Postprint of Transparent Reporting for Early Clinical Evaluation of AI-based Clinical Decision Support Systems

Authors: Lei Fang, Du Liang, Dong Min, Liu Xuemei, Liu Xuemei

Date: 2023-11-01T00:00:00+00:00

Abstract

With the widespread application of artificial intelligence (AI) in the medical field, an increasing number of AI-based clinical decision support systems are being deployed in clinical diagnosis, screening, and other domains. Early clinical evaluation is of significant importance for assessing the clinical performance, safety, and human factors of AI clinical decision support systems, and lays the foundation for further large-scale clinical trials; however, the transparency and completeness of related clinical research reports remain to be improved. In May 2022, the DECIDE-AI guideline for reporting early-stage clinical evaluation of AI-based clinical decision support systems was officially published online. This article will integrate the AI-related reporting items from DECIDE-AI with relevant literature to discuss transparent reporting for early clinical evaluation of AI-based clinical decision support systems, aiming to assist developers and researchers of AI clinical decision support systems in China to better understand and apply the relevant guidelines, thereby improving the transparency of early clinical evaluation reports for AI clinical decision support systems.

Full Text

1 Intended Use of AI Clinical Decision Support Systems

Over the past decade, artificial intelligence (AI) has been widely applied in clinical diagnosis, screening, and other areas to improve healthcare quality and efficiency [1-5]. Although some studies have indicated that certain AI algorithms demonstrate accuracy comparable to clinical experts in preclinical in-silico studies [6], the application of AI clinical decision support systems (CDSS) in clinical decision-making is influenced by complex human and environmental factors. The journey from algorithm to clinical application requires multi-stage evaluation, including in-silico trials, early-stage clinical evaluation, and randomized

controlled trials [7-8]. Early-stage clinical evaluation of AI CDSS aims to validate clinical performance, safety, and human factors within real clinical contexts before large-scale randomized controlled trials, thereby laying the groundwork for subsequent large-scale studies. Consequently, early-stage clinical evaluation of AI CDSS is critically important.

Transparent reporting of early-stage clinical evaluation studies for AI CDSS should appropriately describe the system' s target users, clinical application scenarios, and intended outcomes. AI CDSS integrates interdisciplinary knowledge and technologies, imposing specific requirements on users regarding their profession, training level, and computer literacy. Furthermore, detailed descriptions of clinical application scenarios and intended outcomes—including the system' s scope of application and potential impacts on existing clinical pathways—not only enhance clinicians' and patients' understanding of AI CDSS, thereby improving trust and satisfaction, but also facilitate the dissemination and implementation of these systems.

2 Transparent Reporting of Research Processes

Transparent reporting of clinical research enhances trial reliability and result reproducibility. Moreover, from an ethical perspective, transparency and explainability of AI decision-making are equally crucial [9]. Currently, the transparency and completeness of AI-related clinical research reports remain suboptimal. Systematic reviews have noted that AI-related clinical research reports lack transparency, with some studies exhibiting significant methodological flaws, such as failing to report data preprocessing information, bias risks, model applicability assessments, or external validation [10]. Reporting guidelines for AI-related clinical research can help investigators improve report completeness and transparency. In May 2022, the Developmental and Exploratory Clinical Investigations of Decision support systems driven by Artificial Intelligence (DECIDE-AI) guideline was officially published online [11]. DECIDE-AI specifically targets early-stage clinical evaluation, a phase analogous to Phase 1/2 in drug development or Stage 2a/2b in the Idea, Development, Exploration, Assessment, and Long-term follow-up (IDEAL) framework for surgical innovations/devices [12]. DECIDE-AI comprises 27 reporting items, including 17 AI-specific items (with 28 sub-items) and 10 generic items. The guideline primarily focuses on clinical usability, safety, and human factors, thereby establishing specific requirements for trial evaluation environments, implementation processes, and human factors.

To promote standardized and transparent early-stage clinical research and reporting of AI CDSS in China, this article explores transparent reporting for early-stage clinical evaluation of AI-based decision support systems by integrating AI-specific reporting items from DECIDE-AI with relevant literature. We examine transparent reporting from four perspectives: intended use of AI CDSS, research processes, ethics, and data sharing, aiming to facilitate understanding and application of reporting guidelines and enhance transparency in early-stage clinical evaluation reports.

Although professional associations and organizations have issued policies requiring transparency in AI project development across healthcare and research domains, AI inherently lacks transparency due to its complexity, including unexplainable algorithms, opaque data at all stages, data bias, and insufficient model versioning transparency. AI transparency holds significant meaning for all stakeholders, including developers, users, and regulatory authorities. Below, we elaborate on transparent reporting of research processes across five key dimensions: participants, AI CDSS, implementation, safety and failures, and human factors.

2.1 Participants

Meta-analyses of AI diagnostic tools have revealed high risks of patient selection bias in most studies [13-14]. Patient selection bias not only affects the generalizability of AI diagnostic models but, more importantly, may generate systematic errors—producing incorrect outputs. Therefore, transparent reporting of patient recruitment processes should be strengthened, detailing inclusion and exclusion criteria at both patient and data levels, along with methods for determining recruitment numbers. Participants in AI CDSS studies include not only patients but also users, whose interactions with the system influence system performance and study outcomes. Consequently, researchers should report not only user recruitment processes and inclusion/exclusion criteria but also describe the time and effort required for users to become familiar with the AI CDSS. However, the DECIDE-AI guideline does not recommend excluding users at the data level. Additionally, researchers may preregister their studies on clinical trial registration platforms—a crucial pathway for enhancing research process transparency that also aligns with requirements from the Declaration of Helsinki, WHO, and the International Committee of Medical Journal Editors.

2.2 AI Clinical Decision Support Systems

AI CDSS may undergo multiple iterations and updates. Research indicates that algorithms aligned with decision-support processes can reduce inherent system bias to some extent [15]. Beyond algorithms, data constitute another core component of AI CDSS. Most AI CDSS require large datasets for training and validation, making data acquisition, input, and processing critical determinants of system performance [15-16]. Regarding AI CDSS outputs, both content and presentation format significantly influence user adoption and acceptance [17], while effective human-computer interaction depends on rational user interface design. Therefore, authors should carefully consider and describe algorithm types and versions, input data characteristics, and output content and interface design in both study design and report writing.

2.3 Implementation

Regarding implementation, unlike the Consolidated Standards of Reporting Trials–Artificial Intelligence (CONSORT-AI) and Standard Protocol Items:

Recommendations for Interventional Trials–Artificial Intelligence (SPIRIT-AI), DECIDE-AI requires not only reporting how AI CDSS outputs are applied to clinical decision-making but also clearly describing the evaluation environment, such as healthcare center type and scale, and hardware/software support. These environmental factors may affect data acquisition or measurement, subsequently influencing output results. Researchers must also report how final clinical decisions are formulated, user system utilization patterns, and significant changes to existing diagnostic and care processes, which may reflect the usability and clinical utility of AI CDSS.

2.4 Safety and Failures

Safety is the foremost concern in medical AI, encompassing various legal and ethical issues [18] throughout clinical research and practice. Therefore, safety assessment of AI CDSS before large-scale clinical studies is particularly critical to avoid catastrophic consequences. Naturally, during AI CDSS use, safety risks or failures may stem not only from the system itself but also from hardware/software support and user operation errors. How to identify these failures, assess patient safety risks, and minimize patient harm requires researchers to develop detailed safety risk protocols in advance and report them comprehensively.

2.5 Human Factors

Human factors evaluation primarily focuses on user interactions—namely, usability. Usability is defined as the extent to which a product can be used by specified users to achieve specified goals effectively, efficiently, and satisfactorily in a specified context of use [19], directly influencing user acceptance. Researchers can employ scales and questionnaires to assess satisfaction, task completion efficiency, and workload. For example, the System Usability Scale (SUS), widely used in usability research, enables rapid usability assessment through 10 items scored 0-4 points each [20]. Alternatively, the Net Promoter Score (NPS) can be used, assessing user satisfaction with a single question [21]. Such NPS surveys are common in everyday life (e.g., "How likely are you to recommend product XX to a friend?"). Regarding method selection, relevant guidance documents may be consulted [22].

3 Ethics

While AI CDSS promote healthcare quality and productivity improvements, they also introduce ethical challenges, including patient data privacy, system accuracy, and algorithmic bias. Although healthcare datasets are typically de-identified before use by developers, software may still re-identify individuals through cross-referencing with other data. In March 2022, China' s first national-level guidance document on ethics governance in science and technology, "Opinions on Strengthening Ethics Governance in Science and Technol-

ogy," proposed that ethics in science and technology represent the values and behavioral norms that must be followed in scientific research and technological development, serving as an important safeguard for healthy technological advancement [23]. In February 2023, the "Ethical Review Measures for Life Sciences and Medical Research Involving Human Subjects," jointly issued by the National Health Commission, Ministry of Education, Ministry of Science and Technology, and National Administration of Traditional Chinese Medicine, emphasized the importance of ethical review, stating that "life sciences and medical research involving human subjects should respect participants, follow principles of beneficence, non-maleficence, and justice, and protect privacy rights and personal information" [24]. The principle of "ethics before technology" must be consistently upheld in AI CDSS development and application, with appropriate technical or ethical methods employed to mitigate relevant ethical risks when necessary (e.g., privacy-preserving computation technologies that enable data to be "usable but invisible," preventing privacy breaches). AI-related ethical issues are not entirely consistent with traditional medical ethics, necessitating strengthened advocacy and review of AI-related medical ethics. Academic journals, as important platforms for disseminating research findings, should fulfill their advocacy and review roles effectively. However, according to our investigation, few domestic or international medical journals have updated their submission guidelines or instructions for authors to include AI-related medical ethics requirements [25].

4 Data Sharing

As previously noted, algorithms and data constitute the two core components of AI CDSS. Sharing relevant data enables readers to verify research findings while promoting communication and algorithmic improvement, aligning with principles of transparency and reproducibility. Data access is essential for validating and sharing scientific knowledge generated by AI. Data sharing not only advances scientific development but also enhances scientific rigor. Given the inseparability of AI research from code and data, greater emphasis should be placed on code and data sharing to deepen the value of AI clinical trials and prevent errors, fraud, or bias. Therefore, beyond declaring conflicts of interest, authors should state whether relevant data and code are accessible and how they can be obtained. With the advent of the open science era, data sharing will become increasingly important.

Conclusion

With the widespread application of AI in medicine, corresponding reporting guidelines are continuously being developed to enhance research report completeness and transparency. DECIDE-AI provides a framework for reporting AI CDSS research, helping developers and investigators document critical information throughout development, implementation, and evaluation. The guideline focuses on three aspects of early-stage clinical evaluation: clinical utility

in small-scale use, safety, and human factors. To further promote DECIDE-AI dissemination and strengthen report standardization, completeness, and transparency—while providing reference standards for authors in study design, protocol development, and registration—we recommend that journals incorporate DECIDE-AI into their author guidelines or instructions for authors. However, since AI CDSS may encounter diverse problems and challenges in clinical application, researchers should adapt and refine the guideline according to actual circumstances when applying it. Additionally, as most experts who developed the guideline are from the UK, certain limitations may exist. Nevertheless, DECIDE-AI provides valuable direction for AI CDSS-related research and reporting, contributing to the sustainable development of AI in healthcare.

References

[1] KERMANY D S, GOLDBAUM M, CAI W J, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning[J]. Cell, 2018, 172(5): 1122-1131.e9. DOI: 10.1016/j.cell.2018.02.010.

[2] TANG A, TAM R, CADRIN-CHÊNEVERT A, et al. Canadian association of radiologists white paper on artificial intelligence in radiology[J]. Can Assoc Radiol J, 2018, 69(2): 120-135. DOI: 10.1016/j.carj.2018.02.002.

[3] Pelcyger Ben. Artificial intelligence in healthcare: Babylon Health & IBM Watson take the lead[EB/OL]. [2023-09-19]. https://prescouter.com/2017/12/artificial-intelligence-healthcare/.

[4] SLOMKA P J, DEY D, SITEK A, et al. Cardiac imaging: working towards fully-automated machine analysis & interpretation[J]. Expert Rev Med Devices, 2017, 14(3): 197-212. DOI: 10.1080/17434440.2017.1300057.

[5] DE FAUW J, LEDSAM J R, ROMERA-PAREDES B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease[J]. Nat Med, 2018, 24(9): 1342-1350. DOI: 10.1038/s41591-018-0107-6.

[6] HANNUN A Y, RAJPURKAR P, HAGHPANAHI M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network[J]. Nat Med, 2019, 25(1): 65-69. DOI: 10.1038/s41591-018-0268-3.

[7] VASEY B, NOVAK A, ATHER S, et al. DECIDE-AI: a new reporting guideline and its relevance to artificial intelligence studies in radiology[J]. Clin Radiol, 2023, 78(2): 130-136. DOI: 10.1016/j.crad.2022.09.131.

[8] PARK Y, JACKSON G P, FOREMAN M A, et al. Evaluating artificial intelligence in medicine: phases of clinical research[J]. JAMIA Open, 2020, 3(3): 326-331. DOI: 10.1093/jamiaopen/ooaa033.

[9] PLOUG T, SUNDBY A, MOESLUND T B, et al. Population preferences for performance and explainability of artificial intelligence in health care: choice-

based conjoint survey[J]. J Med Internet Res, 2021, 23(12): e26611. DOI: 10.2196/26611.

[10] TORNERO-COSTA R, MARTINEZ-MILLANA A, AZZOPARDI-MUSCAT N, et al. Methodological and quality flaws in the use of artificial intelligence in mental health research: systematic review[J]. JMIR Ment Health, 2023, 10: e42045. DOI: 10.2196/42045.

[11] VASEY B, NAGENDRAN M, CAMPBELL B, et al. Publisher Correction: reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI[J]. Nat Med, 2022, 28(10): 2218. DOI: 10.1038/s41591-022-01951-8.

[12] HIRST A, PHILIPPOU Y, BLAZEBY J, et al. No surgical innovation without evaluation: evolution and further development of the IDEAL framework and recommendations[J]. Ann Surg, 2019, 269(2): 211-220. DOI: 10.1097/SLA.0000000000002794.

[13] PARKASH O, SIDDIQUI A T S, JIWANI U, et al. Diagnostic accuracy of artificial intelligence for detecting gastrointestinal luminal pathologies: a systematic review and meta-analysis[J]. Front Med, 2022, 9: 1018937. DOI: 10.3389/fmed.2022.1018937.

[14] LI X M, GAO X Y, TSE G, et al. Electrocardiogram-based artificial intelligence for the diagnosis of heart failure: a systematic review and meta-analysis[J]. J Geriatr Cardiol, 2022, 19(12): 970-980. DOI: 10.11909/j.issn.1671-5411.2022.12.002.

[15] GURUPUR V, WAN T T H. Inherent bias in artificial intelligence-based decision support systems for healthcare[J]. Medicina, 2020, 56(3): 141. DOI: 10.3390/medicina56030141.

[16] FAGHANI S, KHOSRAVI B, ZHANG K, et al. Mitigating bias in radiology machine learning: 3. performance metrics[J]. Radiol Artif Intell, 2022, 4(5): e220061. DOI: 10.1148/ryai.220061.

[17] FUJIMORI R, LIU K, SOENO S, et al. Acceptance, barriers, and facilitators to implementing artificial intelligence-based decision support systems in emergency departments: quantitative and qualitative evaluation[J]. JMIR Form Res, 2022, 6(6): e36501. DOI: 10.2196/36501.

[18] O' SULLIVAN S, NEVEJANS N, ALLEN C, et al. Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery[J]. Int J Med Robot, 2019, 15(1): e1968. DOI: 10.1002/rcs.1968.

[19] International Organization for Standardization. ISO 9241-11:2018 ergonomics of human-system interaction—Part 11: usability: definitions and concepts[EB/OL]. [2023-09-19]. https://www.iso.org/standard/63500.html.

[20] Brooke, J. SUS−a quick and dirty usability scale[EB/OL]. [2023-09-19]. https://hell.meiert.org/core/pdf/sus.pdf.

[21] REICHHELD F F. The one number You need to grow[J]. Harv Bus Rev, 2003, 81(12): 46-54, 124.

[22] British Standards Institution. BS EN 62366-1:2015+A1:2020 Application of usability engineering to medical devices[EB/OL]. [2023-09-19]. https://www.bsigroup.com/en-GB/standards/bs-en-62366-12015a12020/.

[23] General Office of the Communist Party of China Central Committee, General Office of the State Council. Opinions on Strengthening Ethics Governance in Science and Technology[A/OL]. (2022-03-20)[2023-09-19]. https://www.gov.cn/zhengce/2022-03/20/content_{5680105}.htm.

[24] National Health Commission, Ministry of Education, Ministry of Science and Technology, et al. Ethical Review Measures for Life Sciences and Medical Research Involving Human Subjects[A/OL]. (2023-02-27)[2023-09-19]. http://www.nhc.gov.cn/qjjys/s7946/202302/c3374c180dc5489d85f95df5b46afaf5.shtml.

[25] LEI F, DU L, DONG M, et al. Reflections on how medical journals should address medical ethics issues in the context of artificial intelligence[J]. Acta Editologica, 2023, 35(3): 263-267. DOI: 10.16811/j.cnki.1001-4314.2023.03.005.

Note: Figure translations are in progress. See original paper for figures.