
NISKANEN C E N T E R

Regulatory Comment

Comments submitted to the Food and Drug Administration in the Matter of:

THE BENEFITS AND RISKS OF SOFTWARE AS A MEDICAL DEVICE: A Response to a Request for Input RE: “Development of 21st Century Cures Act Section 3060 Required Report”

Dr. Anastasia Greenberg
Technology Policy Fellow
Niskanen Center

Ryan Hagemann
Senior Director for Policy
Niskanen Center

Submitted: June 27, 2018
Docket Number: FDA-2018-N-1910

INTRODUCTION

The ongoing buildup of pressure on the U.S. health care system is a call to action for regulators and policymakers to foster innovation in the health sector. Domestic spending on health care continues to rise, and in 2016 alone grew 4.3 percent to \$3.3 trillion, accounting for 17.9 percent of U.S. GDP.¹ Over the next eight years, the annual growth rate of domestic health care spending is expected to average about 5.5 percent; by 2026, annual health care spending will rise to \$5.7 trillion — almost 20 percent of GDP.² Meanwhile, approximately 12 million Americans are misdiagnosed annually, and the situation will likely deteriorate further as the number of elderly in the United States stands to grow by over 50 percent in the next 15 years.³ However, there are some hopeful prospects for absorbing and effectively responding to this expected shock to the system. Coupled with innovative physical assistive devices and increased home-based care, personalized rehabilitation could help dampen the pressure on traditional health care services. These and other innovations in precision medicine will be driven in no small part by digital health solutions, particularly those that utilize artificial intelligence (AI).

AI, and more specifically machine learning (ML), has gained increased attention in recent years for its promising applications in health care, enabled by the availability of large patient-data troves and increased computing power. This technology has the potential to democratize medicine through the creation of low-cost software devices that can aid health care professionals in diagnostic and therapeutic decision-making. User-friendly AI software can also allow patients to get access to real-time personalized medical care and decisions, reducing the burden on health care professionals while increasing access to high-quality medical advice or information.

Many of these forecasted benefits will be driven by venture capital investments in AI, which reached \$794 million in 2016.⁴ The top areas of expected applications for AI in health care are: intelligent diagnostics, patient-provider data management, drug discovery, medical devices and robotics, and home health. Given the magnitude of the investments and potential benefits at stake, government regulators need to clarify how they plan to address the unique concerns arising from emerging AI diagnostic tools and the broader regulatory implications of Software as Medical Devices (SaMD).

To address some of these uncertainties, Congress passed the 21st Century Cures Act (hereafter “Cures Act”), which, in part, amended the Federal Food, Drug, and Cosmetic Act (FD&C Act) to exempt certain software functions from the definition of a “medical device.” In response, and in accordance with section 3060(b) of the Cures Act, the U.S. Food and Drug Administration (FDA) has asked for public input on its recent guidance documents that deal with the section 520(o)(1) amendments to the FD&C Act (the section that provides the criteria for determining which software functions qualify for the device exemption).⁵ Under the

¹ Centers for Medicare and Medicaid Services, Department of Health and Human Services, “National Health Expenditure Fact Sheet,” <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet.html>.

² *Ibid.*

³ Peter Stone et. al., *Artificial Intelligence and Life in 2030*, One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University (Stanford, CA, Sept. 2016), <http://ai100.stanford.edu/2016-report>.

⁴ *Ibid.* (Among others, these VC and corporate investment firms include Accel Partners, Andreessen Horowitz, Google Ventures, and IBM.)

⁵ In the interest of conforming to the language used in the FDA’s guidance documents, these comments will refer to the relevant sections of the FD&C Act — added and amended by section 3060(a) the 21st Century Cures Act (hereafter “Cures Act”) — rather than the United States Code. Any reference to section 520(o)(1) of the FD&C Act can also be read to refer to 21 U.S.C. § 360j(o).

leadership of Scott Gottlieb, the FDA has expressed interest in actively promoting innovation in the digital health space, with these recent guidance documents prepared in that spirit.⁶ In a speech delivered on April 26, 2018, Commissioner Gottlieb stated:

In issuing this guidance, FDA is taking additional steps to clarify what technologies won't fall under avoidable regulation. Our goal is to allow developers to efficiently incorporate into their products the latest advances in technology, while focusing FDA's review on the safety and effectiveness of the higher-risk medical device functions that diagnose or treat patients. We believe this approach will encourage more innovation in this important field.⁷

We concur with the spirit of these remarks, and share the same goal of ensuring ongoing innovation and progress in the application of emerging technologies to meeting the challenges and demands of the domestic health care market. Given the invaluable role that AI is already playing in digital health innovation, particularly in diagnostic and medical decision-making applications, these comments are meant to provide a detailed and systematic overview of how the FDA's current regulatory framework should be adapted to help advance revolutionary AI-based medical software innovations.

The structure of these comments is as follows: Part I focuses on the implications of the section 520(o)(1) exclusions on AI diagnostic software. Part II discusses the broader issues with the current FDA regulatory pathway for new AI diagnostic software. Part III offers a summary of recommendations that can foster increased innovation in AI digital health while maintaining appropriate standards of safety.

PART I: EXCLUSIONS OF SOFTWARE FUNCTIONS FROM DEVICE DEFINITION

The FDA has asked stakeholders to provide input on the risks and benefits to health potentially resulting from implementation of section 3060(a) of the Cures Act.⁸ This section amends the FD&C Act by adding section 520(o), which provides for the exclusion of certain software functions from the definition of "medical device," as defined under section 201(h) of the FD&C Act. Specifically, section 520(o)(1) describes certain software functions that will not be considered a medical device for purposes of regulatory approval or oversight. In order to qualify for that exemption, the software device must be intended:

1. For administrative support of a health care facility;
2. For maintaining or encouraging a healthy lifestyle;
3. To serve as electronic patient records;
4. For transferring, storing, converting formats, or displaying data; or
5. To provide limited clinical decision support.

⁶ Scott Gottlieb, M.D., Commissioner of Food and Drugs, "Transforming FDA's Approach to Digital Health," *Remarks presented at the Academy Health's 2018 Health Datapalooza* (Washington, D.C., 26 Apr. 2018), <https://www.fda.gov/NewsEvents/Speeches/ucm605697.htm>.

⁷ Gottlieb, "Transforming FDA's Approach to Digital Health."

⁸ Pub. L. 114-255 (Dec. 13, 2016), <https://www.gpo.gov/fdsys/pkg/PLAW-114publ255/content-detail.html>.

In this section, we discuss issues with the FD&C amendments and the corresponding guidance describing the agency’s thinking on their implementation, focusing on issues of both innovation and safety in the context of AI software development.

Software Functions Unrelated to Clinical Decision Support

Sections 520(o)(1)(A), (B), (C), and (D) are focused on excluding very particular software functions from the definition of a medical device — in particular, functions that have very narrow and indirect health-related applications. As discussed *supra*, these provisions focus on software intended for administrative support, maintaining electronic records, or data storage — functions that are not immediately related to patient health outcomes. Section 520(1)(B), for example, mentions software functions that encourage a healthy lifestyle, reiterating that any software qualifying for exemption must be “unrelated to the diagnosis, cure, mitigation, prevention, or treatment of a disease or condition.” As confirmed by the *Changes to Existing Medical Software Policies Resulting from Section 3060 of the 21st Century Cures Act* guidance, wellness-related software functions that are involved in the mitigation or prevention of a disease or condition would not be excluded from the definition of a device under section 201(h).⁹

Given these limitations, the exclusions provided for in sections 520(o)(1)(A), (B), (C), and (D) would not open the door to true innovation in the digital health care space in ways that would have a significantly positive impact on patient health outcomes. For the same reason, the risks to patient health and safety from these regulatory exemptions are likely to be negligible.¹⁰ We therefore turn our attention on the implications of section 520(o)(1)(E), which has direct relevance for patient health and innovation in the medical marketplace.

Software Functions for Clinical Decision Support

In the context of AI diagnostic and decision-support systems, section 520(o)(1)(E) has clear implications that may affect both innovation and safety and efficacy in relation to patient outcomes. Section 520(o)(1)(E)(i) excludes from the definition of device those software functions that are “intended for the purpose of displaying, analyzing, or printing medical information.” As noted in the *Clinical and Patient Decision Support Software* guidance, the “FDA interprets this to include software functions that display, analyze, or print patient-specific information, such as demographic information, symptoms, and test results, and/or medical information, such as clinical practice guidelines, peer-reviewed clinical studies, textbooks, approved drug labeling, and government agency recommendations.”¹¹ As stated in section 520(o)(1)(E)(ii), software functions that qualify for an exemption must also be “intended for the purpose of supporting or providing recommendations to a health care professional about *prevention, diagnosis, or treatment of a disease or condition*” [emphasis added].

⁹ *Changes to Existing Medical Software Policies Resulting from Section 3060 of the 21st Century Cures Act*, Draft Guidance for Industry and Food and Drug Administration Staff (U.S. FDA, 8 Dec. 2017), <https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM587820.pdf>.

¹⁰ When considering an explicit reference to emerging AI diagnostic and decision support devices, the FDA is unfortunately silent on whether any provisions of the section 520(o)(1) exemptions might apply. Based on our reading of the statute, as well as the FDA’s guidance, there does not appear to be a route to exemption, as none of the five intended function provisions implicate software that possesses the capability to predict health outcomes or propose treatments. However, the agency should consider making its thinking on this matter more explicit in future guidance, as the lack of any mention of AI diagnostic tools creates lingering regulatory uncertainty for innovators and investors.

¹¹ *Clinical and Patient Decision Support Software*, Draft Guidance for Industry and Food and Drug Administration Staff, (U.S. FDA, 8 Dec. 2017), p. 7, <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm587819.pdf>.

Thus, the software functions that would be excluded from the definition of a device — and exempt from more stringent FDA oversight — are functions that may have serious, direct consequences on patient outcomes, since they would be intended for aiding health care professionals in making critical clinical decisions. It is important to consider how carving out these specific exemptions may (or may not) incentivize digital health innovation and the associated safety and efficacy concerns.

Importantly, as per section 520(o)(1)(E), none of the above-mentioned exclusions would apply to functions that are “intended to acquire, process, or analyze a medical image or a signal from an in vitro diagnostic device or a pattern or signal from a signal acquisition system.”¹² A straightforward interpretation of this provision suggests that the only available data inputs to these exempt software functions would have to be higher-level patient health data, such as data contained in electronic health records (e.g., patient medical history, age, sex, demographic information, blood pressure, etc.). Physiological data obtained directly from the body (e.g., X-ray images, MRI scans, cardiac signals, etc.) would not be available to these exempt devices. These exempt software functions would therefore be operating on data that is much less rich and informative than data contained in physiological images/signals. It is therefore questionable whether these narrow exemptions would spur key innovation in the digital health care space.

More pertinent, however, are the potential safety and efficacy concerns raised by the legislative decision to provide for such narrow software-function exemptions for the purpose of aiding critical clinical decision-making. Section 520(o)(1)(E)(iii) states that these excluded software functions must enable health care professionals “to independently review the basis for such recommendations that such software presents...” The corresponding FDA guidance document — *Clinical and Patient Decision Support Software* — clarifies this provision to mean that the software function must provide the “rationale or support for the recommendation,”¹³ and that exempt software would use “rule-based tools” to make recommendations to health care professionals.¹⁴ The FDA’s concern with allowing exemptions for software only if such software has what may be considered a high level of “explainability” would imply that any software making use of ML algorithms would be categorically prohibited under such exemptions.

ML algorithms learn complex patterns from large datasets and cannot provide direct causal rationale for their output(s). However, they are powerful tools for making accurate predictions from data, despite not being able to offer an “explanation” that is always interpretable by a human. The more numerous and complex the variables used by a ML algorithm, the more difficult it becomes for a human to understand how the algorithm arrived at a given decision. Linear ML models, which can only represent simple relationships between variables, are relatively simple to interpret, while more complex ML methods, such as support vector machines and artificial neural networks, are difficult to interpret. There is a clear trade-off between algorithmic accuracy and explainability; and the balance the FDA strikes between them will have profound implications for innovation in the SaMD/AI/ML ecosystem.¹⁵

Taken together, the software functions that may benefit from the device definition exclusion would have to (1) use only certain types of health data, excluding rich and complex physiological data, and (2) use hard-coded “rule-based” analysis methods that do not take advantage of the most accurate computing potential — namely, AI diagnostic software. These restrictions hold algorithmic decisions to a standard that human

¹² *Ibid.*

¹³ *Ibid.*

¹⁴ *Ibid.*

¹⁵ Bryce Goodman and Seth Flaxman, “EU Regulations on Algorithmic Decision-Making and a ‘Right to Explanation,’” (New York: ICML Workshop on Human Interpretability in ML, 2016), 27, <https://arxiv.org/abs/1606.08813>.

beings — including health care professionals — are not held to.¹⁶ While a physician may be able to provide a justification for a given decision, there is no guarantee that this justification takes into account all of the unconscious biases and assumptions that human beings operate on.

This scenario may incentivize developers who want to avoid FDA regulatory burdens to create less-accurate decision-support products than they otherwise might, providing little added benefit or functionality compared to products already on the market. The result would be a lower quality in the software available to health care professionals to make clinical decisions, which in turn has direct impact on patient health outcomes. Therefore, we suggest the FDA update its section 520(o)(1) guidance — particularly, *Clinical and Patient Decision Support Software* — to explicitly allow ML algorithms to be implemented when designing software that meets one of the intended purposes enumerated in sections 520(o)(1)(A), (B), (C), (D), or (E). The FDA should focus on software device outcomes for patient health, not on the development process of the software itself.

PART II: FDA REGULATORY PATHWAYS FOR AI DIAGNOSTIC SOFTWARE DEVICES

Given the potential for AI diagnostic and medical decision-support software to revolutionize diagnostic tools, and because section 520(o)(1) exemptions appear to offer no path to market for such devices (especially those intended for analyzing physiological data), it is imperative that the FDA consider and clarify alternative or additional regulatory pathways that may be available to an AI diagnostic software developer. Recently, the agency approved marketing of the first medical device utilizing AI to analyze images of the retina to help diagnose diabetic retinopathy,¹⁷ which was granted Breakthrough Device Designation¹⁸ as well as a De Novo marketing request.¹⁹ Similarly, in May 2018, the FDA permitted marketing of a SaMD that uses AI to detect and diagnose wrist fractures.²⁰ This device was also granted a De Novo marketing request. Despite these positive developments, there remains much uncertainty regarding the regulatory pathway that would be best suited for AI diagnostics developers.

A major issue is that there is no FDA guidance for industry that specifically discusses the use of AI in diagnostic devices. The most relevant guidance document currently available specifically deals with Computer-Assisted Detection (CADe) devices, while explicitly leaving out Computer-Assisted Diagnostic (CADx) devices.²¹ CADe devices, as described by the guidance (hereafter “CADe Guidance Document”) are those that may use ML algorithms for pattern-recognition and other data analysis capabilities to detect or

¹⁶ See Joshua New and Daniel Castro, *How Policymakers Can Foster Algorithmic Accountability*, (Center for Data Innovation, 21 May 2018), <http://www2.datainnovation.org/2018-algorithmic-accountability.pdf>; See also Curt Levey and Ryan Hagemann, “Algorithms With Minds of Their Own,” *The Wall Street Journal*, 12 Nov. 2017, <https://www.wsj.com/articles/algorithms-with-minds-of-their-own-1510521093>.

¹⁷ “FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems,” FDA News Release (U.S. FDA, 11 Apr. 2018), <https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm604357>.

¹⁸ 21 U.S.C. § 360e-2(d)(1) (2018), <https://www.law.cornell.edu/uscode/text/21/360e>; FD&C Act § 515(B)(d)(1).

¹⁹ 21 U.S.C. § 360c(a)(1) (2018), <https://www.law.cornell.edu/uscode/text/21/360c>; FD&C Act § 513(f)(2).

²⁰ “FDA permits marketing of artificial intelligence algorithm for aiding providers in detecting wrist fractures,” FDA News Release (U.S. FDA, 24 May 2018), <https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm608833.htm>.

²¹ *Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data – Premarket Notification [510(k)] Submissions*, Guidance for Industry and Food and Drug Administration Staff, 74 F.R. 54053 (U.S. FDA, 3 July 2012) (hereafter “CADe Guidance Document”), <https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm187294.pdf>.

identify physiological abnormalities, such as liver lesions or lung nodules.²² On the other hand, CADx devices — those explicitly not addressed in the CADE Guidance Document — go beyond detection by providing diagnostic insights, such as providing the probability that a given disease or condition is present in the body, or offering treatment options and further prognosis.

In the absence of any other guidance, it would appear that AI diagnostic devices would fit within the CADx device category. Unfortunately, as previously indicated, the FDA has yet to explicitly and formally describe its thinking on these particular devices, leaving a significant hole in the agency’s library of device guidance. We would therefore urge the FDA to clarify what, specifically, would be considered a CADx device, and further specify how AI and ML software capabilities would fit into that definition. Doing so would provide vital clarity for innovators and market actors driving the advancements in AI health care technology that, as discussed *supra*, will undoubtedly offer immense benefits in patient health outcomes, health care cost savings, and improvements in precision medicine offerings in the years to come.

Any future guidance will need to take into account the pressing need for fostering AI digital health innovation based on least burdensome principles, while maintaining high standards for safety and efficacy. To that end, the next section will provide some recommendations on specific elements the FDA ought to consider in detailing its thinking on AI diagnostic tools.

Choosing a Regulatory Pathway for AI Diagnostic Software

FDA guidance addressing AI diagnostic devices (whether updates/expansions specific to the CADx category or otherwise) should clarify how a developer may proceed in determining which regulatory pathway is best-suited for a given device. Thus far, there are few potential AI diagnostic predicate devices to base a 510(k) premarket review process on, so it would be highly likely that a De Novo review may be most relevant.²³ However, as more AI diagnostic devices enter the market, guidance should provide clarity on how industry should decide whether to make a De Novo request or proceed through a 510(k) premarket review based on substantial equivalence to a predicate AI-based device.

The following hypothetical questions highlight some of the potential issues that are unique to AI software development, and may help guide the agency’s thinking on future guidance on this matter:

1. If a developer created an AI diagnostic device for the same intended purpose as a potential “predicate” AI device, using a similar or identical class of ML algorithms to train the data, but with a different dataset used for training, would this scenario be more appropriate for a 510(k) or a De Novo pathway?
2. If a developer created an AI diagnostic device for the same intended purpose as a potential “predicate” AI device, and trained on an identical dataset as a potential predicate device, but employing a different ML algorithm, would this scenario be more appropriate for a 510(k) or a De Novo pathway? Ultimately, employing a different algorithm may still result in substantially identical classification accuracy.

²² *Ibid.*

²³ Although the two AI diagnostic devices discussed *supra* succeeded under the De Novo review process, we wish to emphasize (once again) that the lack of clarity provided by the FDA can only lead us to presume that this pathway, rather than the 510(k) premarket review process, makes more sense for an AI diagnostic device.

3. When deciding on a regulatory pathway, should an AI software developer compare a new device to existing AI-based devices by focusing on the input data and features, the method of labeling “output” data, the specific class of ML algorithms used, and/or the final classification accuracy?

These are some of the questions the FDA will need to address in order to provide clarity to software developers and investors working on AI software diagnostic tools. We recommend that substantial equivalence based on a predicate AI diagnostic device be focused on the intended purpose of the device (e.g., diagnosing malignancy in breast biopsies) and on an appropriate and similar threshold level of classification accuracy (e.g., 80 percent), but not on (1) the specific input data/features, (2) type of ML algorithm, or (3) training parameters. Ultimately, the FDA should focus its regulatory rules and decisions on the intended use, safety, and efficacy of a device — not on the process of ML model building.²⁴

Lastly, within the context of a risk-benefit analysis, any new guidance should provide clear examples of AI software devices that the FDA foresees would be unlikely to fit under Class I or II designation. Put another way, the agency should explicitly describe the traits and features of what it would consider a Class III device, which would necessitate submitting to a premarket approval process (PMA). We would, however, recommend that the FDA acknowledge that most AI diagnostic software that are within the ambit of SaMDs are unlikely to be classified as Class III, given that most of these devices would not pose immediate risks to patient health, unless they were integrated with hardware that support or sustain human life. AI-based software on its own, which does not control or interact with hardware that poses serious a risk to life, should not require a PMA pathway.

As a general matter, then, any determination of Class III designation for an AI software device should require an affirmative justification by the FDA, detailing its specific reason(s) for necessitating premarket approval, including a risk-benefit analysis. The agency’s default presumption should weigh in favor of the innovator, and designate AI software devices as either Class I or Class II.

Breakthrough Device Designation

The FDA’s Breakthrough Device Designation allows for priority review, flexible clinical design, and an iterative review process for devices that: (1) provide for more effective treatment or diagnosis of life-threatening diseases and conditions, and (2) offer significant advantages over existing alternatives. Such advantages include “reduc[ing] or eliminat[ing] the need for hospitalization, improv[ing] patient quality of life, facilitat[ing] patients’ ability to manage their own care (such as through self-directed personal assistance), or establish[ing] long-term clinical efficiencies.”²⁵ The idea that appears embedded in the criteria for Breakthrough Device Designation is that these devices provide personalized-medicine solutions and reduce the burden on the traditional health care system — an idea that is perfectly concordant with the promise of AI diagnostics. It is no surprise, then, that the recent AI-based retinopathy diagnostic software was granted such a designation by the FDA. We therefore recommend that guidance for CADx or other AI-based software devices discuss the potential for new AI diagnostic SaMDs to apply for the Breakthrough Device Designation.

²⁴ The FDA’s experience in approving medical devices lies primarily in determining safety and effectiveness, while the latter arena (determining how the AI is built) is one in which the FDA has no substantive expertise or historical familiarity in addressing. As such, the agency is better equipped and more appropriately-situated to focus its attention on regulating based on principles that reflect a recognition of its institutional capabilities (and limitations).

²⁵ *Breakthrough Devices Program*, Draft Guidance for Industry and Food and Drug Administration Staff, Docket No. FDA-2017-D-5966-0001 (U.S. FDA, 25 Oct. 2017), <https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM581664.pdf>.

Updates to AI Diagnostic Software

There is no FDA guidance that specifically addresses software updates for AI diagnostic devices, and the currently available CADe Guidance Document and 510(k) submission guidelines only contribute to uncertainty and confusion. So, too, do existing rules in the Code of Federal Regulations.

Under 21 C.F.R. § 807.81(a)(3), for example, a device manufacturer would need to submit a new 510(k) application when “a change or modification in the device ... could significantly affect the safety or effectiveness of the device.” In the context of software updates, the FDA’s guidance interprets this provision to mean that a new 510(k) submission would be required when a change is “made with intent to significantly affect safety or effectiveness of a device,” including changes that are meant to improve clinical outcomes or mitigate risks.²⁶ It further states that a 510(k) submission is likely required for an in vitro diagnostic device that “includes a change that could have clinically significant impact in terms of clinical decision-making.”²⁷ This suggests that if an AI diagnostic developer were to update the software with new field data to improve the accuracy of the software, this change would necessitate the submission of a new 510(k). Furthermore, the CADe Guidance Document, which deals directly with devices that may employ ML algorithms for “detection” purposes, states:

We may consider a change or modification in a CADe algorithm to significantly affect the safety or effectiveness of the device ... if ... a change has been made to the *image processing components, features, classification algorithms, training methods, training data sets, or algorithm parameters* [emphasis added].²⁸

As a result, it is difficult to imagine any meaningful software update to an AI diagnostic device that would be exempt from a 510(k) submission. Since ML algorithms have the ability to learn and improve in accuracy with added data — sometimes even with the addition of a single new data input — there would be a burdensome process by which an AI developer would be required to constantly submit new 510(k) applications, creating a large backlog of updates and preventing patients and health care professionals from accessing the most up-to-date, and therefore safest and most accurate, diagnostic tools.²⁹

There is an urgent need to process and analyze real-world patient data on an ongoing basis, and update diagnostic and clinical-decision support software in a timely manner.³⁰ Accordingly, the FDA should provide guidance on how an AI software developer can make critical software updates based on new patient data without the need to proceed through a burdensome 510(k) process for each software modification.

²⁶ *Deciding When to Submit a 510(k) for a Software Change to an Existing Device*, Guidance for Industry and Food and Drug Administration Staff, Docket No. FDA-2016-D-2021 (U.S. FDA, 25 Oct. 2017), <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm514737.pdf>.

²⁷ *Ibid.*

²⁸ *CADe Guidance Document*.

²⁹ Indeed, research suggests that when making medical predictions, the relevance of clinical data decays with a half-life of approximately 4 months. Every day of regulatory delay can actually be a net loss to the mission of promoting public health, and leave patient health outcomes worse off than they might otherwise be. See J. H. Chen and S.M. Asch, “ML and prediction in medicine-beyond the peak of inflated expectations,” *New England Journal of Medicine* 376, no. 26, (June 2017): 2507-2509.

³⁰ As AI diagnostic devices begin to enter the market, they will gather large amounts of new patient data that could be used to improve the performance of the devices in future versions. ML algorithms are known to be data-hungry; feeding them larger, more representative input data will likely lead to better final output accuracy, and therefore improved patient health outcomes. Therefore, the FDA should encourage developers to use real-world data that is gathered following initial device release to continuously update software to improve diagnostic accuracy.

Safety and Efficacy Considerations for AI Diagnostic Software

Although AI diagnostic devices are projected to revolutionize health care by improving patient outcomes and reducing the economic burden on the U.S. health care system, potential safety concerns should not be disregarded. A study by Caruana et. al., for example, assessed one such algorithm that made treatment suggestions for patients with pneumonia. Unexpectedly, the algorithm identified patients who showed symptoms of both pneumonia and asthma as having lower mortality risks than those with only pneumonia.³¹ After a detailed examination of the patient data, the study found that patients with both pneumonia and asthma were more likely to be initially placed in intensive care units, thereby reducing the potential for further complications.³² While the algorithm accurately identified those with both pneumonia and asthma as exhibiting a lower risk of mortality, a deeper understanding of how the algorithm likely arrived at that decision shows that a treatment recommendation based on this “artificial” lower-risk of mortality would not be reasonable.

Although ML algorithms are robust prediction tools, making future clinical decisions based solely on algorithmic outputs may result in serious negative consequences. The FDA can ameliorate potential concerns such as these by providing recommendations for developers on best practices for ensuring safety and efficacy of AI software. To that end, the following list offers suggestions for information the agency might wish to consider requesting in a future classification application:

1. **Descriptive statistics of all features (input variables) that were used to build the classifier.** This would allow for a rough assessment of any potential problematic features that include biased or non-representative data.
2. **Explanation of reasons for including each feature as a variable for training the algorithm.** A strong theoretical justification for how various patient data are related to the diagnosis/treatment of a given disease/condition can help to prevent unintended consequences of complex ML models.
3. **Description of how data was split into training, validation, and test sets.** It is imperative that the test data and training data do not overlap to avoid biasing the final model. Building a model that is generalizable to new data relies on splitting the data into these three groups. It is essential to test models on truly independent datasets, from different populations/times that played no role in model development
4. **Description of how the initial dataset was labeled for training.** This includes an explanation of how a “ground truth”³³ was established for labeling the data in supervised models. For example, was the ground truth established based on clinical data? Was multiple observer averaging used to make a final labeling determination?
5. **Description of any preprocessing and/or data cleaning steps that were performed.** It is important to consider any manipulation(s) performed on the dataset, as preprocessing/cleaning may alter the input data in unintended ways.

³¹ R. Caruana et. al, “Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission,” *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Sydney, Australia: Aug. 2015): 1721-1730.

³² *Ibid.*

³³ Referring to the statistical process of accumulating falsifiable/observable data for the purpose of validating or falsifying a hypothesis (i.e., actual measurements).

6. **Learning curves.** An inclusion of the learning curves during model training allows for identification of major model building issues, such as models that show high bias (i.e., underfitting) or high variance (i.e., overfitting). High bias or high variance is an indication that the model will not generalize well on new data.
7. **Accuracy scores.** It is important to include the final model accuracy scores. Such scores must be reported based on calculations that are appropriate for a given ML situation. For example, F scores should be included for models trained on input data with rare events (i.e., low positive-to-negative ratio in the input data — presence of disease is very rare) to avoid misleading accuracy scores.
8. **Model updates.** Once new data is obtained after initial market release, the model should be tested on the new data and the model should be retrained as necessary. Ground truth can be established again on newly acquired data when feasible and the model accuracy should be compared against the new data. New data should also be used to improve model accuracy and generalizability.

These data could help demonstrate systematic verification checks for generalizability and accuracy when building ML models for diagnostic and clinical decision support. In the event that no formal premarketing submissions are required, these recommendations can nonetheless serve as general guidelines and best practices for AI software developers.³⁴

Precertification Program for AI Diagnostic Software

In June, the FDA released the second version of its proposed Software Precertification Program,³⁵ which would convene a pilot test project allowing software developers to be precertified based on a proven track record of excellent quality in medical-device manufacturing and an adherence to specific standards and best practices. This program would allow precertified firms to either market new devices without further FDA review or through a streamlined premarket review process. The focus is on a tailored review system, decision-making based on real-world performance data, iterative review and interactions between the developer/manufacturer and the FDA, and postmarket validation.

We support the FDA's ongoing efforts at developing an implementation plan for this pilot program and urge the agency to expand the program as soon as is feasible. In this way, the test program will allow for a larger number of industry entities, beyond the small list of currently identified candidates, to participate. Such a program is key for spurring innovation in the digital health care space by providing FDA the necessary real-world "governance learning"³⁶ to adopt a flexible and adaptive regulatory approval process for emerging AI diagnostic devices, as well as other SaMDs.

³⁴ It may also be valuable for the FDA to consult with the National Institute of Standards and Technology (NIST) on these guidelines, as well as request an audit of suggested data types as technology advances.

³⁵ *Developing Software Precertification Program: Working Model – Version 0.2*, Docket No. FDA-2017-N-4301-0001 (U.S. FDA, June 2018),

<https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/UCM611103.pdf>.

³⁶ Ryan Hagemann, "New Rules for New Frontiers: Regulating Emerging Technologies in an Era of Soft Law," *Washburn Law Journal* 57, no. 2, (Spring 2018): 244-255, p. 250, <http://washburnlaw.edu/publications/wlj/issues/57-2.html>. ("Just as the private sector relies on trial-and-error experimentation in developing new technologies, so too does governance of those technologies require a more adaptive and flexible approach to crafting rules. This embrace of 'governance learning' is one of the results of multistakeholder processes, as one of the primary challenges for modern regulators is balancing the need for increasingly adaptive and flexible rules and an inherent institutional aversion to risk, given their statutory focus on ensuring public safety.")

PART III: SUMMARY OF RECOMMENDATIONS

We are happy to share in Commissioner Gottlieb’s avowed goal of promoting regulatory policies that help foster innovation in the digital health care space.³⁷ To that end, we offer the following summary of recommendations discussed *supra*:

1. Embrace a technologically neutral approach to regulating software in medical devices.

The FDA should primarily focus its oversight and enforcement on the outcomes resulting from software incorporated into medical devices, not restricting the processes by which developers actually develop the software (e.g., using ML). Rather than focus on the need for “algorithmic explainability” or “algorithmic transparency,” the agency should direct its limited time and resources towards oversight of higher-risk devices (e.g., implantable diagnostics) with much clearer potential for harm. In vitro diagnostic systems, even those employing AI and ML systems, should be judged based on their outcomes rather than the technologies such devices employ.

2. Emphasize flexibility and adaptability in development standards.

The guidance documents dealing with section 520(o)(1) should be updated to incorporate greater flexibility for developers designing software intended for the purposes described in sections 520(o)(1)(A), (B), (C), (D), and (E) of the FD&C Act. Specifically, the FDA should update its *Clinical and Patient Decision Support Software* guidance to allow AI/ML algorithms to be used incorporated in software functions that meet one of the requisite intended purposes described in sections 520(o)(1)(A), (B), (C), (D), or (E) of the FD&C Act.

3. Prioritize new guidance describing and clarifying FDA’s thinking on AI diagnostic devices.

Building on the work already done in the CADe Guidance Document, the FDA should clarify the rules and pathways to regulatory approval for AI diagnostic devices (i.e., CADx devices) for industry and innovators. This guidance should consider including the following recommendations for AI software developers:

- Advice on how to choose a regulatory path for a new AI software device, with an explanation as to when and whether a De Novo process or a 510(k) process may be appropriate. A 510(k) substantial equivalence determination should be based on the intended purpose of a predicate AI diagnostic device and on safety, not on the process of model development;
- A description of elements likely to necessitate a new 510(k) or De Novo application for a software update. These should be narrowly constructed and applicable to specific situations. Most software updates should not require a new submission for re-approval;
- Step-by-step guidance for how a developer can submit a Breakthrough Device Designation application for a new AI software device; and
- A detailed explanation of requirements to be included in a premarket application — in particular, those data and elements that the agency deems necessary to assess the unique safety and efficacy concerns associated with ML model development.

³⁷ Gottlieb, “Transforming FDA’s Approach to Digital Health.”

4. Expand the Software Precertification Program to include AI medical software developers.

The ongoing development of FDA’s Software Precertification Program is a positive sign of things to come. Unfortunately, the current working model (version 0.2) makes almost no mention of AI, and has yet to describe a clear path forward on software devices that make use of AI/ML technology.³⁸ In order to capture the full benefits of AI-based diagnostic devices, FDA will need to expand its focus beyond SaMDs; a good first step would be to commit to expanding the pilot program to include developers working on AI medical devices and diagnostic tools. By explicitly clearing the way for AI researchers and innovators to participate in the pilot program, the FDA will move toward a “governance learning” approach to regulation that will help actualize the benefits of these technologies for all Americans.³⁹

CONCLUSION

In drug development, it is the toxicity of new molecules and their efficacy as a potential treatment option, not the mechanism of action for a given compound, which determines value. In fact, we have very little understanding of the mechanism of action for the vast majority of molecular compounds on the market today. By comparison, unlike a compound that is ingested, an AI diagnostic software device poses considerably less risk to individuals; it only provides clinical advice — advice that can be verified by a health care professional. And indeed, it is that merging of human and computer comparative advantages that is most likely to benefit the health care ecosystem in the years ahead.⁴⁰

In order to get to that healthier future, however, the FDA must first adopt a clear and nuanced regulatory framework for AI diagnostic software. Such a framework should be rooted in the least burdensome principles, permitting the greatest level of flexibility and efficiency in device approval standards while maintaining high but reasonable standards of safety and effectiveness. In the same way that researchers look to results, so too must the FDA start turning the bulk of its attention to how new developments in software and AI devices can produce net beneficial outcomes, rather than focusing on the specific technological means by which those gains are achieved.

³⁸ To the agency’s credit, the second version of the Software Precertification Program Working Model does make an explicit call for input and comments regarding how best to clarify standards by which AI and ML technologies could potentially be included in future iterations of the program. (“Additionally, we seek comment on how to further clarify these elements and the associated domains to provide a least burdensome approach for software organizations to identify their processes/activities and outcomes. We also seek comments on elements or domains critical to evaluating the development of software functions using artificial intelligence and ML algorithms.”) *Developing Software Precertification Program*, p. 15.

³⁹ Although the focus of these comments is on the FDA’s draft interpretations of the changes to medical software device definitions, as described under the Cures Act, the importance of AI and ML to this space cannot be overstated. As such, we will simply note our intention to expand upon this recommendation in forthcoming comments specifically directed at the most recent version of the Software Precertification Program.

⁴⁰ A recent study showed that when presented with images of lymph node biopsies for detecting metastatic cancer, AI was better at finding true positives while a human physician was better at rejecting false positives (correctly rejecting a cancer diagnosis). When the human and AI skills were combined, correct metastatic cancer detection rose to an astounding 99.5 percent accuracy. See D. Wang et. al., “Deep learning for identifying metastatic breast cancer,” *arXiv preprint* (2016), <https://arxiv.org/pdf/1606.05718.pdf>.

The ongoing fusion of big data and AI will continue transforming the global economy.⁴¹ The health care industry has the potential to witness the most exciting and immediate gains from the application of AI — from improved individual patient health outcomes to the social benefits of reduced financial strains on our domestic health care system. Before AI can begin having this significant impact, however, the FDA needs to set the stage for the 21st century of personalized-medicine. When deciding how to regulate this broad category of emerging devices, the agency must take a technologically neutral approach, focusing its scrutiny on device outcomes, not on a quixotic quest to attempt to peer into the black box.⁴²

We would like to thank the FDA for the opportunity to comment on this issue and look forward to continued engagement on this and other topics.

⁴¹ Tom Standage, “The return of the machinery question,” *The Economist*, 25 June 2016, <https://www.economist.com/special-report/2016/06/25/the-return-of-the-machinery-question>.

⁴² Vijay Pande, “Artificial Intelligence’s ‘Black Box’ Is Nothing to Fear,” *The New York Times* (25 Jan. 2018), <https://www.nytimes.com/2018/01/25/opinion/artificial-intelligence-black-box.html>.