# Toward Transparency: Implications and Future Directions of Artificial Intelligence Prediction Model Reporting in Healthcare

Ryan Turlip, Ali K Ozturk, Jang Won Yoon, William C Welch, Mert Marcel Dagli

Submitted to: JMIR AI
on: April 26, 2024

# *Table of Contents*

# Toward Transparency: Implications and Future Directions of Artificial Intelligence Prediction Model Reporting in Healthcare

Ryan Turlip[1]; Ali K Ozturk[1] MD; Jang Won Yoon[1] MD, MSc, FAANS; William C Welch[1] MD; Mert Marcel Dagli[1] MD

[1]Department of Neurosurgery Perelman School of Medicine University of Pennsylvania Philadelphia US

**Corresponding Author:**
Mert Marcel Dagli MD
Department of Neurosurgery
Perelman School of Medicine
University of Pennsylvania
801 Spruce Street
Philadelphia
US

## *Abstract*

The rapid integration of Artificial Intelligence (AI) in healthcare emphasizes the transformative potential it holds for improving patient outcomes through data-driven decision-making. There is a drive toward implementing more complex predictive algorithms for disease diagnosis and prognosis. However, there are unique implications of AI that limit its clinical applicability and validity. To address these challenges, the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) group released an extension and update of its 2015 reporting guideline, the TRIPOD+AI, to enhance transparency and methodological rigor in regression and AI prediction model studies. The TRIPOD+AI framework encompasses an expanded scope, incorporating new domains such as fairness, open scientific practices, and patient and public engagement. It is anticipated that these augmented guidelines will facilitate the rigorous evaluation and subsequent adoption of artificial intelligence tools across diverse healthcare settings.

(JMIR Preprints 26/04/2024:59947)
DOI: https://doi.org/10.2196/preprints.59947



### Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

# Original Manuscript

# Title: Toward Transparency: Implications and Future Directions of Artificial Intelligence Prediction Model Reporting in Healthcare

**Authors:** Turlip, Ryan[1]; Ozturk, Ali K MD[1]; Yoon, Jang W MD[1]; Welch, William C MD[1]; Dagli, Mert Marcel MD[1]

**Author Affiliations:**

1. Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, US

**Corresponding Author:**

Mert Marcel Dagli, MD

Department of Neurosurgery

Perelman School of Medicine, University of Pennsylvania

801 Spruce Street, Philadelphia, 19107 PA, USA

Email: marcel.dagli@pennmedicine.upenn.edu

Phone: (267) 230-6493

Fax: (215) 615-3701

Keywords: artificial intelligence; machine learning; neurosurgery; transparent reporting; TRIPOD; prediction model

Abstract word count: 106

Manuscript text word count: 877

Number of references: 5

Number of tables and/or figures: 0

Number of videos: 0

Previous presentations: Nil

Financial and competing interests disclosure: The authors declare no conflict of interests.

Funding: none

**Abstract**

The rapid integration of Artificial Intelligence (AI) in healthcare emphasizes the transformative potential it holds for improving patient outcomes through data-driven decision-making. There is a drive toward implementing more complex predictive algorithms for disease diagnosis and prognosis. However, there are unique implications of AI that limit its clinical applicability and validity. To address these challenges, the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) group released an extension and update of its 2015 reporting guideline, the TRIPOD+AI, to enhance transparency and methodological rigor in regression and AI prediction model studies. The TRIPOD+AI framework encompasses an expanded scope, incorporating new domains such as fairness, open scientific practices, and patient and public engagement. It is anticipated that these augmented guidelines will facilitate the rigorous evaluation and subsequent adoption of artificial intelligence tools across diverse healthcare settings.

## Introduction

The integration of Artificial Intelligence (AI) within healthcare represents a transformative paradigm shift, ushering an era of unprecedented progress in healthcare decision making and data-driven analytics to improve patient outcomes. This advancement is significantly driven by machine learning (ML), a subset of AI where algorithms learn from data to develop predictions.[1] Traditional ML statistical models, such as regression, predict outcomes by discerning relationships between independent predictor variables and dependent outcome variables of interest, while adjusting for confounders. ML offers many additional models, such as Support Vector Machines (SVM), Ensemble-Based Methods (EBM), and Artificial Neural Networks (ANN).[1] In healthcare, the shift toward more complex ML algorithms for nuanced datasets holds promise in enhancing the predictive capabilities for patient outcomes. However, the rapid adoption of AI prediction models has outpaced the development of proper clinical and research guidelines, raising concerns about reliability, validity, reproducibility, data security, and potential biases. Addressing these challenges is crucial to ensure the effective and trustworthy integration of AI tools into clinical practice. Hence, in this perspective, we highlight the unique implications and challenges posed by AI prediction models and explore future directions for reporting guidelines tailored to AI in healthcare.

## Implications and Challenges of AI Prediction Models

AI prediction models have unique features such as opaqueness, validation frameworks, and clinical applicability that make creating clinical and research guidelines challenging.[2,3] Opaqueness, often referred to as the "black box" problem, arises when the decision-making process of AI algorithms is not transparent or interpretable to users.[1,2] The complexity of how these models derive their predictions can hinder clinicians' ability to trust and effectively integrate AI recommendations into patient care. Addressing opaqueness is crucial for ensuring that AI tools are not only technically sophisticated but also clinically relevant and understandable. Mitigating this challenge involves

creating interpretable machine learning models within the data analytics pipeline, beyond merely showcasing the top-performing model. This approach advocates for presenting a range of models, from explainable to complex, where the simpler models offer insights into the significance of predictors. Such insights contribute to the scientific body of knowledge, even when a more complex model, due to its superior performance, might be preferred in practical applications.[4]

External validation frameworks are critical in assessing generalizability and reliability.[2,3] These frameworks test AI algorithms on independent data sets that were not used during the model's training phase, offering a rigorous evaluation of the model's performance in real-world scenarios. The importance of external validation lies in its ability to expose and mitigate overfitting, where a model performs exceptionally well on its training data but poorly on unseen data. By employing these frameworks, researchers and clinicians can ensure that AI tools maintain their predictive accuracy and clinical relevance across diverse patient populations and healthcare settings. Despite the importance of external validation, such studies are rare and often hindered by inadequate reporting and data sharing at the time of publication and thereafter.

The clinical applicability of AI in healthcare hinges on the seamless integration of AI models into the existing healthcare infrastructure, ensuring that these technologies can be effectively utilized in real-world patient care settings. This entails not only the technical compatibility of AI systems with clinical workflows but also the models' ability to produce actionable insights relevant to patient-specific conditions and treatment plans. Achieving clinical applicability requires rigorous testing and validation to confirm that AI tools are reliable, accurate, and enhance decision-making processes. Furthermore, it necessitates collaboration between engineers, data scientists, and clinicians to tailor AI solutions to the nuanced demands of healthcare. For example, electronic health records have already started integrating natural language processing and AI prediction models into their systems to

enhance patient care and clinical decision-making.[3]

**Future Directions of AI Prediction Model Reporting in Healthcare**

Although widely adopted, AI models lack standardization and rigorous reporting practices, compromising their reliability and validity. To address these deficiencies, the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement was introduced in 2015 and offered a 22-item checklist to aid in transparent reporting of prediction model development, validation, and updating.[5] In 2021, the TRIPOD statement then published a protocol for the development of an extended reporting guideline (TRIPOD+AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies that applied ML techniques.[4] Acknowledging the pressing need for robust guidelines in the rapidly evolving landscape of AI in healthcare, the recently published TRIPOD+AI guidelines, now a 27-item checklist, represent a significant advancement in the standardization of AI prediction models in healthcare.[6]

Designed to ensure transparency, validity, and utility in AI prediction studies, TRIPOD+AI provides an extended structural framework that aids academic institutions, researchers, journal editors, peer reviewers, funders, patients, public, and study participants, systematic reviewers and meta-researchers, policy makers, regulators, technology and medical device manufacturers, and healthcare professionals in evaluating AI prediction studies more rigorously.[6] It particularly stresses the importance of fairness, the promotion of open scientific practices, and the engagement of the public and patients in the research process, ensuring that these models serve a broad and diverse population effectively.

These guidelines are designed to be applicable across a wide range of healthcare settings, including public health, primary care, and nursing homes. They cater to both prognostic and diagnostic models,

addressing the unique challenges posed by AI in healthcare and ensuring that the benefits of AI are accessible across various medical and patient contexts.[4] While the TRIPOD+AI guidelines did not specifically consider large language models, such the increasingly prominent ChatGPT, during its initial development, focusing instead on non-generative models, their principles remain relevant and translatable, and can substantially advance transparency in the development and evaluation of generative AI within healthcare.[4,6] To ensure these guidelines continue to be applicable amidst the rapid evolution of AI technologies, it will be imperative to periodically update them, taking into account the latest advancements, including those in generative modeling.

Despite the accessibility of the TRIPOD statement and evidence of improved reporting in a pre-post analysis, substantial deficiencies in reporting standards for multivariable prediction model studies remain prevalent and need to be acknowledged. As the TRIPOD+AI statement gains traction, however, its widespread adoption holds the promise of bolstering methodological standards in regression and AI prediction studies, thereby fostering greater reliability, reproducibility, and ultimately, improved patient outcomes.[6]

## References


1. Dagli MM, Rajesh A, Asaad M, Butler CE. The use of artificial intelligence and machine learning in surgery: a comprehensive literature review. *The American Surgeon™*. 2023;89(5):1980-1988.
2. Durán JM, Jongsma KR. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. *Journal of Medical Ethics*. 2021;47(5):329-335.
3. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine*. 2019;17:1-9.
4. Collins GS, Dhiman P, Navarro CLA, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ open*. 2021;11(7):e048008.
5. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. *Circulation*. 2015;131(2):211-219.
6. Collins GS, Moons KG, Dhiman P, et al. TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. bmj. 2024;385