Journal of Law and Health

Volume 35 | Issue 3 Article

10-15-2022

Machine Learning-Based Medical Devices: the FDA's Regulation, Requirements, and Restrictions

Charli Beam

Machine Learning-Based Medical Devices: the FDA's Regulation, Requirements, and Restrictions

Charli Beam[1]

[1] Thanks to my professor, Michael Froomkin, for his exceptional patience and guidance throughout the process of writing this paper, and to my dad for his love and support.

TABLE OF CONTENTS

I. INTRODUCTION

The FDA should develop regulations that require evaluation of and increase transparency about potential bias in medical machine learning. Machine learning (ML) relies on data sets, which may have hidden biases, resulting in medical devices developed with ML algorithms having a risk of faulty results. All machine learning has a potential bias issue when using biased data sets. Medical devices using machine learning are no exception. Underrepresentation or bias in medical data is a problem because "diverse participation [in clinical trials and biobanking for medical research] is necessary to identify the most effective treatments in different groups."[2] The FDA should address these potential bias issues by requiring machine learning medical device manufacturers to disclose the demographics of the data sets that trained the algorithm to the FDA. This includes disclosure of underrepresented populations in the data or populations the data set did not include at all. The FDA should use its authority to require manufacturers to monitor and evaluate the safety, effectiveness, and reliability of the device after approval, including any negative effect on sub-populations.[3] The FDA should also increase support for regulatory science efforts focused on developing methods to evaluate ML-based medical software and eliminating algorithmic bias.

There is no simple solution to algorithmic bias. Every data set is unique and presents possible bias. Without a way to accurately measure the amount of bias in

[2] Terry C. Davis et al., A Qualitative Study Exploring Barriers and Facilitators of Enrolling Underrepresented Populations in Clinical Trials and Biobanking, 7 FRONTIERS CELL AND DEV. BIOLOGY 1, 1 (2019).
[3] 21 C.F.R § 814.82 (2012).

an algorithm, the FDA must require manufacturers to have accurate reports available about the data sets used to train the algorithm. The slogan "garbage in, garbage out" applies to machine learning practices, as does the related phrase "bias in, bias out."[4] Tracking the data is important because "completeness of metadata provides information about the population, disease, and data types on which the algorithm was trained or validated, which is essential to extrapolating assumptions of generalizability of algorithm performance to other populations."[5] Increased attention to the early stages of data collection and algorithm development may allow manufacturers to create less-biased data sets. In addition, the FDA should continue working with universities and industry experts to improve evaluation and development of ML devices.

II. MEDICAL DATA SETS AND MACHINE LEARNING INTERSECT IN DIAGNOSTIC HEALTHCARE TOOLS

Machine learning is "an automated process of discovering correlations (sometimes alternatively referred to as relationships or patterns) between variables in a dataset, often to make predictions or estimates of some outcome."[6] Once an individual or organization creates a machine learning algorithm, it can process a huge amount of data, such as electronic health records, in an extremely short

[4] SOLON BAROCAS ET AL., FAIRNESS AND MACHINE LEARNING: LIMITATIONS AND OPPORTUNITIES, at 236 (2022) (ebook).
[5] David Wen et al., *Characteristics of Publicly Available Skin Cancer Image Datasets: a Systematic Review*, 4 LANCET DIGIT HEALTH e64, e65 (2022).
[6] David Lehr & Paul Ohm, Playing with the Data: What Legal Scholars Should Learn About Machine Learning, 51 U.C. DAVIS L. REV. 653, 671 (2017).

amount of time. Some machine learning models can learn from real-world use and improve performance.[7] Other models use new data to upgrade or modify but do not learn and change on the fly.

The healthcare industry uses machine learning medical devices for both treatment and diagnosis.[8] This paper will focus on diagnostic ML medical devices rather than both because most ML devices approved by the FDA are diagnostic.[9] Properly trained artificial intelligence (AI), which includes ML, "has the potential to dramatically improve diagnosis. [AI's] potential deserves emphasis, given that diagnostic errors effect five percent of U.S. outpatients annually, accounting for between six and 17 percent of adverse events."[10]

A. SOURCES OF BIAS IN MEDICAL DATA SETS

ML medical device training data sets rely heavily on patient data and health records. Patient data is an individual's medical information and includes "past and current health or illness, treatment history, lifestyle choices and genetic data."[11] A

[7] U.S. FOOD & DRUG ADMIN., Proposed Regulatory Framework for Modifications To Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD), at 2 (2022), https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf [hereinafter *Discussion Paper*].
[8] David Lyell et al., *How Machine Learning is Embedded to Support Clinician Decision Making: An Analysis of FDA-Approved Medical Devices*, 28 BMJ HEALTH & CARE INFORMATICS 1, 1 (2021).
[9] IQVIA, *FDA Publishes Approved List of AI/ML-enabled Medical Devices*, https://www.iqvia.com/locations/united-states/blogs/2021/10/fda-publishes-approved-list-of-ai-ml-enabled-medical-devices (Oct. 29, 2021); *see also*, U.S. FOOD & DRUG ADMIN., Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices, https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices (Sept. 22, 2021).
[10] Nicolas Terry, *Of Regulating Healthcare AI and Robots*, 21 YALE J.L. & TECH. 133, 146 (2019).
[11] *What is Patient Data and How is it Used?*, GENETIC ALLIANCE UK, (May 3, 2016), https://geneticalliance.org.uk/information/research-and-innovation/what-is-patient-data-and-how-is-it-used/.

health record is a collection of patient data that stores the information in one spot.[12] The healthcare industry's use of existing healthcare technologies, such as electronic health records, have left behind minority and low socioeconomical status populations in the past and contributed to health disparities.[13] Bias in medical data often occurs because of health disparities, which are differences in health and healthcare between groups that stem from broader inequalities.[14] These disparities may adversely affect groups based on their racial or ethnic groups, gender, age, disabilities, sexual orientation, or other characteristics historically linked to discrimination or exclusion.[15]

U.S. data sets "may inadequately reflect all groups in society, or may under-include women, and overrepresent persons of European ancestry, causing the software to provide unreliable or unsafe recommendations for minorities."[16] In addition, "machine learning algorithms used for medical image classification [may] underperform on images collected from populations independent to those on which the algorithms were trained."[17] Under-enrollment of minorities in clinical studies

[12] *Personal Health Records and Patient Portals*, MAYO CLINIC, June 4, 2022), https://www.mayoclinic.org/healthy-lifestyle/consumer-health/in-depth/personal-health-record/art-20047273.
[13] Xinzhi Zhang et al., *Big Data Science: Opportunities and Challenges to Address Minority Health and Health Disparities in the 21st Century*, 27 ETHNICITY & DISEASE 95, 97 (2017).
[14] Nambi Ndugga & Samantha Artiga, *Disparities in Health and Health Care: 5 Key Questions and Answers*, KAISER FAM. FOUND. (May 11, 2021), https://www.kff.org/racial-equity-and-health-policy/issue-brief/disparities-in-health-and-health-care-5-key-question-and-answers/.
[15] *Id*.
[16] Barbara J. Evans & Frank Pasquale, *Product Liability Suits for FDA-Regulated AI/ML Software*, BROOKLYN LAW SCHOOL, Legal Studies Paper No. 656, (forthcoming 2022) (manuscript at 8), https://ssrn.com/abstract=3719407.
[17] David Wen et al., *supra* note 4, at 70.

due to distrust, limited access to health care, and provider perceptions lead to underrepresentation of minorities in data sets.[18]

B. THE EFFECT OF BIAS IN MEDICAL DATA SETS

Algorithms trained on biased data sets can produce unjustly prejudicial results. In decision-making, fairness is described as "the absence of any prejudice or favoritism toward an individual or group based on their inherent or acquired characteristics."[19] Representation in the development of "AI-driven healthcare products is critical to help provide benefit to all people and avoid systematically disadvantaging minority populations – or worse, actively reinforcing that discrimination."[20] If the data sets reflect existing bias against minorities or other vulnerable populations, the algorithm will adopt and reproduce this bias in the results. If training data sets are "too small, inappropriate, inaccurate, or biased and non-representative of patients the software later will analyze, then the software, by its design, [likely] cannot provide accurate recommendations for their care."[21] In some situations, a small, inaccurate, or biased data set may still provide correct recommendations. However, without a way to know if this is the case for each data set, it is important to have representative data.

---

[18] Jill Fisher & Corey Kalbaugh, *Challenging Assumptions About Minority Participation in US Clinical Research*, 101 AM. J. PUB. HEALTH, at 2217, 2218-19 (2011).
[19] Ninareh Mehrabi et al., *A Survey on Bias and Fairness in Machine Learning*, 54 ASS'N. FOR COMPUTING MACH. SURV., 1, 11 (2022).
[20] Karl Surmacz et al., *Fairness in AI: How can we Avoid Bias and Disparities in Orthopedic Applications of Artificial Intelligence?*, 4 J. ORTHOPEDIC EXPERIENCE & INNOVATION, 1, 2 (2021), https://journaloei.scholasticahq.com/article/25901-fairness-in-ai-how-can-we-avoid-bias-and-disparities-in-orthopedic-applications-of-artificial-intelligence.
[21] Evans & Pasquale, *supra* note 15, at 13.

A 2019 study found a widely used, seemingly effective algorithm, which used health costs as a proxy for risk and reduced the number of Black patients identified for extra care by more than half.[22] Less money is spent on Black patients who have the same level of need as their White counterparts, which led the algorithm to incorrectly conclude the Black patients had fewer health issues.[23] The study found that Black patients generate lower costs than White patients, despite having a higher number of chronic illnesses. There are two likely explanations for the disparity in cost between the races. First, "poor patients face substantial barriers to accessing healthcare, even when enrolled in insurance plans."[24] Second, race may affect costs through discrimination, changes to doctor-patient relationships, or other factors.[25] The effect of these factors "is to lower health spending substantially for Black patients, conditional on need."[26] Looking at the data of healthcare spending alone does not account for these health disparities. Healthcare spending, combined with other socioeconomic factors, resulted in the algorithm missing patients with a greater need.

The delivery of healthcare is "known to vary by factors such as race, ethnicity, and socio-economic status; therefore, it is possible that biases present in our health care system may be inadvertently introduced into the algorithms."[27]

---

[22] Ziad Obermeyer et al., Dissecting Racial Bias in an Algorithm used to Manage the Health of Populations, 366 SCI. 447, 447 (2019).
[23] *Id.* at 447.
[24] *Id.* at 450.
[25] *Id.*
[26] *Id*.
[27] U.S FOOD & DRUG ADMIN., ARTIFICIAL INTELLIGENCE/MACHINE LEARNING (AI/ML)-BASED SOFTWARE AS A MEDICAL DEVICE (SAMD) ACTION PLAN, at 5 (2021) https://www.fda.gov/media/145022/download [hereinafter *Action Plan*].

Underrepresentation of populations in training data sets may lead to poorer real-world application of machine learning. For example, "if algorithms for reading chest X-rays are trained with data from primarily male patients, the results are not as accurate when applied to chest X-rays of female patients."[28] In dermatology, ML-based software as a medical device (SaMD) trained on a data set with a disproportionally high number of fair skinned images may have a poorer diagnosis rate for patients with darker skin.[29]

III. HOW THE CURRENT FDA REGULATION OF MACHINE LEARNING FAILS TO ADDRESS BIAS ISSUES

The FDA defines a locked algorithm as an algorithm that provides the same result each time the same input is applied to it and does not change with use.[30] Between 2012 and 2020, the FDA approved 161 of these locked devices using the agency's current pre-marketing approval process.[31] This process requires manufacturers to prepare and submit a premarket submission that corresponds to the class of the medical device.[32] The FDA assigns devices to one of three classes "based on the level of control necessary to assure the safety and effectiveness of

[28] *Uncovering and Removing Data Bias in Healthcare*, HEALTHCARE INFO. & MGMT. SYS. SOC'Y., (Apr. 6, 2021), https://www.himss.org/resources/uncovering-and-removing-data-bias-healthcare.
[29] Stephanie Chan et al., *Machine Learning in Dermatology: Current Applications, Opportunities, and Limitations*, 10 DERMATOL THER (HEIDELB) 365, 380 (2020).
[30] *Discussion Paper*, supra, note v6, at 3.
[31] Casey Ross, *As the FDA clears a flood of AI tools, missing data raise troubling questions on safety and fairness*, STAT NEWS (Feb. 3, 2021), https://www.statnews.com/2021/02/03/fda-clearances-artificial-intelligence-data/.
[32] *Overview of Device Regulation*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/overview-device-regulation(Sept. 4, 2020).

the device."[33] The manufacturer, not the software, controls all algorithm updates when it is locked. Once the medical device is approved, the FDA expects the manufacturer to comply with applicable regulatory controls, including establishment registration and device listing. This process has worked for previous locked medical devices because "in a locked algorithm, the same input will always produce the same result unless the developer updates the program."[34]

In contrast, a machine learning medical device with an unlocked algorithm, which can modify its behavior with changes in the environment or information, will not always have the same input. In a 2019 paper titled *Developing A Software Precertification Program: a Working Model*, the FDA stated that the agency's "traditional approach for the regulation of hardware-based medical devices is not well-suited for the faster, iterative design and development, and type of validation used for software device functions, including SaMD."[35]

A. RELEASE OF AI/ML ACTION PLAN

The FDA released a 2021 AI/ML Action Plan that outlined the agency's intended steps regarding regulation of machine learning medical devices.[36] The Action Plan included details of what the FDA called a Predetermined Change Control Plan. The Predetermined Change Control Plan included an SaMD Pre-

[33] *Classify Your Medical Device*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/medical-devices/overview-device-regulation/classify-your-medical-device (Feb. 7, 2020).
[34] *How FDA Regulates Artificial Intelligence in Medical Products*, THE PEW CHARITABLE TRUSTS, (Aug. 5, 2021), https://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2021/08/how-fda-regulates-artificial-intelligence-in-medical-products.
[35] U.S. FOOD & DRUG ADMIN., DEVELOPING A SOFTWARE PRE-CERTIFICATION PROGRAM: A WORKING MODEL, at 6 (2019) https://www.fda.gov/media/119722/download.
[36] *Action Plan*, *supra* note 26, 1-7.

Specifications section and an Algorithm Change Protocol. Under SaMD Pre-Specifications, the FDA would require manufacturers to describe what aspects the medical device manufacturer intended to change through algorithmic learning. The Algorithm Change Protocol would require an ML medical device manufacturer to explain the intended changes and adaptations of the algorithm to the FDA. The FDA would essentially review the planned future changes before the changes went into effect.

B. FDA RELIANCE ON MANUFACTURERS FOR DEVELOPMENT AND EVALUATION

The FDA will rely on ML medical device manufacturers following Good Machine Learning Practices (GMLP) in the development and evaluation of the manufacturers' medical devices.[37] Along with Health Canada and the United Kingdom's Medicines and Healthcare products Regulatory Agency, the FDA identified ten guiding principles to "inform development of GMLP."[38] The FDA continues to develop these standards but intends for the standards to guide "responsible innovations in [the AI/ML] area."[39] Public feedback about the Proposed Regulatory Framework, a discussion paper released by the FDA detailing the agency's potential future regulation of AI/ML devices, stated the need for manufacturers to "clearly describe the data used to train the algorithm…the role

[37] U.S Food & Drug Admin., *Good Machine Learning Practice for Medical Device Development: Guiding Principles*, https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles (Oct. 27, 2021).
[38] *Id.*
[39] *Id.*

intended to be served by its output, and the evidence of the device's performance."[40] In response, the FDA said it intended to consider the public input in order to determine what types of information the agency would recommend manufacturers disclose to users.[41]

Although this is a step in the right direction, the FDA needs to do more than recommend information for manufacturers to provide. The FDA should require manufacturers to report to the FDA the data used to train the algorithm, the intended role of the output, and evidence of the device's performance. The FDA identified the importance of real-world performance monitoring.[42] This data would allow the FDA to "understand how their products are being used, identify opportunities for improvements, and respond proactively to safety or usability concerns."[43]

IV. IMPROVING PRE AND POST MARKET REQUIREMENTS FOR MANUFACTURERS CREATING MACHINE LEARNING DEVICES

The Federal Food, Drug, & Cosmetics Act (FD&C Act) allows the FDA to regulate medical devices. The FDA must make decisions about requirements for manufacturers that are broad enough to encompass the many types of devices created while promoting safe and effective development. Using the agency's current regulatory tools as a starting framework, the FDA must implement changes that focus on the health disparities and potential bias during pre-market approval and require higher review during post market to properly monitor the devices.

[40] *Action Plan*, *supra* note 26, at 5.
[41] *Id.*
[42] *Id.* at 6.
[43] *Id.*

A. REQUIRED LABELING OF MACHINE LEARNING MEDICAL DEVICES

To increase transparency and communicate risk to consumers of ML medical devices, the FDA should require ML medical devices to include a label describing the data used to train the algorithm. Manufacturers may place the label on the device itself or include the label or other written, printed, or graphic matter upon any containers or wrappers, or accompanying the device while the device is held for sale after shipment or delivery for shipment in interstate commerce.[44] If manufacturers fail to do so, the FDA could find the manufacturer in violation of §502 of the FD&C Act and halt or withdraw marketing approval of the device. §502 states a device is misbranded if the device's labeling "is false or misleading."[45] Under §502(a), the FDA can find a label in violation of the law if it considers the label deceptive.

"It is not a necessary condition that the labeling [is] flatly and badly false; the word 'misleading' in the [FD&C Act] means that the labeling is deceptive if it [creates] or leads to a false impression in the mind of the reader. A 'false impression' may result not only from a false or deceptive statement, but may also be instilled in the mind of the purchaser by ambiguity or misdirection. It may also be caused by failure to inform the consumer of facts that are relevant to those statements actually made. In other words, the label that remains silent as to certain consequences may be as deceptive as the label that contains extravagant clauses."[46]

---

[44] U.S Food & Drug Admin., *Device Labeling*, https://www.fda.gov/medical-devices/overview-device-regulation/device-labeling (Oct. 23, 2020).
[45] Federal Food, Drug, and Cosmetic Act (FD&C) § 502, 21 U.S.C 352(a) (2010).
[46] *Id*.

The FDA may issue warning letters or seize FDA-regulated products that are adulterated or misbranded within the meaning of the FD&C Act. A seizure removes the products in violation from commerce. Of the 161 locked AI products approved by the FDA between 2012 and 2020, "only 73 disclosed the amount of patient data used to validate the performance of their device in public devices,…only seven reported the racial makeup of their study population, and just 13 provided a gender breakdown."[47] It is important that physicians and medical professionals using these devices in practice understand the limitations and possible adverse outcomes on certain populations. For example, an algorithm developed in a large academic medical center but used by a small, rural hospital "may recommend treatments that are not available or appropriate in a facility with less access to specialists and cutting-edge technology."[48]

B. REPORTING AND PERFORMANCE REQUIRMENTS

The FDA must implement reporting and performance requirements for manufacturers of ML medical devices. Because the FDA plans to rely on manufacturers to describe how the algorithms in the manufacturers' devices will change and adapt, the agency must deny and withdraw approval of medical devices at any stage of the process if necessary. The current approval process for fixed devices requires manufacturers to stay in compliance with regulatory controls. The FDA should have the same requirements for manufacturers of unlocked algorithms.

[47] Casey Ross, *supra* note 30.
[48] *The Pew Charitable* Trusts, *supra* note 33.

The FDA must require ML medical device manufacturers to provide the agency with clinical data prior to approval. Under the current pre-market requirements, developers of Class III devices, which are devices the FDA considers high risk, must submit clinical evidence that the product's benefits outweigh the risks.[49] This application, called a Premarket Approval (PMA), must meet data requirements to show the safety and effectiveness of the device. Clinical investigations include "study protocols, safety and effectiveness data, adverse reactions and complications, device failure and replacements, patient information, patient complaints, tabulations of data from all individual subjects, and results of statistical evidence."[50] The FDA requires devices seeking PMA to meet post approval requirements, including "continuing evaluation and periodic reporting on the safety, effectiveness, and reliability of the device for its intended use. When reporting is required, the FDA will state in the PMA approval order the reason or purpose for such requirement, the number of patients to evaluate, and the required reports to submit.[51] The agency should adopt a similar standard for all ML medical devices, regardless of the risk category.

Currently, manufacturers may choose to conduct studies for clinical validation after their devices are approved by the FDA. This is often done by working with providers in certain hospitals to test the products.[52] The FDA should

---

[49] *Id*.
[50] U.S Food & Drug Admin., *Premarket Approval*, https://www.fda.gov/medical-devices/premarket-submissions-selecting-and-preparing-correct-submission/premarket-approval-pma (May 16, 2019).
[51] 21 C.F.R. § 814.82 (2012).
[52] Casey Ross, *Illuminating the black box: Five ways the FDA could build transparency into AI devices*, STAT NEWS (Oct. 18, 2021), https://www.statnews.com/2021/10/18/fda-artificial-intelligence-data-transparency/.

require manufacturers of machine learning medical devices to conduct a clinical study to review the device's performance following the agency's approval.[53] §519(a)(1)(B) of the FD&C Act requires a device manufacturer to report to the Secretary if it becomes aware of information suggesting that one of the manufacturer's devices has malfunctioned, and that "such device... would be likely to cause or contribute to a death or serious injury if the malfunction were to recur."[54] §519 also requires that "every person who is a manufacturer or importer of a device intended for human use shall establish and maintain such records, make such reports, and provide such information, as the Secretary may by regulation reasonably require to assure that such device is not adulterated or misbranded and to otherwise assure its safety and effectiveness."[55]

The FDA's Good Machine Learning Practices guidance document aligns with the statutory language of §519 requiring records and reports. Post-market real world performance reporting is encompassed by the requirement of §519(a)(1)(B) to report to the Secretary any information of device malfunctions. For ML medical devices, device malfunctions should include errors in expected behaviors and disparate results for certain populations in real-world use. Under §519, the Secretary could then use the report to assess the safety and effectiveness of the device in real-world use.

---

[53] Jeremy Jordan, *Effective testing for machine learning systems*, JEREMY JORDAN (Aug. 19, 2020), https://www.jeremyjordan.me/testing-ml/.
[54] Federal Food, Drug, and Cosmetic Act (FD&C) § 502, 21 U.S.C. § 519(a)(1)(B) (2010).
[55] *Id*.

V. CONCLUSION

The FDA should create functional regulations for the growing number of machine learning medical devices. The healthcare system is increasingly using these devices for diagnosis. Machine learning devices trained on biased data sets are susceptible to furthering certain types of bias and generating flawed outcomes. The FDA should require ML medical devices to include a label that describes the demographics of the tested population. If manufacturers fail to include this information, the FDA could determine the label false or misleading under §502 of the FD&C Act and stop sales of the device. After approval, the FDA should use §814.89(2) and §519 to require manufacturers to report and evaluate the real-world performance of their approved devices. These reviews should include studies for clinical validation or model evaluation and model testing. While addressing bias in diagnostic medical machine learning devices will take more than the FDA, the agency should continue to support efforts to find an effective way to mitigate and measure bias.