https://doi.org/10.1038/s41591-023-02367-8

Patients should be informed when AI systems are used in clinical trials

Artificial intelligence (AI) systems are increasingly being investigated in clinical trials. Trials that use AI must be held to the same ethical standards for risk assessment and disclosure as all human participant studies. All clinical AI systems, especially those under active investigation, have new risks, including human–machine interactions, interpretability and data limitations. The full magnitude and scope of these risks is not yet known because clinical AI integration is still in its infancy. We argue that in light of these risks and the uncertainty therein, disclosure is a minimal standard when patients' data are being used in an AI clinical trial that may affect clinical decisions, even if written informed consent is not required.

The US Food and Drug Administration (FDA) classifies AI systems as 'Software as a Medical Device' (SaMD), with a lineage from simpler decision support tools such as risk calculators[1,2]. However, AI algorithms have unique risk considerations (Table 1), and the full effects of emergent risks remain incompletely understood.

Some AI systems can make human-like predictions and recommendations, and an important considerations for risk assessment is therefore the extent of clinical oversight. At present, most AI systems are implemented in decision-support capacities, in which a human always makes the final clinical decision[3]. Implementation as decision-support may mitigate risk by requiring a human check, but this introduces human–machine interaction issues. For example, overreliance on the AI recommendation may lead to clinician de-skilling and inattention to AI errors, which could affect the performance of the human–machine system in unknown ways[4]. Another challenge to the ability of human oversight to mitigate risks is that AI systems are often not transparent or interpretable. Developers and clinicians are not always privy to how AI-generated outputs are calculated and so a clinician-user would not be able to explain outputs to a patient or be able to easily identify or understand AI errors[5].

Table 1 | AI system clinical trials

Potential risks	Key information for disclosure
Clinical oversight	• Clear declaration of whether the human or model will make the final clinical decision • Contingency plans if model cannot be accessed or unexpectedly fails
Human–machine interaction	• Explanation that the interaction between humans and AI models is new and may introduce unforeseen risk • Whether the clinician will base their decision only on the model, or on the model in addition to other clinical factors • Whether, and how, patients will receive the predictions of the model • Whether the clinician can understand how the model arrived at its prediction • Whether measures of confidence will be provided along with predictions
Data limitations and algorithmic bias	• The population that the model was developed on • Whether it is reasonable to expect that the model will perform similarly in the trial population • Key clinical and population-level implications of biased predictions
Performance shifts	• Whether the model is fixed or continually learning • Plans to communicate observed changes in performance with patients • Plans to stop the trial early in case of performance deterioration
Data security and privacy	• Whether the model will be updated with patients' data during the trial • Where and with whom patient data will be shared • Where and with whom the model will be shared • Whether patient data will be de-identified • Security protections for patient data and model sharing/storage

AI systems are inherently reliant on the data used to train them, which creates new challenges for risk assessment. The provenance and quality of the training data used to develop an AI model are arguably the most important determinants of system performance, yet training and validation datasets do not equally or accurately reflect all populations[6]. As a result, the ability to use existing evidence to extrapolate risk is imperfect if the clinical population of the trial differs from those in previous studies of the system. Relatedly, a consequence of the data limitations of AI systems is that its use might perpetuate racism, sexism, ageism, and/or other biases[7]. Inappropriate development and application of AI models may inadvertently worsen disparities, amplifying the population-level risk of such systems.

AI systems have complex patient privacy and security risk considerations, because they often entail sharing data and predictions at unprecedented scale. During an AI clinical trial, patient data may be transferred between one or more institutions where participants receive care, as well as a different institution that houses the AI system. Data security of the AI system itself, each institution involved in sharing and/or storing patient data, and data transfer practices all impact the risk of unintentional or nefarious threats to patient privacy and security.

A final uncertainty for risk assessment is that system performance may change over time owing to performance drift, shifting the risk-to-benefit ratio. Performance drift is when data, environments, and/or clinical paradigms change from those used for model development, worsening performance compared with initially reported results[8]. Drift is expected to occur in most models and could lead to unintended harms if not rapidly identified[8]. However, the best methods to avoid these harms, such as ongoing drift monitoring and correction in operationalized systems, are not yet known.

Studies of AI systems should follow existing standards for disclosure requirements, which are based on the risk to participants. In general, research involving human participants requires prospective informed consent

to protect patients' right to autonomy. However, revised federal regulations, including changes to the Common Rule and the passage of the 21st Century Cures Act, allow for waiver of prospective informed consent for selected studies[9]. These waivers are contingent on several conditions, including that: participation poses minimal risk to patients; investigations could not otherwise be carried out; waiver or alteration of informed consent does not adversely affect the rights or welfare of participants; and additional pertinent information is provided to participants when appropriate.

Risks associated with AI systems can vary widely and depend on the severity of the medical condition it is designed for, how its output is used to inform care, the patient population it is intended to be used in, and the existing level of evidence for the system[3]. Risk determination, and therefore whether an informed consent waiver is appropriate, will need to be on a case-by-case basis for trials of AI systems.

Many AI trials will obtain a waiver of informed consent, such as those that are non-interventional, use pre-existing data, and do not make suggestions beyond standard-of-care. Furthermore, a strength of many AI trials is that they can include huge numbers of patients at a scale that may not be feasible with informed consent. When AI trials obtain a waiver of informed consent, patients may be unaware that they are in an AI trial that affects their care. Indeed, in an FDA patient engagement meeting, patient advocates expressed a desire to be notified about any use of an AI product in their care[10]. Researchers carrying out a minimal risk trial of an AI system still have ethical and regulatory duty to notify participants when appropriate. Given this desire from patients, coupled with the known and unknown risks of AI, use of AI systems that inform clinical actions should be disclosed, even if the overall risk is minimal.

Participation in an AI clinical trial without notification can infringe on patients' right to self-determine who and what is involved in their care. At present, it is reasonable for patients to assume that only humans, not AI systems, are involved in making their healthcare decisions. However, many AI systems can now make human-like decisions that patients may reasonably expect to be made by clinicians. Disclosure as a minimal standard ensures patients retain their ability to determine how, and by whom, their healthcare decisions are made.

Broad disclosures may be sufficient for AI trials that share risk consideration, and patients should be able to easily access more detailed information about the investigational AI systems used as part of their care. Disclosure as a minimal standard for AI trials will enable self-determination and engender trust, two key ingredients to the successful and safe integration of AI systems into healthcare.

Subha Perni[1,2,3], **Lisa Soleymani Lehmann**[2,4] & **Danielle S. Bitterman**[1,2] ✉

[1]Artificial Intelligence in Medicine Program, Mass General Brigham, Harvard Medical School, Boston, MA, USA. [2]Department of Radiation Oncology, Brigham and Women's Hospital/Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [3]MD Anderson Cancer Center, Houston, TX, USA. [4]Harvard T.H. Chan School of Public Health, Harvard Medical School, Boston, MA, USA.

✉e-mail: Danielle_Bitterman@dfci.harvard.edu

Published online: 23 May 2023

References

1. US Food and Drug Administration. https://go.nature.com/41xD0em (2022).
2. US Food and Drug Administration. https://go.nature.com/3zefINL (2019).
3. Bitterman, D. S. et al. *Lancet Digit. Health* **2**, e447–e449 (2020).
4. He, J. et al. *Nat. Med.* **25**, 30–36 (2019).
5. Reyes, M. et al. *Radiol. Artif. Intell.* **2**, e190043 (2020).
6. Kaushal, A. et al. *JAMA* **324**, 1212–1213 (2020).
7. Panch, T. et al. *J. Glob. Health* **9**, 010318 (2019).
8. Finlayson, S. G. et al. *N. Engl. J. Med.* **385**, 283–286 (2021).
9. US Department of Health and Human Services. 45 CFR 46 https://go.nature.com/3yl1Rq4 (2018).
10. The Pew Charitable Trusts. https://go.nature.com/3I4ZuwJ (2021).

Competing interests

D.S.B. is an associate editor of *Radiation Oncology*, HemOnc.org, unrelated to this work, for which she receives no financial compensation; and funding from American Association for Cancer Research, unrelated to this work.