Journal of the American Medical Informatics Association, 28(6), 2021, 1353–1355
doi: 10.1093/jamia/ocab035
Advance Access Publication Date: 1 March 2021
Perspective

Perspective

Conflicting information from the Food and Drug Administration: Missed opportunity to lead standards for safe and effective medical artificial intelligence solutions

Tina Hernandez-Boussard [1], Matthew P. Lundgren [2] and Nigam Shah [1]

[1]Department of Medicine, Stanford University, Stanford, California, USA, and [2]Department of Radiology, Stanford University, Stanford, California, USA

Corresponding Author: Tina Hernandez-Boussard, Department of Medicine, Stanford University Medical School, 1265 Welch Road #245, Stanford, CA 94305, USA (boussard@stanford.edu)

Received 1 February 2021; Editorial Decision 7 February 2021; Accepted 10 February 2021

ABSTRACT

The Food & Drug Administration (FDA) is considering the permanent exemption of premarket notification requirements for several Class I and II medical device products, including several artificial Intelligence (AI)–driven devices. The exemption is based on the need to rapidly more quickly disseminate devices to the public, estimated cost-savings, a lack of documented adverse events reported to the FDA's database. However, this ignores emerging issues related to AI-based devices, including utility, reproducibility and bias that may not only affect an individual but entire populations. We urge the FDA to reinforce the messaging on safety and effectiveness regulations of AI-based Software as a Medical Device products to better promote fair AI-driven clinical decision tools and for preventing harm to the patients we serve.

Key words: artificial intelligence, bias, regulation, clinical decision support, reporting standards

FOOD AND DRUG ADMINISTRATION REGULATORY ENVIRONMENT

At the start of the COVID-19 (coronavirus disease 2019) Public Health Emergency, the Food and Drug Administration (FDA) Center for Devices and Radiological Health (CDRH) temporarily relaxed several regulatory requirements to enable rapid access to COVID-related devices, including premarket notification requirements for software intended for medical purposes, known as Software as a Medical Device (SaMD) products. SaMD products are increasingly driven by artificial intelligence (AI) and are designed to treat or diagnose, drive clinical management, or inform clinical management for nonserious, serious, and critical healthcare conditions. SaMDs fall into 4 risk categories based on the product's use-case definition, with Category I being the lowest risk and Category IV being the highest impact, for example, an SaMD that performs diagnostic image analysis for making treatment decisions in patients with acute stroke.[1]

On January 12, 2021, the CDRH permanently exempted 7 Category I SaMD products from premarket regulatory review and proposed to exempt an additional 83 Class II products from premarket review.[2] There are 2 forms of FDA premarket review, an approval for high-risk categories (similar to that for a new drug) and a premarket notification or 501(k) in which equivalence to a similar, already marketed, product can be established. Among the 83 Class II products, several are AI driven, including monitoring devices and image analyses products. The exemption means that developers of these devices will no longer be required to provide reasonable assurance of safety and effectiveness of these devices prior to marketing. This decision is based on the need to rapidly disseminate these devices to the public, estimated cost savings, and the "complete lack of or de minimis number" of adverse events related to these devices reported to the FDA's MAUDE (Manufacturer and User Facility Device Experience) database since the start of the Public Health

Emergency. The exemptions provided for and proposed under this Notice for these 91 device classes could eliminate anywhere from $9.1 to $364 million in startup costs if there were one new entrant into each device market.

USEFUL AI-DRIVEN SAMDS

While MAUDE is a valuable source of information related to person-level adverse events, this system provides limited information on the utility, reliability, and biases of a product, as noted by the FDA (https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm). This limitation is particularly concerning for AI-based SaMDs, which may present population-level harm if the product has deficits in any of these categories. Useful AI products are those that lead to a favorable change in clinical decision making, as measured by improved patient outcomes or lower costs, consistently and across all populations to which they are applied. However, emerging evidence suggests that many AI-based SaMDs may not useful, reliable, or fair across all populations and may instead perpetuate biases in historically disadvantaged groups.[3] Such biases have been identified for models that use only medical imaging data. Therefore, a permanent exemption from premarket evaluation for AI-based SaMDs would potentially undermine the safety and effectiveness of upcoming products for several population subgroups.

A NEED FOR STANDARDS

The need for standardized evaluations of AI-driven SaMDs has led to an explosion of reporting guidelines and practice recommendations. Reporting guidelines, such as MINIMAR (Minimum Information for Medical AI Reporting),[4] propose a checklist of items to provide information on cohort selection and training data, model development, and performance, as well as data processing procedures. Other guidelines focus on end user needs and external validation including practice recommendations that suggest processes for model training and evaluation, including external validation and subgroup evaluation. There are also recommendations for examining models to understand the basis of their predictions and for evaluating their downstream impact and utility. These recommendations note the undue focus on technical performance metrics and a lack of consensus around best practices for model fairness, predictability, repeatability, explainability, and evaluations of risks and benefits of model use.[5] Efforts beyond checklists call for randomized controlled trials for models as well as guidelines for such trials bridging the CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) and SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence) proposals. There are calls for reporting, similar to that for drug adverse events and postmarket surveillance, including allowing for model recalls, as well as proposals for regulatory guidance prior to clinical use. The medical informatics community at large is grasping for an overarching framework to synthesize these proposed standards and promote the safe and effective design, development, and deployment of AI-based SaMDs as well as other algorithmic innovations to assist screening, diagnosis, or prognosis.

In early January, the CDHR Digital Health Center of Excellence division at the FDA released a report calling for a more thorough review of SaMDs than currently in place per the AI-based SaMD Action Plan,[6] effectively contradicting the proposed relaxed regulations for SaMDs. The Action Plan arose from stakeholder feedback, on a 2019 FDA white paper proposing regulatory frameworks for SaMDs, calling for a strict regulatory framework covering the total SaMDs product life cycle with 5 actions: guidance for regulating "learning" machine learning algorithms, support for Good Machine Learning Practice, support for patient-centered devices and transparency, development of methods to improve machine learning algorithms, and support for real-world monitoring to address bias and fairness. In the same vein, the American Medical Informatics Association recently released a position paper on AI-based SaMDs regulation calling for "new, coordinated initiatives and oversight" to both establish and support the expanding use of adaptive clinical decision support (as part of AI-based SaMD products) at point of care.[7] The crux of the proposed frameworks centers on thorough, continuous evaluation and monitoring of SaMDs to ensure that the safety and effectiveness of a deployed device is maintained over time so that marketed products are equitable across populations and remain robust against shifts in clinical practice, target populations, and dataset drifts over time.

A CALL TO ACTION

The exemption of the FDA premarket evaluation for AI-based SaMDs is counter to the multiple calls from the community for more guidance as well as oversight on the evaluation and monitoring of AI-based SaMD. Therefore, we urge the FDA to provide consistent and coordinated guidance in the evaluations of AI-based SaMD products going on the market. If an unregulated product scales across healthcare systems, that may harm not only individuals, but also entire populations and can further perpetuate biases in already vulnerable populations.[8] Without systematic guidance on reporting of the safety and effectiveness of these devices—beyond MAUDE—it is not clear how care providers should evaluate the utility and applicability of a product to a particular population prior to use. If this obligation is at the manufacturer's discretion to disclose, how useful and comprehensive can we expect this information to be?

Adoption of new AI solutions requires trust across stakeholders, including patients, as well as clear standards for evaluating medical AI solutions. First calling for more rigorous evaluations of AI-based SaMDs and then, less than 1 week later, suggesting permanent exemption from premarket notifications for many products seems counterproductive. We urge the FDA to clarify their messaging on safety and effectiveness regulations of AI-based SaMDs—clear guidance is necessary to promote fair AI-driven clinical decision tools and for preventing harm to the patients we serve.

AUTHOR CONTRIBUTIONS

All authors equally contributed to the concept and writing of the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to declare.

DATA AVAILABILITY

There are no new data associated with this article.

REFERENCES

1. U.S. Food and Drug Administration. "Proposed Regulatory Framework for Modifications to Artificial Intelligence." Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD); 2019.
2. U.S. Department of Health and Human Services, U.S. Food and Drug Administration. 86 FR 4088 - Making Permanent Regulatory Flexibilities Provided During the COVID-19 Public Health Emergency by Exempting Certain Medical Devices From Premarket Notification Requirements; Request for Information, Research, Analysis, and Public Comment on Opportunities for Further Science and Evidence-Based Reform of Section 510(k) Program. *Fed Regist* 2021; 86 (10): 4088–98.
3. Pierson E, Cutler DM, Leskovec J, Mullainathan S, Obermeyer Z. An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nat Med* 2021; 27 (1): 136–40. doi:10.1038/s41591-020-01192-7
4. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. *J Am Med Inform Assoc* 2020; 27 (12): 2011–5.
5. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform* 2021; 113: 103655.
6. U.S. Food and Drug Administration. *Artificial Intelligence/Machine Learning (AI/ML) Software as a Medical Device Action Plan*. Washington, DC: U.S. Department of Health and Human Services; 2021.
7. Petersen C, Smith J, Freimuth RR, *et al*. Recommendations for the safe, effective use of adaptive CDS in the US healthcare system: an AMIA position paper. *J Am Med Inform Assoc* 2021. doi: 10.1093/jamia/ocaa319.
8. Röösli E, Rice B, Hernandez-Boussard T. Bias at Warp Speed: How AI may Contribute to the Disparities Gap in the Time of COVID-19. *J Am Med Inform Assoc* 2021; 28 (1): 190–2.