Quality and Safety in the Literature: June 2024

Quality, Safety, and Artificial Intelligence

Dr Tayana Soukup PhD CPsychol FRSPH[a]✉ Prof Bryony Dean Franklin [b]

[a] Department of Surgery and Cancer, Imperial College London, London, UK

[b] Imperial College Healthcare NHS Trust and UCL School of Pharmacy, London, UK

CORRESPONDING AUTHOR

✉Dr Tayana Soukup PhD CPsychol FRSPH, t.soukup@imperial.ac.uk

Word count: 2,868

KEY POINTS

- ⇨ Global perspectives from 964 artificial intelligence (AI) and cancer researchers into the future impact of AI on cancer care highlight AI's potential to improve cancer grading, classification, diagnostic accuracy, and follow-up, while also identifying significant barriers to its integration into clinical practice and the need for standardization in cancer health data. *Current Oncology, 16 March 2023.*
- ⇨ Additional information (from human peers or AI) can have a strong influence on prescribing decisions made by intensive care doctors, particularly for AI recommendations, but the presence of an additional simple explanation did not significantly further increase adherence to AI suggestions. *Digital Medicine, 07 November 2023*
- ⇨ A methodical approach to assessing and improving the safety of AI systems in a clinical setting is important, as demonstrated through a case study with the AI Clinician for sepsis, providing a concrete example of AI's potential impact and the complexities involved in safely implementing AI in healthcare. *BMJ Health & Care Informatics, 04 June 2022*

THE FUTURE OF ARTIFICIAL INTELLIGENCE APPLICATIONS IN CANCER CARE

Current Oncology, 16 March 2023

This seminal study by Cabral et al[1] delves into the transformative potential of artificial intelligence (AI) in oncology, highlighting its pivotal role in enhancing healthcare quality and safety. The study aligns with the broader discourse on AI's capacity to revolutionize healthcare outcomes, drawing from insights previously proposed on the synergy between human expertise and AI across various medical disciplines.[2]

The study’s goal was to explore future applications of AI in cancer care (including reducing screening costs, diagnostics, grading and classifying cancer stages, improving follow-up services, aiding drug discovery, and improving prognostics), identify potential challenges, and assess the level of optimism among researchers in the field. The authors sought to use a thorough methodology to ensure collection of meaningful data from a global sample of AI and cancer research experts.

Cabral et al’s survey targeted authors of articles on AI and cancer, published between 20 September 2020 and 20 September 2022, indexed in Web of Science SCI-EXPANDED. After refining the contact list to 25,000, the questionnaire was validated through a pilot study, and 1,030 researchers agreed to participate; 881 (87.57% of total valid responses) were ultimately included in the analysis after excluding those who did not consent and those reporting no knowledge of the subject. giving a 4.1% response rate. Despite the low response rate, the study achieved a high absolute number of respondents, offering sufficient statistical power within the study’s specified margin of error (5%) and 95% confidence interval. Despite the potential for non-response bias, the participants, mainly affiliated with universities or research organizations and holding doctoral degrees, enrich our understanding by offering a broad spectrum of opinions and expectations about AI's application in cancer care that span Europe (42.47% of respondents), Asia (28.92%), and North America (including Central America and the Caribbean; 17.19%). The findings suggest a consensus on AI's beneficial impact on cancer care, particularly in diagnostic precision, grading and classification, reaffirming AI's potential in refining diagnostic accuracy not just in oncology, but across the healthcare spectrum.[4] However, the study also reveals critical barriers to AI's clinical adoption, including challenges with integration of AI algorithms and systems with existing healthcare systems and workflows due to the lack of standardization in cancer-related health data. It also highlights an urgent need for data standardization—a concern that resonates with regulatory insights on the impact of data protection laws on medical research.[5]

Emphasizing the need for empirical research to validate the clinical applications of AI, Cabral et al align with previous sentiments[6] on the need for rigorous testing of AI's effectiveness in real-world settings. Such research is essential to navigate the dual challenges of maximizing benefits in patient care while addressing clinical safety alongside potential biases and ethical dilemmas, highlighting ethical considerations such as patient privacy, data security, and explainable AI, a sentiment echoed elsewhere.[7-8]

Cabral et al conclude with a call for increased investment in AI research and the establishment of healthcare standards to guide AI implementation. This perspective is bolstered by the broader implications of AI's role in healthcare, as highlighted by Topol[2] and Jha and Topol,[9] who highlight the importance of collaborative efforts in overcoming integration barriers and ensuring AI's equitable and effective deployment across various medical specialties. Their insights suggest that by addressing these challenges collectively, the AI holds the potential to enhance patient care, improve diagnostic accuracy and optimise healthcare workflows.[2,9]

However, Cabral et al also acknowledge the study's limitations, including (a) the potential optimism bias of respondents, given their involvement in AI and cancer research, which could lead to more favourable views on the technology's potential, and (b) limited diversity of respondents, primarily consisting of researchers and academics, which might not fully capture the broader range of perspectives and concerns that exist among other stakeholders i.e. clinicians, policymakers, and importantly, patients, regarding AI's role in cancer care. Additionally, the low response rate, while not uncommon in survey research, introduces a potential for non-response bias and the focus on a specific demographic (primarily academia and research), which may limit the generalizability of the results.

In synthesizing these insights, the study showcases AI's promising future in oncology but also exposes the practical challenges and ethical considerations that must be navigated to realize its full potential. Drawing further insights from elsewhere,[8-9] there is a need to develop AI technologies with a foundational commitment to patient safety and quality care. This involves rigorous validation of AI against diverse datasets and comparisons with human judgment to minimize biases and errors, thereby ensuring these systems can perform at or above the level of experienced professionals. It is also critical to implement robust oversight mechanisms that monitor AI systems' performance in clinical settings. This effort must be complemented by continued empirical research and collaborative policymaking. Together, these strategies ensure that AI not only meets the current standards of care but also evolves responsively to enhance patient outcomes and healthcare quality continuously. Such efforts must prioritize the development of AI applications that not only enhance diagnostic accuracy and treatment efficacy but also actively contribute to advancing the overall quality and safety of healthcare, ensuring that every technological advancement directly benefits patients.

QUANTIFYING THE IMPACT OF ARTIFICIAL INTELLIGENCE RECOMMENDATIONS WITH EXPLANATIONS ON PRESCRIBING DECISION MAKING

Digital Medicine, 07 November 2023

Nagendran et al.'s recent study[10] investigates how AI-driven decision support recommendations affect prescribing decisions in intensive care settings, particularly focusing on sepsis resuscitation. Their research examines the role of additional information—either from human peers or AI. They also assess whether inclusion of 'explainable AI' influences clinicians' decision-making, where explainable AI aims to provide not only recommendations but also justification for those recommendations, a common demand from clinicians, AI researchers and regulators.

To assess the impact of different information sources on prescribing, the authors utilised a reinforcement-learning based AI Clinician system, trained on data from the Medical Information Mart for Intensive Care (MIMIC-III) database, covering 17,083 intensive care unit patients with sepsis. Using a modified between-subjects design, they involved 86 intensive care doctors who were presented with patient cases and asked to make dosing decisions for two drugs across 16 scenarios (i.e., trials). The first four trials served as the pre-training period and the subsequent 12 as the main experiment. The experimental setup comprised four arms: a control group (doctors asked to prescribe doses based on their judgment alone), a peer human recommendation group (doctors shown what doses had been prescribed by other doctors with similar patients), an AI suggestion group (doctors provided with AI Clinician system's suggested doses), and an AI suggestion with explanation ('XAI') group (doctors were shown the AI Clinician system's suggested doses along with an explanation based on simple feature importance).

The study found that additional explanation had a strong influence on prescription decisions, significantly for the AI suggestion group, but not for peers. Attitudes towards AI and clinical experience did not have a significant association with adherence to AI suggestions. This indicates that doctors may report that they found the explanation useful, but it may not actually have influenced their decision-making process. While this casts doubt on the usefulness of self-report as a robust metric for assessing explanations in clinical experts, it also highlights the need for alternative methodologies that assess the natural behaviour of clinicians when interacting with decision support tools in order to provide more objective and granular markers for assessing the usefulness of explanations. The study also highlights the need for further research to understand the design and deployment of AI-based medical decision support tools.

Nagendran et al's research highlights significant safety and quality implications for the integration of AI in clinical decision-making. By demonstrating that AI recommendations can shift prescribing behaviours more so than peer recommendations, it highlights the potential for AI to enhance patient treatment strategies. There is a translation gap in the real-world clinical environments with a few AI systems currently active, despite AI-driven Clinical Decision Support System (CDSS having the potential to have a major impact on medical care due to their theoretically superhuman performance.[10,11,12] The integration of AI into healthcare is further supported by its potential to refine clinical decision-making and patient outcomes through the convergence of human and artificial intelligence, as emphasized in recent discourses.[13,14] Furthermore, the study highlights the potential for AI to reduce variation in prescribing, particularly given how clinician prescribing decisions were influenced by whether the doses were recommended by the AI system compared to the baseline practices of the clinicians, (baseline was defined as the experimental condition in which clinicians were provided with only the patient data, without any additional information from peers or AI). When the AI dosing recommendation was higher than the baseline practices of the clinicians, the prescriptions of doctors in the AI arm were more variable across doctors. On the other hand, when the AI dosing recommendation was lower than the baseline, prescriptions were less variable. This suggests that AI systems can have a mixed impact on practice variation, indicating that a thoughtful approach is required in their deployment to address this variation and promote standardized care..

However, the study's findings that explanations do not additionally affect decisions challenge the perceived value of XAI, suggesting that the presence of AI, rather than its interpretability, is what influences clinician behaviour. This aligns with ongoing discourse on the necessity of developing interpretable machine learning models that healthcare professionals can trust and understand, underlining the challenges in balancing model complexity with interpretability,[15] while also highlighting the need for further research into how explanations and AI recommendations are presented to clinicians. In particular, the design and presentation of XAI systems need careful consideration to ensure their effectiveness in influencing clinical decisions and improving patient safety.

Moreover, the study did not find any correlation between clinicians' attitudes toward AI or their years of clinical experience and their adherence to AI suggestions. This implies that AI acceptance and adherence may not be influenced by individual factors such as attitudes or experience. Instead, the impact of AI on decision-making may be more dependent on the specific patient scenario and the quality of the AI recommendation itself.

The study acknowledges limitations, including the simplicity of the XAI modality used, and the scenarios' low fidelity, emphasizing the need for future research to develop more comprehensive XAI that aligns with clinicians' cognitive processes. Low sample size, and a lack of depth into the clinicians' decision-making rationale was also acknowledged. Nonetheless, this study contributes to the broader research on AI and XAI in healthcare by highlighting the nuanced impact of AI on clinical decision-making and the challenges in effectively integrating XAI.

The study recommends further research into developing more sophisticated, contextually relevant forms of XAI that align with clinicians' cognitive processes and decision-making needs. It highlights the importance of integrating AI tools in a manner that respects and enhances clinical judgment, ultimately aiming to improve patient outcomes through more informed and efficient treatment strategies. Additionally, the integration of AI tools in clinical settings must not only focus on technological capabilities but also on fostering an effective human-AI team dynamic, utilising transactive memory systems and open communication to enhance team effectiveness and decision-making.[16] It also raises ethical considerations, emphasizing the need for a combination of principles, practices, and governance structures to navigate the ethical complexities of AI deployment.[17,18] While this study suggests that AI recommendations can have a positive impact on service quality and patient safety by influencing prescribing decisions and reducing practice variation, the design and implementation of the XAI needs to be carefully tailored to ensure effectiveness and reliability in clinical practice.[10]

ASSURING THE SAFETY OF ARTIFICAL INTELLIGENCE-BASED CLINICAL DECISION SUPPORT SYSTEMS: A CASE STUDY OF THE 'AI CLINICIAN' FOR SEPSIS TREATMENT

BMJ Health & Care Informatics, 04 June 2022

Festor et al[19] delve into issue of ensuring the safety of AI-based clinical decision support systems (CDSS), with a specific focus on the AI Clinician, a reinforcement learning-based treatment recommendation system for sepsis treatment. They emphasize the need for systematic safety assurance before such systems gain clinical deployment and regulatory approval, particularly for those with increasing autonomy in decision-making.

Through the application of the Assurance of Machine Learning in Autonomous Systems (AMLAS) methodology, Festor et al aimed to systematically ensure the safety of the AI Clinician for sepsis treatment. This involved defining safety constraints to mitigate clinical hazards associated with sepsis resuscitation. Specifically, the AMLAS methodology is a safety assessment framework used to establish traceable links between system-level hazards, risks, and safety requirements in machine learning-based autonomous systems. It takes a whole system approach to safety assurance and aims to ensure that the safety requirements of machine learning components are satisfied. AMLAS is modular and iterative, allowing for the refinement of safety requirements and the evaluation of safety evidence throughout the development and deployment lifecycle of the system. It complements other safety assessment methodologies and can be applied to various domains, including healthcare. While robust, it requires significant interdisciplinary collaboration, highlighting the importance of integrating technical, ethical, clinical and regulatory expertise in the development and implementation of AI-based CDSS.[20,21]

Utilizing a subset of the Medical Information Mart for Intensive Care (MIMIC-III) database,[22] a freely accessible critical care database with data from over 40,000 patients (of which 17,083 were patients with sepsis), Festor et al analyzed intensive care unit data on patients with sepsis. The research team, comprising clinical experts and safety engineers, identified four clinical hazards in sepsis resuscitation and outlined unsafe scenarios to limit the AI agent's action space (i.e., the range of possible actions that can be taken by the AI Clinician or human clinicians in the treatment of sepsis, such as for example, the administration of fluids and vasopressors to patients with sepsis). Based on retrospective data analysis, they compared the frequency of hazardous decisions made by the AI Clinician against those made by human clinicians for each scenario. They used a z-test to test the null hypothesis that the underlying Bernoulli distribution parameters were equal for human clinicians and the AI Clinician. The z-test statistic was calculated using the observed proportions of unsafe decisions and the sample size. The use of Bernoulli random variables was important because these allow quantification and comparison of the likelihood of different outcomes. In this case, they were used to compare the proportion of unsafe decisions made by human clinicians and the AI Clinician in each scenario.

The AI Clinician demonstrated a statistically lower frequency of hazardous decisions compared to human clinicians in three of four predefined scenarios. This means that in these scenarios, the AI Clinician made safer treatment recommendations than human clinicians. Additionally, by modifying the AI's reward function to satisfy safety constraints and retraining the AI Clinician, the study demonstrated enhanced safety without negatively affecting model performance. Specifically, the authors made changes to the way the AI Clinician was rewarded for its actions, penalizing instances where harmful decisions were taken by clinicians. They then retrained the AI model using these modified rewards, resulting in an AI Clinician that was 12% less likely to suggest hazardous decisions than human clinicians, demonstrating an improvement in safety and highlighting AI's potential to reduce unsafe clinical decisions and possibly improve patient outcomes.

The study points to the importance of integrating safety considerations early in the AI design process to proactively generate safety evidence for CDSS, in line with current discourse.[5] This proactive approach is essential in a field where the consequences of errors can be lifethreatening. By embedding safety into the design process, AI developers can anticipate and mitigate potential risks before they manifest in clinical settings.[23]

However, the study acknowledges specific limitations, such as the challenge of defining safety constraints in the absence of clinical consensus and their reliance on retrospective data for training the AI. These point to a broader issue in healthcare AI: the need for developing standardized frameworks for evaluating and ensuring the safety of AI applications; and the importance of careful consideration and validation of AI models before clinical implementation.[24,25,26] These limitations highlight the need for ongoing collaboration between AI developers and clinical practitioners to refine AI-based CDSS.

While the study marks a significant step towards assuring the safety of AI-based CDSS for sepsis treatment, it also raises questions about the implementation of such systems in real-world clinical settings. Future research should address these challenges, exploring methods to refine safety constraints further, integrating more clinical considerations into the AI learning process and evaluating the impact of AI-based recommendations on patient outcomes.[27] This future research is vital for moving beyond theoretical safety assurances to practical, measurable improvements in patient care, crucial for bridging the gap between research and practice.[27,28] Implementing and assessing AI-based recommendations in live clinical environments will be crucial in ensuring that AI technologies can safely and effectively support healthcare providers.[28]

The study’s methodical approach to assessing and improving the safety of AI systems in a clinical setting, through integration of safety constraints early in the AI design process and the subsequent modification of the AI model’s reward function, represents significant contributions to regulatory practices and the development of safety assurance methodologies for AI in healthcare. The insights provided could be valuable across various healthcare applications, demonstrating a concrete example of AI’s potential impact and aiding the understanding of the complexities involved in implementing AI in healthcare safely.[19,21,23,28] This nuanced perspective on the integration of AI into healthcare highlights the essential balance between innovation and patient safety, advocating for a future where AI technologies support healthcare providers in delivering safer, high quality care.

ADDITIONAL INFORMATION

Funding

Infrastructure support for this research for TS was provided by the NIHR London Medtech and In vitro diagnostic Co-operative. BDF is supported by supported by the National Institute for Health Research (NIHR) North West London Patient Safety Research Collaboration. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Competing interests

TS received funding from Cancer Alliances and NHS England for training MDTs in assessment and quality improvement methods in the United Kingdom; and honoraria for public speaking from Parsek, and consultancy fees from Roche Diagnostics, Parsek and Salutare.

Ethics Approval

Not applicable.

Provenance and peer review

Commissioned; internally peer reviewed.

ORCID ID

Tayana Soukup 0000-0003-0203-7264

Bryony Dean Franklin 0000-0002-2892-1245

REFERENCES

1. Cabral BP, Braga LAM, Syed-Abdul S, Mota FB. Future of Artificial Intelligence Applications in Cancer Care: A Global Cross-Sectional Survey of Researchers. Curr Oncol. 2023;30:3432–46.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019 Jan;25(1):44-56.
3. Parikh RB, Obermeyer Z, Navathe AS. Regulation of predictive analytics in medicine. Science. 2019 Feb;363(6429):810-812.
4. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. A guide to deep learning in healthcare. Nat Med. 2019 Jan;25(1):24-29.
5. Rumbold JMM, Pierscionek B. The effect of the General Data Protection Regulation on medical research. J Med Internet Res. 2017 Feb;19(2):e47.
6. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019 Apr;380(14):1347-1358.
7. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. PLoS Med. 2018 Nov;15(11):e1002689.
8. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf. 2019;28(3):231-7.
9. Jha AK, Topol EJ. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. JAMA. 2016 Dec;316(22):2353-2354.
10. Nagendran M, Festor P, Komorowski M, Gordon AC, Faisal AA. Quantifying the impact of AI recommendations with explanations on prescription decision making. npj Digital Med. 2023;6:206. doi:10.1038/s41746-023-00955-z.
11. van de Sande, D., van Genderen, M. E., Huiskens, J., Gommers, D. & van Bommel, J. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. Intensive Care Med. 47, 750–760 (2021).
12. Olaye, I. M. & Seixas, A. A. The Gap Between AI and Bedside: Participatory Workshop on the Barriers to the Integration, Translation, and Adoption of Digital Health Care and AI Startup Technology Into Clinical Practice. J. Med. Internet Res. e32962 (2023).
13. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA. 2018;320(21):2199-2200.

14. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56.
15. Rudin C, Ustun B. Optimized risk scores. J Mach Learn Res. 2019;20(150):1-75.
16. Bienefeld N, Kolbe M, Camen G, Huser D, Buehler PK. Human-AI teaming: leveraging transactive memory and speaking up for enhanced team effectiveness. Front Psychol. 2023;14:1208019.
17. Char DS, Shah NH, Magnus D. Implementing machine learning in health care — addressing ethical challenges. N Engl J Med. 2018;378(11):981-983.
18. Mittelstadt B. Principles alone cannot guarantee ethical AI. Nat Mach Intell. 2019;1(11):501-507.
19. Festor P, Jia Y, Gordon AC, Faisal AA, Habli I, Komorowski M. Assuring the safety of AI-based clinical decision support systems: a case study of the AI Clinician for sepsis treatment. BMJ Health Care Inform. 2022;29(1):e100549.
20. Picardi C, Hawkins R, Paterson C, Habli I, Lawton T, Porter Z. Assurance of Machine Learning for Use in Autonomous Systems: A Survey. IEEE Access. 2020;8:111326-111345.
21. Kusters R, Misevic D, Berry H, Cully A, Cunff YL, Dandoy L, Díaz-Rodríguez N, Ficher M, Grizou J, Othmani A, Palpanas T, Komorowski M, Loiseau P, Moulin Frier C, Nanini S, Quercia D, Sebag M, Soulié Fogelman F, Taleb S, Tupikina L, Sahu V, Vie JJ, Wehbi F. Interdisciplinary Research in Artificial Intelligence: Challenges and Opportunities. Front Big Data. 2020;3:577974. DOI: 10.3389/fdata.2020.577974.
22. Johnson AEW, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. MIMIC-III, a freely accessible critical care database. Sci Data. 2016 May 24;3:160035.
23. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310.
24. Wang Y, Li N, Chen L, Wu M, Meng S, Dai Z, Zhang Y, Clarke M. Guidelines, Consensus Statements, and Standards for the Use of Artificial Intelligence in Medicine: Systematic Review. J Med Internet Res. 2023;25:e46089. DOI: 10.2196/46089.

25. Crossnohere NL, Elsaid M, Paskett J, Bose-Brill S, Bridges JF. Guidelines for Artificial Intelligence in Medicine: Literature Review and Content Analysis of Frameworks. J Med Internet Res. 2022;24(8):e36823. DOI: 10.2196/36823
26. Chen JH, Asch SM. Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations. N Engl J Med. 2017;376(26):2507-2509.
27. Prosperi M, Min JS, Bian J, Modave F. Big data hurdles in precision medicine and precision public health. BMC Med Inform Decis Mak. 2018;18(1):139.
28. Obermeyer Z, Emanuel EJ. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016;375(13):1216-1219.