical journals as search words and selected newspaper stories:

- Printed in 2000 in the four above-mentioned quality and four non-quality newspapers.
- That quoted articles that were published in the above-mentioned 40 journals.
- In which we could identify the original medical journal article.
- That mentioned the effectiveness, recommendation of treatment or prevention at that time.

Pairs of independent investigators (AT, YO, NT, YH and NI) selected eligible newspaper articles for analysis. Disagreements were resolved through discussions between the two investigators and, when necessary, in consultation with a third author (TAF). We found the original articles quoted in these newspapers. When two or more articles were quoted in a newspaper story, we selected all the articles. When the number of eligible studies was greater than 100, 100 studies were randomly selected. Original articles were classified into the following categories:

- Animal or laboratory study.
- Clinical study.
  - Case reports or case series.
  - OS.
  - RCT.
  - Systematic review (SR) of OSs with or without meta-analysis.
  - SR of RCTs with or without meta-analysis.
  - Other reviews (eg, narrative reviews).
- Others (eg, comment, letter).

We excluded studies in which specific clinical questions were not identifiable (eg, health economics studies) because we could not search for corresponding subsequent studies in the next step.

Selection of subsequent studies on the same clinical questions

For each original article, we searched for subsequent studies that examined the same clinical questions using 'stronger' research designs. The evidence levels of all the studies were classified according to the following hierarchy:

1. SR of RCTs.
2. Single RCT.
3. SR of OSs/single OS.
4. Case series/a case study.

The characteristics of 'stronger design' are as follows[5 13]:

- The subsequent study used a design with a higher level of evidence hierarchy than the original study.
- If studies had the same level of evidence hierarchy, a study with a larger sample size constituted stronger evidence.

- If the design of the original study was an SR of an RCTs, we searched for the latest SR for the RCTs.
- If the design of the original study was the SR of OSs or other reviews, we searched for the largest RCT or the latest meta-analysis of RCTs. If we could not find these studies, we searched for the latest OS meta-analyses.
- If the original study was an animal or laboratory study, we searched for the most appropriate clinical study asking the same clinical question according to the evidence hierarchy.

First, two authors (AT, YaT, AO, YuT and SF) independently searched the Web of Science for potential new papers in which the original paper was cited through December 2021. Subsequently, to prevent search omissions, AT conducted a PubMed search through June 2022 to search for anything more valid than the candidates' new articles on the Web of Science. If new candidate papers were found, the authors discussed them in pairs to identify the new papers. The PubMed search was conducted using the most comprehensive terms possible, and the search formula was documented.

Comparisons of original and subsequent studies

We extracted the data when the original study authors presented their primary outcomes. If the authors failed to designate their primary outcome(s), the outcome described first was considered the primary outcome. Next, we extracted the outcomes of the subsequent studies, which were as similar as possible to those of the original studies.

We conducted the following two-step comparison. First, we compared the effectiveness of the original studies with that of newer studies and classified each comparison into one of three categories: 'unchallenged', 'contradicted' or 'confirmed'.[5 13]

- Unchallenged: when there was no subsequent study with a higher level of evidence.
- Contradicted: when a subsequent study denied the effectiveness of the original study.
- Confirmed: The original and subsequent studies concluded that the intervention was effective, regardless of the effect size difference.

When we could not compare these outcomes, we compared the benefits and applicability of both studies and made qualitative judgements.

Furthermore, among 'confirmed' cases, when the outcomes of both original and subsequent studies were exactly comparable (ie, when a new paper was a meta-analysis, the original paper was included in the funnel plot of the new paper, and accurate effect size comparison was possible), we compared the effect sizes of both studies. Outcomes were extracted as continuous or dichotomous data such as standardised mean difference (SMD), OR, risk ratio (RR) or HR. We gave preference to continuous data. We compared these values when the SMD was shown in the subsequent meta-analysis, and when the SMD of the original paper was shown in that study. When studies showed effectiveness using only dichotomous data, the OR was calculated first. We then converted OR into SMD using the following formula[14]:

$$\mathrm{SMD} = \frac{\sqrt{3}}{\pi} \ln \mathrm{OR}$$

We classified 'confirmed' cases into one of two categories: 'initially stronger effects' or 'replicated'.[13]

- Initially stronger effects: when the point estimate of the original study was not included in the 95% CI of the SMD of the subsequent study or the SMD of the original study was 0.2 SD units or greater than that of the subsequent study (0.2 SD units would signify a small effect difference according to Cohen's rule of thumb).[15]
- Replicated: when the point estimate of the original study was included in the 95% CI of the SMD of the subsequent study, and the two SMDs were within 0.2 SD units apart, or the effect size of the subsequent study was larger than that of the original study.

When the SMD could not be calculated from the RR or the study showed only the HR, as it could not be converted into SMD, we directly compared only the RRs or HRs. Their 95% CI was presented in the papers without considering the difference of 0.2 SD units of SMDs.

Outcomes

Primary outcome

We defined the primary outcome, 'the proportion of confirmed studies', as follows:

$$\text{Proportion of confirmed studies} = \frac{\text{Confirmed studies}}{\text{Total studies} - \text{Unchallenged studies}} \times 100\ (\%)$$

Secondary outcomes

We classified the original studies according to their research design and medical fields and examined the differences between quality and non-quality papers. The proportion of confirmed studies in each subgroup was calculated.

Analyses

Statistical analyses were performed using STATA V.17.0. Statistical differences among subgroup categories were tested using the χ^2 test, and SMD was compared using the Wilcoxon signed-rank test. The level of significance was set at $p<0.05$ (two tailed).

Patient and public involvement

No patients or public members were involved in conducting this research.

RESULTS

Characteristics of newspaper stories, original studies and subsequent studies

Figure 1 illustrates the details of the search. The eight newspapers selected were the *New York Times* (USA, quality), *Washington Post* (USA, quality), *Daily Telegraph* (UK, quality), *Times* (UK, quality), *USA Today* (USA,

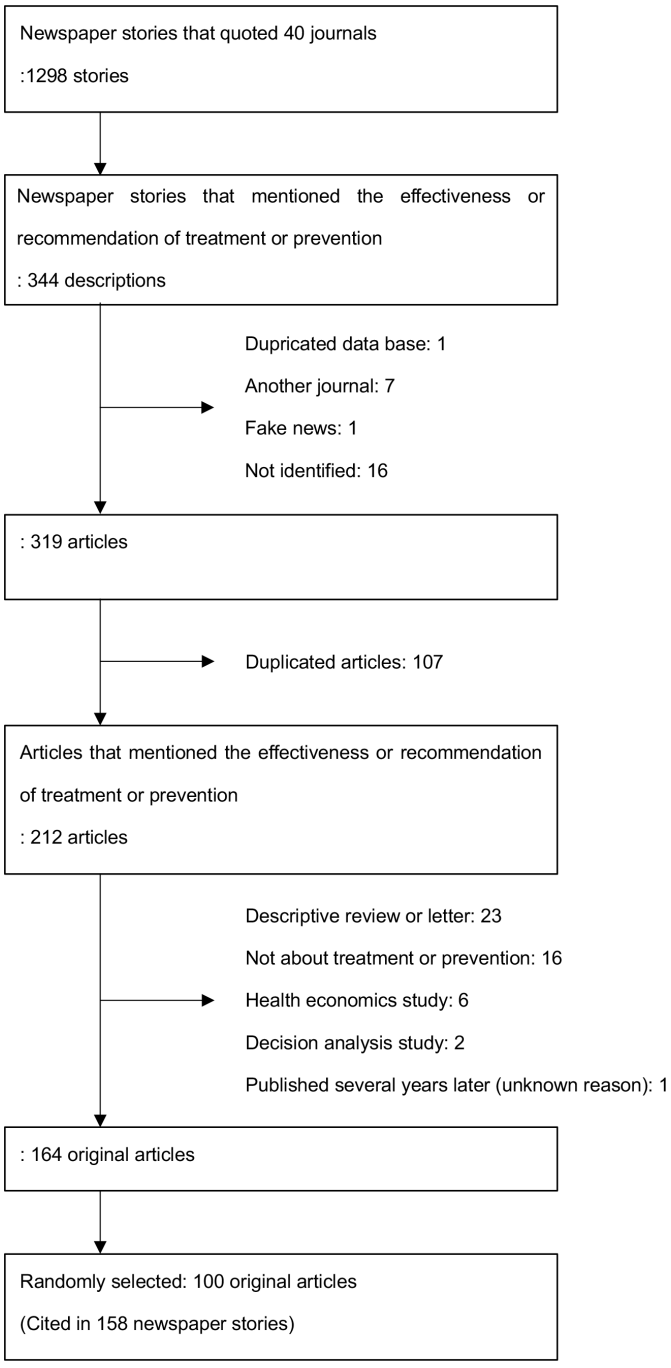

Figure 1 Flow chart of original study identification process.

non-quality), *Daily News* (USA, non-quality), *Daily Mail* (UK, non-quality) and *Daily Mirror* (UK, non-quality). When searching for journal names in newspaper stories, we found 1298 newspaper stories, of which 344 described the effectiveness of or recommended certain treatments or preventive measures (kappa=0.73) (table 1). Online supplemental eTable 1 lists the names of 40 medical journals.

A total of 344 newspaper stories were referred to in 319 scientific journal articles. After excluding duplicates, we identified 212 articles that mentioned the effectiveness of the recommended treatment or prevention. We excluded 48 articles because the research questions could not be identified. Finally, we identified 164 original articles and randomly selected 100 of them. These were cited in 158 newspaper articles. The journals in which the 100 original articles were published were as follows: *New England Journal of Medicine* (*NEJM*), 39; *Journal of the American Medical Association* (*JAMA*), 21; *Lancet*, 16; *British Medical Journal* (*BMJ*), 9; *Archives of Internal Medicine*, 8; *Annals of Internal Medicine*, 3; *American Journal of Epidemiology*, 1; *American Journal of Public Health*, 1; *Infection Control and Hospital Epidemiology*, 1; and *Mayo Clinic Proceedings*, 1. Approximately three-quarters of these articles were published in three major journals (*NEJM, JAMA, Lancet*).

Of the 100 articles, 58 were RCTs and 31 OSs. A few other designs corresponded to various ICD-10 categories. Of the 158 newspaper stories, two-thirds were in quality papers and the rest in non-quality.

For four of the 100 original studies, the newspapers stated their effectiveness, but the primary outcome of those studies did not indicate their effectiveness. Therefore, these were excluded from this study. In the remaining 96 studies, 104 effective treatments were identified. Subsequent studies on each treatment were searched. We identified relevant subsequent studies for 86 of these 104 treatments. The 18 others remained unchallenged (table 2). Of the 86 subsequent studies, 83 were SR (SR of RCTs, n=45; SR of OSs, n=23; SR of RCTs and OSs, n=15), followed by RCT (n=2) and OS (n=1). The PubMed search formulae are listed in online supplemental eTable 2.

Comparisons of original and subsequent studies

Table 2 shows the proportions of the confirmed studies. A total of 69% (59/86) (95% CI 58.1 to 77.5) of the original studies were confirmed in subsequent studies. Furthermore, of the 59 confirmed original studies, 16 were comparable to subsequent studies in terms of effect size. Among these 16, 13 were replicated and three reported effect sizes larger than the corresponding subsequent studies. Of these 16 studies, 11 compared SMDs. The median SMDs of the original and subsequent studies were 0.23 (0.18, 0.45) and 0.25 (0.15, 0.32), respectively (p=0.34, Wilcoxon signed-rank test). However, for the remaining 43 studies, strict comparisons of effect sizes were not possible because the outcomes were not fully matched between the original and subsequent studies. Details of the original and subsequent studies are presented in online supplemental eTable 3.

We conducted subgroup analyses on the proportions of confirmed studies for each research design in the original articles (online supplemental eTable 4). The proportions of confirmed OS and RCT studies (of which there was a relatively large number) were 61.3% (19/31) and 70.5% (31/44), respectively. Other designs included fewer studies, and we found no significant differences in the research design (p=0.74, χ^2 test). For the ICD-10 categories, the differences according to disease were not significant (p=0.67, χ^2 test). The proportion of confirmed studies cited in quality papers (56/88, 63.6%) was lower

Table 1 Characteristics of included newspaper stories

Newspaper	Country	Newspaper type	Newspaper stories that quoted 20 general and internal medicine journals	Newspaper stories that quoted 20 public, environmental and occupational health journals	Total
New York Times	USA	Quality	258	13	271
Washington Post	USA	Quality	279	22	301
Daily Telegraph	UK	Quality	28	5	33
Times	UK	Quality	191	18	209
USA Today	USA	Non-quality	122	11	133
Daily News	USA	Non-quality	65	7	72
Daily Mail	UK	Non-quality	173	9	182
Daily Mirror	UK	Non-quality	91	6	97
Total			1207	91	1298

than that in non-quality papers (31/44, 70.5%); however, the difference was not statistically significant (p=0.42, χ^2 test).

Example 1: contradicted

A prospective cohort study published in *BMJ* in 2000, covered by *Daily Mail*, suggested that drinking fluoridated water significantly reduced hip fractures.[16] Neither the subsequent matching study, meta-analysis of 14 observational studies, nor the original study[17] found any significant risk reduction in hip fractures.

Example 2: confirmed

One RCT published in the *JAMA* in 2000 and covered by the *Washington Post* suggested that sertraline was more effective than a placebo in patients with post-traumatic stress disorder (PTSD). The subsequent matching study was a meta-analysis comparing pharmacotherapies for PTSD, published in 2022.[18] In the subgroup analysis, which included the original RCT, sertraline was compared with placebo. The authors concluded that sertraline was effective. Therefore, the effectiveness reported in the original study was confirmed in a subsequent study. Furthermore, the point estimate of the original study's RR described in the subsequent study's forest plot was 0.70, and the point estimate and 95% CI of the RR of the new article was 0.68 (0.56 to 0.81). After calculating the SMD from these values, the original study had an SMD of 0.26, and the new study had a value of 0.27 (95% CI 0.15 to 0.40). We categorised this finding as not only 'confirmed' but also 'replicated'.

Example 3: unchallenged

Examples included in the unchallenged studies are as follows: Most studies have investigated unique interventions (eg, short nails for preventing infection, anti-digoxin fab for cardiac arrhythmia, horse chestnut seed extract for chronic venous insufficiency, beta-sheet breaker peptides for prion-related disorders, the Krukenberg procedure for double-hand amputees and yoga for carpal tunnel syndrome), and several studies have examined the effects of special drug use (eg, ondansetron for bulimia nervosa, growth hormone for Crohn's disease and combination therapy with old antidepressants, nefazodone and psychotherapy for chronic depression). However, these findings are difficult to validate using well-designed studies. The details are shown in online supplemental eTable 3.

DISCUSSION

This is the first study to examine a 20-year course of treatment or prevention recommended by newspaper articles in various medical fields. We selected newspaper stories that recommended certain treatments or preventions in 2000 and compared their results with those in the original research articles and compared the original studies with newer ones with better-controlled designs. Sixty-nine per cent (59/86) of the original studies were confirmed by subsequent studies. Among the confirmed studies, 13 of the 16 studies replicated both the direction and magnitude of the treatment effect. In studies in which the effects were confirmed, the effect sizes were relatively stable. However, the results of the remaining 43 studies were not comparable.

Table 2 Main analyses of the proportion of confirmed studies

	Total	Unchallenged	Contradicted	Confirmed	Proportion of confirmed studies, 95% CI (%)
Original studies	104*	18	27	59	68.6 (58.1 to 77.5)

*104 comparisons from 96 original studies (including duplicates).

As far as we know, few studies investigated the replicability of articles quoted in daily newspapers.[2 19] One is about attention deficit hyperactivity disorder studies, and the other is about risk factor studies; the proportions of 'confirmed' studies according to their definitions were 20% and 49%, respectively. The proportion of confirmed cases in our study (68.6%) was higher than those in these studies. The reasons for this may be as follows. Previous studies have not focused on treatment or prevention. Therefore, these proportions could not be compared. Furthermore, the definition of 'confirmed' in these studies was stricter than in our study. However, even in well-known newspapers, one-third of the stories may have been overturned by subsequent studies. Several studies have reported that the reporting standard in quality newspapers is significantly higher than that in non-quality papers.[9 20 21] In this study, the proportion of confirmed studies in quality newspapers was slightly lower than that in non-quality newspapers; however, this difference was not statistically significant. There may not be much of a difference between highly circulated quality papers and low-quality papers.

This study had some limitations. First, newspaper story authors often do not provide details about their information sources. It is often claimed that the best journalists are those with the most sources'.[22] In these cases, we could not find any articles quoted in newspapers. Therefore, for convenience, we used the journal names as search words. Consequently, only better-quality newspaper stories, in which journal names were written, were included. This may have led to the discovery of higher quality stories. Consequently, the proportion of quoted RCT may be higher than that of other standard newspaper stories. The credibility of studies cited in newspaper articles that do not list the sources of citations remains unclear. Second, an increasing number of SRs have been published in recent years, and several similar SRs can often be found on any research topic. Therefore, it is difficult to select the most appropriate option. To find the optimal subsequent study, two independent researchers checked the full paper and selected the best study from among several candidates. This reduced the number of arbitrary choices as much as possible. Third, we assumed that most subsequent study designs would be SR. Therefore, we searched the Web of Science for new studies that cited the original paper, and compared them with the effect sizes shown in the forest plot. However, the authors of subsequent SRs did not always cite the original articles for various reasons (eg, subtle differences in the type of outcome or timing of measurement). If cited, they were excluded from forest plots. Only 11 studies compared SMDs and 43 studies, although found to be effective, were unable to compare effect sizes. It is possible that the original studies reported a very large effect size, while the subsequent studies were only marginally significant. Based on these results, it is impossible to determine whether the SMDs are stable. Future studies should rigorously compare effect sizes by aligning outcomes. Fourth, 18 unchallenged studies focused on unique topics. Our definition of primary outcome excluded these numbers from the denominator, which makes the proportion of confirmed studies appear higher than it is. If these were included in the denominator, the proportion of confirmed cases would have been much lower.

However, this study has several strengths. This is the first study to examine the veracity of newspaper stories on treatment and prevention in various medical fields. Second, we followed up on each treatment over a 20-year period and took relevant subsequent studies with stronger designs as the gold standard. Although we cannot rule out the possibility that the results of subsequent studies may be reversed in the future, we believe that the results obtained over the past 20 years are generally robust. Third, to find the most appropriate subsequent study, we reviewed and discussed many SRs using the Web of Science and PubMed. We spent a lot of time carefully going through this process to make sure we did not miss any relevant papers.

CONCLUSION

The results for clinical research articles were relatively stable for papers in which the citation source was properly listed in newspaper articles. Journalists should provide information on the source studies to enable researchers to identify them. However, the results of approximately one-third of these studies were overturned over the following two decades. Readers should be aware that more than a few claims made in highly circulated newspapers based on high-profile journal articles may be overturned in subsequent studies.

Author affiliations
[1]Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan
[2]Oku Medical Clinic, Osaka, Japan
[3]Department of Advanced Medicine for Rheumatic diseases, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[4]Department of Emergency Medicine, National Hospital Organization Mito Medical Center, Ibaraki, Japan
[5]Department of Human Health Science, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[6]Department of Healthcare Epidemiology, School of Public Health in the Graduate School of Medicine, Kyoto University, Kyoto, Japan

Contributors AT and TAF conceived the study. AT and TAF designed the study. AT, YaT, AO, YuT, SF, YO, NT, YH and NI did the literature search and extracted the data. AT did the analyses. AT and TAF wrote the first draft of the manuscript. All authors contributed to the interpretation of the findings and subsequent edits of the manuscript. TAF provided overall supervision to the project. AT is the guarantor and accepts full responsibility for the work and the conduct of the study, had access to the data and controlled the decision to publish.

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 AT received lecture fees from Sumitomo Dainippon Pharma, Eisai, Janssen Pharmaceutical, Meiji-Seika Pharma, Mitsubishi Tanabe Pharma, Otsuka and Takeda Pharmaceutical. AO received research grants and/or speaker fees from Pfizer, Bristol-Myers Squibb, Advantest, Asahi Kasei Pharma, Chugai Pharmaceutical, Eli Lilly Japan K. K., Ono Pharmaceutical Co., UCB Japan Co., Mitsubishi Tanabe Pharma Co., Eisai, Abbvie, Takeda Pharmaceutical, and Daiichi

Sankyo. SF received a research grant from JSPS KAKENHI (Grant Number JP 20K18964), the KDDI Foundation and the Pfizer Health Research Foundation. TAF reports personal fees from Boehringer-Ingelheim, DT Axis, Kyoto University Original, Shionogi, and SONY, and a grant from Shionogi outside the submitted work; TAF has patents 2020-548587 and 2022-082495 pending, and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

ORCID iDs
Aran Tajika http://orcid.org/0000-0003-3926-8867
Yusuke Tsutsumi http://orcid.org/0000-0002-9160-0241
Satoshi Funada http://orcid.org/0000-0002-8925-2348
Toshi A Furukawa http://orcid.org/0000-0003-2159-3776

REFERENCES

1 Woloshin S, Schwartz LM, Kramer BS. Promoting healthy skepticism in the news: helping journalists get it right. *J Natl Cancer Inst* 2009;101:1596–9.
2 Gonon F, Konsman J-P, Cohen D, *et al*. Why most biomedical findings echoed by newspapers turn out to be false: the case of attention deficit hyperactivity disorder. *PLoS One* 2012;7:e44275.
3 Selvaraj S, Borkar DS, Prasad V, *et al*. Media coverage of medical journals: do the best articles make the news *PLoS ONE* 2014;9:e85355.
4 Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:e124.
5 Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005;294:218.
6 Pereira TV, Horwitz RI, Ioannidis JPA. Empirical evaluation of very large treatment effects of medical interventions. *JAMA* 2012;308:1676.
7 Circulations ABO. n.d. Available: http://www.auditbureau.org
8 Media AfA. n.d. Available: http://www.auditedmedia.com
9 Wilson A, Bonevski B, Jones A, *et al*. Media reporting of health interventions: signs of improvement, but major problems persist. *PLoS One* 2009;4:e4831.
10 Bell L, Seale C. The reporting of cervical cancer in the mass media: a study of UK newspapers. *Eur J Cancer Care (Engl)* 2011;20:389–94.
11 Dean M. Tabloid campaign forces UK to reconsider sex-offence laws. *Lancet* 2000;356:745.
12 Lexisnexis. n.d. Available: https://www.lexisnexis.com/en-us/home.page
13 Tajika A, Ogawa Y, Takeshima N, *et al*. Replication and contradiction of highly cited research papers in psychiatry: 10-year follow-up. *Br J Psychiatry* 2015;207:357–62.
14 Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000;19:3127–31.
15 Cohen J. *Statistical power analysis in the behavioral science*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
16 Phipps KR, Orwoll ES, Mason JD, *et al*. Community water fluoridation, bone mineral density, and fractures: prospective study of effects in older women. *BMJ* 2000;321:860–4.
17 Yin X-H, Huang G-L, Lin D-R, *et al*. Exposure to fluoride in drinking water and hip fracture risk: a meta-analysis of observational studies. *PLoS One* 2015;10:e0126488.
18 Williams T, Phillips NJ, Stein DJ, *et al*. Pharmacotherapy for post traumatic stress disorder (PTSD). *Cochrane Database Syst Rev* 2022;3:CD002795.
19 Dumas-Mallet E, Smith A, Boraud T, *et al*. Poor replication validity of biomedical association studies reported by newspapers. *PLoS One* 2017;12:e0172650.
20 Cooper BEJ, Lee WE, Goldacre BM, *et al*. The quality of the evidence for dietary advice given in UK national newspapers. *Public Underst Sci* 2012;21:664–73.
21 Robinson A, Coutinho A, Bryden A, *et al*. Analysis of health stories in daily newspapers in the UK. *Public Health* 2013;127:39–45.
22 Weitkamp E. British newspapers privilege health and medicine topics over other science news. *Public Relations Review* 2003;29:321–33.

BMJ Health & Care Informatics

Validation framework for the use of AI in healthcare: overview of the new British standard BS30440

Mark Sujan ,[1,2] Cassius Smith-Frazer,[3] Christina Malamateniou,[4] Joseph Connor,[5] Allison Gardner,[6] Harriet Unsworth,[7] Haider Husain[3]

To cite: Sujan M, Smith-Frazer C, Malamateniou C, *et al*. Validation framework for the use of AI in healthcare: overview of the new British standard BS30440. *BMJ Health Care Inform* 2023;**30**:e100749. doi:10.1136/bmjhci-2023-100749

Received 13 February 2023
Accepted 09 May 2023

[1]Human Factors Everywhere, Woking, UK
[2]Education, Healthcare Safety Investigation Branch, Reading, UK
[3]Healthinnova Ltd, Bath, UK
[4]School of Health and Psychological Sciences, City University London, London, UK
[5]CarefulAI, Cwmbran, UK
[6]National Institute for Health and Care Excellence, London, UK
[7]Wellcome Trust, London, UK

Correspondence to
Dr Mark Sujan;
mark.sujan@gmail.com

INTRODUCTION

The British standard 'BS30440: Validation Framework for the Use of AI in Healthcare' will be published in the second quarter of 2023.[1] It details the evidence required by technology developers to assess and validate products using artificial intelligence (AI) in healthcare settings. Healthcare providers can mandate that their suppliers' products be certified against BS30440 to assure themselves and their service users that the AI product is effective, fair and safe.

For a decade now, there has been growing interest in healthcare AI, especially applications using machine learning approaches, such as deep neural networks.[2] This interest has grown exponentially over the past 5 years, with government bodies and regulatory authorities, non-governmental think tanks, professional associations and academic institutions developing a multitude of relevant guidance to address their local contexts.[3] In the United Kingdom (UK) this includes, for example, the National Institute for Health and Care Excellence Evidence Standards Framework for digital health technologies, NHSX guidance on 'Artificial Intelligence: how to get it right', and guidance on algorithmic impact assessment published by the Ada Lovelace Institute. In addition, there are several international reporting guidelines, including SPIRIT-AI[4] (The Standard Protocol Items: Recommendations for Interventional Trials - Artificial Intelligence) and CONSORT-AI[5] (Consolidated Standards of Reporting Trials - Artificial Intelligence) for clinical trials of healthcare AI technologies.

As a result, the landscape of guidance on how to develop safe and effective AI systems for healthcare is fragmented across hundreds of documents, largely with a focus on products that would be regulated as medical devices. This has led to a lack of formalised guidance for healthcare AI technologies that are out of remit of medical device regulations, such as those with a focus on healthcare resource planning, logistics or general health and well-being support. While regional regulations for AI (such as the European Union AI act) are in development, and national regulators (eg, the UK Medicines and Healthcare Products Regulatory Agency) develop their own regulatory strategies, there is a clear space for well designed and auditable standards to ensure safety, effectiveness and equity. Such standards do not replace legislation but can form the basis for novel regulatory approaches.

Against this backdrop of a multitude of guidance and frameworks, BS30440 is unique in two ways. First, BS30440 has been developed from an extensive review, which synthesises the fragmented healthcare AI landscape into a single, comprehensive framework. It has received additional input from a multidisciplinary panel of experts, two rounds of public consultations, as well as a community and patient engagement panel.

Second, BS30440 represents a fully auditable standard for the assessment of healthcare AI products. Auditing is critical to ensure that healthcare AI products offer demonstrable clinical benefits, that they reach sufficient levels of performance, that they successfully and safely integrate into the health and care environment, and that they deliver inclusive outcomes for all patients, service users and practitioners. Any healthcare AI product that is successfully certified against BS30440, has passed a broad and substantial evaluation across these properties.

This thorough process of synthesis and stakeholder consultation, coupled with the introduction of clear assessment criteria for

auditing, offers significant potential to suppliers who wish to navigate the complex AI guidance landscape by complying with a single framework.

STRUCTURE AND ASSESSMENT CRITERIA

BS30440 is structured around a product life-cycle for healthcare AI, described in five phases: inception, development, validation, deployment and monitoring. For each phase of the product life-cycle, a set of assessment criteria has been defined. The life-cycle within the framework is not intended to be prescriptive or to be thought of in a necessarily linear fashion. However, all the assessment criteria should be addressed during the product life-cycle.

The assessment criteria were developed through literature reviews and in consultation with a committee of subject matter experts from academia, governmental bodies, healthcare institutions and standards organisations. Patient and public representatives were involved to inform the development and review of the assessment criteria through written contribution and as part of a focus group to ensure diverse and inclusive input.

The standard includes carbon impact criteria because of the anticipated expansion of AI across the sector, which has the potential to result in significant environmental impact if not managed. Feedback from the public consultation on this topic was overwhelmingly positive. The importance of equity and fairness is highlighted as a core criterion for the development of ethical AI products,[6] both in ensuring engagement with the target audience, but also in terms of diversity and inclusiveness of decision-making and development.

Consideration is given to the inclusion of human factors and ergonomics, which runs across all life-cycle phases. The importance of human factors and ergonomics in the healthcare AI product life-cycle is increasingly being recognised,[7 8] and this is reflected in the standard.

In total, BS30440 includes 18 assessment criteria. Each assessment criterion is specified through auditable clauses against which an AI product can be assessed for conformity. An overview is provided in figure 1.

INTENDED AUDIENCE, AUDITING AND COMPLIANCE

The assessment criteria specified in BS30440 are intended to provide assurance of the safety, quality and performance of healthcare AI products. Patients and the public are the main beneficiaries of BS30440 as recipients of healthcare services, but they are not expected to engage directly with the standard.

BS30440 can support healthcare organisations in the procurement and assessment of AI products. Healthcare providers can adopt the standard as a requirement for their suppliers, similar to conformance to other standards such as ISO9001. For a given AI product to be certified against the standard, the product must have been developed and validated following a process aligned to the assessment criteria specified in BS30440. The developer must document evidence for the assessment criteria, which will be evaluated by a competent external auditor. This can provide reassurance to clinicians and staff working with AI products and to patients and their families. It also adheres to core principles of ethical AI in terms of transparency and accountability in providing clarity as to the chain of responsibility and evidence throughout the product life-cycle.

Developers are encouraged to begin with a self-assessment of their current development processes against each of the assessment criteria to establish their current level of conformity, to identify gaps in their development process and documentation, and to decide where they might need to improve their development processes. Developers should create an action plan for how to address any identified gaps to achieve certification and gather evidence as they design and develop their AI product. It is recommended that internal auditors or quality assurance managers work alongside the development team to collate and present the evidence in a systematic and standardised way and to minimise potential rework. Service

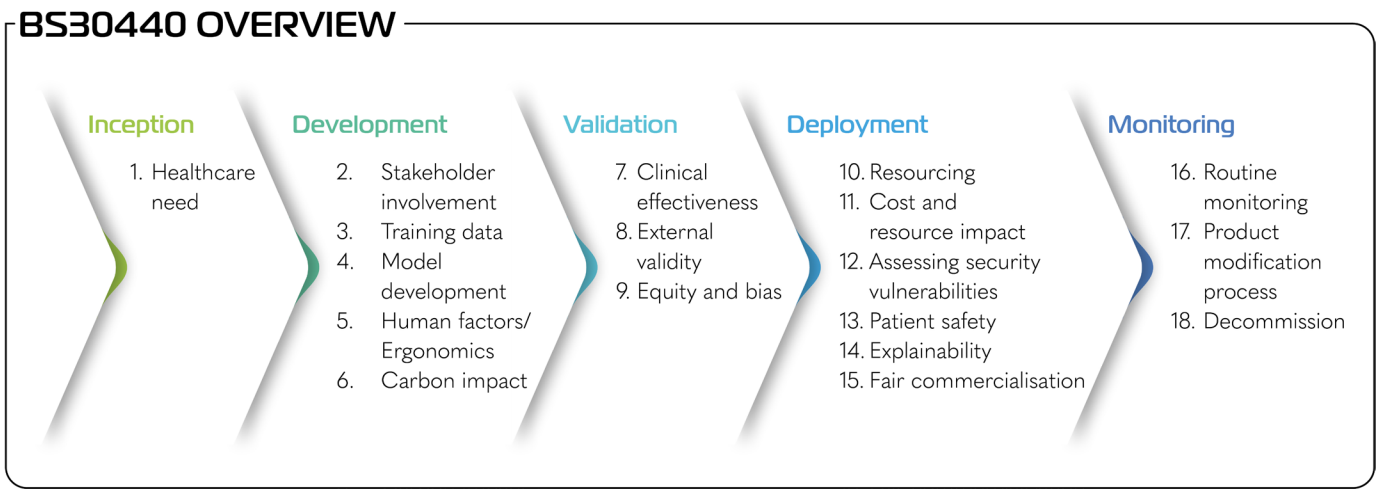

Figure 1 BS30440 structure and assessment criteria.

users and key stakeholders should be involved at all stages of the product life-cycle.

CONCLUSION

BS30440 provides within a single resource an actionable, comprehensive and auditable validation framework for healthcare AI. Conformity with the standard can provide assurance to developers, deploying healthcare organisations and patients.

There is a degree of overlap between BS30440 and other relevant forms of assessment and regulation of healthcare information technology, including medical device regulations and the NHS Digital clinical safety standards (DCB 0129 and DCB 0160). However, BS30440 covers specifically AI products, including those which are not included in current medical device regulations. In this way, such healthcare AI technologies can still be subjected to a process of assessment and certification to ensure a minimum standard across relevant assessment criteria. The standard has been shared with and received input from a wide range of organisations and stakeholders, and, as such, the evidence provided for the assessment criteria should facilitate any necessary regulatory approvals.

BS30440 applies across all development and use contexts of healthcare AI. However, the standard might be especially valuable in contexts where developers and deploying organisations have limited prior experience, knowledge, and resources about suitable healthcare AI development processes, including formal software engineering and assurance processes.

BS30440 assumes suppliers will have knowledge of all relevant design information either because they have developed the algorithm and models themselves or because they can access this information from the developers. This can be problematic in future scenarios, where potentially suppliers might make use of generic AI products, such as the increasingly popular large language models applications. In these situations, the supplier will not have designed the model and they might be unable to explain its origin. In that case, suppliers would not be compliant unless they are able to design an assurance wrapper around the generic model. This is not yet addressed in the standard.

BS30440 has been developed as a national initiative. While international committees including International Organization for Standardization (ISO) / International Electrotechnical Commission Subcommittee (IEC SC) 42 and European Committee for Standardization (CEN) / European Committee for Electrotechnical Standardization (CENELECT) Joint Technical Committee 21 (JTC21) have published standards and are in the process of developing their future work programmes, these initiatives are not specific to healthcare AI. The publication and use of BS30440 can serve as a first testbed to inform subsequent international standardisation activities for healthcare AI.

Twitter Mark Sujan @MarkSujan and Christina Malamateniou @CMalamatenioi

Acknowledgements We are grateful to the experts on the BS30440 panel who developed the standard: Ele Harwich, Danny Ruta, Shuaib Haris, Arup Paul, Tina Woods, Bilal Mateen, Michelle Williams, Pritesh Mistry, Lorna Harper, Russel Pearson, Shakir Laher, Magda Bucholc, Nina Wilson, Martha Martin, Alexander Deng, Rob Vigneault, Paul Cuddeford, Abdul Sayeed, Paul Davies, Mat Rawsthorne, Melissa Wood, Peter Bloomfield, Craig York, Nathan Hill, Matthew Fenech. We are also grateful to the reviewers who provided feedback during two public consultations.

Contributors MS: conceptualisation, writing—original Draft; CS-F: writing—review and editing; CM: writing—review and editing; JC: writing—review and editing; AG: writing—review and editing; HU: writing—review and editing; HH: writing—review and editing.

Funding MS received funding from the Assuring Autonomy International Programme, which is a joint initiative between the University of York and Lloyd's Register Foundation.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Mark Sujan http://orcid.org/0000-0001-6895-946X

REFERENCES

1 British Standards Institution. *BS 30440 validation framework for the use of AI within healthcare*. BSI, 2023. Available: https://standardsdevelopment.bsigroup.com/projects/2021-00605#/section
2 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
3 Malamateniou C, McFadden S, McQuinlan Y, *et al*. Artificial intelligence: guidance for clinical imaging and therapeutic radiography professionals, a summary by the society of Radiographers AI working group. *Radiography (Lond)* 2021;27:1192–202.
4 Liu X, Rivera SC, Moher D, *et al*. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *BMJ* 2020;370:m3164.
5 Liu X, Rivera SC, Moher D, *et al*. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *BMJ* 2020;370:m3164.
6 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.
7 Sujan M, Furniss D, Grundy K, *et al*. Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ Health Care Inform* 2019;26:e100081.
8 Vasey B, Nagendran M, Campbell B, *et al*. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ* 2022;377:e070904.

BMJ Health & Care Informatics

TransFAIR study: a European multicentre experimental comparison of EHR2EDC technology to the usual manual method for eCRF data collection

Nadir Ammour,[1] Nicolas Griffon,[2,3] Juliette Djadi-Prat ,[4] Gilles Chatellier,[5,6] Martine Lewi,[7] Marija Todorovic,[7] Augustín Gómez de la Cámara,[8] Maria Teresa García Morales,[8] Sara Testoni,[9] Oriana Nanni,[9] Christoph Schindler,[10] Mats Sundgren,[11] Almenia Garvey,[12] Tomothy Victor,[12] Manon Cariou ,[1] Christel Daniel[2,3]

To cite: Ammour N, Griffon N, Djadi-Prat J, *et al*. TransFAIR study: a European multicentre experimental comparison of EHR2EDC technology to the usual manual method for eCRF data collection. *BMJ Health Care Inform* 2023;**30**:e100602. doi:10.1136/bmjhci-2022-100602

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2022-100602).

Received 22 July 2022
Accepted 11 May 2023

For numbered affiliations see end of article.

Correspondence to
Dr Nadir Ammour;
nadir.ammour@sanofi.com

ABSTRACT

Purpose Regulatory authorities including the Food and Drug Administration and the European Medicines Agency are encouraging to conduct clinical trials using routinely collected data. The aim of the TransFAIR experimental comparison was to evaluate, within real-life conditions, the ability of the Electronic Health Records to Electronic Data Capture (EHR2EDC) module to accurately transfer from EHRs to EDC systems patients' data of clinical studies in various therapeutic areas.

Methods A prospective study including six clinical trials from three different sponsors running in three hospitals across Europe has been conducted. The same data from the six studies were collected using both traditional manual data entry and the EHR2EDC module. The outcome variable was the percentage of data accurately transferred using the EHR2EDC technology. This percentage was calculated considering all collected data and the data in four domains: demographics (DM), vital signs (VS), laboratories (LB) and concomitant medications (CM).

Results Overall, 6143 data points (39.6% of the data in the scope of the TransFAIR study and 16.9% when considering all data) were accurately transferred using the platform. LB data represented 65.4% of the data transferred; VS data, 30.8%; DM data, 0.7% and CM data, 3.1%.

Conclusions The objective of accurately transferring at least 15% of the manually entered trial datapoints using the EHR2EDC module was achieved. Collaboration and codesign by hospitals, industry, technology company, supported by the Institute of Innovation through Health Data was a success factor in accomplishing these results. Further work should focus on the harmonisation of data standards and improved interoperability to extend the scope of transferable EHR data.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Several articles reported on use of electronic health records (EHR) as eSource for clinical trials, however, they were performed in single centre, with a single EHR system, a single electronic data capture (EDC) system and most often not in an actual clinical study.

WHAT THIS STUDY ADDS

⇒ This is the first study that proved the ability to use EHRs data as eSource in actual studies conducted by different sponsors at different sites using different EHRs systems, in different countries in Europe.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The results provide practical insights to enable use of EHR2EDC technologies in actual clinical trials and help policy-makers to promote regulations to encourage adoption of EHRs as eSource in clinical trials.

INTRODUCTION

Clinical trials complexity increased over the last decade, leading to a growing amount of data to be collected. Meantime hospitals transitioned from paper records to electronic health records (EHRs), making it possible for reuse in clinical research. Previous studies reported that 13%–75% of the trial data points are redundantly captured in EHR and the electronic data capture (EDC) system and might sometimes be present in a third paper copy.[1 2] This results in time-consuming redundant data entry, data cleaning and source data verification, leading to an increase burden and costs.

For almost a decade, in addition to regulators, industry forums are recommending the broad implementation of EHRs as eSource in clinical trials.[3–13] A recent literature review identified attempts to use EHR data as an eSource through direct electronic transfer into EDC systems.[14 15] Most of the EHR-EDC integration initiatives are usually

one-time-only, not scalable solutions limited to a single site, single vendor, single pharmaceutical company context, not using standards for data representation.[16–18]

Several obstacles require to be addressed to enable use of EHR data as source data in multicentric clinical trials. The main obstacles are the lack of integrated workflow between care and clinical research conducted in silos and of intersystem interoperability. Other barriers include resistance to change, and poor quality of EHR data that could influence assessment of outcomes. To improve the transparency and completeness of publications of the results of clinical trials conducted using cohorts or routinely collected data, a reporting guideline, the CONSORT-ROUTINE (extension for the reporting of randomised controlled trials conducted using cohorts and routinely collected data), has been recently developed, including a checklist to facilitate the compliance.[19]

A widely acceptable and cost-effective approach to interoperability between EHRs and clinical research systems operating under different legal frameworks across Europe[1 20 21] was developed by the Innovative Medicines Initiative EHRs for Clinical Research (EHR4CR) project conducted between 2011 and 2016.

The EHR2EDC project, which is a continuation of EHR4CR, is a public–private partnership, funded by the European Institute of Innovation and Technology (EIT) Health involved in improving European healthcare systems. This initiative was led by Sanofi and included three other pharmaceutical companies (AstraZeneca, Janssen, UCB Pharma), a clinical research organisation (ICON), a health data technology company (InSite network platform, Custodix a TriNetX company), four European hospital organisations (Assistance Publique-Hôpitaux de Paris (AP-HP) in Paris, France; Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST) in Meldola, Italy; Medizinische Hochschule Hannover (MHH) in Hannover, Germany and Hospital Universitario 12 de Octubre, (12 de Octubre) in Spain) and the European Institute for Innovation through Health Data (a non-for-profit organisation). The aim of this project was to design, develop and evaluate a technology enabling use of EHRs as eSource in clinical trials.[22]

The objective of the EHR2EDC consortium was to prove that at least 15% of data entered in the EDC can be semiautomatically transferred from its source EHRs. To evaluate this the TransFAIR study was designed, within relevant context of use, by including six different clinical studies across three research sites in Europe. The primary endpoint was the ability to achieve 15% of correct and accurate data transfer from EHRs to study EDC. This percentage was agreed as a consensus, and based on published work on this subject, such as the RE-USE project.[1]

METHODS

Study design

The TransFAIR study consisted in the experimental comparison of two data collection methods: the EHR2EDC module implementing a semiautomatic transfer of EHR data to an EDC system versus the usual manual data collection (protocol available in online supplemental material). We included real ongoing clinical trials (support CT). Selected trials were conducted according to their protocol and were not affected by the TransFAIR study. FAIR refers to the FAIR principles: Findability, Accessibility, Interoperability and Reuse of data assets guided the design of the EHR2EDC module.[23]

Data were shared between partners according to the European Union General Data Protection Regulation. The interoperability implementation and data flow were performed within a solution compliant with data privacy and good clinical practice regulations.

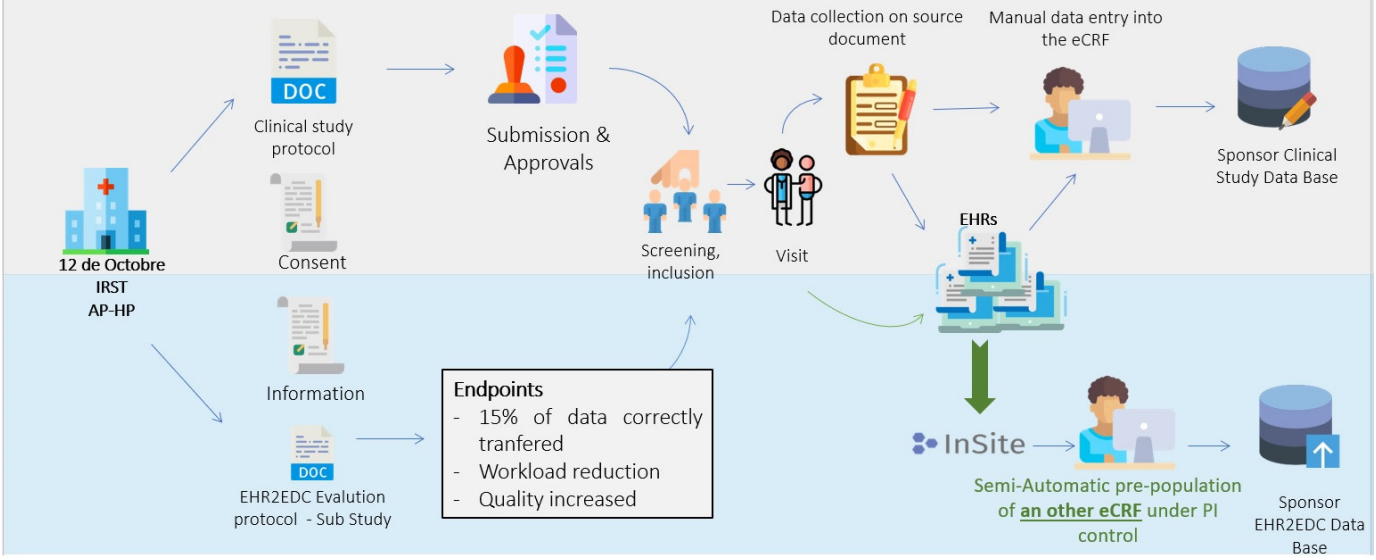

Figure 1 General organisation of the TransFAIR study. AP-HP, Assistance Publique-Hôpitaux de Paris; EHR2EDC, Electronic Health Records to Electronic Data Capture; PI, principal investigator ; eCRF, electronic Case Report Form.

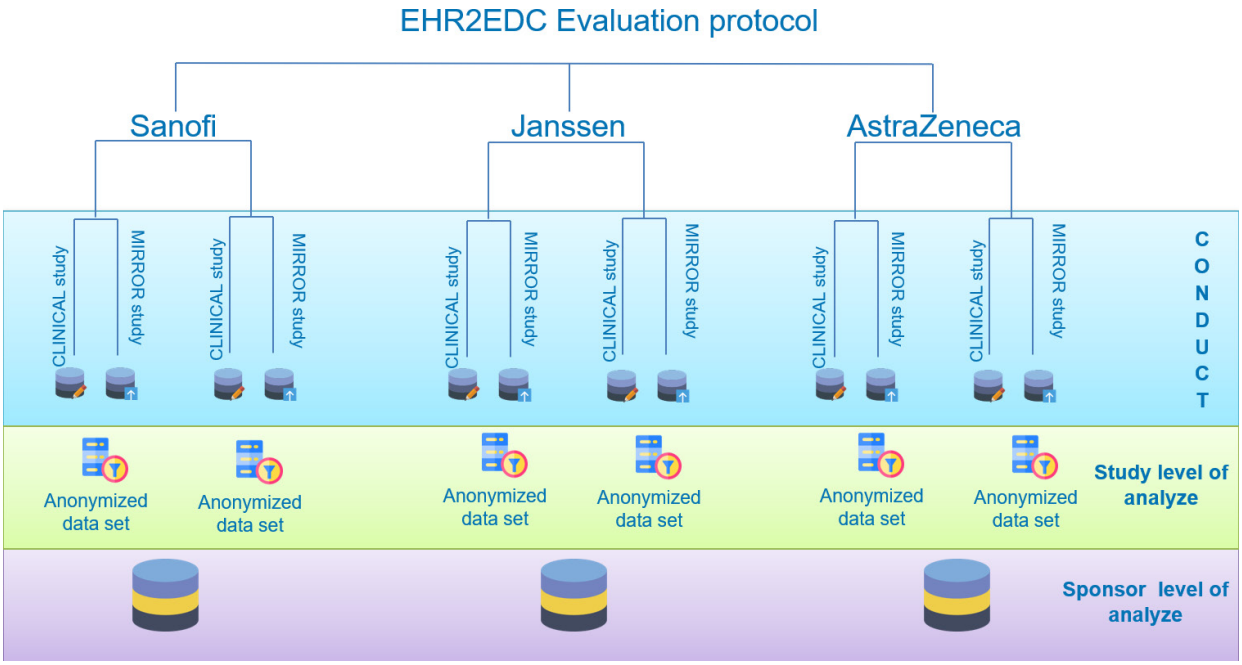

Figure 2 Mirror study representation. EHR2EDC, Electronic Health Records to Electronic Data Capture.

EHR2EDC data in scope, module setup, study and patient selection

The data domains of interest were selected based on the frequency of data types collected in a large pool of studies (N=120) and present across multiple therapeutic areas. The results were reviewed by members of the project experts in clinical data standards with extensive experience in designing study eCRFs, including experts from the clinical research organisation (CRO) ICON, for their experience across sponsors and therapeutic areas. The four data domains selected are: demographics (DM), vital signs (VS), laboratory (LB) and concomitant medication (CM), from where a core set of 48 Clinical Data Interchange Standards Consortium (CDISC) data elements was identified and the 20 associated CDISC code lists were mapped to selected terminologies (International Classification of Diseases, 10th Revision (ICD-10), Logical Observation Identifiers Names and Codes (LOINC), Anatomical Therapeutic Chemical (ATC) Classification and Systematised Nomenclature of Medicine Clinical Terms (SNOMED-CT)). CDISC standard is the destination format selected as it is used by pharmaceutical companies or CRO for their eCRF. The semantic mappings developed for this project is accessible at the following site:

It covers four CDISC domains: DM, LB analysis, VS and CM. LOINC is the main reference terminology used on hospital side, however, it has sometimes been necessary to use other terminologies.

Four Fast Healthcare Interoperability Resources (FHIR) profiles associated with a list of standardised value sets were defined to support data extraction specification and guide mappings done by hospitals terminology experts.

The EHR2EDC module, from the InSite platform has been installed successfully in: AP-HP, 12 de Octubre, IRST and MHH.

Six studies from three different Sponsors (AstraZeneca, Janssen and Sanofi) were selected by the consortium according to the following criteria: support CT had to be conducted in a hospital partner with principal investigators agreeing to support the TransFAIR study, it had to include patients during the evaluation period (July to December 2019) and preferably collecting a large number of LB data.

The selected studies were conducted in three hospitals: AP-HP, 12 de Octubre and IRST. MHH only started to map on SNOMED-CT, with weekly data refresh from the clinical live systems, hence was not included.

Data collection and management

For each clinical trial selected for the TransFAIR study, a mirrored EDC database, replicating the study specific EDC database, was set up and connected to the EHR2EDC module of the InSite platform installed at each site. The mirror EDC database represents the 'experimental' database while the original database was used as a 'control' (figure 1). The data collected in each EDC system of participating clinical trials were captured in the study eCRF (Medidata Classic Rave V.2020.2.0) using traditional manual data entry by a study coordinator or an investigator. In the mirrored database, the same data were collected through the InSite platform (figure 2). Once connected to the InSite platform data, the study coordinator/investigator selects a clinical trial, a subject and a visit (as defined in the protocol). Then he/she must associate the visit to the actual date of the patient's visit. The

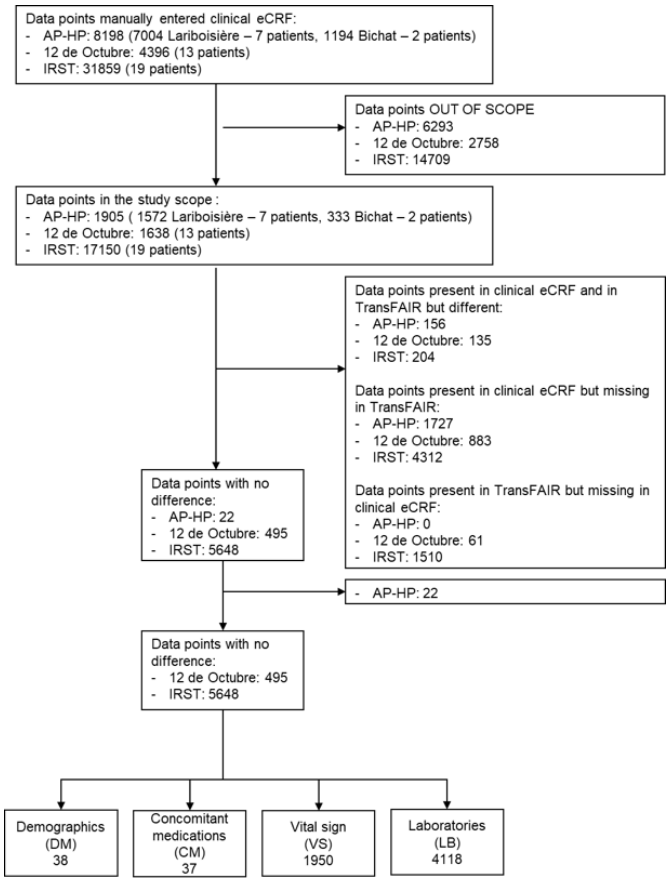

Figure 3 Study flow chart. AP-HP, Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori; IRST, Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori; eCRF, electronic Case Report Form.

platform provides an interface, with fields prefilled with EHR data (required by study protocol) at the selected date. The study coordinator/investigator can, therefore, review and validate data before their transfer to the mirror EDC.

Several patients were included, and visits completed before the TransFAIR study started. Data were transferred, retrospectively for completed visits and for new visits. Investigators supervised the automated data collection by reviewing, validating and transferring data to the experimental database. Experimental and control databases were then reconciled by the sponsor to identify discrepancies. An absence of difference between data points collected in both databases was classified as OK, while a difference was classified as NOK. Each discrepancy was investigated by the investigator by checking source documents to verify the actual value of the data point for which a discrepancy was identified and to document the reason.

Study endpoints

The primary endpoints are the percentage of data points accurately processed.

- Per individual studies.
- Across studies.

The secondary endpoints are the percentage of data points in scope accurately processed.

- Per individual study.
- Pooled across studies.
- Per data domain pooled across studies.

Statistical analysis

Since the TransFAIR was a proof-of-concept study, neither a sample size calculation nor a power consideration was performed. The only hypothesis to be tested was that at least 15% of the datapoints could be semiautomatically and accurately transferred by the EHR2EDC module. Results were analysed individually, for each study and pooled together to be presented across studies.

The percentage of data accurately transferred was calculated as the number of data correctly transferred in the experimental database divided by the total number of data manually entered into the control database.

The hypothesis of transferring at least 15% of the data was tested using a one-sided exact binomial test. An estimate of proportions with their 95% CI was provided. The exact calculation method was used if the approximation of the Normal law was not possible. Subgroup analyses were planned on the following variables: study site and data domain (DM, VS, LB and CM).

The statistical significance level was set at $p<0.05$ (two sided). The global statistical analysis was carried out with the R software (release V.3.6.3; R Foundation for Statistical Computing, Vienna, Austria), by the Clinical Trial Unit of each site and by ICON.

RESULTS

Presentation of the studies and patient data

The EHR2EDC transfer module of the InSite platform was active from 20 September 2019 to 30 November 2019. The analysis included the data points of five of the six selected studies: AZ D169CC, PCR3001 and TED14856 at 12 de Octubre in Madrid, BCL30003 and D19BC at IRST. The data from the EFC14875 study at AP-HP were excluded from the overall analysis. Most data collected for that study, at that site, were captured using paper as a source.

The data from the five studies databases were pooled and represented a total of 41 424 data points. The subset of data in the scope of the study (ie, DM, VS, LB and CM) represented 19 240 data points, 46.4% of total data collected (figure 3).

Primary endpoint: percentage of data accurately transferred (all data)

- Per individual studies

Studies TED14856 and AZ D19BC had reached higher results than set objective of 15%. They achieved, respectively, 26.5% (one-sided 95% CI 24.0%) and 22.8% (one-sided 95% CI 22.2%) (table 1).

Table 1 Percentage of accurately transferred data, overall and by study

Hospital and study	No of data accurately transferred (n)	% of accurately transferred data	
		In the scope of the TransFAIR study % (95% CI lower limit)	For the whole study % (95% CI lower limit)
12 de Octubre	495	32.7% (30.3%)	11.3% (10.3%)
AZ D169CC (AstraZeneca, NCT03619213)	143	26.2.% (22.4%)	7.8% (6.6%)
PCR3001 (Janssen, NCT02257736)	35	25.6% (18.5%)	2.6% (1.8%)
TED14856 (Sanofi, NCT03284957)	317	35.0% (31.9%)	26.5% (24.0%)
IRST	5648	55.6% (54.6%)	17.7% (17.3%)
BCL30003 (Janssen, NCT03390504)	400	60.3% (56.5%);	6.7% (6.0%)
AZ D19BC (AstraZeneca, NCT02516241)	5248	55.2% (52.4%)	22.8% (22.2%)
Total	6143	39.6%	16.9% (16.6%)

IRST, Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori.

Other studies achieved less than 10.0% of correctly processed data.

- Across studies

The EHR2EDC module was able to transfer accurately 16.9% of data points across studies, (one-sided 95% CI 16.6%) and represents 6143 data points.

Secondary endpoints (data in scope)

- Results per individual study varies between 26.6% and 60.3% (table 1).

The AZ D19BC trials and TED14856 trial both had a majority of VS and LB data (table 2).

- Results pooled across studies: The EHR2EDC module was able to process accurately 39.6% (p<0.0001) of data points in scope (N=6143 data points) (table 3).
- Results per data domain pooled across studies: Within each data domain in scope, the percentage of data correctly processed varies. The highest results are observed for VS (40.9%), LB (40.6%) and for DM (34.2%). Data from CM have the lowest percentage: 7.7% (table 4).

DISCUSSION

The concept of mirror study has proven to be an effective method for validation of a novel technology to support data collection, in a relevant context of use: different EHRs, investigation sites, sponsors and studies.

The primary objective of the study was successfully met, with over 15% (16.9%) of the data points entered in the e-CRF correctly processed from EHR source records.

The four domains DM, VS, LB and CM selected by the consortium represent 46.4% of the data collected through the five trials in scope, this results validates the consortium choice.

A per study analysis demonstrates the major contribution of the local LB data followed to a lesser degree by the VS data to achieve an acceptable proportion of transferable data. This suggests that studies in oncology (ex: TED14856 and the AZ D19BC), with high volume of local LB data are best candidates for the early use of this digital data collection technology in the near future.[24]

The two domains LB and VS covers around 40% of the data in scope and represent more than 96% of accurately transferred data. This reflects the availability and good quality of these data at the hospitals EHRs.

The interoperability challenge has been successfully addressed through the implementation within the EHR2EDC module of a core list of data elements and its associated library of terminology mappings. The

Table 2 Data transferred per domain and per study

Hospitals and studies	Patients (N)	Break down of data points by data domain and by study				
		CM	DM	VS	LB	Total
12 de Octubre	13	385	90	564	599	1638
AZ D169CC (AstraZeneca, NCT03619213)	8	126	28	394	0	548
PCR3001 (Janssen, NCT02257736)	4	72	52	61	0	185
TED14856 (Sanofi, NCT03284957)	1	187	10	144	564	905
IRST	19	156	41	4541	12412	17150
BCL30003 (Janssen, NCT03390504)	2	156	41	40	449	686
AZ D19BC (AstraZeneca, NCT02516241)	17	0	0	4501	11963	16464

CM, concomitant medications; DM, demographics; IRST, Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori; LB, laboratories; VS, vital signs.

Table 3 Number of data points per data domain

N (%)	DM	CM	VS	LB	Total
No difference	38 (0.2)	37 (0.2)	1950 (12.6)	4118 (26.6)	6143 (39.6)
Missing in TransFAIR	73 (0.5)	432 (2.8)	2566 (16.6)	4373 (28. 2)	7444 (48.0)
Different	0 (0.0)	4 (0.0)	150 (0.1)	185 (1.2)	339 (2.2)
Missing in clinical eCRF*	0 (0.0)	10 (0.1)	104 (0.7)	1457 (9.4)	1571 (10.1)
Total	111 (0.7)	483 (3.1)	4770 (30.8)	10 133 (65.4)	15 497 (100.0)

eCRF, electronic Case Report Form.
*Excluded from total.
CM, concomitant medications; DM, demographics; LB, laboratories; VS, vital signs.

EHR2EDC module has been efficiently deployed in the four hospitals and the different users trained. The mapping and its implementation were designed to be reusable across studies, with limited (re)verification activities, to provide operational efficiencies, both for the sponsor and for site staff.

The limitations on the results for data in scope highlight a combination of factors affecting the ability to achieve higher performance. Among those factors, we have identified several root causes with possible remediations:

Regulations

For DM data (DM domain), legal limitations in collecting ethnicity in Europe produces an artefact as this information is collected during trials. When analysing only legally acceptable DM data, the result was 100%. This suggests that calculation methods and possible automatic quality controls must consider local regulations to be accurate.

Case report form design

The primary cause of missing data for the VS and LB domains arises for specific data points collected in study eCRFs to document the execution of the procedure. Most of the empty fields expect a 'yes' value for the question 'Has the test been performed?'/'Was the blood sample taken?'. This could be resolved by using auto populated fields (updated to a 'yes' value if results are present).

Local investigator's team practices

Unlike IRST, other hospitals did not routinely train their staff to fill-in structured forms of the EHRs, and so the proportion of data accurately transferred was adversely affected by the proportion of data collected in EHR as free text or in paper source documents when running a clinical trial.

Special attention should be focused on staff using EHRs to collect patient data associated with a clinical study for preventing free text data entry or paper source. This includes training hospital staff in data quality standards, upgrading quality assurance measures and strengthening data governance activities, to enable EHR data to be trustworthy reused in research.

In the TransFAIR study, the low percentage of CM data correctly transferred reflects that they are more often recorded as free text, for example, in unstructured documents (eg, doctor's letters) and a large part is prescribed outside of the investigational site and is consequently not captured in the EHR.

Table 4 Proportion of data collected and not collected for the four domains in the TransFAIR study scope

Data domain	% of data correctly transferred	% of missing data
Demographics	34.2	65.8
Laboratories	40.6	59.4
Vital signs	40.9	59.1
Concomitant medications	7.7	92.3

Clinical site maturity/readiness

Other factors influencing the level of performance include the site maturity in using their EHRs for clinical trials activities. Site organisational capabilities, best practices (EHR data quality assurance, use of EHRs as eSource in clinical trials, just-in-time data flow), skilled staff (data integration, data management) are essential to benefit from this new method of digital data collection.

Guided work effort is needed to augment the proportion of data recorded as eSource in EHRs to be collected using EHR2EDC solutions. Initial focus would expand transferability of structured data in EHRs, and work at rendering unstructured data to be collected. We envision this effort to be made possible through the development of consensus on 'high-value data sets', representing the data most commonly collected in clinical trials.

Nevertheless, not all data collected in clinical trials has its correspondence in patients' EHRs sources. For example, specific forms in eCRFs collect data in relation with the management and evaluation of investigational medicinal products (tracking, patient's compliance, pharmacokinetic data, etc).

CONCLUSION

Overall, a 16.9% successful transfer rate was achieved across the five trials included in the TransFAIR study. A

transfer rate of 26.5% of data used as eSource EHRs was achieved in one of the trials.

Clinical investigational sites, CRO staff and sponsor personnel involved in the planning and the execution of trials, as well as those involved in the management of EHR, EDC and EHR2EDC technologies must join forces for success. It is recommended to promote coordination and synchronisation of all actors to align, not only on the European EHR technology standards, but also on addressing the following different dimensions: change management, and new roles, needed to achieve routine use of EHR data as eSource in clinical trials.

A roadmap to transition use EHR2EDC in clinical trials would include the following recommendations: (1) Sponsors should further develop sets of high value data, combining structured and unstructured data to help guide and prioritise the efforts needed for scalability. (2) Clinical sites should initially focus on structured data, such as LB, DM, VS and CM using common data models, for example, HL7 FHIR, increasingly implemented in clinical research[25–27] and reference terminologies for example, ICD10, LOINC, ATC, SNOMED, etc. (3) Clinical sites should develop capabilities to leverage data from unstructured format (free text, clinical documents, images), not standardised data, using natural language processing technologies and efforts to enhance both data interoperability and data quality controls. and (4) Collaborative effort at the ecosystem level should be encouraged to create the right incentives to develop and grow the market with technology providers to offer EHR2EDC services to sponsors' organisations.

Author affiliations
[1]Clinical Innovation Office, Sanofi SA Recherche & Developpement, Paris, France
[2]DSI-WIND, Assistance Publique - Hopitaux de Paris, Paris, France
[3]LIMICS, INSERM U1142, Paris, France
[4]Unité de Recherche Clinique, AP-HP, Hôpital Européen Georges Pompidou, Paris, France
[5]Unité de Recherche Clinique, Assistance Publique - Hopitaux de Paris, Paris, France
[6]Université de Paris, Paris, France
[7]Janssen Pharmaceutica NV, Beerse, Belgium
[8]Instituto de Investigacion, Hospital 12 de Octubre, Madrid, Spain
[9]Biostatistics and Clinical Trials, IRCCS Istituto Romagnolo per lo Studio dei Tumori Dino Amadori, Meldola, Italy
[10]Medizinische Hochschule Hannover, Hannover, Germany
[11]Data sciences AI, Biopharmaceuticals RD, AstraZeneca FoU Göteborg, Goteborg, Sweden
[12]ICON plc, Dublin, Ireland

Acknowledgements Aurèle N'Dja, Christina Mueller, David Voets, Elias Neri, Frédéric Billois, Greet Byttebier, Gregory Navarro, Jacques Hilbey, Jean Charlet, Karen Fanouillere, Laurent Luttenauer, Khalid Benhammou, Lars Fransson, Mikael Forsby, Nicola Gentili, Noelle Moly, Pablo Serrano, Pascal Coorevits, Philippe Rocolle, Prathima Surabhi, Sabri Miled, Trisha Simpson, Yannick Girardeau, Khalil Lechheb.

Contributors NA contributed to the launch of the consortium and acted as the project director, contributed to the TransFAIR study protocol design, supervised the execution of the study and is guarantor. NG contributed to the data analysis. JD-P wrote the statistical analysis plan, contributed to the study submission, execution, and the data analysis at AP-HP. GC contributed to the protocol design. MT coordinated the study execution. AGdIC contributed to the study submission, execution at 12 de Octubre. MTGM contributed to the study submission, execution, the data analysis at 12 de Octubre. ST contributed to the study submission, execution, the data analysis at IRST. CS reviewed the protocol. MS contributed to the protocol desig. AG reviewed the manuscript. TV performed the statistical analysis. MC contributed to the protocol design, coordinated the study execution, performed the statistical analysis. CD contributed to the protocol design.

Funding Funded by a grant from the EU (EIT Health Activity No 18269).

Competing interests The EITheallth funds the EHR2EDC project in which the experimentation described in the article took place. EIT reimburse hospitals for the expenses related to the project.And Sanofi, Janssen and ICON provided partial financial support to hospital consortium partners to attend workshops, covering travel and meal related-costs.

Patient consent for publication Consent obtained directly from patient(s)

Ethics approval A TransFAIR study protocol was submitted and received approval form each site's ethics committee (approval numbers: AP-HP: CSE-EDS no 19-16; 12de0: 19/324; IRST: 5802/2019). Patients included in the selected support CTs were asked for their oral or written consent to process their data for the TransFAIR study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. Data are not available as sharing the data as public acess was not part of the inform consent signed by patients.

ORCID iDs
Juliette Djadi-Prat http://orcid.org/0000-0003-4949-9474
Manon Cariou http://orcid.org/0000-0003-3212-3446

REFERENCES

1 El Fadly A, Rance B, Lucas N, *et al*. Integrating clinical research with the healthcare enterprise: from the RE-USE project to the EHR4CR platform. *J Biomed Inform* 2011;44 Suppl 1:S94–102.
2 Nordo A, Eisenstein EL, Garza M, *et al*. Evaluative outcomes in direct extraction and use of EHR data in clinical trials. *Stud Health Technol Inform* 2019;257:333–40.
3 Food and Drug Administration (FDA). Guidance for industry: electronic source data in clinical investigations. Rockville, MD FDA; 2013. Available: https://www.fda.gov/media/85183/download
4 Food and Drug Administration (FDA. Use of electronic health record data in clinical investigations guidance for industry. Rockville, MD FDA; 2018. Available: https://www.fda.gov/media/97567/download
5 European Commission. Commission recommendation on a European electronic health record exchange format (C(2019)800); 2019. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019H0243&rid=2
6 European Medicines Agency (EMA). Reflection paper on expectations for electronic source data and data transcribed to electronic data collection tools in clinical trials. London EMA; 2010. Available: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/reflection-paper-expectations-electronic-source-data-data-transcribed-electronic-data-collection_en.pdf
7 Pharmaceuticals and Medical Devices Agency (PMDA). Technical conformance guide on electronic study data submissions. Tokyo PMDA; 2015. Available: https://www.pmda.go.jp/files/000215352.pdf
8 Medicines and Healthcare products Regulatory Agency (MHRA). MHRA position statement and guidance electronic health records. London MHRA; 2015. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/470228/Electronic_Health_Records_MHRA_Position_Statement.pdf

9 eClinical Forum. EDC site survey: investigational site perspectives on clinical trial information systems. 2012. Available: https://eclinicalforum.org/DesktopModules/EasyDNNNews/DocumentDownload.ashx?portalid=0&moduleid=433&articleid=26&documentid=56
10 eClinical Forum EDC Hosting Task Force. Checklist for electronic data capture systems in clinical trials using service providers. 2015. Available: https://eclinicalforum.org/downloads/white-paper-checklist-and-task-ownership-matrix-for-electronic-data-capture-systems-in-clinical-trials-using-service-providers
11 eClinical Forum. The future vision of electronic health records as eSource for clinical research. 2006. Available: https://eclinicalforum.org/DesktopModules/EasyDNNNews/DocumentDownload.ashx?portalid=0&moduleid=433&articleid=9&documentid=75
12 Using health records to help clinical research [Innovative Medicines Initiative web site]. 2017. Available: https://www.imi.europa.eu/projects-results/success-stories-projects/using-health-records-help-clinical-research
13 Parab AA, Mehta P, Vattikola A, *et al*. Accelerating the adoption of eSource in clinical research: a transcelerate point of view. *Ther Innov Regul Sci* 2020;54:1141–51.
14 Garza M, Myneni S, Nordo A, *et al*. Nordo A et al. eSource for standardized health information exchange in clinical research: a systematic review. *Stud Health Technol Inform* 2019;257:115–24.
15 El Fadly A, Lucas N, Rance B, *et al*. The REUSE project: EHR as single datasource for biomedical research. *Stud Health Technol Inform* 2010;160:1324–8.
16 Nordo AH, Eisenstein EL, Hawley J, *et al*. A comparative effectiveness study of eSource used for data capture for a clinical research registry. *Int J Med Inform* 2017;103:89–94.
17 Wang B, Hao X, Yan X, *et al*. Evaluation of the clinical application effect of eSource record tools for clinical research. *BMC Med Inform Decis Mak* 2022;22.
18 Garza MY, Rutherford M, *et al*. Evaluating site-level Implementations of the HI7 FHIR standard to support eSource data exchange in clinical research. *AMIA Annu Symp Proc* 2021;2020:472–81.
19 Kwakkenbos L, Imran M, McCall SJ, *et al*. CONSORT extension for the reporting of randomised controlled trials conducted using cohorts and routinely collected data (CONSORT-ROUTINE): checklist with explanation and elaboration. *BMJ* 2021;373:n857.
20 De Moor G, Sundgren M, Kalra D, *et al*. Using electronic health records for clinical research: the case of the EHR4CR project. *J Biomed Inform* 2015;53:162–73.
21 Coorevits P, Sundgren M, Klein GO, *et al*. Electronic health records: new opportunities for clinical research. *J Intern Med* 2013;274:547–60.
22 Griffon N, Pereira H, Djadi-Prat J, *et al*. Performances of a solution to semi-automatically fill eCRF with data from the electronic health record: protocol for a prospective individual participant data meta-analysis. *Stud Health Technol Inform* 2020;270:367–71.
23 FAIR principles. n.d. Available: https://www.go-fair.org/fair-principles/
24 Vattikola A, Dai H, Buckley M, *et al*. Direct data extraction and Exchange of local LABS for clinical research protocols: a partnership with sites, Biopharmaceutical firms, and clinical research organizations. *JSCDM* 2021;1.
25 Cheng AC, Duda SN, Taylor R, *et al*. Redcap on FHIR: clinical data interoperability services. *J Biomed Inform* 2021;121:103871.
26 Vorisek CN, Lehne M, Klopfenstein SAI, *et al*. Fast healthcare Interoperability resources (FHIR) for interoperability in health research: systematic review. *JMIR Med Inform* 2022;10:e35724.
27 Duda SN, Kennedy N, Conway D, *et al*. HI7 FHIR-based tools and initiatives to support clinical research: a scoping review. *J Am Med Inform Assoc* 2022;29:1642–53.