2021

Health AI for Good Rather Than Evil? The Need for a New Regulatory Framework for AI-Based Medical Devices

Sara Gerke
Penn State Dickinson Law, sgerke@psu.edu

Health AI for Good Rather Than Evil? The Need for a New Regulatory Framework for AI-Based Medical Devices

Sara Gerke[*]

Abstract:

Artificial intelligence (AI), especially its subset machine learning, has tremendous potential to improve health care. However, health AI also raises new regulatory challenges. In this Article, I argue that there is a need for a new regulatory framework for AI-based medical devices in the U.S. that ensures that such devices are reasonably safe and effective when placed on the market and will remain so throughout their life cycle. I advocate for U.S. Food and Drug Administration (FDA) and congressional actions. I focus on how the FDA could—with additional statutory authority—regulate AI-based medical devices. I show that the FDA incompletely regulates health AI-based products, which may jeopardize patient safety and undermine public trust. For example, the medical device definition is too narrow, and several risky health AI-based products are not subject to FDA regulation. Moreover, I show that most AI-based medical devices available on the U.S. market are 510(k)-cleared. However, the 510(k) pathway raises significant safety and effectiveness concerns. I thus propose a future regulatory framework for premarket review of medical devices, including AI-based ones. Further, I discuss two problems that are related to specific AI-based medical devices, namely opaque ("black-box") algorithms and adaptive algorithms that can continuously learn, and I make suggestions on how to address them. Finally, I encourage the FDA to broaden its view and consider AI-based medical devices as systems, not just devices, and focus more on the environment in which they are deployed.

[*] Assistant Professor of Law, Penn State Dickinson Law, Carlisle, PA; Leadership Team Member, The Project on Precision Medicine, Artificial Intelligence, and the Law (PMAIL), The Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School. I would like to thank Glenn Cohen, Nicholson Price, Carmel Shachar, and anonymous reviewers for helpful comments and feedback on an earlier version of this paper. I also wish to thank the organizers Mark Lemley, David L. Schwartz, and Christopher S. Yoo, commentators Brian Litt and Theodore Ruger, and attendees of the 2021 Junior Faculty Forum for Law and STEM for their feedback. All errors are my own.

INTRODUCTION

Artificial Intelligence (AI) is rapidly entering health care and may fundamentally change the way physicians practice medicine in the future. AI, especially its subset Machine Learning (ML), shows great potential to improve health care by enabling precision medicine, where patients receive better diagnoses and treatment recommendations tailored to their individual needs. The United States (U.S.) Food and Drug Administration (FDA) has already permitted marketing of over 340 AI/ML-based medical devices.[1]

According to one recent estimate, the global health AI market size is expected to increase more than nine-fold, from $6.9 billion in 2021 to $67.4 billion by 2027.[2] The COVID-19 pandemic has also hastened the adoption of health AI.[3] The enormous venture capital investment in the U.S. indicates the rising deployment of AI in the health care market.[4] In 2020, the U.S. accounted for the largest health AI market share in North America as it is home to several giant technology companies that are investing strongly in the development of health AI-based products, such as Microsoft, Google, and IBM.[5]

---

1 U.S. Food & Drug Admin., *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices* (Sept. 22, 2021), https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices. For non-FDA resources, see, for example, Stan Benjamens et al., *The State of Artificial Intelligence-Based FDA-Approved Medical Devices and Algorithms: An Online Database*, 3 NPJ DIGIT. MED., no. 118, at 2 (2020) (identifying 64 AI/ML-based medical devices that received FDA marketing authorization); Urs J. Muehlematter et al., *Approval of Artificial Intelligence and Machine Learning-Based Medical Devices in the USA and Europe (2015–20): A Comparative Analysis*, 3 LANCET DIGIT. HEALTH e195 (2021) (identifying 222 AI/ML-based medical devices that received FDA marketing authorization); Eric Wu et al., *How Medical AI Devices Are Evaluated: Limitations and Recommendations From an Analysis of FDA Approvals*, 27 NATURE MED. 582 (2021) (aggregating 141 FDA-approved AI devices); Am. Coll. Radiology, *Data Science Institute AI Central*, https://models.acrdsi.org (providing a database for FDA-cleared, AI-based medical devices in medical imaging) (last visited Mar. 19, 2022); *FDA-Approved A.I.-Based Algorithms*, MED. FUTURIST, https://medicalfuturist.com/fda-approved-ai-based-algorithms (last visited Mar. 19, 2022); and Casey Ross, *As the FDA Clears a Flood of AI Tools, Missing Data Raise Troubling Questions on Safety and Fairness*, STAT (Feb. 3, 2021), https://www.statnews.com/2021/02/03/fda-clearances-artificial-intelligence-data.

2 *Artificial Intelligence in Healthcare Market*, MARKETSANDMARKETS (Oct. 2021), https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-healthcare-market-54679303.html.

3 *Id.*; *see also* Sara Gerke et al., *Regulatory, Safety, and Privacy Concerns of Home Monitoring Technologies During COVID-19*, 26 NATURE MED. 1176 (2020) (raising concerns about the hasty adoption of home monitoring technologies); Carmel Shachar et al., *AI Surveillance during Pandemics: Ethical Implementation Imperatives*, 50 HASTINGS CTR. REP. 18 (2020) (discussing ethical implementation imperatives for AI surveillance during a pandemic).

4 Jason Schoettler, *Investors Are Piling Into Healthcare AI Start-Ups; There's Room for More*, VENTURE CAP. J. (Sept. 28, 2021), https://www.venturecapitaljournal.com/investors-are-piling-into-healthcare-ai-start-ups-theres-room-for-more.

5 *Artificial Intelligence in Healthcare Market*, *supra* note 2.

Health AI also poses new legal challenges, including ensuring the products' safety and effectiveness[6], obtaining informed consent[7], providing an adequate level of privacy protection[8], and comprehending and resolving liability issues[9]. As SpaceX and Tesla CEO/founder Elon Musk warned about AI in 2014 at the Massachusetts Institute of Technology's AeroAstro Centennial Symposium:

> I'm increasingly inclined to think that there should be some regulatory oversight, maybe at the national and international level, just to make sure that we don't do something very foolish. I mean with artificial intelligence we're summoning the demon.[10]

But how does one ensure that AI is good rather than evil? As Elon Musk correctly pointed out, the world needs proper regulatory oversight, and this starts at the national level. Such oversight is especially essential in health care to ensure that AI does not leave behind the most vulnerable populations, such as racial and ethnic minorities or people with disabilities, and benefits all patients. In particular,

---

6 *See, e.g.*, Boris Babic et al., *Direct-To-Consumer Medical Machine Learning and Artificial Intelligence Applications*, 3 NATURE MACH. INTEL. 283 (2021); W. Nicholson Price II, *Artificial Intelligence in Health Care: Applications and Legal Implications*, 14 SCITECH LAW. 10 (2017); W. Nicholson Price II, *Regulating Black-Box Medicine*, 116 MICH. L. REV. 421 (2017).

7 *See, e.g.*, Sara Gerke et al., *Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare*, *in* ARTIFICIAL INTELLIGENCE IN HEALTHCARE 295, 301 (Adam Bohr & Kaveh Memarzadeh eds., 1st ed. 2020); I. Glenn Cohen, *Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?* 108 GEO. L.J. 1425 (2020).

8 *See, e.g.*, Nathan Cortez, *Substantiating Big Data in Health Care*, 14 I/S: J.L. & POL'Y INFO. SOC'Y 61 (2017); Roger Allan Ford & W. Nicholson Price II, *Privacy and Accountability in Black-Box Medicine*, 23 MICH. TELECOMM. & TECH. L. REV. 1 (2016); Sara Gerke et al., *Ethical and Legal Aspects of Ambient Intelligence in Hospitals*, 323 JAMA 601 (2020); W. Nicholson Price II et al., *Shadow Health Records Meet New Data Privacy Laws*, 363 SCIENCE 448 (2019); W. Nicholson Price II & I. Glenn Cohen, *Privacy in the Age of Medical Big Data*, 25 NATURE MED. 37 (2019); Charlotte A. Tschider, *Regulating the Internet of Things: Discrimination, Privacy, and Cybersecurity in the Artificial Intelligence Age*, 96 DENV. L. REV. 87 (2018).

9 *See, e.g.*, W. Nicholson Price II, *Medical Malpractice and Black-Box Medicine*, *in* BIG DATA, HEALTH LAW, AND BIOETHICS 295 (I. Glenn Cohen et al., eds., 1st ed. 2018); A. Michael Froomkin et al., *When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning*, 61 ARIZ. L. REV. 33 (2019); George Maliha et al., *Artificial Intelligence and Liability in Medicine: Balancing Safety and Innovation*, 99 MILBANK Q. 629 (2021); W. Nicholson Price II et al., *How Much Can Potential Jurors Tell Us about Liability for Medical AI?* 62 J. NUCLEAR MED. 15 (2021); W. Nicholson Price II et al., *Potential Liability for Physicians Using Artificial Intelligence*, 322 JAMA 1765 (2019); Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. 1315 (2020); Kevin Tobia et al., *When Does Physician Use of AI Increase Liability?* 62 J. NUCLEAR MED. 17 (2020).

10 Massachusetts Institute of Technology, *Elon Musk at the MIT AeroAstro Centennial Symposium*, YOUTUBE, at 01:07:58 (July 2, 2015), https://www.youtube.com/watch?v=4DUbiCQpw_4&ab_channel=ElonMuskSoundBites.

regulators like the FDA need to reconsider the current regulatory paradigm to ensure that AI-based products classified as medical devices (AI-based medical devices) are reasonably safe and effective when placed on the market and will remain so throughout their life cycle. In this regard, several regulatory issues need to be thoroughly examined and have not received enough attention in the legal literature.[11] This Article endeavors to start to remedy that by focusing on unresolved regulatory issues of AI-based medical devices in the U.S. and proposing solutions.

In this Article, I advocate for FDA and congressional actions. I focus on how the FDA could—with additional statutory authority—regulate AI-based medical devices. The current regulatory framework for AI-based medical devices is not only complex and opaque at various points, but there are also recent developments in this area, which makes it even more difficult to keep track of the applicable framework. I go beyond the current literature[12] by unraveling, inter alia, the complex network of relevant provisions in the Federal Food, Drug, and Cosmetic Act (FDCA) and (draft) guidance documents related to AI-based medical devices, and thereby creating transparency in the field. Only by thoroughly cataloguing and analyzing the applicable framework can one identify loopholes and flaws, make suggestions, and thus refashion the discourse and move forward. I also discuss new regulatory proposals in the field and suggest ways to strengthen them. For many of my suggestions, the FDA will need to request additional statutory authority. Once the FDA has acquired enough information to design a new premarket and postmarket regulatory framework for AI-based medical devices that would ensure that such devices would be reasonably safe and effective throughout their life cycle, Congress should enact legislation to enable the FDA to fully implement its new framework. With the additional statutory authority and its new Digital Health Center of Excellence,[13] the FDA would have the necessary resources to tackle the regulatory challenges raised by AI.

I argue that the FDA incompletely regulates health AI-based products, which may jeopardize patient safety and undermine public trust. For example, the medical

11 A few regulatory issues have been discussed by, for example, Nathan Cortez, *Digital Health and Regulatory Experimentation at the FDA*, 21 YALE J.L. & TECH. 4 (2019); Barbara Evans & Pilar Ossorio, *The Challenge of Regulating Clinical Decision Support Software after 21st Century Cures*, 44 AM. J.L. & MED. 388 (2018); Price, *Artificial Intelligence in Health Care*, *supra* note 6; Price, *Regulating Black-Box Medicine*, *supra* note 6; and Nicolas P. Terry, *Assessing the Thin Regulation of Consumer-Facing Health Technologies*, 48 J.L. MED. & ETHICS 94 (2020).

12 *See, e.g.*, sources cited *supra* note 11.

13 The new Digital Health Center of Excellence is located in the Center for Devices and Radiological Health and is primarily focused on helping stakeholders get high-quality digital health products to patients. For more information, see Press Release, U.S. Food and Drug Admin., FDA Launches the Digital Health Center of Excellence (Sept. 22, 2020), https://www.fda.gov/news-events/press-announcements/fda-launches-digital-health-center-excellence.

device definition is too narrow, leaving out several risky health AI-based products that consequently are not subject to FDA regulation. Moreover, I show that although the 510(k)[14] premarket notification is the most frequently used type of premarket submission for AI-based medical devices, that pathway may not be sufficient to identify safety and effectiveness concerns. Hence, I propose a future regulatory framework for premarket review of medical devices, including AI-based ones, that would better ensure that devices are reasonably safe and effective when placed on the market. Further, I discuss two problems that are related to specific AI-based medical devices, namely opaque ("black-box") algorithms and "adaptive" algorithms that can continuously learn, and I suggest ways to address them. I also encourage the FDA to broaden its view and consider AI-based medical devices as systems, not *just* devices, and focus more on the environment in which they are deployed. This system view is essential to ensure that AI-based medical devices are reasonably safe and effective and benefit patients.

This Article proceeds in five Parts. Part I briefly explains relevant terms in computer science. It also provides an overview of the potential benefits of health AI-based products.

Part II establishes that the current medical device definition, FDCA section 201(h)(1),[15] is too narrow. I argue that several risky health AI-based products currently fall outside of the FDA's jurisdiction, such as certain clinical decision support (CDS) software functions. I propose that all CDS should be considered a priori medical devices under FDCA section 201(h)(1), and thus that Congress should consider amending the FDCA accordingly by deleting FDCA section 520(o)(1)(E).[16] I also suggest that Congress could amend the medical device definition to clearly include AI-based mortality prediction models and other models that are intended for use in the prediction or prognosis of disease or other conditions.

Part III shows that the FDA cleared most AI-based medical devices currently available on the U.S. market via the 510(k) pathway, raising significant safety and effectiveness concerns. It also examines the new 510(k) reforms. In particular, I argue that the new Safety and Performance Based Pathway likely will not apply to AI-based medical devices in the next few years. Even if it were applicable, the new pathway is voluntary and thus manufacturers would still have the option to submit a Traditional, Special, or Abbreviated 510(k) instead. I therefore propose a future regulatory framework for premarket review of medical devices, including AI-based medical devices. If the new Safety and Performance Based Pathway proves to be effective, it should replace the other 510(k) pathways and become the only

14 Federal Food, Drug, and Cosmetic Act (FDCA) § 510(k), 21 U.S.C. § 360(k).
15 FDCA § 201(h)(1), 21 U.S.C. § 321(h)(1).
16 FDCA § 520(o)(1)(E), 21 U.S.C § 360j(o)(1)(E).

available 510(k) pathway. In addition, my proposal includes modifying the De Novo pathway to also cover low to moderate risk medical devices that have a predicate but would not be eligible for the new 510(k) pathway. Finally, I argue that the FDA's new Software Pre-Cert Program—envisioned by the agency as a voluntary pathway for precertified companies that develop Software as a Medical Device (SaMD)—comes with its own challenges.

Part IV focuses on issues related to specific AI-based medical devices. First, I discuss the problem of AI/ML-based medical devices that are inherently black boxes and explainable versus interpretable AI/ML. I argue that the FDA should demand AI/ML makers use an interpretable AI/ML system if a white-box model performs better than or as well as a black-box model. I also show that the focus on explainable AI/ML in health care is deceptive and argue that regulators like the FDA should instead focus on ensuring safety and effectiveness. This goal can be achieved, for example, by requiring at least clinical trials for AI/ML-based medical devices that have a higher risk level. However, for AI/ML-based medical devices intended to be used to allocate scarce resources, such as organs or ventilators, the FDA should demand AI/ML makers use interpretable AI/ML systems rather than black boxes.

Second, I focus on what I call the "update problem." AI/ML-based medical devices can only fully realize their potential if they continuously learn and adapt to novel situations. But how should regulators like the FDA make sure that these devices remain safe and effective throughout their life cycle and do not compromise patient safety? I argue that the FDA could implement a monitoring system, such as Sentinel, that continuously monitors AI/ML-based medical devices.

Part V discusses two aspects of the system view: (1) considering human-AI interaction and (2) improving patient outcomes. The FDA could require rigorous human factors testing for all AI-based medical devices that require premarket submission to demonstrate that users can read the labeling and use such devices correctly. The agency could also more frequently require AI makers to set up a training program with instructions on how to use their device and/or to include a detailed description of the recommended user training in the device labeling. Further, AI-based medical devices should not only be safe but should actually improve patient outcomes. This could be demonstrated by comparative studies that the FDA could require, where appropriate, either as a premarket or postmarket requirement, depending on whether the AI-based medical device in question is urgently needed on the market.

Finally, I conclude that much more thinking and work needs to be done to realize the potential of health AI and ensure that such products are reasonably safe and effective.

I. THE POTENTIAL OF HEALTH AI-BASED PRODUCTS

The term "artificial intelligence" (AI) was first coined in 1955 when the four computer scientists John McCarthy, Marvin Minsky, Claude Shannon, and Nathaniel Rochester applied for funding from the Rockefeller Foundation for a two-month, ten-man study of AI to be carried out in 1956 at Dartmouth College in Hanover, New Hampshire, in the U.S.[17] Since then, the term "AI" has been widely used with different meanings. For example, in a 2004 Article, McCarthy defined AI as follows:

> It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.[18]

The FDA refers to John McCarthy's definition.[19] There is no universal definition of AI to date, but the term is often used as an umbrella term that encompasses several subsets. In particular, its subset Machine Learning (ML) has become one of the most promising fields of computer science in recent years. ML uses algorithms to detect patterns in data.[20] Deep learning is a subset of ML that identifies data patterns by employing artificial neural networks with several layers.[21] Advances within deep learning are also major reasons for the success of health AI in recent years.

Many AI/ML algorithms are "black boxes," meaning that the estimated function relating inputs to outputs is difficult or impossible for humans to understand.[22] For example, algorithms labeled as "deep learning" are considered

17 JOHN MCCARTHY ET AL., A PROPOSAL FOR A DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE (Aug. 31, 1955), http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf.

18 John McCarthy, What Is Artificial Intelligence? 2 (Nov. 24, 2004) (unpublished manuscript), https://homes.di.unimi.it/borghese/Teaching/AdvancedIntelligentSystems/Old/IntelligentSystems_2008_2009/Old/IntelligentSystems_2005_2006/Documents/Symbolic/04_McCarthy_whatisai.pdf.

19 U.S. FOOD & DRUG ADMIN., PROPOSED REGULATORY FRAMEWORK FOR MODIFICATIONS TO ARTIFICIAL INTELLIGENCE/MACHINE LEARNING (AI/ML)-BASED SOFTWARE AS A MEDICAL DEVICE (SAMD): DISCUSSION PAPER AND REQUEST FOR FEEDBACK 4 (2019), https://www.fda.gov/media/122535/download.

20 Kun-Hsing Yu et al., *Artificial Intelligence in Healthcare*, 2 NATURE BIOMED. ENG'G 719, 720 (2018).

21 *Id.*

22 *See* Boris Babic et al., *Beware Explanations from AI in Health Care*, 373 SCIENCE 284, 284 (2021); Boris Babic & Sara Gerke, *Explaining Medical AI Is Easier Said Than Done*, STAT (July 21, 2021), https://www.statnews.com/2021/07/21/explainable-medical-ai-easier-said-than-done.

black-box AI/ML models.[23] The term "black boxes" can also refer to models that are not too complex to be understood by humans, but that are deliberately kept opaque by AI companies for intellectual property reasons.[24]

Most AI/ML algorithms are "adaptive"—they continuously learn and adapt to new conditions.[25] It is also possible to "lock" AI/ML algorithms in such a way that they do not change with use and provide the same outcome each time the same input data is applied to them.[26]

Computer vision is also a vital subset of AI that focuses on developing autonomous systems that can perform particular tasks that the human visual system can carry out, and in some cases even surpass the human system's ability to do so.[27] Computer vision is essential for the growth of augmented reality, a technology that is often associated with mobile games such as Pokémon Go and blends digital and physical environments.[28] Robotics is a branch of technology that deals with the development and design of physical robots.[29] Sometimes robotics is also considered a subset of AI, but experts in the robotic world find it more appropriate to see AI and robotics as separate fields that overlap in cases of artificially intelligent robots.[30]

Health AI-based products are already in use in the U.S., and many more products are expected to be developed and enter the market in the coming years. In particular, it is anticipated that health AI will be applied not only in clinics but also outside the traditional clinical setting.[31]

---

23 *See* sources cited *supra* note 22.

24 *See* GREGORY DANIEL ET AL., DUKE MARGOLIS CTR. FOR HEALTH POL'Y, CURRENT STATE AND NEAR-TERM PRIORITIES FOR AI-ENABLED DIAGNOSTIC SUPPORT SOFTWARE IN HEALTH CARE 14 (2019), https://healthpolicy.duke.edu/sites/default/files/2019-11/dukemargolisaienableddxss.pdf; Price, *Regulating Black-Box Medicine*, *supra* note 6, at 430.

25 *See* U.S. FOOD & DRUG ADMIN., *supra* note 19, at 3; Boris Babic et al., *Algorithms on Regulatory Lockdown in Medicine*, 366 SCI. 1202, 1203 (2019).

26 *See* sources cited *supra* note 25.

27 T. S. Huang, Computer Vision: Evolution and Promise (Sept. 13, 1996) (unpublished manuscript), http://cds.cern.ch/record/400313/files/p21.pdf; Priya Dialani, *Five Important Subsets of Artificial Intelligence*, ANALYTICS INSIGHT (May 14, 2020), https://www.analyticsinsight.net/five-important-subsets-of-artificial-intelligence.

28 Jameson Toole, *Combining Artificial Intelligence and Augmented Reality in Mobile Apps*, HEARTBEAT (June 7, 2019), https://heartbeat.fritz.ai/combining-artificial-intelligence-and-augmented-reality-in-mobile-apps-e0e0ad2cfddc; *Fourth Workshop on Computer Vision for AR/VR*, XR @ CORNELL (June 15, 2020), https://xr.cornell.edu/workshop/2020.

29 Dialani, *supra* note 27; Alex Owen-Hill, *What's the Difference Between Robotics and Artificial Intelligence?* ROBOTIQ BLOG (Mar. 11, 2020), https://blog.robotiq.com/whats-the-difference-between-robotics-and-artificial-intelligence.

30 *See* sources cited *supra* note 29.

31 *See, e.g.*, FROST & SULLIVAN, TRANSFORMING HEALTHCARE THROUGH ARTIFICIAL INTELLIGENCE SYSTEMS (2016), https://docplayer.net/36848717-Transforming-healthcare-through-artificial-intelligence-systems.html.

A. Clinical Application

Health AI-based products are already used by U.S. health care providers and are expected to be implemented more frequently in the clinical setting in the future. Health AI shows great promise in medical imaging and disease diagnostics. For example, Digital Diagnostic's AI-based medical device, IDx-DR, detects greater than mild levels of diabetic retinopathy in diabetic patients ages twenty-two and older.[32] The system includes a special camera used by primary care physicians to take images of patient retinas and upload them to a cloud server.[33] The system is considered "autonomous," meaning that its decision—either to refer the patient to an eye doctor or to rescreen in twelve months—does not need to be checked by the primary care physician who uses the system.[34] IDx-DR has been used in clinical care at over twenty sites across the U.S.[35] Another example is Imagen's OsteoDetect, a computer-aided diagnosis and detection software powered by AI that helps providers to detect wrist fractures.[36]

The hope is that health AI-based products will increasingly help health care providers to detect diseases earlier and make more accurate diagnoses. Alongside health AI, robotics is expected to experience a boom in the coming years.[37] According to one recent estimate, the global medical robots market accounted for $5.9 billion in 2020 and is expected to reach $12.7 billion by 2025, and the U.S. is

32 U.S. FOOD & DRUG ADMIN., DE NOVO SUMMARY CLASSIFICATION REQUEST FOR IDX-DR 1, 5 (2018), https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN180001.pdf; *FDA News Release: FDA Permits Marketing of Artificial Intelligence-Based Device to Detect Certain Diabetes-Related Eye Problems*, U.S. FOOD & DRUG ADMIN. (Apr. 11, 2018), https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye. For more information on IDx-DR, see *IDx-DR*, DIGITAL DIAGNOSTICS, https://dxs.ai/products/idx-dr/idx-dr-overview (last visited Mar. 19, 2022).

33 *FDA News Release: FDA Permits Marketing of Artificial Intelligence-Based Device to Detect Certain Diabetes-Related Eye Problems*, *supra* note 32.

34 *Id.*; *IDx-DR*, *supra* note 32.

35 Jack Carfagno, *IDx-DR, the First FDA-Approved AI System, is Growing Rapidly*, DOCWIRE NEWS (Nov. 12, 2019), https://www.docwirenews.com/docwire-pick/future-of-medicine-picks/idx-dr-the-first-fda-approved-ai-system-is-growing-rapidly.

36 U.S. FOOD & DRUG ADMIN., EVALUATION OF AUTOMATIC CLASS III DESIGNATION FOR OSTEODETECT: DECISION SUMMARY 1, 2 (2018), https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN180005.pdf; *FDA News Release: FDA Permits Marketing of Artificial Intelligence Algorithm for Aiding Providers in Detecting Wrist Fractures*, U.S. FOOD & DRUG ADMIN. (May 24, 2018), https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-algorithm-aiding-providers-detecting-wrist-fractures.

37 PWC, WHAT DOCTOR? WHY AI AND ROBOTICS WILL DEFINE NEW HEALTH 22 (2017), https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health/ai-robotics-new-health.pdf.

a key market player.[38] In particular, increased implementation of AI-assisted surgery appears likely in the future.[39] The use of autonomous systems as robot surgeons is also not far from reality. Considerable research resources are being invested in the development of smart surgical robots with different degrees of autonomy to perform technical tasks, such as suturing, localizing wounds, and removing tumors.[40] These innovations promise better results and wider access to specialized procedures for patients.[41]

Augmented reality is also anticipated to experience a strong upswing in the health care market in the next few years.[42] For example, the California-based company, EchoPixel, developed True 3D, an FDA-cleared augmented reality device software that provides an environment where health care professionals can view patient-specific holographic-like images of organs and tissues.[43] Medical imaging and diagnostics, alongside robotics and augmented reality, are just the beginning of many more potential clinical AI applications that may significantly change the way health care providers practice medicine.

B. Outside the Clinical Setting

In the 21st century, large amounts of health data are gathered from individuals not only in clinical settings but also in daily life, such as through the internet, health applications (apps), Fitbits, and other products. For example, a recent study predicts that the total amount of data created worldwide will grow from 79 zettabytes in 2021 to 181 zettabytes in 2025.[44] The use of big data, coupled with enhanced computing power, suggests that health AI will likely have rising

38 *Medical Robots Market*, MARKETSANDMARKETS (2021), https://www.marketsandmarkets.com/PressReleases/medical-robotic-systems.asp.

39 *See, e.g.*, Sebastian Bodenstedt et al., *Artificial Intelligence-Assisted Surgery: Potential and Challenges*, 36 VISCERAL MED. 450 (2020); Tom J. M. van Mulken et al., *First-In-Human Robotic Supermicrosurgery Using a Dedicated Microsurgical Robot for Treating Breast Cancer-Related Lymphedema: A Randomized Pilot Trial*, 11 NATURE COMMC'NS 757 (2020); *see also* Phil Britt, *How AI-Assisted Surgery Is Improving Surgical Outcomes*, ROBOTIC BUS. REV. (June 19, 2018), https://www.roboticsbusinessreview.com/health-medical/ai-assisted-surgery-improves-patient-outcomes (discussing the promise of AI-assisted surgery to improve surgical outcomes).

40 Elizabeth Svoboda, *Your Robot Surgeon Will See You Now*, 573 NATURE S110, S110 (2019).

41 *See id*.

42 *Augmented Reality (AR) In Healthcare Market - Forecasts from 2020 to 2025*, RES. & MKTS. (2020), https://www.researchandmarkets.com/reports/4801765/augmented-reality-ar-in-healthcare-market; Toole, *supra* note 28.

43 *See* Letter from Robert Ochs, Dir. Div. of Radiological Health, U.S. Food & Drug Admin. to Mark Job, EchoPixel Inc. (Mar. 3, 2017), https://www.accessdata.fda.gov/cdrh_docs/pdf17/K170167.pdf; ECHOPIXEL, https://www.echopixeltech.com (last visited Mar. 19, 2022).

44 Arne von See, *Volume of Data/Information Created, Captured, Copied, and Consumed Worldwide From 2010 to 2025 (in Zettabytes)*, STATISTA (June 7, 2021), https://www.statista.com/statistics/871513/worldwide-data-created.

importance in the future. Already today, the range of direct-to-consumer health AI-based apps and chatbots, on topics from diet guidance to psychological advice, is immense and is expected to increase even more in the next years.[45] For example, the health AI-powered chatbot, Ada, assesses users' most likely conditions based on their symptoms and recommends the next steps to seek appropriate care.[46] Another example is the pocket AI therapist, Youper, a self-help app designed by a San Francisco-based company that supports mental health.[47]

Wearable health care products such as smartwatches, patches, and fitness trackers are also in high demand, and the global market is expected to almost double from $16.2 billion in 2021 to $30.1 billion by 2026.[48] For example, in September 2018, the FDA permitted marketing of Apple's electrocardiogram (ECG) app, a consumer-facing medical device intended for use with the Apple Watch by people ages twenty-two and older that can create, store, record, display, and transfer a single channel ECG.[49] The FDA also authorized Apple's irregular rhythm notification feature, an app that is also intended for use with the Apple Watch and for notifying the user of possible atrial fibrillation (AFib).[50] Several companies are also working on the next future-of-health AI-based fitness products where virtual trainers plan a user's workout based on their individual preferences and needs, motivate the user to complete their workout, and recommend healthy eating.[51]

The boundaries between hospitals and homes are also becoming increasingly porous. The American population is aging, and with this demographic shift comes

45 Boris Babic et al., *supra* note 6, at 283; Gerke et al., *supra* note 7, at 301; Remy Franklin, *11 Surprising Mobile Health Statistics*, MOBIUS MD (Oct. 25, 2021), https://mobius.md/2021/10/25/11-mobile-health-statistics.

46 Ada Health, *Ada – Check Your Health*, APPLE APP STORE PREVIEW (2022), https://apps.apple.com/app/id1099986434?mt=8; Leontina Postelnicu, *ADA Health's Chief Medical Officer on AI and Building Trust in Digital Health Tools*, MOBIHEALTHNEWS (Nov. 29, 2019), https://www.mobihealthnews.com/news/emea/ada-health-s-chief-medical-officer-ai-and-building-trust-digital-health-tools.

47 YOUPER (2021), https://www.youper.ai (last visited Mar. 19, 2022); Youper, Inc., *Youper Online Therapy*, APPLE APP STORE PREVIEW (2022), https://apps.apple.com/us/app/youper-online-therapy/id1060691513.

48 *Wearable Healthcare Devices Market*, MARKETSANDMARKETS (2021), https://www.marketsandmarkets.com/Market-Reports/wearable-medical-device-market-81753973.html.

49 Letter from Angela C. Krueger, Dep. Dir., Eng'g & Sci. Rev., Off. Device Evaluation, U.S. Food & Drug Admin. to Donna-Bea Tillman, Apple Inc. (Sept. 11, 2018), https://www.accessdata.fda.gov/cdrh_docs/pdf18/DEN180044.pdf.

50 Letter from Angela C. Krueger, Dep. Dir., Eng'g & Sci. Rev., Off. Device Evaluation, U.S. Food & Drug Admin. to Donna-Bea Tillman, Apple Inc. DEN180042 (Sept. 11, 2018), https://www.accessdata.fda.gov/cdrh_docs/pdf18/DEN180042.pdf.

51 Corey Lewis, *How AI Fitness Technology Will Take Your Health to The Next Level*, 1AND1 LIFE (Oct. 17, 2019), https://www.1and1life.com/blog/ai-fitness-technology.

the need to develop new digital health products that enable individuals to live an independent and healthy life at home as long as possible.[52] Computer vision-driven ambient intelligence systems use video capture to gather and interpret physical activity data.[53] These systems will likely be increasingly used not only in hospitals but also in patients' homes in the future. Remote patient monitoring is predicted to experience a boom in the next few years.[54] Such products, including those powered by AI, can help physicians to remotely monitor their patients' health conditions, such as diabetes, asthma, and cardiovascular disease, while improving clinical efficiency and reducing costs.[55] For example, the start-up Current Health offers an AI-powered wireless device worn on a patient's upper arm that continuously tracks vital signs, such as pulse, respiratory rate, and temperature.[56]

Home monitoring technologies have also been increasingly used during the COVID-19 pandemic to reduce personal contacts and thus exposure to the virus.[57] Further, robots can be helpful assistants in the COVID-19 pandemic. For example, the San Francisco-based company, RobotLAB, developed a self-driving, humanoid robot, Cruzr, that is designed to be used in schools. Cruzr can measure the body temperature of up to sixty people in a minute and detect people who do not wear a face mask and alert the staff.[58]

II. NARROW MEDICAL DEVICE DEFINITION

A. Device Software Functions

Are health AI-based products classified as medical devices under U.S. law? This is a crucial question for manufacturers in particular, since medical devices usually must meet medical device requirements under the FDCA and are regulated by the FDA.[59] The term "medical device" is defined in FDCA section 201(h)(1) as follows:

> an instrument, apparatus, implement, machine, contrivance,

52 Gerke et al., *supra* note 3.

53 Gerke et al., *supra* note 8.

54 *Remote Patient Monitoring (RPM) Market*, MARKETSANDMARKETS (2021), https://www.marketsandmarkets.com/Market-Reports/remote-patient-monitoring-market-77155492.html; *The State of the Remote Patient Monitoring Market in 2019*, DEFINITIVE HEALTHCARE (2021), https://blog.definitivehc.com/remote-patient-monitoring-market-2019.

55 *The State of the Remote Patient Monitoring Market in 2019*, *supra* note 54.

56 CURRENT HEALTH, https://www.currenthealth.com/products-page/product-remote-patient-monitoring (last visited Mar. 19, 2022).

57 Gerke et al., *supra* note 3.

58 *Keeping Schools Virus-Safe Is Not an Easy Feat*, ROBOTLAB (2020), https://www.robotlab.com/pandemic-covid19-health-robots.

59 *See infra* Section II.C. and Section III.A.

> implant, in vitro reagent, or other similar or related article, including any component, part, or accessory, which is—
>
> (A) recognized in the official National Formulary, or the United States Pharmacopeia, or any supplement to them,
>
> (B) intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or
>
> (C) intended to affect the structure or any function of the body of man or other animals, and
>
> which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of its primary intended purposes. The term "device" does not include software functions excluded pursuant to section 520(o).[60]

In the context of health AI, it is particularly relevant whether software functions are classified as medical devices (device software functions). The FDA distinguishes between two relevant types of software functions related to medical devices: "Software in a Medical Device" (SiMD) and "Software as a Medical Device" (SaMD). SiMD is software that is integral to a medical device.[61] In contrast, SaMD is standalone software that is, on its own, a medical device.[62] In 2013, the International Medical Device Regulators Forum (IMDRF)—a volunteer group of medical device regulators from across the world, including the U.S., whose goal is to accelerate international medical device regulatory harmonization—recognized the increasing importance of software in health care and published a document on SaMD in which it defines the term as "software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device."[63] The FDA embraced this

60 FDCA § 201(h)(1), 21 U.S.C. § 321(h)(1).

61 *Software as a Medical Device (SaMD)*, U.S. FOOD & DRUG ADMIN. (Dec. 4, 2018), https://www.fda.gov/medical-devices/digital-health/software-medical-device-samd.

62 *Id.*

63 INT'L MED. DEVICE REGULS. F., SOFTWARE AS A MEDICAL DEVICE (SAMD): KEY DEFINITIONS 6 (2013), http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf. The IMDRF SaMD Working Group also published two other guidance documents related to SaMD. *See* INT'L MED. DEVICE REGULS. F., "SOFTWARE AS A MEDICAL DEVICE": POSSIBLE

definition and further clarified that it defines medical purposes "as those purposes that are intended to treat, diagnose, cure, mitigate, or prevent disease or other conditions."[64] Apple's irregular rhythm notification Apple Watch feature is an example of an AI/ML-based SaMD because it is standalone software intended for a medical purpose.[65] Another example of an AI/ML-based SaMD is IDx-DR, standalone software intended to be used to diagnose a medical condition, namely detecting greater than mild levels of diabetic retinopathy in diabetic adults.[66]

B. Non-Device Software Functions

To assess whether the FDA adequately regulates health AI-based products, it is important to look at the agency's statutory authority. Only by analyzing the law in-depth can one identify legal gaps that may jeopardize patient safety and undermine public trust.

FDCA section 201(h)(1) clarifies that there are certain software functions that do not fall under the medical device definition (non-device software functions) and are thus not subject to FDA regulation. FDCA section 520(o)(1)(A)–(E), added by the 21st Century Cures Act,[67] contains five categories of software functions that usually are *not* considered to be medical devices, namely software functions intended

> (A) for administrative support of a health care facility . . .;
>
> (B) for maintaining or encouraging a healthy lifestyle . . .;
>
> (C) to serve as electronic patient records . . .;
>
> . . . .
>
> (D) for transferring, storing, converting formats, or displaying

FRAMEWORK FOR RISK CATEGORIZATION AND CORRESPONDING CONSIDERATIONS (2014), http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf; INT'L MED. DEVICE REGULS. F., SOFTWARE AS A MEDICAL DEVICE (SAMD): APPLICATION OF QUALITY MANAGEMENT SYSTEM (2015), http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-151002-samd-qms.pdf. In addition, the FDA published guidance on SaMD in which the agency adopts another 2017 guidance by the IMDRF SaMD Working Group on the clinical evaluation of SaMD. *See* U.S. FOOD & DRUG ADMIN., SOFTWARE AS A MEDICAL DEVICE (SAMD): CLINICAL EVALUATION (2017), https://www.fda.gov/media/100714/download. For further information on the IMDRF, see INT'L MED. DEVICE REGULS. F. (2021), http://www.imdrf.org.

64 U.S. FOOD & DRUG ADMIN., *supra* note 19, at 2.

65 For more information on Apple's irregular rhythm notification feature, see *supra* Section I.B.

66 For more information on IDx-DR, see *supra* Section I.A.

67 21st Century Cures Act, Pub. L. No. 114–255, § 3060(a), 130 Stat. 1033, 1130-31 (2016) (codified at 21 U.S.C. § 360j(o)).

clinical laboratory test or other device data and results . . .; [and]

(E) [to support certain clinical decisions.][68]

The second and fifth categories are particularly relevant for health AI.

1. Software Functions Intended for Maintaining or Encouraging a Healthy Lifestyle

Under FDCA section 520(o)(1)(B), a software function is generally not covered by the term "medical device" in FDCA section 201(h)(1) if it is intended "for maintaining or encouraging a healthy lifestyle and is unrelated to the diagnosis, cure, mitigation, prevention, or treatment of a disease or condition . . . ."[69]

In September 2019, the FDA issued the guidance "Changes to Existing Medical Software Policies Resulting from Section 3060 of the 21st Century Cures Act" (Cures Act Guidance) in which the agency provides its current thinking and non-binding recommendations on FDCA section 520(o)(1)(B).[70] In particular, the FDA clarifies that its updated non-binding guidance "General Wellness: Policy for Low Risk Devices" (General Wellness Guidance) helps interpret FDCA section

68 *Id.* For exceptions, see FDCA § 520(o)(3), 21 U.S.C. § 360j(o)(3) ("Notwithstanding paragraph (1), a software function described in subparagraph (C), (D), or (E) of paragraph (1) shall not be excluded from the definition of device under section 201(h) if . . . (i) the Secretary makes a finding that use of such software function would be reasonably likely to have serious adverse health consequences . . .); and FDCA § 520(o)(4)(B)-(C), 21 U.S.C. § 360j(o)(4)(B)-(C) ("Nothing in this subsection shall be construed as limiting the authority of the Secretary to . . . (B) regulate software used in the manufacture and transfusion of blood and blood components to assist in the prevention of disease in humans; or (C) regulate software as a device under this Act if such software meets the criteria under section 513(a)(1)(C) [for Class III classification]"). But these exceptions are only for certain individual software functions. FDCA § 520(o)(2), 21 U.S.C. § 360j(o)(2) regulates products with multiple functions that contain at least one function that is not a medical device and one that meets the definition of a medical device. The FDA issued guidance for such products. *See* U.S. FOOD & DRUG ADMIN., MULTIPLE FUNCTION DEVICE PRODUCTS: POLICY AND CONSIDERATIONS—GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2020), https://www.fda.gov/media/112671/download.

69 FDCA § 520(o)(1)(B), 21 U.S.C. § 360j(o)(1)(B). For some of the problems of the FDA's regulatory framework for mobile health technologies before the 21st Century Cures Act, see Nathan G. Cortez et al., *FDA Regulation of Mobile Health Technologies*, 371 NEW ENG. J. MED. 372 (2014); and Nathan Cortez, *The Mobile Health Revolution?*, 47 U.C. DAVIS L. REV. 1137 (2014), who argues that without proper oversight, users will be flooded with mobile technologies that are unsafe and ineffective. A similar discussion is also continued after the 21st Century Cures Act. *See, e.g.*, Babic et al., *supra* note 6.

70 U.S. FOOD & DRUG ADMIN., CHANGES TO EXISTING MEDICAL SOFTWARE POLICIES RESULTING FROM SECTION 3060 OF THE 21ST CENTURY CURES ACT—GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2019), https://www.fda.gov/media/109622/download.

520(o)(1)(B).[71]

Under the General Wellness Guidance, wellness products are products that present a low risk to users' and other individuals' safety and are intended for general wellness use only.[72] The FDA defines two different categories of general wellness intended uses:

> (1) an intended use that relates to maintaining or encouraging a general state of health or a healthy activity, or
>
> (2) an intended use that relates the role of healthy lifestyle with helping to reduce the risk or impact of certain chronic diseases or conditions and where it is well understood and accepted that healthy lifestyle choices may play an important role in health outcomes for the disease or condition.[73]

The FDA explains in its Cures Act Guidance that products that are intended "for maintaining or encouraging a healthy lifestyle" under FDCA section 520(o)(1)(B) means products that fall within the first category of general wellness intended uses.[74] Thus, FDCA section 520(o)(1)(B) is fulfilled in cases where software functions maintain or encourage "a general state of health or a healthy activity" (e.g., physical fitness, sleep management, relaxation and stress management, weight management, self-esteem, mental acuity, or sexual function) and are "unrelated to the diagnosis, cure, mitigation, prevention, or treatment of a disease or condition."[75] For example, an AI-based mobile app that plays music to relax and soothe a user and to manage stress and an AI-based mobile app that actively monitors and trends exercise activity are covered by FDCA section 520(o)(1)(B) and thus are not considered to be medical devices.[76]

---

71 *Id*. at 4–5; U.S. FOOD & DRUG ADMIN., GENERAL WELLNESS: POLICY FOR LOW RISK DEVICES—GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2019), https://www.fda.gov/media/90652/download.

72 U.S. FOOD & DRUG ADMIN., *supra* note 71, at 2.

73 *Id.* at 3.

74 U.S. FOOD & DRUG ADMIN., *supra* note 70, at 4–5.

75 *Id.* at 5; *see also* U.S. FOOD & DRUG ADMIN., *supra* note 71, at 3–4 (explaining the first category of general wellness intended uses).

76 *See* U.S. FOOD & DRUG ADMIN., *supra* note 70, at 6–7; U.S. FOOD & DRUG ADMIN., *supra* note 71, at 7. The FDA defines the term "mobile app" as "a software application that can be executed (run) on a mobile platform (i.e., a handheld commercial off-the-shelf computing platform, with or without wireless connectivity), or a web-based software application that is tailored to a mobile platform but is executed on a server." Mobile platforms are "commercial off-the-shelf (COTS) computing platforms, with or without wireless connectivity, that are handheld in nature. Examples of these mobile platforms include mobile computers such as smart phones, tablet computers, or other portable computers;" *see* U.S. FOOD & DRUG ADMIN., POLICY FOR DEVICE SOFTWARE FUNCTIONS AND

There is a fine line between the first and second categories of general wellness intended uses since both categories involve claims about or related to "sustaining or offering general improvement to functions associated with a general state of health."[77] The difference is that the second category references diseases or conditions, while the first category does not.[78]

The second category of general wellness claims consists of two subcategories: "intended uses to promote, track, and/or encourage choice(s), which, as part of a healthy lifestyle, **may help to reduce the risk of**" or "**may help living well with** certain chronic diseases or conditions . . . ."[79] The claims should be generally accepted—i.e., the associations are described in official statements made by health care professional organizations, such as the American Heart Association, American Medical Association, and American College of Rheumatology, or in peer-reviewed scientific publications.[80]

In contrast to products that fall within the first category of general wellness intended uses, products that fall within the second category do not meet the requirements under FDCA section 520(o)(1)(B) since they relate to the prevention or mitigation of a disease or condition and are thus medical devices under FDCA section 201(h)(1).[81] An example is a health AI/ML-based SaMD that facilitates making healthy lifestyle choices such as eating a balanced diet that may help living well with the chronic disease type 2 diabetes.[82] Consequently, manufacturers need to think carefully about the intended use(s) of their health AI-based product, as this determines whether the product is classified as a medical device. The intended use may be shown, for example, by advertising materials, labeling claims, or manufacturers' or their representatives' written or oral statements.[83]

2. Clinical Decision Support Software

Under FDCA section 520(o)(1)(E), certain clinical decision support (CDS) software functions are excluded from the medical device definition in FDCA section 201(h)(1). FDCA section 520(o)(1)(E) reads:

> The term device, as defined in section 201(h), shall not include a

---

MOBILE MEDICAL APPLICATIONS: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF 4, 18 (2019), https://www.fda.gov/media/80958/download.

77 U.S. FOOD & DRUG ADMIN, *supra* note 71, at 3–4.

78 *Id.*

79 *Id.* at 4 (emphasis in original).

80 *Id.* at 5.

81 U.S. FOOD & DRUG ADMIN., *supra* note 70, at 5–6.

82 *See* U.S. FOOD & DRUG ADMIN., *supra* note 71, at 5.

83 U.S. FOOD & DRUG ADMIN., *supra* note 76, at 5.

> software function that is intended—
>
> . . . .
>
> (E) unless the function is 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, for the purpose of [criterion (1)]—
>
> (i) displaying, analyzing, or printing medical information about a patient or other medical information (such as peer-reviewed clinical studies and clinical practice guidelines) [criterion (2)];
>
> (ii) supporting or providing recommendations to a health care professional about prevention, diagnosis, or treatment of a disease or condition [criterion (3)]; and
>
> (iii) enabling such health care professional to independently review the basis for such recommendations that such software presents so that it is not the intent that such health care professional rely primarily on any of such recommendations to make a clinical diagnosis or treatment decision regarding an individual patient [criterion (4)].[84]

The FDA issued a draft guidance in September 2019 that intends to describe the agency's approach to CDS software functions (CDS draft guidance).[85] A software function is *CDS* under this guidance if the following criteria are met:

- Not intended to acquire, process, or analyze [criterion (1)];
- Intended for the purpose of displaying, analyzing, or printing medical information [criterion (2)]; and
- Intended for the purpose of supporting or providing recommendations [part of criterion (3)].[86]

84 FDCA § 520(o)(1)(E), 21 U.S.C. 360j(o)(1)(E).

85 U.S. FOOD & DRUG ADMIN., CLINICAL DECISION SUPPORT SOFTWARE: DRAFT GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF 5 (2019), https://www.fda.gov/media/109618/download.

86 *Id.* at 8 (alterations in original); *see infra* Figure 1.

CDS can be Device CDS or Non-Device CDS. Device CDS fails to meet part of criterion (3) ("to a health care professional) and/or all or part of criterion (4) ("enabling such health care professional to independently review the basis for such recommendations") and thus is a medical device.[87] *Non-Device CDS* meets all four criteria in FDCA section 520(o)(1)(E) and thus is not a medical device.[88]

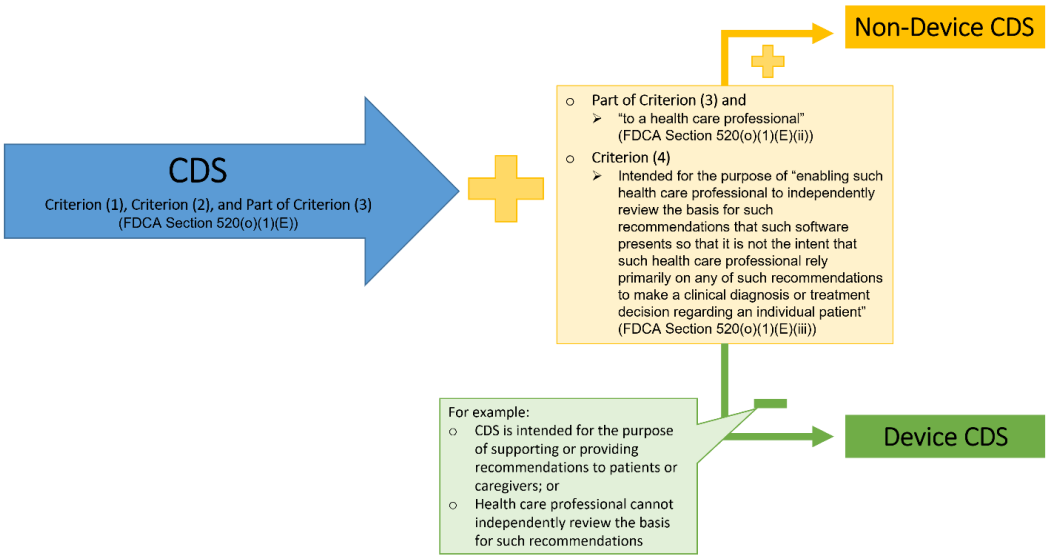

Figure 1: Device and Non-Device CDS

Blue shows the criteria—i.e., criterion (1), criterion (2), and part of criterion (3)—that software functions need to meet to be classified as CDS. Orange shows the criteria—i.e., part of criterion (3) and criterion (4)—that CDS need to additionally fulfill to be considered Non-Device CDS. Green shows Device CDS—i.e., they meet all criteria in the blue box but fail to fulfill part of criterion (3) and/or all or part of criterion (4) in the orange box.

The FDA describes in its CDS draft guidance, among other things, its current interpretation regarding criterion (4). In particular, the agency asks manufacturers of Non-Device CDS to describe—in plain language—their software functions as follows:

1) The purpose or intended use of the software function;

87 *See* U.S. FOOD & DRUG ADMIN., *supra* note 85, at 6–9.
88 *Id.*; *see infra* Figure 1.

2) The intended user (e.g., ultrasound technicians, vascular surgeons);

3) The inputs used to generate the recommendation (e.g., patient age and sex); and

4) The basis for rendering a recommendation.[89]

To describe the basis for a recommendation, irrespective of whether or not the software is proprietary and of the complexity of the software, the FDA clarifies that software developers "should describe the underlying data used to develop the algorithm and should include plain language descriptions of the logic or rationale used by an algorithm to render a recommendation."[90] The agency also explains that the sources underlying the basis of the recommendation or the sources supporting the recommendation should be identified, available to, and understandable by the intended health care professional user.[91] Examples of identified and available sources include published literature, clinical practice guidelines with the version or date, or information the CDS developer has provided to the intended health care professional user.[92] Understandable sources include data points, for example, the meaning of which is well understood by the intended health care professional user.[93] However, criterion (4) is not fulfilled in cases where the meaning of the information on which the recommendation is based cannot "be expected to be independently understood by the intended . . . user."[94] For example, if the inputs used to generate the recommendation were not identified, a health care professional would be unable "to independently review the basis for such recommendation that such software presents" and thus would be relying primarily upon it.[95]

3. The Problem of Health AI-Based Products

The FDA's CDS draft guidance indicates that AI-based CDS are *not* a priori Device CDS and can be considered Non-Device CDS as long as they are intended for "a health care professional" (criterion (3)) and for the purpose of "enabling such health care professional to independently review the basis for such recommendation that such software presents so that it is not the intent that such health care professional rely primarily on any of such recommendations to make a

89 U.S. FOOD & DRUG ADMIN., *supra* note 85, at 12.
90 *Id.*
91 *Id.*
92 *Id.*
93 *Id.*
94 *Id.*
95 *Id.*

clinical diagnosis or treatment decision regarding an individual patient" (criterion (4)). Two issues should be highlighted here. First, the term "health care professional" is important to distinguish between Device CDS and Non-Device CDS. The FDA does not define this term in its CDS draft guidance, but at least clarifies that CDS intended for the purpose of supporting or providing recommendations to patients or caregivers are Device CDS (and thus that patients and caregivers are *not* health care professionals).[96] Second, the FDA's current thinking suggests that health care professionals will likely be unable "to independently review the basis for such recommendation" in cases where the AI systems rely on algorithms that are "black boxes."[97] It will be challenging, or even impossible, for software developers of black-box AI/ML models, typically those that are labeled as "deep learning,"[98] to describe the basis for rendering a recommendation, such as the logic and rationale used by the algorithms. Manufacturers that keep their models opaque due to intellectual property reasons may also hesitate to describe the underlying data used to develop the algorithms. Thus, AI/ML algorithms, for which the inputs and logic are not explained, are Device CDS.[99]

But is criterion (4) ("independently review the basis") convincing enough to draw the line between Device CDS and Non-Device CDS? The FDA uses a risk-based approach to its regulation of Device CDS by applying the IMDRF framework for risk categorization of SaMD.[100]

State of the health care situation or condition	Significance of the information provided by the SaMD to the health care decision		
	Treat or diagnose	Drive clinical management	Inform clinical management
Critical	IV	III	II
Serious	III	II	I
Non-serious	II	I	I

Figure 2: SaMD Risk Categories Developed by the IMDRF[101]

96 *Id.* at 11.

97 For a definition of "black boxes," see *supra* Part I. For more information on black-box AI/ML models, see *infra* Part IV.

98 *Id.*

99 *See* U.S. FOOD & DRUG ADMIN., *supra* note 85, at 21, 23.

100 *Id.* at 6; *see infra* Figure 2.

101 INT'L MED. DEVICE REGULS. F., "SOFTWARE AS A MEDICAL DEVICE": POSSIBLE FRAMEWORK

Device CDS inform clinical management. The FDA intends to focus its regulatory oversight on those Device CDS that fall within the two red boxes. The agency does not currently intend to enforce applicable medical device requirements for some Device CDS that fall within the orange box.

The IMDRF framework in Figure 2 above explains two factors that are essential for the risk categorization of SaMD, which are (1) significance of the information provided by the SaMD to the health care decision and (2) state of the health care situation or condition. The first factor is divided into three categories—i.e., treat or diagnose, drive clinical management, and inform clinical management. The second factor is also divided into three categories—i.e., critical, serious, and non-serious.[102] There are four risk levels: level I (lowest risk) to level IV (highest risk).

The right column in Figure 2 is relevant for Device CDS. The IMDRF interprets the category *inform clinical management* as follows:

> Informing clinical management infers that the information provided by the SaMD will not trigger an immediate or near term action:
>
> - To inform of options for treating, diagnosing, preventing, or mitigating a disease or condition.
>
> - To provide clinical information by aggregating relevant information (e.g., disease, condition, drugs, medical devices, population, etc.).[103]

Thus, Device CDS exclusively fall within this category and "inform clinical management" since they are intended for the purpose of "supporting or providing recommendations . . . about prevention, diagnosis, or treatment of a disease or condition . . . ."[104] Device CDS intended to provide information, such as diagnostic or treatment options or aggregating relevant clinical information, may support

---

FOR RISK CATEGORIZATION AND CORRESPONDING CONSIDERATIONS, *supra* note 63, at 14 (Figure 2 has been slightly modified from its original form).

102 For more information on the IMDRF's interpretation of these terms, see INT'L MED. DEVICE REGULS. F., "SOFTWARE AS A MEDICAL DEVICE": POSSIBLE FRAMEWORK FOR RISK CATEGORIZATION AND CORRESPONDING CONSIDERATIONS, *supra* note 63, at 10–12.

103 *Id.* at 11.

104 FDCA § 520(o)(1)(E)(ii), 21 U.S.C. § 360j(o)(1)(E)(ii). Device CDS do not "drive clinical management" or "treat or diagnose," *see supra* Figure 2 columns two and three, since both categories refer to SaMD that go beyond "supporting or providing recommendations," *see* U.S. FOOD & DRUG ADMIN., *supra* note 85, at 14.

recommendations to health care professionals, caregivers, or patients.[105] They provide information that will not trigger a near term or immediate action—unlike SaMD that diagnose, screen, or detect a disease or condition.[106]

The FDA intends to focus its regulatory oversight on those Device CDS that inform clinical management for "critical" or "serious" health care situations or conditions, shown in the red boxes in Figure 2 above.[107] The agency does not currently intend to enforce applicable medical device requirements of the FDCA for some Device CDS that inform clinical management for "non-serious" health care situations or conditions, represented by the orange box in Figure 2.[108]

The IMDRF framework for risk categorization[109] is developed for SaMD but could also easily be applied to products that are not considered to be medical devices. Thus, criterion (4) of FDCA section 520(o)(1)(E) ("independently review the basis") would only be convincing to draw the line between Device CDS and Non-Device CDS if it ensured that at least all risk level I and level II products that inform clinical management for "critical" or "serious" health care situations or conditions (compare the red boxes in Figure 2) were classified as medical devices under the FDCA and were thus subject to FDA regulation. However, unfortunately, this is not the case. It is easy to imagine AI-based CDS that, under current law, are considered Non-Device CDS but inform clinical management for "critical" or "serious" health care situations or conditions and thus could pose a risk to the safety of patients if they were not to function as intended.

As an example, consider Watson for Oncology developed by IBM.[110] Watson for Oncology is CDS that assesses information from a patient's medical record and uses AI algorithms to provide physicians with individualized cancer treatment recommendations.[111] Watson did not undergo FDA review since it is considered Non-Device CDS that is intended for health care professionals who are able to "independently review the basis" for its recommendations.[112] However, the

105 U.S. FOOD & DRUG ADMIN., *supra* note 85, at 7, 13–14.

106 *Id.* at 14.

107 *Id.* at 17.

108 *See infra* Section II.C.

109 *See supra* Figure 2.

110 IBM has recently sold main parts of its Watson Health business to Francisco Partners. *See* Casey Ross, *The Sale of Watson Health Assets Ends a Dark Chapter for IBM. For Its Buyer, the Opportunity Looks Brighter*, STAT (Jan. 21, 2022), https://www.statnews.com/2022/01/21/ibm-watson-health-francisco-partners.

111 *See* Gerke et al., *supra* note 7, at 301; *IBM Watson for Oncology*, IBM (2021), https://www.ibm.com/products/clinical-decision-support-oncology.

112 Jacqueline Mulryne et al., *What's the Deal With Watson? Artificial Intelligence Systems and Medical Software Regulation in the U.S. and EU*, MONDAQ (Feb. 27, 2017), https://www.mondaq.com/unitedstates/healthcare/571712/what39s-the-deal-with-watson-artificial-intelligence-systems-

supercomputer came under criticism in 2018 because of a STAT report that alleged it recommended "unsafe and incorrect" cancer treatments.[113] To IBM's credit, the erroneous recommendations were apparently corrected by the company before the release of the product and its use on real patients.[114] Nevertheless, in light of patient safety, one would like to see Watson and similar products classified as medical devices (i.e., Device CDS) under the FDCA and subject to FDA regulation so that manufacturers must provide reasonable assurance of their safety and effectiveness. STAT also reported previously that the 21st Century Cures Act was hoped to be the impetus for the FDA to fully regulate medical advisory tools like Watson.[115] But IBM reportedly had an extensive team of lobbyists pushing hard for proposals to vitiate regulatory obstacles facing health software.[116] Perhaps as a result of this lobbying, the 21st Century Cures Act introduced FDCA section 520(o) that excludes certain categories of software functions, including several CDS, from the medical device definition.[117]

If one applied the SaMD risk categories established in the IMDRF framework[118] to Watson for Oncology, the AI-based product would probably be classified as a risk level II product: Watson informs clinical management by providing cancer treatment recommendations to physicians, and the state of a cancer patient's health care situation or condition would be critical since accurate and timely diagnosis and treatment action would be vital to avoid death.[119] Thus, Watson and similar products are exactly the kinds of products that the FDA usually intends to focus its regulatory oversight on. However, such products currently slip

---

and-medical-software-regulation-in-the-us-and-eu; David D. Luxton, *Should Watson Be Consulted for a Second Opinion?*, 21 AMA J. ETHICS E131 (2019).

113 Casey Ross & Ike Swetlitz, *IBM's Watson Supercomputer Recommended "Unsafe and Incorrect" Cancer Treatments, Internal Documents Show*, STAT (July 25, 2018), https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments.

114 *See* Jo Cavallo, *Confronting the Criticisms Facing Watson for Oncology*, ASCO POST (Sept. 10, 2019), https://www.ascopost.com/issues/september-10-2019/confronting-the-criticisms-facing-watson-for-oncology; Ross & Swetlitz, *supra* note 113.

115 Casey Ross & Ike Swetlitz, *IBM to Congress: Watson Will Transform Health Care, So Keep Your Hands off Our Supercomputer*, STAT (Oct. 4, 2017), https://www.statnews.com/2017/10/04/ibm-watson-regulation-fda-congress; *see also* Gerke et al., *supra* note 7, at 307 (discussing the Watson scandal).

116 *See* sources cited *supra* note 115.

117 21st Century Cures Act, Pub. L. No. 114–255, § 3060(a), 130 Stat. 1033 (2016) (codified at 21 U.S.C. § 360j); *see* Gerke et al., *supra* note 7, at 307; Ross & Swetlitz, *supra* note 115 ("The company's fingerprints are all over legislation passed last year that exempted several types of health software from FDA jurisdiction. A former IBM executive helped draft the blueprint for the law.").

118 *See supra* Figure 2.

119 Critical situations or conditions are "situations or conditions where accurate and/or timely diagnosis or treatment action is vital to avoid death, long-term disability or other serious deterioration of health of an individual patient or to mitigating impact to public health." *See* INT'L MED. DEVICE REGULS. F., "SOFTWARE AS A MEDICAL DEVICE": POSSIBLE FRAMEWORK FOR RISK CATEGORIZATION AND CORRESPONDING CONSIDERATIONS, *supra* note 63, at 11.

off of the agency's radar due to the fact that they fulfill all four criteria of FDCA section 520(o)(1)(E) and are thus classified as Non-Device CDS.[120] Consequently, under this analysis, criterion (4) seems insufficient to draw the line between Device CDS and Non-Device CDS.

Another problem is AI-based prediction/prognosis models that are intended to aid health care professionals in their decision-making. Are such models CDS? Imagine, for instance, an AI-based model that leverages data from electronic health records—without analyzing medical images—for predicting the development of hospital-acquired pressure injuries among surgical critical care patients.[121] Based on its prediction, the AI-based model provides recommendations to clinicians as to which patient should be assigned a specialty bed—which cannot be given to all patients for cost reasons[122]—and which patient should receive in-depth skin assessments.

In this example, it seems relatively straightforward to determine the answer to the question of whether the software is CDS. Criterion (1) of FDCA section 520(o)(1)(E) is fulfilled since the AI-based prediction tool is not "intended to . . . analyze a medical image" for predicting the development of pressure injuries.[123] Criterion (2) is also fulfilled since the tool is intended for the purpose of "analyzing . . . medical information about a patient . . . ."[124] The AI-based prediction model is also intended to provide recommendations to clinicians as to which patient should be assigned a specialty bed to prevent the development of hospital-acquired pressure injuries and which patient should receive in-depth skin assessments to detect such injuries early and treat them at a reversible stage.[125] Hence, criterion (3) is also met since the tool is intended for the purpose of "supporting or providing recommendations to a health care professional about prevention, diagnosis, or treatment" of a hospital-acquired pressure injury.[126] Consequently, this AI-based prediction tool is considered CDS. Under current law, it would be classified as Device CDS only if the health care professional could not "independently review the basis for" its recommendations.[127]

Now consider an AI-based model that leverages data from electronic health

120 The FDCA has a few regulatory safeguards in place. *See*, *e.g.*, FDCA § 520(o)(3), (4)(B)-(C), 21 U.S.C. § 360j(o)(3), (4)(B)-(C). However, such exceptions are limited to particular software functions only.

121 *See, e.g.*, Jenny Alderden et al., *Predicting Pressure Injury in Critical Care Patients: A Machine-Learning Model*, 27 AM. J. CRITICAL CARE 461 (2018).

122 *See id.* at 461.

123 FDCA § 520(o)(1)(E), 21 U.S.C. § 360j(o)(1)(E).

124 FDCA § 520(o)(1)(E)(i), 21 U.S.C. § 360j(o)(1)(E)(i).

125 *See* Alderden et al., *supra* note 121, at 461.

126 FDCA § 520(o)(1)(E)(ii), 21 U.S.C. § 360j(o)(1)(E)(ii).

127 FDCA § 520(o)(1)(E)(iii), 21 U.S.C. § 360j(o)(1)(E)(iii); *see supra* Figure 1.

records—without analyzing medical images—for predicting six-month mortality among cancer patients.[128] Is this model CDS under FDCA section 520(o)(1)(E)? Criteria (1) and (2) are fulfilled since the prediction model is not "intended to . . . analyze a medical image" for predicting mortality, but it is intended for the purpose of "analyzing . . . medical information about a patient."[129] However, is this software also intended for the purpose of "supporting or providing recommendations . . . about prevention, diagnosis, or treatment of a disease or condition . . . [?]"[130]

This question is much more difficult to answer. The algorithm predicts whether a cancer patient is at high or low risk of dying within the next six months. The patient has already developed cancer, and thus the software is not intended for the purpose of supporting or providing recommendations about *prevention* of cancer.

The AI-based model is also not intended for the purpose of supporting or providing recommendations about *diagnosis* of a disease or condition since cancer has already been diagnosed in the patient. Instead, the model predicts that the patient could die within the next six months. Death may be the consequence of a disease or condition or several diseases or conditions but is *not* a disease or condition itself.

Further, one may argue that the output of the AI-based model may initiate a conversation between the physician and the patient about cancer treatment, and thus the software is at least indirectly intended for the purpose of supporting or providing recommendations about *treatment* of a disease. However, one may argue as well—probably much more convincingly—that the AI-based model's prediction is intended to initiate early end-of-life discussions between physicians and cancer patients at high risk of dying within the next six months. If one accepts the latter argument, then the software would not be intended for the purpose of supporting or providing recommendations about treatment of a disease or condition but rather the opposite—i.e., to stop treatment, cut costs, and start palliative care. Consequently, it is unclear whether part of criterion (3) is fulfilled, and thus whether the AI-based mortality prediction model is CDS.

If one assumes that such a model is intended for the purpose of "supporting or providing recommendations to a health care professional about prevention, diagnosis, or treatment of a disease or condition" and thus that it is CDS, then the classification of a Device or Non-Device CDS depends on whether the model is intended to enable a "health care professional to independently review the basis

128 *See, e.g.*, Ravi B. Parikh et al., *Machine Learning Approaches to Predict 6-Month Mortality Among Patients With Cancer*, 2 JAMA NETWORK OPEN e1915997 (2019).

129 FDCA § 520(o)(1)(E)(i), 21 U.S.C. § 360j(o)(1)(E)(i).

130 FDCA § 520(o)(1)(E)(ii), 21 U.S.C. § 360j(o)(1)(E)(ii).

for" its recommendations.[131]

However, if one assumes that the model is *not* intended for the purpose of "supporting or providing recommendations to a health care professional about prevention, diagnosis, or treatment of a disease or condition,"[132] then the model may already not be considered a medical device under FDCA section 201(h)(1). The software is not "intended for use in the diagnosis of disease or other conditions, or in the . . . treatment, or prevention of disease . . . ."[133] The software is then also not intended for use in the "cure" of cancer, but rather for identifying patients at high risk of dying within the next six months and thus for enabling an early end-of-life discussion between the physician and that patient. One may argue that the model is at least indirectly intended for use in the "mitigation" of disease since it may contribute to the start of palliative care and thus may support a patient's dying without pain. A convincing counterargument may be that the model only indirectly mitigates the symptoms of cancer (i.e., the pain) but not the disease itself. As a result, it is highly unclear whether mortality prediction models are medical devices under current law, and thus whether software developers need to comply with device requirements of the FDCA.

4. Amending Proposals

I have argued above[134] that criterion (4) of FDCA section 520(o)(1)(E)[135] is not convincing to draw the line between Device CDS and Non-Device CDS because it does not ensure that at least all risk level I and level II products that inform clinical management for "critical" or "serious" health care situations or conditions are classified as medical devices under the FDCA and are subject to FDA regulation. It is easy to imagine AI-based CDS that are considered Non-Device CDS, although they inform clinical management for "critical" or "serious" health care situations or conditions.[136] Such Non-Device CDS could pose a risk to the safety of patients if they were not to function as intended. I therefore propose that—irrespective of whether CDS is intended to enable health care professionals "to independently review the basis for such recommendations that such software presents"—all CDS should be considered a priori medical devices under FDCA section 201(h)(1). Congress should consider amending the FDCA accordingly by deleting FDCA section 520(o)(1)(E).[137]

131 FDCA § 520(o)(1)(E)(ii)-(iii), 21 U.S.C. § 360j(o)(1)(E)(ii)-(iii).
132 FDCA § 520(o)(1)(E)(ii), 21 U.S.C. § 360j(o)(1)(E)(ii).
133 FDCA § 201(h)(1)(B), 21 U.S.C. § 321(h)(1)(B).
134 *See supra* Section II.B.3.
135 *See supra* Figure 1.
136 *See supra* Figure 2.
137 FDCA § 520(o)(1)(E), 21 U.S.C. § 360j(o)(1)(E).

This proposal would promote patient safety since it would ensure that all risk level I and level II products that inform clinical management for "critical" or "serious" health care situations or conditions would be classified as medical devices under FDCA section 201(h)(1) and thus would be subject to FDA regulation. It would also eradicate the current regulatory gray zone of whether a particular CDS is or is not a medical device. Criterion (4) is too vague to draw the line between Device CDS and Non-Device CDS. AI companies are trying very hard not to fall under the medical device definition, arguing that their CDS is intended for health care professionals who are able to "independently review the basis" for its recommendations.[138] A proper premarket review can also be seen as a safeguard against "automation bias." Studies of human-computer interaction demonstrate that people tend to trust the machine, even if they have a reason to question it.[139] This is especially a danger in medicine as physicians are very busy.[140] So is it the physician who is currently the captain of the ship, or is it the CDS that is actually steering the ship? Furthermore, the proposal to classify all CDS as medical devices would simplify the current regulatory landscape and facilitate more transparency. Finally, the FDA could continue to focus its regulatory oversight on those Device CDS that inform clinical management for "critical" or "serious" health care situations or conditions and exercise its enforcement discretion for some Device CDS that inform clinical management for "non-serious" health care situations or conditions.[141]

For example, following this proposal, the AI-based CDS that leverages data from electronic health records for predicting the development of hospital-acquired pressure injuries among surgical critical care patients would be classified as a medical device, irrespective of whether the CDS is intended to enable the health care professional "to independently review the basis for" its recommendations.[142] It would be likely categorized as a risk level I SaMD since it informs clinical management for a "serious" health care situation or condition.[143] If patients' hospital-acquired pressure injuries are not detected and treated early, they can

138 Evans & Ossorio, *supra* note 11, at 390, 394 (arguing correctly that statements of intend by manufacturers or their representatives tend to be dispositive); *see also* Cortez, *supra* note 11, at 11 (arguing that the line between Device CDS and Non-Device CDS remains murky, as it has for decades).

139 Cortez, *supra* note 11, at 24. A recent FDA report also says, "Medical informatics experts expressed concern that providers may rely too heavily on CDS software to determine appropriate treatments." U.S. FOOD & DRUG ADMIN., REPORT ON RISKS AND BENEFITS TO HEALTH OF NON-DEVICE SOFTWARE FUNCTIONS (2020), https://www.fda.gov/media/143795/download.

140 *Id.*

141 *See supra* Figure 2 (orange box); *see also infra* Section II.C (discussing the FDA's enforcement discretion).

142 For this particular example, see *supra* Section II.B.3.

143 *See supra* Figure 2.

become irreversible and may require costly interventions (e.g., skin biopsies).[144]

In addition, the uncertainty of whether AI-based mortality prediction models are medical devices under current law must be addressed immediately since more and more hospitals are using them.[145] Such models are likely risk level II products since they inform clinical management for "critical" health care situations or conditions—i.e., the respective disease, such as cancer, or condition is likely life-threatening and timely and accurate diagnosis and treatment action is vital to avoid death or other serious deterioration of a patient's health.[146] Thus, AI-based mortality prediction models may pose a risk to the safety of patients if they were not to function as intended. For example, a model could lead to the cessation of a patient's treatment if it incorrectly predicts the patient's early death. Consequently, AI-based mortality prediction models should be clearly classified as medical devices under FDCA section 201(h)(1) and subject to FDA regulation.

As a result, in addition to deleting FDCA section 520(o)(1)(E) in the form of an amendment, Congress could amend FDCA section 201(h)(1)(B)[147] as follows:

> intended for use in the diagnosis of disease or other conditions, *or in the prediction or prognosis of disease or other conditions or mortality*, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or

This broad definition would ensure that not only AI-based mortality prediction models but also other models that are intended for use in the prediction or prognosis of disease or other conditions would be clearly covered by the medical device definition. This proposal would promote patient safety and would also enable the FDA to continue focusing its regulatory oversight on those prediction/prognosis devices that may pose a moderate to high risk to patients and exercise enforcement discretion over those that are low risk.[148] A clear medical device definition would also help clarify the outer boundaries of the arena within

144 Serious situations or conditions are "situations or conditions where accurate diagnosis or treatment is of vital importance to avoid unnecessary interventions (e.g., biopsy) or timely interventions are important to mitigate long term irreversible consequences on an individual patient's health condition or public health." INT'L MED. DEVICE REGULS. F., *supra* note 63, at 11, 12.

145 *See, e.g.*, Rebecca Robbins, *An Experiment in End-Of-Life Care: Tapping AI's Cold Calculus to Nudge the Most Human of Conversations*, STAT (July 1, 2020), https://www.statnews.com/2020/07/01/end-of-life-artificial-intelligence.

146 Critical situations or conditions are "situations or conditions where accurate and/or timely diagnosis or treatment action is vital to avoid death, long-term disability or other serious deterioration of health of an individual patient or to mitigating impact to public health." INT'L MED. DEVICE REGULS. F., *supra* note 63, at 11.

147 FDCA § 201(h)(1)(B), 21 U.S.C. § 321(h)(1)(B).

148 For further discussion, see *infra* Section II.C.

which the FDA operates.[149]

Finally, Congress could amend FDCA section 520(o)(1)(B)[150] accordingly to reflect the previous change. The new version could read:

> for maintaining or encouraging a healthy lifestyle and is unrelated to the diagnosis, cure, mitigation, prevention, or treatment of a disease or condition *or to the prediction or prognosis of a disease or condition or mortality*;

C. Enforcement Discretion

1. The FDA's Current Approach

The FDA currently intends to exercise enforcement discretion over many health AI-based products. The agency follows a risk-based approach and aims to focus its regulatory oversight exclusively on those device software functions whose functionality might pose a risk to the safety of patients if they were not to function as intended.[151] The FDA does not at present intend to enforce compliance with the regulatory requirements of the FDCA for software functions that are low risk and are medical devices or may meet the medical device definition.[152] For example, the FDA intends to exercise enforcement discretion over AI-based wellness products that are medical devices—i.e., low risk products that fall within the second category of general wellness intended uses.[153] Another example is AI-based mobile apps that may meet the medical device definition but pose a low risk to patients, such as an AI-based mobile app that uses GPS location data to alert people with asthma of environmental conditions that may cause symptoms.[154]

The agency also at this time considers two types of Device CDS that inform clinical management for "non-serious" health care situations or conditions[155] as low risk and thus the FDA does not intend to enforce compliance with the applicable medical device requirements of the FDCA.[156] The first type is Device CDS that is intended for the purpose of supporting or providing recommendations to a *caregiver or a patient* to inform clinical management for a "non-serious" health care situation or condition, as long as the medical device is intended for the caregiver or patient to be able "to independently review the basis for such

149 PETER BARTON HUTT ET AL., FOOD AND DRUG LAW 77 (4th ed. 2014).
150 FDCA § 520(o)(1)(B), 21 U.S.C. § 360j(o)(1)(B).
151 U.S. FOOD & DRUG ADMIN., *supra* note 76, at 2, 10.
152 *See id.*, at 2, 9, 12.
153 *See* U.S. FOOD & DRUG ADMIN., *supra* note 71, at 7, 8; *supra* Section II.B.1.
154 *See* U.S. FOOD & DRUG ADMIN., *supra* note 76, at 9, 22.
155 *See supra* Figure 2 (orange box).
156 U.S. FOOD & DRUG ADMIN., *supra* note 85, at 16.

recommendations that such software presents . . . ."[157]

The second type is Device CDS that is intended for the purpose of supporting or providing recommendations to a *health care professional* to inform clinical management for a "non-serious" health care situation or condition.[158] This Device CDS is *not* intended to enable the health care professional "to independently review the basis" of its recommendations, and thus the health care professional relies primarily upon it.[159]

In contrast, the FDA currently intends to focus its regulatory oversight on such Device CDS that is intended for *a caregiver or patient* to inform clinical management for a "non-serious" health care situation or condition and is *not* intended for the caregiver or patient to be able "to independently review the basis" of its recommendations.[160] Thus, the FDA considers "opaque" ("black-box") Device CDS that are intended for the purpose of supporting or providing recommendations to *caregivers or patients* to inform clinical management for "non-serious" health care situations or conditions as riskier than similar Device CDS that are intended for *health care professionals*.[161] This distinction is convincing since health care professionals are usually clinically more experienced than patients and caregivers and thus may better manage the use of "opaque" Device CDS and will likely rely on additional sources to make a clinical diagnosis or treatment decision.

2. Proposal for a Regulatory Policy

If FDCA section 520(o)(1)(E) were deleted and FDCA section 201(h)(1)(B) and FDCA section 520(o)(1)(B) were amended by Congress as suggested,[162] the medical device definition would comprehensively include all CDS, AI-based mortality prediction models, and other models that are intended for use in the prediction or prognosis of disease or other conditions. These amending proposals would still enable the FDA to exercise its enforcement discretion over lower risk software functions that are medical devices or may meet the medical device definition. For example, the agency could exercise its enforcement discretion over low-risk prediction/prognosis devices and focus its regulatory oversight on those that pose a moderate to high risk to patients.

Concerning Device CDS, the FDA could decide not to enforce compliance

157 *Id. Compare* FDCA § 520(o)(1)(E)(ii), 21 U.S.C. § 360j(o)(1)(E)(ii) (criterion 3), *with* FDCA § 520(o)(1)(E)(iii), 21 U.S.C. § 360j(o)(1)(E)(iii) (criterion 4).

158 U.S. FOOD & DRUG ADMIN., *supra* note 85, at 16.

159 *Id*; *see* FDCA § 520(o)(1)(E)(iii), 21 U.S.C. § 360j(o)(1)(E)(iii) (criterion 4).

160 *Id.* at 17.

161 *See supra* Section II.B.3.

162 *See supra* Section II.B.4.

with the applicable medical device requirements of the FDCA for two types of Device CDS. First, the agency could exercise enforcement discretion over those Device CDS that are intended for a health care professional to inform clinical management for *non-serious* health care situations or conditions—irrespective of whether such Device CDS are intended to enable the health care professional to independently review the basis of their recommendations.[163]

Second, the FDA could also exercise enforcement discretion over those Device CDS that are intended for a caregiver or patient to inform clinical management for *non-serious* health care situations or conditions and are intended to enable the caregiver or patient to independently review the basis of their recommendations.[164] The risk of harm is relatively low in this scenario because independent review by the caregiver or patient of the basis of those Device CDS' recommendations would likely reveal at least obviously flawed ones at relatively minimal consequences of error.

Thus, in this way, the FDA could focus its regulatory oversight on those Device CDS that inform clinical management for *critical or serious* health care situations or conditions, and those Device CDS that are intended for a caregiver or patient to inform clinical management for *non-serious* health care situations or conditions but that are *not* intended to enable the caregiver or patient to independently review the basis of their recommendations.[165]

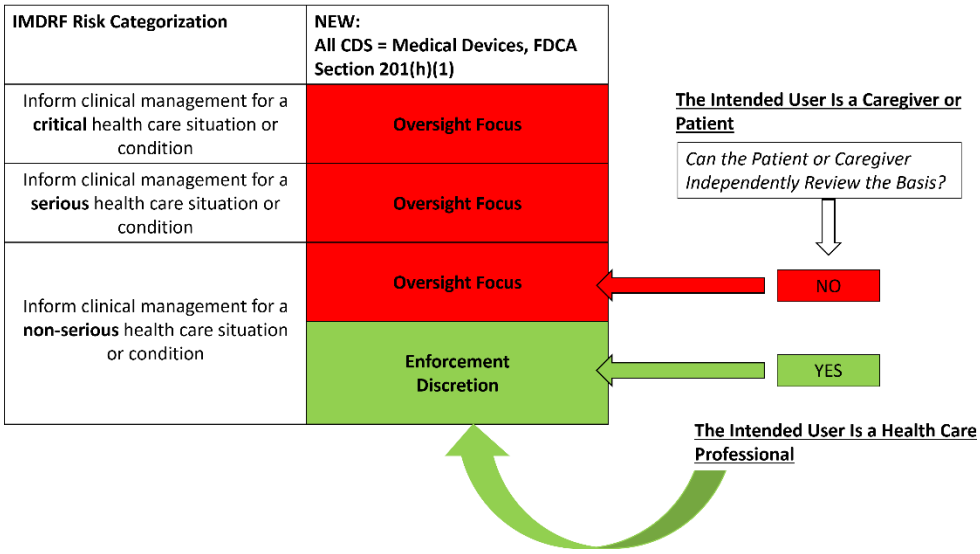

Figure 3: Proposal for a Regulatory Policy for Device CDS

163 *See infra* Figure 3.
164 *Id.*
165 *Id.*

"Oversight Focus" means that the FDA would focus its regulatory oversight on those Device CDS. "Enforcement Discretion" means that the FDA would not intend to enforce compliance with the applicable device requirements of the FDCA.

III. SAFETY AND EFFECTIVENESS CONCERNS OF 510(K) CLEARANCES

A. 510(k) Premarket Notification and Other Premarket Pathways

Manufacturers intending to bring an AI-based medical device on the market should follow four steps:

(1) discern the classification of the medical device and understand the applicable controls,

(2) choose and prepare the proper premarket submission,

(3) send the submission to the FDA and interact with the agency during its review, and

(4) comply with the applicable controls.[166]

The first step contains a prerequisite that manufacturers find out whether their health AI-based product is considered to be a medical device under FDCA section 201(h)(1) and, if so, whether the FDA intends to exercise enforcement discretion over their medical device.[167] If the health AI-based product is a medical device under the FDCA and the FDA intends to focus its regulatory oversight on such a device, manufacturers then need to figure out how the agency has classified their medical device.[168] Medical devices, including device software functions, are categorized into three classes based on their risk degree: Class I (lowest risk), Class II (moderate risk), and Class III (highest risk).[169] The correct classification of the medical device is essential to understand the applicable controls.[170] In general, Class I medical devices are subject to general controls, Class II medical devices are additionally subject to special controls, and Class III medical devices

166 *See How to Study and Market Your Device*, U.S. FOOD & DRUG ADMIN. (Oct. 14, 2020), https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/how-study-and-market-your-device.

167 *See supra* Section II.C.

168 *How to Study and Market Your Device*, *supra* note 166.

169 *Id.*; *see also* U.S. FOOD & DRUG ADMIN., *supra* note 76, at 10 (clarifying that device software functions can be categorized into the three classes of medical devices).

170 *How to Study and Market Your Device*, *supra* note 166.

are subject to general controls and premarket approval.[171] Examples of general controls include labeling requirements,[172] medical device reporting,[173] establishment registration and medical device listing,[174] and quality system regulation.[175]

As a second step, manufacturers need to choose and prepare the correct premarket submission. The class of the particular medical device determines the submission type. There are four common types of premarket submissions:

(1) 510(k) premarket notification,

(2) Premarket Approval (PMA),

(3) De Novo classification request, and

(4) Humanitarian Device Exemption (HDE).[176]

Class I and Class II medical devices, for which a PMA is not required, require a 510(k) unless they are exempt.[177] Sponsors must demonstrate in a 510(k) that their medical device is "substantially equivalent" to a legally marketed device (predicate device) that is not subject to PMA.[178] The term "substantially equivalent" or "substantial equivalence" is defined in FDCA section 513(i)(1)(A) as follows:

> the term "substantially equivalent" or "substantial equivalence" means, with respect to a device being compared to a predicate device, that the device has *the same intended use as the predicate device and* that the Secretary by order has found that the device—
>
> > (i) has the *same technological characteristics* as the predicate device, *or*

171 FDCA § 513(a)(1), 21 U.S.C. § 360c(a)(1); 21 C.F.R. pt. 814 (2022).
172 21 C.F.R. pt. 801 (2022).
173 *Id.* pt. 803.
174 *Id.* pt. 807.
175 *Id.* pt. 820.
176 *How to Study and Market Your Device*, *supra* note 166.
177 *Premarket Notification 510(k)*, U.S. FOOD & DRUG ADMIN. (Mar. 13, 2020), https://www.fda.gov/medical-devices/premarket-submissions/premarket-notification-510k. There are also a few Class III preamendment medical devices that may require a 510(k). *See Premarket Approval (PMA)*, U.S. FOOD & DRUG ADMIN. (May 16, 2019), https://www.fda.gov/medical-devices/premarket-submissions/premarket-approval-pma.
178 *Id.*; *see also How to Find and Effectively Use Predicate Devices*, U.S. FOOD & DRUG ADMIN. (Sept. 4, 2018), https://www.fda.gov/medical-devices/premarket-notification-510k/how-find-and-effectively-use-predicate-devices.

> (ii)(I) has *different technological characteristics* and the information submitted that the device is substantially equivalent to the predicate device contains information, including appropriate clinical or scientific data if deemed necessary by the Secretary or a person accredited under section 523, that *demonstrates that the device is as safe and effective as a legally marketed device, and* (II) *does not raise different questions of safety and effectiveness* than the predicate device.[179]

The FDA defines the term "intended use" for purposes of substantial equivalence as "the general purpose of the device or its function, and encompasses the indications for use."[180] The term "different technological characteristics" means "that there is a significant change in the materials, design, energy source, or other features of the device from those of the predicate device."[181]

A medical device cannot be launched on the market until the FDA has issued a letter that states that the medical device is "substantially equivalent" to the predicate device and thus has "cleared" the device for commercial distribution.[182] The submitter of a 510(k) has several options for selecting a predicate. Examples for a predicate include a preamendment device—a medical device that was legally marketed before May 28, 1976—a medical device that has been cleared via the 510(k) pathway, a medical device that was initially launched on the market as a Class III medical device and was later reclassified to a Class I or II, or a medical device that received marketing authorization through the De Novo pathway and that is not exempt from the premarket notification requirements.[183]

There are three 510(k) Programs: (1) Traditional, (2) Special, and (3) Abbreviated. The Traditional 510(k) Program can be used under all circumstances.[184] In contrast, the Special and Abbreviated 510(k) Programs were developed in 1998 to facilitate the 510(k) review process for particular types of submissions.[185] The Special 510(k) Program is an optional pathway and applicable

179 FDCA § 513(i)(1)(A), 21 U.S.C. § 360c(i)(1)(A) (emphasis added).

180 U.S. FOOD & DRUG ADMIN., THE 510(K) PROGRAM: EVALUATING SUBSTANTIAL EQUIVALENCE IN PREMARKET NOTIFICATIONS [510(K)]—GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF 16 (2014), https://www.fda.gov/media/82395/download.

181 FDCA § 513(i)(1)(B), 21 U.S.C. § 360c(i)(1)(B).

182 *Premarket Notification 510(k)*, *supra* note 177.

183 *See How to Find and Effectively Use Predicate Devices*, *supra* note 178; *Premarket Notification 510(k)*, *supra* note 177.

184 *How to Prepare a Traditional 510(k)*, U.S. FOOD & DRUG ADMIN. (Sept. 12, 2019), https://www.fda.gov/medical-devices/premarket-notification-510k/how-prepare-traditional-510k.

185 *Id.*; *Safety and Performance Based Pathway*, U.S. FOOD & DRUG ADMIN. (Aug. 27, 2021),

for certain well-defined changes by the manufacturer to an already legally marketed predicate.[186] The Abbreviated 510(k) Program is also optional and intended for submissions that rely on the use of special controls, guidance documents, and/or voluntary consensus standards.[187]

However, the majority of Class I medical devices and some Class II medical devices are exempt from the 510(k) premarket notification requirement.[188] Even if a medical device is exempt and the second and third steps—i.e., prepare and submit a 510(k) to the FDA and receive marketing clearance—are not required, manufacturers still need to comply with other general controls (fourth step), such as establishment registration and medical device listing.[189]

Class III medical devices usually require the most stringent type of premarket

---

https://www.fda.gov/medical-devices/premarket-notification-510k/framework-safety-and-performance-based-pathway. The FDA also issued nonbinding guidance for the Special 510(k) Program and Abbreviated 510(k) Program. *See* U.S. FOOD & DRUG ADMIN., THE ABBREVIATED 510(K) PROGRAM: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2019), https://www.fda.gov/media/72646/download; U.S. FOOD & DRUG ADMIN., THE SPECIAL 510(K) PROGRAM: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2019), https://www.fda.gov/media/116418/download. The FDA also issued nonbinding guidance for the content of premarket submissions for software devices, including stand-alone software; *see* U.S. FOOD & DRUG ADMIN., GUIDANCE FOR INDUSTRY AND FDA STAFF: GUIDANCE FOR THE CONTENT OF PREMARKET SUBMISSIONS FOR SOFTWARE CONTAINED IN MEDICAL DEVICES (2005), https://www.fda.gov/media/73065/download. This guidance document will soon be superseded by new guidance when final; the FDA has recently issued draft guidance on the premarket submissions' content of device software functions. *See* U.S. FOOD & DRUG ADMIN., CONTENT OF PREMARKET SUBMISSIONS FOR DEVICE SOFTWARE FUNCTIONS (2021), https://www.fda.gov/media/155022/download. The FDA has also issued nonbinding guidance for off-the-shelf software use in medical devices. *See* U.S. FOOD & DRUG ADMIN., OFF-THE-SHELF SOFTWARE USE IN MEDICAL DEVICES: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2019), https://www.fda.gov/media/71794/download.

186 *How To Prepare A Special 510(k)*, U.S. FOOD & DRUG ADMIN. (May 22, 2020), https://www.fda.gov/medical-devices/premarket-notification-510k/how-prepare-special-510k; U.S. FOOD & DRUG ADMIN, THE SPECIAL 510(K) PROGRAM: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2019), *supra* note 185, at 4.

187 U.S. FOOD & DRUG ADMIN., THE ABBREVIATED 510(K) PROGRAM: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF, *supra* note 185, at 3; *How to Prepare an Abbreviated 510(k)*, U.S. FOOD & DRUG ADMIN. (Jan. 23, 2019), https://www.fda.gov/medical-devices/premarket-notification-510k/how-prepare-abbreviated-510k.

188 *How to Study and Market Your Device*, *supra* note 168; *see also Class I and Class II Device Exemptions*, U.S. FOOD & DRUG ADMIN. (July 1, 2019), https://www.fda.gov/medical-devices/classify-your-medical-device/class-i-ii-exemptions (providing information on Class I and Class II device exemptions).

189 21 C.F.R. pt. 807 (2022); *Class I and Class II Device Exemptions*, *supra* note 188; *Device Classification Panels*, U.S. FOOD & DRUG ADMIN. (Aug. 31, 2018), https://www.fda.gov/medical-devices/classify-your-medical-device/device-classification-panels; *see also Medical Device Exemptions 510(k) and GMP Requirements*, U.S. FOOD & DRUG ADMIN. (2021), https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpcd/315.cfm (listing Class I and Class II exempt devices).

submission: a PMA.[190] To receive FDA PMA approval, the sponsor needs to provide valid scientific evidence that reasonably assures that the medical device is safe and effective for its intended use.[191] The FDA considers "valid scientific evidence," for example, to be evidence from partially controlled studies, well-controlled investigations, studies and objective trials without matched controls, or well-documented case histories carried out by qualified experts.[192]

The De Novo classification request is for novel medical devices of low to moderate risk, for which there is no predicate device.[193] The FDA will carry out a risk-based assessment for classification of such novel medical devices into Class I or II.[194] Novel medical devices that are classified into Class I or II via the De Novo pathway may also be marketed and used as predicate devices for prospective 510(k) submissions.[195] Originally, the manufacturer needed to submit a 510(k) and receive a "not substantially equivalent" determination from the FDA before being eligible for the De Novo pathway.[196] This was changed in July 2012, and manufacturers who determine that there is no predicate now also have the option directly to submit a De Novo classification request.[197] Thus, the new De Novo pathway is more efficient and less time-consuming. The FDA has also recently issued a final rule, effective since January 3, 2022, to establish regulations for the De Novo pathway that shall contribute greater clarity and transparency to the process, including the submission requirements and criteria for granting,

190 FDCA § 513(a)(1)(C), 21 U.S.C. § 360c(a)(1)(C); *How to Study and Market Your Device*, *supra* note 166; *see also* U.S. FOOD & DRUG ADMIN., *Premarket Approval (PMA)*, *supra* note 177 (explaining when a PMA is required).

191 U.S. FOOD & DRUG ADMIN., *Premarket Approval (PMA)*, *supra* note 177.

192 *PMA Clinical Studies*, U.S. FOOD & DRUG ADMIN. (May 22, 2020), https://www.fda.gov/medical-devices/premarket-approval-pma/pma-clinical-studies.

193 FDCA § 513(f)(1)-(2), 21 U.S.C. § 360c(f)(1)-(2); *De Novo Classification Request*, U.S. FOOD & DRUG ADMIN. (Jan. 7, 2022), https://www.fda.gov/medical-devices/premarket-submissions/de-novo-classification-request; *How to Study and Market Your Device*, *supra* note 166.

194 *De Novo Classification Request*, *supra* note 193.

195 *Id*. For an example (Proteus's wearable sensor), see Sara Gerke et al., *Ethical and Legal Issues of Ingestible Electronic Sensors*, 2 NATURE ELECS. 329, 331 (2019). The FDA also issued several guidances related to the De Novo classification process. *See, e.g.*, U.S. FOOD & DRUG ADMIN., ACCEPTANCE REVIEW FOR DE NOVO CLASSIFICATION REQUESTS: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2019), https://www.fda.gov/media/116945/download; U.S. FOOD & DRUG ADMIN., DE NOVO CLASSIFICATION PROCESS (EVALUATION OF AUTOMATIC CLASS III DESIGNATION): GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2021), https://www.fda.gov/media/72674/download.

196 *Evaluation of Automatic Class III Designation (De Novo) Summaries*, U.S. FOOD & DRUG ADMIN. (Jan. 25, 2022), https://www.fda.gov/about-fda/cdrh-transparency/evaluation-automatic-class-iii-designation-de-novo-summaries.

197 *See* FDCA § 513(f)(2), 21 U.S.C. 360c(f)(2); *Evaluation of Automatic Class III Designation (De Novo) Summaries*, *supra* note 196.

accepting, withdrawing, or declining a De Novo request.[198] The hope is that more manufacturers take advantage of the De Novo pathway for new technologies.[199] Finally, HDE is for Class III medical devices that are intended to help patients with rare diseases or conditions.[200]

B. Safety and Effectiveness Concerns

The FDA has already permitted marketing of over 340 AI/ML-based medical devices.[201] However, most AI-based medical devices currently available on the U.S. market were cleared via the 510(k) pathway. According to a new list of AI/ML-based medical devices marketed in the U.S., created by the FDA in September 2021, only 16 of 343 devices were authorized via the De Novo pathway, such as IDx-DR and OsteoDetect.[202] Only one device, QVCAD System for detecting mammography-occult lesions,[203] has so far received PMA approval. All other 326 AI/ML-based medical devices were 510(k)-cleared. For example, in January 2017, the FDA cleared Arterys Cardio DL as the first device software function that uses deep learning to analyze cardiovascular images captured by magnetic resonance scanners.[204] The device is intended to help radiologists, cardiologists, and other health care practitioners in making clinical decisions.[205] Another example is Viz.ai's notification-only, parallel workflow tool, Viz ICH, which the FDA cleared in March 2021.[206] Viz ICH uses an AI algorithm to analyze computed tomography (CT) images of the brain obtained in the acute setting and notifies a neurosurgical or neurovascular specialist where a suspected intracranial

198 *See* Medical Device De Novo Classification Process, 86 Fed. Reg. 54826 (Oct. 5, 2021); *De Novo Classification Request*, *supra* note 193.

199 *FDA in Brief: FDA Proposes Improvements to the De Novo Pathway for Novel Medical Devices to Advance Safe, Effective, and Innovative Treatments for Patients*, U.S. FOOD & DRUG ADMIN. (Dec. 4, 2018), https://www.fda.gov/news-events/fda-brief/fda-brief-fda-proposes-improvements-de-novo-pathway-novel-medical-devices-advance-safe-effective-and.

200 *How to Study and Market Your Device*, *supra* note 166.

201 *See* U.S. Food & Drug Admin., *supra* note 1.

202 *Id*. For more information on these two devices, *see supra* Section I.A.

203 Letter from Robert Ochs, Dir., Div. Radiological Health, U.S. Food & Drug Admin. to Robert M. Foley, Vice Pres., Regul. Affs., QView Medical, Inc. (Nov. 9, 2016), https://www.accessdata.fda.gov/cdrh_docs/pdf15/P150043A.pdf.

204 Letter from Robert Ochs, Dir., Div. Radiological Health, U.S. Food & Drug Admin. to Golnaz Moeini, Dir. Quality & Reg., Arterys Inc. (Jan. 5, 2017), https://www.accessdata.fda.gov/cdrh_docs/pdf16/K163253.pdf; *see also Cardio AI*, ARTERYS, https://arterys.com/clinicalapp/cardioapp (last visited Mar. 19, 2022) (providing more information about Cardio AI).

205 *See* Letter from Robert Ochs to Golnaz Moeini, *supra* note 204, at 16.

206 Letter from Thalia T. Mills, Dir., Div. Radiological Health, U.S. Food & Drug Admin. to Gregory Ramina, Dir. Regul. Affs., Viz.ai, Inc. (Mar. 23, 2021), https://www.accessdata.fda.gov/cdrh_docs/pdf19/K193658.pdf. For more information on Viz ICH, *see also Viz ICH*, VIZ.AI (2022), https://www.viz.ai/viz-ich.

hemorrhage has been detected.[207]

The fact that most AI/ML-based medical devices currently available on the U.S. market were 510(k)-cleared also reflects the general picture that 510(k) is the most frequently used type of premarket submissions. For example, in 2017, over 3000 medical devices received 510(k) clearances, representing over 80% of all cleared or approved medical devices.[208] Some Class I or III medical devices are cleared through the 510(k) pathway, but the majority of 510(k)-cleared medical devices are classified as Class II devices, and thus are of moderate risk.[209] For example, Arterys Cardio DL and Viz.ai's Viz ICH were both FDA cleared as Class II medical devices. However, this statistic is concerning since the 510(k) pathway has already been under criticism for a long time due to safety and effectiveness concerns.

1. The Institute of Medicine Report

The Institute of Medicine (IOM) published a report on the FDA 510(k) clearance process in 2011.[210] In its report, the IOM came to the following conclusion, among other things:

> The 510(k) clearance process is not intended to evaluate the safety and effectiveness of medical devices with some exceptions. The 510(k) process cannot be transformed into a premarket evaluation of safety and effectiveness as long as the standard for clearance is substantial equivalence to any previously cleared device.[211]

The IOM clearly communicates that "clearance" does *not* mean that the FDA "determined that the device is actually safe and effective . . . ."[212] The agency only confirms with a 510(k) clearance that the medical device is "substantially

207 Letter from Thalia T. Mills to Gregory Ramina, *supra* note 206.

208 *See* FDA Statement: Statement From FDA Commissioner Scott Gottlieb, M.D. and Jeff Shuren, M.D., Director of the Center for Devices and Radiological Health, on Transformative New Steps to Modernize FDA's 510(k) Program to Advance the Review of the Safety and Effectiveness of Medical Devices, (Nov. 26, 2018), https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-and-jeff-shuren-md-director-center-devices-and.

209 Thomas Sullivan, *Institute of Medicine Report Medical Devices and the Public's Health: The FDA 510(k) Clearance Process at 35 Years*, POL'Y & MED. (May 6, 2018), https://www.policymed.com/2011/07/institute-of-medicine-report-medical-devices-and-the-publics-health-the-fda-510k-clearance-process-a.html.

210 INST. MED., MEDICAL DEVICES AND THE PUBLIC'S HEALTH: THE FDA 510(K) CLEARANCE PROCESS AT 35 YEARS (2011).

211 *Id.* at 5.

212 *Id.*

equivalent" to, and thus as safe and effective as, the predicate.[213] However, the classification of preamendment devices, for example, did not comprise an assessment of whether an individual device was safe and effective.[214] Thus, many old predicates were never individually assessed for safety and effectiveness.[215] Moreover, data show that a considerable number of manufacturers still rely on old predicates today. Nearly 20% of all current 510(k) clearances are based on predicates that are older than 10 years.[216] For example, Arterys Oncology DL uses a deep learning algorithm to assist with lung and liver cancer diagnosis.[217] This device was FDA cleared in 2018, although it relied on a medical diagnostic application for manipulation, viewing, comparison, and 3-D visualization of medical images as a predicate to demonstrate "substantial equivalence," which in turn relied on another predicate, and so on, up to the reliance on preamendment devices marketed before May 28, 1976.[218]

It is important for users such as health care professionals and patients to understand that "clearance" does not mean "approval." As discussed above,[219] PMA approval is based on a successful demonstration of reasonable assurance of the safety and effectiveness of the medical device. This needs to be provided by valid scientific evidence—i.e., usually by clinical studies. However, according to the list published on the FDA's website, only one AI/ML-based medical device has received PMA approval so far.[220] In contrast, as mentioned, a 510(k) clearance only confirms that the medical device is "substantially equivalent" to the predicate. The 510(k) pathway usually does not require clinical evidence. In fact, the FDA generally requests clinical evidence for fewer than 10% of 510(k) submissions for moderate risk devices.[221] Thus, the agency often does not require AI makers to systematically document how the AI-based medical device was created, including the validation of its performance with another dataset than the training dataset.[222]

213 *Id.* at 5, 6.

214 *Id.* at 6.

215 *See id.* at 6.

216 FDA Statement: Statement From FDA Commissioner Scott Gottlieb, M.D. and Jeff Shuren, M.D., *supra* note 208.

217 *See* Letter from Robert Ochs, Dir., Div. Radiological Health, U.S. Food & Drug Admin. to John Axerio-Cilies, Chief Operating Officer, Arterys Inc. (Jan. 25, 2018), https://www.accessdata.fda.gov/cdrh_docs/pdf17/K173542.pdf; Thomas J. Hwang et al., *Lifecycle Regulation of Artificial Intelligence– and Machine Learning–Based Software Devices in Medicine*, 322 JAMA 2285 (2019).

218 *See* Letter from Robert Ochs to John Axerio-Cilies, *supra* note 217; Letter from Donald J. St. Pierre, Acting Dir., Div. Radiological Devices, U.S. Food & Drug Admin. to Casey Conry, Sr. Project Engineer, Siemens Medical Solutions USA, Inc. (Aug. 16, 2010), https://www.accessdata.fda.gov/cdrh_docs/pdf10/K101749.pdf.

219 *See supra* Section III.A.

220 *See* U.S. Food & Drug Admin., *supra* note 1.

221 Vinay K. Rathi & Joseph S. Ross, *Modernizing the FDA's 510(k) Pathway*, 381 NEW ENG. J. MED. 1891, 1892 (2019).

222 *See* Ross, *supra* note 1.

However, this is a critical step to ensure that such devices are safe and effective across various patient populations.[223]

Concerned that the 510(k) clearance process cannot assure safety and effectiveness, the IOM recommended that the FDA explore a new medical device regulatory framework for Class II devices:

> The Food and Drug Administration should obtain adequate information to inform the design of a new medical-device regulatory framework for Class II devices so that the current 510(k) process, in which the standard for clearance is substantial equivalence to previously cleared devices, can be replaced with an integrated premarket and postmarket regulatory framework that effectively provides a reasonable assurance of safety and effectiveness throughout the device life cycle. Once adequate information is available to design an appropriate medical-device regulatory framework, Congress should enact legislation to do so.[224]

The IOM also articulated certain attributes to include in the new framework. The process should be risk-based, clear, straightforward, predictable, fair, self-sustaining, self-improving, and based on sound science.[225] The process should also "facilitate innovation that improves public health by making medical devices available in a timely manner and ensuring their safety and effectiveness throughout their life cycle," and "should apply relevant and appropriate regulatory authorities and standards throughout the life cycle of devices to ensure safety and effectiveness."[226]

Further, the IOM states in its 2011 report that the De Novo process may potentially serve as "a better regulatory model for premarket review of Class II devices."[227] However, the IOM was also of the opinion that the De Novo process in its then-current form "is time-consuming and difficult for both the FDA and manufacturers to navigate."[228] Thus, the IOM recommended the FDA explore a modified De Novo process to assess the safety and effectiveness of Class II medical devices.[229] The IOM also suggested that the FDA "promptly call for PMA

223 *See id.*

224 INST. MED., *supra* note 210, at 8.

225 *Id.* at 9.

226 *Id.*

227 *Id.* at 11.

228 *Id*. Since the IOM's report in 2011, the De Novo Pathway has been changed and is now less time-consuming and more efficient. *See supra* Section III.A.

229 *See* INST. MED., *supra* note 210, at 11.

applications for or reclassify Class III devices that remain eligible for 510(k) clearance."[230] Concerning software, the IOM recommended the FDA "develop procedures that ensure the safety and effectiveness of software used in devices, software used as devices, and software used as a tool in producing devices."[231]

2. The 510(k) Reforms and Critique

To its credit, the FDA has committed to modernizing the 510(k) pathway—even though the agency did not follow the IOM's recommendation of developing a new medical device regulatory framework for Class II devices. In November 2018, the FDA published a statement in which it communicated, among other things, three major goals to ensure that 510(k)-cleared medical devices meet the gold standard for safety and effectiveness:

(1) promoting reliance on more modern predicates,

(2) "up-classifying" medical devices, and

(3) finalizing guidance establishing an alternative 510(k) pathway.[232]

The first goal of the FDA is to promote reliance on more modern predicates.[233] As discussed,[234] nearly one-fifth of all current 510(k) clearances are based on predicates that are more than ten years old. The FDA aims to drive manufacturers to rely on newer predicates that reflect modern technology and thereby promote innovation and improved safety.[235] For this reason, the agency suggested in its November 2018 statement to publish a list on its website of all cleared medical devices that are substantially equivalent to predicates that are older than ten years.[236] This list would intend to promote transparency and make it easier for users to decide between older and newer device type versions.[237] The FDA has not yet published such a list, perhaps due to the received criticism by some manufacturers who called the ten-year threshold "an arbitrary exclusion criterion."[238] While this

230 *Id.* at 13.

231 *Id*.

232 *See* FDA Statement: Statement From FDA Commissioner Scott Gottlieb, M.D. and Jeff Shuren, M.D., *supra* note 208.

233 *See id*.

234 *See supra* Section III.B.1.

235 *See* FDA Statement: Statement From FDA Commissioner Scott Gottlieb, M.D. and Jeff Shuren, M.D., *supra* note 208.

236 *See id*.

237 *See id.*

238 Ana Mulero, *FDA's Proposal to Limit Device Predicates Fails to Garner Industry Support*,

suggestion promotes newer predicates, it likely does not ensure that all newly cleared devices are reasonably safe and effective.

The FDA's second goal is to continue the efforts of "up-classifying" medical devices. "Up-classifying" means that the agency re-assigns a medical device to Class III and requires PMA if the device raises considerable safety concerns.[239] The FDA has already up-classified some previously 510(k)-cleared devices to Class III so that these devices can no longer be put on the market through the 510(k) pathway.[240] Examples include metal-on-metal hip implants, automated external defibrillators, and vaginal mesh for the treatment of pelvic organ prolapse.[241] From 2012 to 2018, the FDA up-classified a total of approximately 1,500 medical devices.[242]

The FDA is aware that up-classifying medical devices is resource- and time-intensive, and thus established a third goal: finalizing guidance establishing an alternative 510(k) pathway.[243] In its Medical Device Safety Action Plan, the FDA discussed the plan to "establish a voluntary, more modern 510(k) pathway for demonstration of safety and effectiveness for certain moderate risk devices."[244] Under this plan, manufacturers of particularly well-understood device types can use objective safety and performance criteria recognized or established by the FDA to demonstrate substantial equivalence.[245] In particular, this new pathway aims to provide more direct evidence of the performance and safety of a medical device.[246]

The agency achieved its goal and finalized its guidance "Safety and Performance Based Pathway" in September 2019.[247] The new pathway is optional and an expansion of the concept of the Abbreviated 510(k) Program for

---

REGUL. FOCUS (May 14, 2019), https://www.raps.org/news-and-articles/news-articles/2019/5/fdas-proposal-to-limit-device-predicates-fails-to.

239 *See* FDA Statement: Statement From FDA Commissioner Scott Gottlieb, M.D. and Jeff Shuren, M.D., *supra* note 208.

240 *See id.*

241 *Id.* For further information on the safety scandal of metal-on-metal hip implants, see Brent M. Ardaugh, *The 510(k) Ancestry of a Metal-on-Metal Hip Implant*, 368 NEW ENG. J. MED. 97 (2013).

242 *See* FDA Statement: Statement From FDA Commissioner Scott Gottlieb, M.D. and Jeff Shuren, M.D., *supra* note 208.

243 *See id.*

244 U.S. FOOD & DRUG ADMIN., MEDICAL DEVICE SAFETY ACTION PLAN: PROTECTING PATIENTS, PROMOTING PUBLIC HEALTH 12 (2018), https://www.fda.gov/media/112497/download.

245 *See id.* at 1; FDA Statement: Statement From FDA Commissioner Scott Gottlieb, M.D. and Jeff Shuren, M.D., *supra* note 208.

246 *See* FDA Statement: Statement From FDA Commissioner Scott Gottlieb, M.D. and Jeff Shuren, M.D., *supra* note 208.

247 U.S. FOOD & DRUG ADMIN., SAFETY AND PERFORMANCE BASED PATHWAY: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION (2019), https://www.fda.gov/media/112691/download.

particularly well-understood device types.[248] The aim is to ensure that new devices' performance characteristics are assessed against a set of transparent, objective, and well-validated performance and safety metrics.[249] The FDA has issued several final and draft guidance documents that identify performance criteria and testing methodologies for particular device types, and more will likely follow in the future.[250] Examples of device types for which the FDA has published final guidance documents are spinal plating systems,[251] conventional Foley catheters,[252] and cutaneous electrodes for recording purposes.[253] Manufacturers have the option to use the performance criteria suggested in the final guidance documents to support "substantial equivalence," rather than directly comparing their medical device with that of a predicate.[254] The new Safety and Performance Based Pathway is applicable to manufacturers who intend to submit a 510(k) when three requirements are simultaneously met:

(1) the device has the same indications for use as the predicate,

(2) the technological characteristics do not raise different questions of safety and effectiveness than the predicate, and

(3) the device meets all the FDA-recognized performance criteria.[255]

The new pathway is certainly laudable and seems promising but raises some issues, especially in the context of health AI. First, it is only available for those device types for which the FDA has identified performance criteria. Although the

248 *See id.* at 4; *Safety and Performance Based Pathway*, *supra* note 185. For more information on the Abbreviated 510(k) Program, *see supra* Section III.A.

249 FDA Statement: Statement From FDA Commissioner Scott Gottlieb, M.D. and Jeff Shuren, M.D., Director of the Center for Devices and Radiological Health, on Latest Steps to Strengthen FDA's 510(k) Program for Premarket Review of Medical Devices (Jan. 22, 2019), https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-and-jeff-shuren-md-director-center-devices-and-4.

250 *Safety and Performance Based Pathway*, *supra* note 185.

251 U.S. FOOD & DRUG ADMIN., SPINAL PLATING SYSTEMS – PERFORMANCE CRITERIA FOR SAFETY AND PERFORMANCE BASED PATHWAY: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2020), https://www.fda.gov/media/130867/download.

252 U.S. FOOD & DRUG ADMIN., CONVENTIONAL FOLEY CATHETERS – PERFORMANCE CRITERIA FOR SAFETY AND PERFORMANCE BASED PATHWAY: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2020), https://www.fda.gov/media/130865/download.

253 U.S. FOOD & DRUG ADMIN., CUTANEOUS ELECTRODES FOR RECORDING PURPOSES – PERFORMANCE CRITERIA FOR SAFETY AND PERFORMANCE BASED PATHWAY: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2020), https://www.fda.gov/media/130864/download.

254 *See, e.g.*, *id.* at 3.

255 *See Safety and Performance Based Pathway*, *supra* note 185; *supra* Section III.A.

FDA aims to publish more guidance documents identifying performance criteria for additional device types, this pathway targets those that are "well-understood." AI-based medical devices are newer products that have only entered the U.S. market in recent years. There remains much to learn about health AI, including the optimal data to use to train the model. Thus, it is unlikely that the FDA will identify performance criteria and publish corresponding guidance documents for AI-based medical device types in the near future. As a result, the new Safety and Performance Based Pathway will likely not be applicable to AI-based medical devices in the next years.

Second, even if such guidance documents for certain well-understood AI-based medical device types were published in the future, this new pathway is voluntary and therefore manufacturers would still have the option to submit a Traditional, Special, or Abbreviated 510(k) instead. Thus, a direct comparison of the performance of the medical device to that of a predicate would still be possible under the Traditional and Special 510(k) without the agency's determination that the device is actually safe and effective.

On January 8, 2021, during the last weeks of Donald Trump's presidency, then Health and Human Services Secretary Alex Azar signed a surprising notice that aimed to make permanent certain regulatory flexibilities provided during the COVID-19 pandemic by exempting particular medical devices from 510(k) premarket notification requirements.[256] This notice, published in the Federal Register on January 15, 2021, exempted seven Class I medical devices, namely different types of gloves, from the 510(k) premarket notification requirement with immediate effect.[257] The notice also suggested to exempt 83 Class II medical devices and one unclassified medical device from the 510(k) premarket notification requirement and requested public comments within sixty days of publication in the Federal Register.[258] Several of the eighty-three medical devices proposed to be exempt from FDA review carry out tasks using AI, such as computer assisted detection software to help identify bone fractures, respiratory illnesses, lesions suspicious for cancer, and other medical issues.[259]

The notice justified these exemptions by stating that the 510(k) premarket

256 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, 86 Fed. Reg. 4088 (Jan. 15, 2021).

257 *Id.* at 4088, 4096.

258 *Id.* at 4088, 4096–98.

259 *See id.* at 4096–98; Casey Ross, *"Slippery Slope Territory": Health Officials Propose Waiving Regulatory Review of Medical AI Tools*, STAT (Jan. 16, 2021), https://www.statnews.com/2021/01/16/slippery-slope-territory-health-officials-propose-waiving-regulatory-review-of-medical-ai-tools.

notification "is no longer necessary to assure the safety and effectiveness of those devices."[260] Apparently such devices listed in the notice were associated with no or only few adverse events.[261] However, adverse events are tricky to detect in many AI-based medical devices since they interact with physicians. It can take time to identify health AI problems, such as hidden biases, and the absence or rarity of reported adverse events does not mean that the devices work as promised.[262] As argued above and below,[263] the FDA needs to tighten, rather than relax, its oversight of health AI to adequately protect patients' health. In addition, this proposal appeared to contradict a newly released Action Plan for AI/ML-based SaMD issued by the FDA's Digital Health Center of Excellence in January 2021.[264]

It was unlikely, however, that the Biden Administration would further pursue this proposal.[265] Indeed, on April 16, 2021, the Department of Health and Human Services and the FDA issued two related notices in the Federal Register. The first notice refers to the seven Class I medical devices (i.e., the different types of gloves).[266] It clarifies that the previous determination that these devices "no longer require premarket notification . . . is flawed" and that it is appropriate to reverse it.[267] The second notice withdraws the proposed exemptions for the eighty-three Class II medical devices and one unclassified medical device from the 510(k) premarket notification requirement.[268] It highlights that the Department of Health and Human Services did not notify the FDA before issuing the January notice and that the proposal by the Trump Administration was made "without adequate

260 Making Permanent Regulatory Flexibilities Provided During the COVID-19 Public Health Emergency by Exempting Certain Medical Devices From Premarket Notification Requirements, 86 Fed. Reg. at 4096.

261 *See id.* at 4096-4098; *see also* Ross, *supra* note 259 (quoting Karandeep Singh, who criticizes the notice).

262 *See, e.g.*, Ross, *supra* note 259.

263 *See, e.g.*, *infra* Section III.B.3.

264 *Id.*; U.S. FOOD & DRUG ADMIN., ARTIFICIAL INTELLIGENCE/MACHINE LEARNING (AI/ML)-BASED SOFTWARE AS A MEDICAL DEVICE (SAMD) ACTION PLAN (Jan. 2021), https://www.fda.gov/media/145022/download. For more information on the new Action Plan, *see infra* Section IV.B.2. This also underscores the question about whether the FDA should become an independent federal agency distinct from the Department of Health and Human Services. *See, e.g.*, Eli Y. Adashi et al., *When Science and Politics Collide: Enhancing the FDA*, 364 SCI. 628, 630 (2019); Holly Fernandez Lynch, Steven Joffe & Matthew S. McCoy, *The Limits of Acceptable Political Influence Over the FDA*, 27 NATURE MED. 188, 189 (2021).

265 *See, e.g.*, Ross, *supra* note 1; Ross, *supra* note 259; Ronald A. Klain, *Regulatory Freeze Pending Review*, WHITE HOUSE (Jan. 20, 2021), https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/regulatory-freeze-pending-review.

266 Medical Devices; Class I Surgeon's and Patient Examination Gloves, 86 Fed. Reg. 20167 (Apr. 16, 2021).

267 *Id.* at 20167, 20170.

268 Making Permanent Regulatory Flexibilities Provided During the COVID-19 Public Health Emergency by Exempting Certain Medical Devices From Premarket Notification Requirements; Withdrawal of Proposed Exemptions, 86 Fed. Reg. 20174 (Apr. 16, 2021).

scientific support."[269] Both April notices are to be welcomed and emphasize the importance of regulation to ensure the safety and effectiveness of medical devices, including those based on AI.

3. Proposal for a Future Regulatory Framework for Premarket Review of Medical Devices, Including AI-Based Medical Devices

If the Safety and Performance Based Pathway is found to be effective, the FDA should replace the Traditional, Special, and Abbreviated 510(k) with the new Safety and Performance Based Pathway entirely, thus making it the only available 510(k) pathway for eligible medical devices, including AI-based medical devices.[270] Having only one 510(k) pathway—alongside the other premarket pathways such as De Novo and PMA—would also make the process more streamlined for manufacturers. In particular, the Abbreviated 510(k) has been used only rarely in the past,[271] and thus keeping it in addition to the new Safety and Performance Based Pathway would only make the process unnecessarily complicated.

Indeed, it seems that the FDA may be open to this proposal. In its November 2018 statement, the FDA mentioned that its goal is to make the Safety and Performance Based Pathway "the primary pathway for devices eligible for 510(k) review."[272] The FDA also said that the agency would like "this efficient new pathway to eventually supplant the practice of manufacturers comparing their new device technologically to a specific, and sometimes old, predicate device."[273]

My proposal to make the new Safety and Performance Based Pathway the only applicable pathway for 510(k)-eligible medical devices, including AI-based medical devices, would also require that the current De Novo pathway be modified. For example, it will probably take several more years for the FDA to identify performance criteria for some (unlikely all) AI-based medical device types, and even if the FDA identified such criteria, some devices would perhaps not be able to meet all of the identified performance criteria. The scope of the De Novo pathway should thus be expanded to also cover those new devices that would not be appropriate for the new Safety and Performance Based Pathway. Consequently, the De Novo pathway could be applicable in two circumstances. First, as is currently the case, for novel medical devices of low to moderate risk,

269 *Id.* at 20176.

270 *See infra* Figure 4.

271 Rathi & Ross, *supra* note 221, at 1893.

272 FDA Statement: Statement From FDA Commissioner Scott Gottlieb, M.D. and Jeff Shuren, M.D., *supra* note 208.

273 *Id.*

for which there is no predicate.[274] Second, for low and moderate risk medical devices that have a predicate, but where the new 510(k) Safety and Performance Based Pathway is not applicable because the FDA has, for example, not identified performance criteria for the respective device type.[275]

The FDA would need to design the exact differentiation criteria between the 510(k) Safety and Performance Based Pathway and the De Novo pathway, such as their precise scope, detailed requirements for submission, etc. As with the current regulatory framework, the majority of Class I medical devices and some Class II medical devices can still be exempt from the 510(k) premarket notification requirement as long as the exemptions are made with adequate scientific support. Congress should also enact legislation so that the suggested new regulatory framework for premarket review of medical devices, including AI-based medical devices, could be implemented.[276]

274 *See supra* Section III.A.
275 *See infra* Figure 4.
276 *See infra* Figure 4.

Traditional Premarket Pathways	Software Pre-Cert Program
(1) 510(k) Premarket Notification ➢ *The Safety and Performance Based Pathway:* Typically for class II medical devices **(2) PMA** ➢ For class III medical devices **(3) De Novo Classification Request** ➢ For novel medical devices of low to moderate risk, for which there is no predicate, or ➢ For low to moderate risk medical devices that have a predicate, but the Safety and Performance Based Pathway is not applicable (e.g., the FDA has not identified performance criteria for the device type) **(4) HDE** ➢ For class III medical devices that are intended to help patients with rare diseases or conditions	➢ **Voluntary** pathway ➢ For **precertified** companies of **SaMD** ➢ Perhaps someday also be expanded to **SiMD** or other **software that are accessories to hardware medical devices**

Figure 4: Proposal for a Future Regulatory Framework for Premarket Review of Medical Devices, Including AI-Based Medical Devices

The left column shows the traditional premarket pathways—i.e., 510(k) Premarket Notification, PMA, De Novo Classification Request, and HDE. The new framework would only have one 510(k) Pathway—i.e., the Safety and Performance Based Pathway. The new modified De Novo pathway would also apply in cases where a low or moderate risk device would have a predicate, but where the 510(k) Safety and Performance Based Pathway would not be applicable due to, for example, lack of FDA-identified performance criteria. The right column shows the Software Pre-Cert Program that would exist alongside the traditional premarket pathways.[277]

277 *See infra* Section III.C.

C. The New Software Pre-Cert Program

1. Overview

The FDA is currently carrying out a nine-company Pilot Program, launched in 2019, to explore how to best establish the so-called "Software Precertification (Pre-Cert) Program."[278] Companies that are involved in the testing phase include Johnson & Johnson, Apple, Roche, Samsung, and Google's sister-company Verily.[279] This Program aims to help the agency develop a future regulatory model for software-based medical devices.[280] The first version of the Software Pre-Cert Program is limited to SaMD. However, if the testing shows that the Program could also be leveraged for SiMD or other software that are accessories to hardware medical devices, the FDA will likely expand the Program.[281]

The Software Pre-Cert Program is designed as a voluntary pathway.[282] It would apply to manufacturers of SaMD that would be "precertified"—i.e., they would have demonstrated a culture of quality and organizational excellence—and would have agreed to monitor the real-world performance of their devices once they are launched on the U.S. market.[283] The new regulatory model aims to provide more efficient and streamlined regulatory oversight of SaMD and to promote innovation of digital health technologies.[284]

A key component of the Software Pre-Cert Program would be that the FDA or an FDA-accredited third-party would perform an Excellence Appraisal.[285] Companies would need to be granted a precertification status before being eligible for this pathway. They would need to demonstrate a culture of quality and organizational excellence.[286] At the moment, the FDA envisions the Excellence Appraisal to be based on five Excellence Principles:

---

278 *Digital Health Software Precertification (Pre-Cert) Program*, U.S. FOOD & DRUG ADMIN. (May 6, 2021), https://www.fda.gov/medical-devices/digital-health/digital-health-software-precertification-pre-cert-program.

279 *Id.*

280 *Id*.

281 *Id*.; U.S. FOOD & DRUG ADMIN., DEVELOPING A SOFTWARE PRECERTIFICATION PROGRAM: A WORKING MODEL 9, 10 (January 2019), https://www.fda.gov/media/119722/download. For the definition of SaMD and SiMD, *see supra* Section II.A.

282 U.S. FOOD & DRUG ADMIN., *supra* note 281, at 6.

283 *Id*. at 6, 37; *Digital Health Software Precertification (Pre-Cert) Program*, *supra* note 278.

284 U.S. FOOD & DRUG ADMIN., *supra* note 281, at 7; *Digital Health Software Precertification (Pre-Cert) Program*, *supra* note 278; *Precertification (Pre-Cert) Pilot Program: Frequently Asked Questions*, U.S. FOOD & DRUG ADMIN. (Sept. 14, 2020), https://www.fda.gov/medical-devices/digital-health-software-precertification-pre-cert-program/precertification-pre-cert-pilot-program-frequently-asked-questions.

285 U.S. FOOD & DRUG ADMIN., *supra* note 281, at 16–24. For more information on non-FDA certifiers, see Cortez, *supra* note 11, at 19 (arguing that it is a genuine innovation at the FDA).

286 U.S. FOOD & DRUG ADMIN., *supra* note 281, at 16–24.

(1) patient safety,

(2) product quality,

(3) clinical responsibility,

(4) cybersecurity responsibility, and

(5) proactive culture.[287]

Companies that demonstrate excellence in product development in all five Excellence Principles would additionally be categorized into one of two precertification levels.[288] Level 1 Pre-Cert would be granted to companies that have limited or no experience in delivering SaMD.[289] Level 2 Pre-Cert would be awarded to companies that have a proven track record in developing, providing, and maintaining safe and effective SaMD.[290]

Once companies are granted precertification status, they would be able to bring their SaMD with a streamlined premarket review or without any premarket review to the U.S. market. Whether a streamlined premarket review would be required would depend on the risk categorization of their SaMD and their precertification level.[291] The FDA is determining the information needed for a streamlined premarket review.[292] The goal is to allow faster market access while simultaneously ensuring safety and effectiveness.[293]

To determine the risk level of the product, the FDA envisions leveraging the IMDRF framework for risk categorization of SaMD.[294] SaMD with a risk level I would *not* need to undergo any FDA premarket review. High risk SaMD with a risk level III or IV would need to undergo a premarket review but a streamlined version. Risk level II SaMD could be brought to market with no premarket review or a streamlined one depending on the precertification level of the respective company. If the company were awarded a Level 1 Pre-Cert, then a streamlined premarket review would be necessary. However, if the company were granted a Level 2 Pre-Cert, then its product would *not* need to undergo any FDA premarket review. Figure 5 gives an overview of which SaMD would need to undergo a streamlined premarket review or no premarket review at all.

287 *Id.* at 11; *Digital Health Software Precertification (Pre-Cert) Program*, *supra* note 278.
288 U.S. FOOD & DRUG ADMIN., *supra* note 281, at 23.
289 *Id.*
290 *Id.*
291 *Id.* at 25.
292 *Id.* at 31–36; *Digital Health Software Precertification (Pre-Cert) Program*, *supra* note 278.
293 U.S. FOOD & DRUG ADMIN., *supra* note 281, at 31.
294 *Id.* at 25–30. For more information on the IMDRF framework, *see supra* Section II.B.3.

State of the health care situation or condition	Significance of the information provided by the SaMD to the health care decision		
	Treat or diagnose	Drive clinical management	Inform clinical management
Critical	IV	III	II
Serious	III	II	I
Non-serious	II	I	I

Figure 5: SaMD Risk Categorization Developed by the IMDRF[295] (modified to reflect whether an SaMD from a precertified company would need to undergo FDA premarket review under the Software Pre-Cert Program)

An SaMD that falls within one of the green boxes would not need to undergo any FDA premarket review. However, a streamlined premarket review would be required for an SaMD that falls within one of the red boxes. An SaMD that falls within one of the orange boxes would need to undergo a streamlined FDA premarket review if the company were Level 1 precertified. In contrast, if the company were Level 2 precertified, an SaMD that falls within the orange boxes would *not* need to undergo FDA premarket review.

The FDA envisions applying a Total Product Lifecycle (TPLC) approach.[296] Once the SaMD were marketed within the U.S., the precertified companies would monitor their real-world performance.[297] The FDA's approach aims to ensure that SaMD are safe and effective during their entire life cycle—from premarket development to postmarket performance.[298]

2. Analysis

The current Pre-Cert Pilot Program is a sensible approach to assess whether the new regulatory model for SaMD assures that the devices are reasonably safe and effective. The Pre-Cert Pilot Program provides the opportunity to fine-tune the Program and to solve many open questions. For example, what would happen if a precertified company were acquired by another company? Already during the

295 INT'L MED. DEVICE REGULS. F., *supra* note 63, at 14.

296 For more information on the TPLC approach, see U.S. FOOD & DRUG ADMIN., *supra* note 281, at 12–14.

297 *Digital Health Software Precertification (Pre-Cert) Program*, *supra* note 278. For more information on real-world performance, see U.S. FOOD & DRUG ADMIN., *supra* note 281, at 37–43.

298 U.S. FOOD & DRUG ADMIN., *supra* note 281, at 13. To the update problem, *see infra* Section IV.B.

testing phase, Fitbit, one of the nine participating companies in the Pilot, was acquired by Google for $2.1 billion.[299] The FDA has indicated that organizational restructuring or acquisition that impacts the assessed quality system and processes might trigger the need for an additional Excellence Appraisal.[300]

It will be interesting to see the Pre-Cert Pilot Program's final results and whether this Program that aims to establish trust and leverage transparency[301] can ensure that SaMD will be reasonably safe and effective throughout their life cycle.[302] This organization-based approach is undoubtedly an experiment with a new focus on assessing companies and products. It may hold valuable lessons for other countries and should be closely watched.[303] One point, however, is certain: It is a complicated endeavor, and the Pilot is already taking longer than initially expected.[304]

Perhaps one of the biggest challenges the agency currently faces is how the Software Pre-Cert Program would fit into the current traditional premarket pathways—i.e., 510(k), PMA, De Novo classification request, and HDE. For the Pilot, the FDA has leveraged the De Novo pathway.[305] The current Pilot is running in parallel with the traditional De Novo pathway. If a precertified company wants to place an SaMD on the U.S. market that is eligible for the De Novo process, it can submit a "Pre-Cert De Novo" during the testing period, and the FDA will run a traditional De Novo pathway in parallel.[306] Thus, the FDA can compare the Pre-Cert De Novo with the traditional De Novo and determine safety and effectiveness.

---

299 Erin Brodwin & Mario Aguilar, *Two Ways Fitbit Could Boost Google's Health Ambitions*, STAT (Jan. 15, 2021), https://www.statnews.com/2021/01/15/google-fitbit-clinical-trials; Rick Osterloh, *Google Completes Fitbit Acquisition*, GOOGLE (Jan. 14, 2021), https://blog.google/products/devices-services/fitbit-acquisition.

300 U.S. FOOD & DRUG ADMIN., *supra* note 281, at 15.

301 *Id.* at 7.

302 *See* Cortez, *supra* note 11, at 20–22 (expressing skepticism of the Software Pre-Cert Program); *see also* Terry, *supra* note 11, at 96 (worrying about the fact that the Software Pre-Cert Program will likely remove more consumer-facing devices from direct regulatory scrutiny).

303 Gerke et al., *supra* note 7, at 310.

304 For an updated timetable, see U.S. FOOD & DRUG ADMIN., DEVELOPING THE SOFTWARE PRECERTIFICATION PROGRAM: SUMMARY OF LEARNINGS AND ONGOING ACTIVITIES 2 (Sept. 2020), https://www.fda.gov/media/142107/download.

305 U.S. FOOD & DRUG ADMIN., SOFTWARE PRECERTIFICATION PROGRAM: REGULATORY FRAMEWORK FOR CONDUCTING THE PILOT PROGRAM WITHIN CURRENT AUTHORITIES (Jan. 2019), https://www.fda.gov/media/119724/download; *see also* U.S. FOOD & DRUG ADMIN., SOFTWARE PRECERTIFICATION PROGRAM: 2019 TEST PLAN (Jan. 2019), https://www.fda.gov/media/119723/download (describing the FDA's 2019 test plan).

306 U.S. FOOD & DRUG. ADMIN., SOFTWARE PRECERTIFICATION PROGRAM: 2019 TEST PLAN, *supra* note 305, at 3, 4; U.S. FOOD & DRUG ADMIN., SOFTWARE PRECERTIFICATION PROGRAM: REGULATORY FRAMEWORK, *supra* note 305, at 3. For more information, see *Software Precertification Program 2019 Mid-Year Update*, U.S. FOOD & DRUG ADMIN. (2019), https://www.fda.gov/media/129047/download.

To date, the Pilot has been restricted to SaMD of low to moderate risk for which there is no predicate, and thus are eligible for the De Novo pathway. SaMD with a predicate are not currently tested, except if they are eligible for 510(k) under a device classification created by the Pre-Cert De Novo. Only in this case could precertified companies submit a "Pre-Cert 510(k)" during the Pilot.[307]

The FDA has already come under criticism for the limited scope of the Pilot.[308] However, it seems that the FDA decided to implement the Pre-Cert Pilot Program under the De Novo pathway because the agency received pushback from Congress regarding its statutory authority to implement such a Program.[309] As a result, the FDA decided to leverage the De Novo pathway in the belief that the agency can test the Program within its current power.[310] However, even with this limited testing format, the FDA has been criticized by scholars and others for exceeding its statutory authority by implementing the Pre-Cert Pilot Program under the De Novo Pathway.[311]

Bakul Patel, the Director of the newly launched FDA's Digital Health Center of Excellence,[312] expects that the FDA will need to ask Congress for statutory authority to fully implement the Software Pre-Cert Program.[313] This statement also finds support in the law: The FDA draws its authority from the FDCA and its

307 U.S. FOOD & DRUG ADMIN., SOFTWARE PRECERTIFICATION PROGRAM: REGULATORY FRAMEWORK FOR CONDUCTING THE PILOT PROGRAM WITHIN CURRENT AUTHORITIES, *supra* note 305, at 3, 4; U.S. FOOD & DRUG ADMIN., SOFTWARE PRECERTIFICATION PROGRAM: 2019 TEST PLAN, *supra* note 305, at 2.

308 *See, e.g.*, David Lim, *FDA Targets De Novo Path to Shepherd Medical Software Through Pre-Cert*, MEDTECH DIVE (Jan. 8, 2019), https://www.medtechdive.com/news/fda-targets-de-novo-path-to-shepherd-medical-software-through-pre-cert/545519.

309 SCOTT THIEL & JASON BROOKE, REGUL. AFFS. PROS. SOC'Y, WILL THE FDA PRECERTIFICATION PILOT PROGRAM WORK? (May 2019), https://guidehouse.com/-/media/www/site/insights/healthcare/2019/raps--fda-precertification-pilot-program--52419.pdf.

310 U.S. FOOD & DRUG ADMIN., SOFTWARE PRECERTIFICATION PROGRAM: 2019 TEST PLAN, *supra* note 305, at 2; U.S. FOOD & DRUG ADMIN., *supra* note 281, at 2.

311 *See, e.g.*, Letter from Sen. Elizabeth Warren, Sen. Tina Smith & Sen. Patty Murray to Norman E. Sharpless, Acting Comm'r, U.S. Food & Drug Admin. & Jeffrey Shuren, Dir., Ctr. Devices & Radiological Health, U.S. Food & Drug Admin. 7, 8 (Oct. 30, 2019), https://www.warren.senate.gov/imo/media/doc/2019.10.30%20Letter%20with%20Senators%20Murray%20and%20Smith%20to%20FDA%20requesting%20additional%20information%20on%20the%20agency's%20software%20pre-certification%20pilot%20program..pdf; David Lim, *Top Democrats Question FDA Pre-Cert Program Safety, Statutory Authority*, MEDTECH DIVE (Oct. 11, 2018), https://www.medtechdive.com/news/top-democrats-question-fda-pre-cert-program-safety-statutory-authority/539389.

312 For more information on the new Digital Health Center of Excellence, *see supra* note 13 and accompanying text.

313 Greg Slabodkin, *FDA Still Trying to Fine-Tune Pre-Cert as Pilot Enters 2020*, MEDTECH DIVE (Mar. 25, 2020), https://www.medtechdive.com/news/fda-pre-cert-software-device-pilot-enters-another-year/574822.

amendments.[314] For any work the FDA wants to pursue outside of the FDCA and its amendments, the agency must obtain Congress's approval in the form of another amendment to the FDCA.[315] Looking at an earlier draft of the 21st Century Cures Act also suggests that the FDA must ask Congress for statutory authority to implement the Program fully. This earlier draft contained a provision that would have amended the FDCA and authorized the FDA to implement a new regulatory framework for health software,[316] but that provision was not incorporated into the final version of the Act.[317]

3. Implementation Proposal

So how could the Software Pre-Cert Program ideally be implemented in the future? It makes sense that the Software Pre-Cert Program would be implemented as a voluntary pathway, as it is currently designed. It is in the nature of things that not every company can be awarded a precertification status based on excellence. However, one needs to see in the long-term how many companies—e.g., a handful or hundreds—would ultimately use this pathway. In particular, the FDA needs to make sure that the Program would not de facto favor larger companies that have the necessary resources to undergo an Excellence Appraisal. The Program should also benefit small- and medium-sized enterprises. In the field of health AI, for example, there are many new start-ups that should also be given a realistic chance to get precertified and benefit from such a Program. Thus, it will be crucial for the FDA to closely watch the potential market effects of implementing the Software Pre-Cert Program. Such a Program could potentially bias the market toward established big players who are able to achieve a precertification status and thereby either quash innovation by new players or possibly over-incentivize intellectual property sales of health AI to precertified players. Thus, it will be crucial that the Software Pre-Cert Program distributes precertification status in a manner that promotes innovation at the same time as safety and effectiveness.

Suppose the FDA establishes the Software Pre-Cert Program's specific details, the Pilot proves to be effective, and the FDA has statutory authority. In that case, the agency theoretically would have two options regarding the Program's implementation. First, the agency could implement it similarly to the Pre-Cert Pilot Program, and even expand its scope so that precertified companies could submit, for example, a Pre-Cert 510(k) without the need for a device classification created by the Pre-Cert De Novo. At a later stage, the FDA could further expand the

314 THIEL & BROOKE, *supra* note 309, at 4.
315 *Id.*
316 *See* 21st Century Cures Act, H.R. 6, 114th Cong. § 2242 (2015).
317 *See* Cortez, *supra* note 11, at 25.

Program for SiMD and other software that are accessories to hardware medical devices. Second, the Software Pre-Cert Program could run completely separate from the traditional premarket pathways as an independent voluntary pathway with its own conditions.[318]

Irrespective of whether the FDA would choose the first or second option, the traditional premarket pathways would continue to be available for those companies that do not receive precertification status. Thus, it will be all the more important that the traditional pathways are robust and ensure that medical devices, including AI-based medical devices, are reasonably safe and effective when placed on the market. Consequently, the FDA needs to address the safety and effectiveness concerns of the traditional premarket pathways as soon as possible and implement—after receiving additional statutory authority—a new regulatory framework, such as the one that I have suggested above.[319]

IV. PROBLEMS RELATED TO SPECIFIC AI-BASED MEDICAL DEVICES

A. Black-Box AI/ML Models and Explainable Versus Interpretable AI/ML

1. The Problem

Another problem that needs to be addressed in the new suggested framework[320] is AI-based medical devices that are "black boxes." As explained above, many high-performing AI/ML systems rely on algorithms that are "black boxes."[321] Black-box algorithms are difficult or impossible for humans to understand.[322] Algorithms typically labeled as "deep learning" are black-box AI/ML models.[323] The term "black boxes" can also refer to algorithms that are deliberately black boxes because, for intellectual property reasons, developers do not want to disclose the details of how these algorithms work.[324] I focus here on the first group of algorithms, namely those that are inherently black boxes.

Noninterpretable black-box models have been shown to perform better than interpretable models in several practicable scenarios.[325] In particular, in health care, black-box AI/ML models often perform better, such as in image

318 *See supra* Figure 4.
319 *See supra* Section III.B.3.
320 *See supra* Figure 4.
321 *See supra* Part I.
322 Babic et al., *supra* note 22, at 284; Babic & Gerke, *supra* note 22.
323 *Id.*
324 Price, *Regulating Black-Box Medicine*, *supra* note 6, at 430.
325 Hongfang Liu et al., *AI Model Development and Validation*, *in* ARTIFICIAL INTELLIGENCE IN HEALTH CARE: THE HOPE, THE HYPE, THE PROMISE, THE PERIL 124 (Michael Matheny et al., eds., 1st ed. 2019).

recognition.[326] However, especially in Europe, there is a movement for explainable AI/ML since various scholars argue that the EU General Data Protection Regulation (2016/679)[327] contains a "right to explanation" of automated decision-making.[328] In contrast, the U.S. follows a more market-driven approach, and the FDA has already permitted marketing of several AI/ML-based medical devices that use noninterpretable black-box models.[329] For example, Imagen's OsteoDetect and Arterys Cardio DL both use deep learning.[330]

So which approach is the right one? Should regulators like the FDA continue to permit marketing of black-box AI/ML systems or only permit marketing of explainable and/or interpretable AI/ML?

One thing should be clear here: It is crucial to understand the difference between interpretable AI/ML and explainable AI/ML. As defined here, interpretable AI/ML uses a "white-box" model (i.e., a transparent system), such as a linear or simple decision tree model, instead of a black box.[331] The advantage of interpretable AI/ML algorithms is that they are open and understandable at a human level with reasonable effort.[332] In contrast, the term "explainable AI/ML" is understood here in connection with a black-box model that is used to make diagnoses or predictions.[333] A second explanatory algorithm—which is itself a white-box model—is developed that closely approximates the outputs of the black box.[334]

The issue with explainable AI/ML, however, is that because the second algorithm is usually not as accurate as the black box, it is normally used to develop

326 Babic et al., *supra* note 22.

327 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L. 119) 1.

328 For more information on this debate, see, for example, Andrew Burt, *Is There a 'Right to Explanation' for Machine Learning in the GDPR?*, IAPP (June 1, 2017), https://iapp.org/news/a/is-there-a-right-to-explanation-for-machine-learning-in-the-gdpr; Bryce Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision Making and a "Right to Explanation"*, 38 AI MAG. 50 (2017); Sandra Wachter et al., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIV. L. 76 (2017); Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189 (2019); Gerke et al., *supra* note 7, at 322.

329 Babic et al., *supra* note 22. *See also* Mark Ratner, *FDA Backs Clinician-Free AI Imaging Diagnostic Tools*, 36 NATURE BIOTEC. 673, 674 (2018) (quoting Eric Perakslis, former chief information officer at the FDA: "You are seeing FDA not just approving these tools, they are accelerating them").

330 For more information on such devices, *see supra* Section I.A and Section III.B.

331 *See* Babic et al., *supra* note 22, at 284; Babic & Gerke, *supra* note 22.

332 *See* sources cited *supra* note 331.

333 *Id.*

334 *Id.*

only *post hoc* explanations for the outputs of the black box and not to make actual predictions.[335] In other words, explainable AI/ML offers post hoc explanations for black-box predictions without necessarily giving the *actual* reasons behind such predictions.[336] For example, imagine a black-box model predicting a patient's high risk of stroke.[337] The second explanatory algorithm might say that the black-box prediction is consistent with a linear model, which relies on one's smoking and blood pressure status.[338] However, this post hoc explanation may not be the actual reason why the black-box model predicted the patient's high risk of stroke. Explainable AI/ML only generates an "ersatz understanding."[339] Many other algorithmically generated explanations are easily conceivable here that are also consistent with the prediction of the black box.[340] For instance, it could also be the case that the patient's high risk of stroke is consistent with a decision tree, which relies on their diabetes and gender status.[341] Hence, in the context of explainable AI/ML, there is a high risk of a false impression that one better understands black-box predictions and thus a false sense of user (over)confidence in the explanations provided.[342]

Consequently, regulators like the FDA need to be cautious about requiring explainable AI/ML as a prerequisite of marketing authorization since its benefits in health care are not what they currently appear to be.[343] The gold standard should be that regulators require AI/ML makers to use an interpretable AI/ML system—if a white-box model performs better than or as well as a black-box AI/ML model—and focus on ensuring the model's safety and effectiveness. However, if there is sufficient proof that a black-box model performs better than a white-box model and is reasonably safe and effective, and the accuracy increase outweighs the loss of model interpretability, then regulators should generally permit marketing of the black-box AI/ML model as such (without requiring explainable AI/ML) to facilitate innovations. To achieve this goal, regulators could reach, at least in some cases, into an already existing toolbox: clinical trials.

2. Clinical Trials

For drugs and vaccines, clinical trials are the standard method to prove that they are reasonably safe and effective for their intended use. There are several steps

335 *Id*.
336 Babic et al., *supra* note 22, at 285; Babic & Gerke, *supra* note 22.
337 Babic & Gerke, *supra* note 22.
338 *Id.*
339 Babic et al., *supra* note 22, at 285; Babic & Gerke, *supra* note 22.
340 Babic & Gerke, *supra* note 22.
341 *Id.*
342 Babic et al., *supra* note 22, at 285; Babic & Gerke, *supra* note 22.
343 Babic et al., *supra* note 22, at 286; Babic & Gerke, *supra* note 22.

involved in the drug and vaccine development process, one of which is clinical research. The FDA typically requires successful completion of three phases before granting marketing approval of a drug or vaccine.[344] For clinical trials of drugs, for example, Phase 1 is typically carried out with 20 to 100 healthy volunteers or people with the disease or condition to test safety and dosage; Phase 2 has up to several hundred people with the disease or condition and aims to evaluate the drug's efficacy and side effects; and Phase 3 is carried out on a large scale with about 300 to 3,000 volunteers who have the disease or condition and is designed to further assess the efficacy and to monitor adverse reactions.[345] In Randomized Clinical Trials (RCTs), participants are randomly allocated to separate groups that compare different treatments/interventions.[346] In this way, RCTs help to mitigate bias and assess efficacy.[347]

For some medical devices the FDA demands clinical studies.[348] These are typically medical devices that require a PMA.[349] Medical device trials are usually smaller than drug and vaccine trials, but they serve a similar purpose: to support a reasonable assurance that the medical device is safe and effective for its intended use.[350]

However, in the field of health AI, clinical trials are nearly nonexistent. As discussed above,[351] most AI-based medical devices that are currently available on the U.S. market received 510(k) clearances, for which the FDA usually does not request any clinical evidence. One example of an exception in the field is Digital Diagnostic's IDx-DR, which received marketing authorization via the De Novo pathway.[352] The AI company carried out a pivotal clinical study with 900 patients to show IDx-DR's performance.[353] However, even IDx-DR did not receive

344 *Step 3: Clinical Research*, U.S. FOOD & DRUG ADMIN. (Jan. 4, 2018), https://www.fda.gov/patients/drug-development-process/step-3-clinical-research; *Vaccine Development – 101*, U.S. FOOD & DRUG ADMIN. (Dec. 14, 2020), https://www.fda.gov/vaccines-blood-biologics/development-approval-process-cber/vaccine-development-101.

345 *Step 3: Clinical Research*, *supra* note 344.

346 *Randomized Clinical Trial*, NAT'L INSTS. HEALTH, https://www.cancer.gov/publications/dictionaries/cancer-terms/def/randomized-clinical-trial.

347 *Id.*

348 *See* OWEN FARIS, CLINICAL TRIALS FOR MEDICAL DEVICES: FDA AND THE IDE PROCESS 9, https://www.fda.gov/media/87603/download; *supra* Section III.A.

349 *How to Study and Market Your Device*, *supra* note 166. For more information on investigational device exemptions, see *Investigational Device Exemption (IDE)*, U.S. FOOD & DRUG ADMIN. (Dec. 13, 2019), https://www.fda.gov/medical-devices/how-study-and-market-your-device/investigational-device-exemption-ide.

350 FARIS, *supra* note 348, at 5.

351 *See supra* Section III.B.

352 For more information about IDx-DR, *see supra* Section I.A.

353 U.S. FOOD & DRUG ADMIN., *supra* note 32; Michael D. Abràmoff et al., *Pivotal Trial of an*

marketing authorization based on RCT evidence that the information provided by the AI-based medical device improved care.[354] A recent study has also shown that between 2011 and 2019 the FDA often permitted marketing of therapeutic medical devices via the De Novo pathway regardless of limited clinical evidence of effectiveness.[355] Moreover, the first two RCTs of AI/ML have only just been published in 2019.[356] By way of example, in one of these RCTs, 536 patients were randomly allocated to standard colonoscopy and 522 patients to colonoscopy with computer-aided diagnosis.[357]

When exploring a new regulatory framework for AI-based medical devices, the FDA should prefer the use of interpretable AI/ML systems in cases where white-box models perform as good as or better than black-box AI/ML models. Of course, the manufacturer must also provide reasonable assurance of the safety and effectiveness of a white box, which may also require the conduct of a clinical trial in a case where the device presents a higher risk level. However, suppose the black-box AI/ML model performs better in a specific case, and the accuracy improvement outweighs the loss of model interpretability. Rather than requiring explainable AI/ML, regulators should generally permit marketing of the black box, as long as the device has been proven to be reasonably safe and effective, such as via a clinical trial. There are drugs available on the U.S. market whose mechanisms of action are still unknown, such as Acetaminophen.[358] Nevertheless, such drugs are widely used since they have been shown to be reasonably safe and effective. Consequently, it seems likely that black-box AI/ML models do not affect the trust of patients and health professionals and thus their use, as long as they function as promised.[359]

Autonomous AI-Based Diagnostic System for Detection of Diabetic Retinopathy in Primary Care Offices, 1 NPJ DIGIT. MED., no. 39 (2018).

354 Derek C. Angus, *Randomized Clinical Trials of Artificial Intelligence*, 323 JAMA 1043 (2020).

355 James L. Johnston et al., *Clinical Evidence Supporting US Food and Drug Administration Clearance of Novel Therapeutic Devices via the De Novo Pathway Between 2011 and 2019*, 180 JAMA INTERNAL MED. 1701 (2020); James L. Johnston et al., *Clinical Evidence Supporting FDA Clearance of First-of-a-Kind Therapeutic Devices via the De Novo Pathway Between 2011 and 2019*, https://www.medrxiv.org/content/10.1101/2020.04.23.20077164v2.

356 Haotian Lin et al., *Diagnostic Efficacy and Therapeutic Decision-Making Capacity of an Artificial Intelligence Platform for Childhood Cataracts in Eye Clinics: A Multicentre Randomized Controlled Trial*, 9 ECLINICALMED. 52 (2019); Pu Wang et al., *Real-Time Automatic Detection System Increases Colonoscopic Polyp and Adenoma Detection Rates: A Prospective Randomised Controlled Study*, 68 GUT 1813 (2019). For more information, see Myura Nagendran et al., *Artificial Intelligence Versus Clinicians: Systematic Review of Design, Reporting Standards, and Claims of Deep Learning Studies*, 368 BMJ m689 (2020).

357 Wang et al., *supra* note 356, at 1813.

358 *See, e.g.*, K. Toussaint et al., *What Do We (Not) Know About How Paracetamol (Acetaminophen) Works?*, 35 J. CLINICAL PHARM. & THERAPEUTICS 617 (2010).

359 *See* Liu et al., *supra* note 325.

Clinical trials can support a reasonable assurance that the AI/ML-based medical device is safe and effective for its intended use. In an ideal world, RCTs would perhaps be desirable for all AI/ML-based medical devices, especially black boxes, but are they really feasible? Clinical trials will work for some but not for all AI/ML models.[360] For example, they will work for those algorithms that divide patients into groups and propose a specific treatment.[361] However, some algorithms are intended to make recommendations that are highly personalized so that clinical trials would be challenging, perhaps even infeasible, and might overwhelm standard RCT designs.[362] Another problem is adaptive algorithms that can continuously learn and adapt to new conditions.[363] These AI/ML systems are not static, and thus the benefit of clinical trials will likely not last long since the algorithms change.[364] This is particularly problematic given that clinical trials are costly and time-consuming. For adaptive algorithms, regulators like the FDA need to focus their efforts especially on continuous risk monitoring.[365]

On the flip side, the lack of reliable evidence may jeopardize patient safety and undermine public trust in the FDA. Some people fear that AI companies live the motto "fail fast and fix it later."[366] If this is true, the risk concerns for black-box AI/ML models are significant since the users cannot look inside the boxes and thus do not know whether their outputs are correct. Nathan Cortez has also correctly pointed out that "the lack of reliable evidence may depress demand and thus adoption of digital health products," including AI.[367] On the other hand, Nicholson Price rightly warns that mandating clinical trials for black-box AI/ML models could "slow or stifle innovation."[368]

This is a dilemma for regulators: An optimal path would be to facilitate innovation while ensuring that AI/ML models, especially black boxes, are reasonably safe and effective. It will be a challenge to juggle the different stakeholder interests. However, for the new regulatory framework for AI-based medical devices, the FDA should, where feasible and in light of patient safety, at least require clinical trials for those AI/ML-based medical devices (i.e.,

360 Price, *Artificial Intelligence in Health Care*, *supra* note 6, at 11.

361 *Id.*

362 *See* Angus, *supra* note 354, at 1044; Price, *Artificial Intelligence in Health Care*, *supra* note 6, at 11.

363 For adaptive algorithms, see *supra* Part I.

364 W. Nicholson Price II, *Black-Box Medicine*, 28 HARV. J.L. & TECH. 419, 460 (2015). For the update problem, *see infra* Section IV.B.

365 *See infra* Section IV.B.3.

366 Liz Szabo, *A Reality Check On Artificial Intelligence: Are Health Care Claims Overblown?* KHN (Dec. 30, 2019), https://khn.org/news/a-reality-check-on-artificial-intelligence-are-health-care-claims-overblown.

367 Cortez, *supra* note 11, at 21.

368 Price, *Artificial Intelligence in Health Care*, *supra* note 6, at 11.

interpretable AI/ML systems and black boxes) that have a higher risk level. The FDA could leverage the IMDRF framework for risk categorization of SaMD[369] to determine whether a clinical trial is needed. The FDA could, for example, require clinical evidence for all AI/ML-based medical devices that would be classified as risk level III or IV devices, and for some black boxes that would be classified as risk level II devices, such as those that fall into the category "treat or diagnose" or "drive clinical management." It is justified to require clinical trials for AI/ML-based medical devices that are black boxes more often than for white boxes, since black boxes raise additional concerns because of their noninterpretability.

There may be also exceptions where one always wants to know *why* an AI/ML-based medical device made a particular recommendation and where the use of a black box would not be sufficient, even with a successful clinical trial that provides valid scientific evidence that the device is reasonably safe and effective for its intended use. For example, imagine a black-box prediction model is used for triage decisions during a pandemic to decide which patient should be prioritized for receiving a ventilator based on the patient's risk of mortality. In such a life-or-death decision, one would like to know for concerns of justice—understood here as concerns about how one should fairly allocate scarce resources[370]—why the model concluded that patient X has a high or low risk of dying and thus should (not) be prioritized over patient Y. Consequently, AI-based mortality prediction models should not only be clearly classified as medical devices under FDCA section 201(h)(1) and subject to FDA regulation—as I have argued above[371]—but the FDA should also require AI makers to use interpretable systems from the outset in cases where their intended use poses concerns of justice. In general, for reasons of procedural fairness, if AI/ML-based medical devices are intended to be used to allocate scarce resources, such as ventilators or organs,[372] it would be appropriate and likely necessary for the FDA to demand the use of interpretable AI/ML systems even if black boxes performed better.

These are certainly not easy waters to navigate. But once the FDA has figured out the details of the new regulatory framework for AI-based medical devices, as

369 *See supra* Figure 2.

370 *See* Babic et al., *supra* note 22, at 286.

371 *See supra* Section II.B.4. In this scenario, the AI-based mortality prediction model would already not be CDS since the model would "drive clinical management," which would go beyond "supporting or providing recommendations. *See supra* Figure 1 and Figure 2; INT'L MED. DEVICE REGULS. F., *supra* note 63, at 11; U.S. FOOD & DRUG ADMIN., *supra* note 85, at 14. Moreover, the model would perhaps already not be considered a medical device under FDCA § 201(h)(1); it is highly unclear whether it would be "intended for use in the . . . treatment . . . of disease . . . ." FDCA § 201(h)(1)(B), 21 U.S.C. § 321(h)(1)(B); *see supra* Section II.B.3.

372 *See* Babic et al., *supra* note 22, at 286; Gali Katznelson & Sara Gerke, *The Need for Health AI Ethics in Medical School Education*, 26 ADVANCES HEALTH SCI. EDUC. 1447, 1453 (2021); Boris Babic et al., *Can AI Fairly Decide Who Gets An Organ Transplant?*, HARV. BUS. REV. (2020).

suggested here, Congress should enact legislation to enable the FDA to implement it.

B. Update Problem

1. Safety Concerns

AI/ML-based SaMD are distinct from other medical devices insofar as they can learn from new data and improve their performance. This distinctive feature, however, poses challenges for regulators like the FDA. At the moment, the FDA typically only clears or approves AI/ML-based SaMD with "locked" algorithms.[373] "Locked" algorithms do not change with use and provide the same outcome each time the same input data is supplied.[374] In cases where an algorithm changes, the AI/ML-based SaMD will likely need to undergo another premarket review.[375] However, the problem is that to fully realize their potential, AI/ML-based SaMD need to constantly learn and thus require frequent updates, many of which involve algorithm architecture changes and retraining with new data sets.[376] But since these updates will likely require another round of premarket review, they may not be carried out. The manufacturer, for example, could be a small start-up that simply cannot afford the costs of one or multiple new premarket submissions.[377] Further, it may well be that a company refrains from carrying out necessary updates to not send the wrong message about the AI/ML's current quality.[378] It could also be that the manufacturer wants to avoid the significant efforts and time involved in preparing a new submission, and thus decides to perform fewer updates than needed or, worse, no updates at all.

Consequently, this "update problem" raises new regulatory challenges for the FDA. An AI/ML-based SaMD that is not frequently updated may pose significant risks to patients. For example, imagine the FDA permits marketing authorization of an AI/ML-based SaMD that analyzes photos taken by the physician of a patient's skin and assesses the risk for certain types of skin cancer, such as melanoma. In the U.S., skin cancer is the most common cancer, and early diagnosis

373 U.S. FOOD & DRUG ADMIN., *supra* note 19, at 3.

374 *Id.* For locked algorithms, see *supra* Part I.

375 U.S. FOOD & DRUG ADMIN., *supra* note 19, at 3, 6. For more information on when to submit a 510(k) for software changes to existing devices, see U.S. FOOD & DRUG ADMIN., DECIDING WHEN TO SUBMIT A 510(K) FOR A SOFTWARE CHANGE TO AN EXISTING DEVICE: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF 16 (2017), https://www.fda.gov/media/99785/download.

376 U.S. FOOD & DRUG ADMIN., *supra* note 19, at 6.

377 Babic et al., *supra* note 25, at 1202.

378 *Id.*

may be essential to avoid death.[379] However, suppose this AI/ML-based SaMD was trained mainly on images of white skin. Thus, this device will likely have high false-positive and false-negative results when used on patients with darker skin. For example, inflammation often appears pink or red on white skin, while it is violaceous or brown on black skin,[380] and there are many more differences related to skin color. In addition, although melanoma, the most serious skin cancer type, is rare in African American people, it is associated with a worse prognosis than in Caucasian people.[381] Thus, if melanoma goes undetected, for example, it can cost lives that could have been saved. However, if the illustrative AI/ML-based SaMD is used more frequently on patients with darker skin and more data are collected, the device can improve its clinical performance and make a more accurate diagnosis if updated. Of course, for an AI/ML-based SaMD like the one in this hypothetical example, the FDA should ensure that it does not receive marketing authorization in the first place and demand training of the algorithm on diverse data sets, including African American patients, to mitigate such bias. Regulators like the FDA could require AI/ML developers to sufficiently diversify training data in order to mitigate biases and ensure that AI/ML-based medical devices are reasonably safe and effective across various subpopulations.[382] However, even then, there is always a chance that a relevant subpopulation is unknown at the time of marketing authorization.[383] Thus, AI/ML-based SaMD with adaptive algorithms that continuously learn and adapt to new conditions could "unlock" the full potential of health AI and enable precision medicine.[384]

As a result, it is important that regulators like the FDA develop a regulatory framework that promotes innovation and updates of AI/ML-based SaMD, while

379 3Derm Systems, Inc., *3Derm Announces Two FDA Breakthrough Device Designations for Autonomous Skin Cancer AI*, PR NEWSWIRE (Jan. 7, 2020), https://www.prnewswire.com/news-releases/3derm-announces-two-fda-breakthrough-device-designations-for-autonomous-skin-cancer-ai-300982072.html.

380 Art Papier, *To Begin Addressing Racial Bias in Medicine, Start With the Skin*, STAT (July 20, 2020), https://www.statnews.com/2020/07/20/to-begin-addressing-racial-bias-in-medicine-start-with-the-skin.

381 Krishnaraj Mahendraraj et al., *Malignant Melanoma in African–Americans*, 96 MED. 1 (2019). For more information about melanoma, see *Melanoma*, MAYO CLINIC (Jan. 20, 2022), https://www.mayoclinic.org/diseases-conditions/melanoma/symptoms-causes/syc-20374884.

382 *See, e.g.*, Ross, *supra* note 1 (criticizing a lack of transparency in the current FDA approach that also seems to be inconsistent); Casey Ross, *Could AI Tools for Breast Cancer Worsen Disparities? Patchy Public Data in FDA Filings Fuel Concern*, STAT (Feb. 11, 2021), https://www.statnews.com/2021/02/11/breast-cancer-disparities-artificial-intelligence-fda.

383 Babic et al., *supra* note 25, at 1202 (providing an example on HIV vaccine studies, where a relevant subpopulation—uncircumcised men who had high titers of preexisting antibodies against Ad5 and who both had sex with men—were unknown ex ante). For more information on immune activation with HIV vaccines, see Anthony S. Fauci et al., *Immune Activation with HIV Vaccines*, 344 SCI. 49 (2014).

384 *See supra* Part I for adaptive algorithms.

ensuring that the devices remain safe and effective throughout their life cycle.

2. The FDA's TPLC Approach and Action Plan

To its credit, the FDA has already spent a considerable amount of time thinking about how to address the update problem. In April 2019, the FDA released a discussion paper in which the agency proposed a regulatory framework for modifications to AI/ML-based SaMD ("discussion paper").[385] As envisioned in its Software Pre-Cert Program, the FDA intends to apply a Total Product Lifecycle (TPLC) approach for AI/ML-based SaMD that would enable such devices to continuously learn and improve while providing adequate safeguards.[386] As discussed above, to fully implement the Pre-Cert TPLC approach, where particular companies would be "precertified," the FDA would need to ask Congress for additional statutory authority.[387]

The TPLC approach for AI/ML-based SaMD suggested in the FDA's discussion paper would apply exclusively to those AI/ML-based SaMD that are subject to premarket submission.[388] AI/ML-based SaMD that are Class I or Class II exempt are not within the scope of this suggested approach.[389] In particular, the TPLC approach would rely on a *predetermined change control plan* that manufacturers could optionally submit during the initial premarket review of their AI/ML-based SaMD.[390] This plan would include *SaMD Pre-Specifications* and an *Algorithm Change Protocol*.[391] SaMD Pre-Specifications delineate the types of anticipated modifications.[392] The Algorithm Change Protocol is the associated methodology that the manufacturer has in place to implement those modifications and to control their risks to patients.[393]

The FDA divides the types of anticipated modifications into three broad categories:

(1) performance,

(2) inputs, and

385 U.S. FOOD & DRUG ADMIN., *supra* note 19.
386 *Id.* at 3, 4; for the Pre-Cert Program, *see supra* Section III.C.
387 *See supra* Section III.C.
388 U.S. FOOD & DRUG ADMIN., *supra* note 19, at 8.
389 *Id.*
390 *Id.* at 10.
391 *Id.*
392 *Id.*
393 *Id.*

(3) intended use.[394]

The first category includes modifications that improve clinical and analytical performance, such as an increased sensitivity of the AI/ML-based SaMD at detecting breast cancer.[395] The second category is modifications that change the inputs used by the algorithm, such as adding different input data types.[396] For the third category, the FDA leverages the IMDRF framework for risk categorization of SaMD.[397] It includes those types of modifications that result in a change in the:

- state of the health care situation or condition (e.g., expanding the intended patient population to include children), and such modifications are explicitly claimed by the manufacturer; or
- intended condition or disease (e.g., expanding the use of an AI/ML-based SaMD to detect a second type of cancer); or
- significance of the information provided by the SaMD (e.g., a change from "drive clinical management" to "treat or diagnose").[398]

According to the FDA's proposal in its discussion paper, a manufacturer of an AI/ML-based SaMD could submit a predetermined change control plan for many scenarios.[399] However, the FDA considers SaMD Pre-Specifications and Algorithm Change Protocols inappropriate in cases where the AI/ML-based SaMD's intended use or risk may significantly change.[400] An example would be a change from a "non-serious" to a "critical" health care situation or condition, such as an AI/ML-based SaMD that initially uses skin images to manage scar healing and is updated to diagnose melanoma.[401]

In its discussion paper, the FDA also highlights that the TPLC approach can only fully be adopted by enabling real-world performance monitoring of AI/ML-

394 *Id.* at 6.
395 *Id.*
396 *Id.* at 7.
397 For more information on the IMDRF framework for risk categorization of SaMD, *see supra* Section II.B.3.
398 *See* U.S. FOOD & DRUG ADMIN., *supra* note 19, at 7. For SaMD risk categories developed by the IMDRF, see *supra* Figure 2.
399 U.S. FOOD & DRUG ADMIN., *supra* note 19, at 7.
400 *Id.*
401 *Id.* For SaMD risk categories developed by the IMDRF, see *supra* Figure 2.

based SaMD and increased user transparency.[402] Manufacturers would be expected to commit to both of these principles.[403] For example, they would need to provide periodic reporting to the FDA on updates that were carried out based on the predetermined change control plan.[404] However, there are still numerous questions unanswered, such as: How much data would have to be provided? How can manufacturers demonstrate transparency about performance improvement, labeling changes, or algorithm updates of AI/ML-based SaMD?[405]

Many details of the FDA's proposed regulatory framework in its discussion paper still need to be figured out.[406] In January 2021, the newly launched FDA's Digital Health Center of Excellence issued an Action Plan for AI/ML-Based SaMD.[407] This Action Plan is a response to stakeholder feedback to the discussion paper and outlines five actions the FDA aims to take:

> (1) Updating the FDA's proposed regulatory framework laid out in its discussion paper, including publishing draft guidance on the predetermined change control plan.
>
> (2) Encouraging the development of Good Machine Learning Practice.
>
> (3) Supporting a patient-centered approach by holding, for example, a public workshop on AI/ML-based medical device labeling to promote transparency to users.
>
> (4) Fostering efforts on the development of methods to assess and improve machine learning algorithms, including to identify and eliminate bias.
>
> (5) Advancing real-world performance pilots together with stakeholders.[408]

3. The Need for Continuous Risk Monitoring

The FDA's vision of relying on SaMD Pre-Specifications and Algorithm Change Protocols in many scenarios is flawed because manufacturers often do not

402 U.S. FOOD & DRUG ADMIN., *supra* note 19, at 14.
403 *Id.*
404 *Id.*
405 *Id.*at 15.
406 *Id.* at 4.
407 U.S. FOOD & DRUG ADMIN., *supra* note 264.
408 *See id.* at 7.

know at the time of the initial premarket review what updates will be required in the future.[409] Only after the marketing authorization and use of the AI/ML-based SaMD in clinical practice do many necessary updates become apparent. Thus, it is especially important for the FDA to focus on continuous risk monitoring once the AI/ML-based SaMD is legally launched on the U.S. market.[410] The agency needs to look out for new risks due to AI/ML features, such as covariate shift, concept drift, and instability.[411]

Covariate shift occurs when the data the algorithm was trained on before marketing authorization is different from the input distribution of new data.[412] For example, an AI/ML-based SaMD may be trained on data from a nursing home with only patients over sixty-five but shall now be deployed in a large municipal hospital with a diverse patient population.

Concept drift exists in cases where there is a change of the true relation between inputs and outputs.[413] Take an AI/ML-based SaMD, for example, that makes recommendations on breast cancer risk by analyzing the results of mammograms. Suppose the device does not track the patient's race. However, the breast density varies between Caucasian women and African American women, and African American women are also more likely to die from malignant tumors than are Caucasian women.[414] Thus, depending on the patient's race, the same image may result in two different probabilistic diagnoses.[415]

Instability describes a situation where an AI/ML-based SaMD does not treat similar patients similarly.[416] For example, an AI/ML-based SaMD that detects lung cancer and classifies medically similar lung lesions entirely differently is unstable.

For continuous monitoring of AI/ML-based SaMD, the FDA could, for example, leverage its national monitoring system Sentinel.[417] The FDA launched

409 Babic et al., *supra* note 25, at 1203-04.

410 *See id.* at 1204.

411 *Id.* at 1203-04.

412 *Id.* at 1203. For more information on covariate shift, see, for example, Steffen Bickel et al., *Discriminative Learning for Differing Training and Test Distributions* (2007) (unpublished manuscript), https://icml.cc/imls/conferences/2007/proceedings/papers/303.pdf.

413 Babic et al., *supra* note 25, at 1203.

414 *Id.* at 1202; Amrita Khalid, *Google's AI for Mammograms Doesn't Account for Racial Differences*, QUARTZ (Jan. 9, 2020), https://qz.com/1781123/googles-ai-for-mammograms-doesnt-account-for-race. For statistics on breast cancer, see KAISER FAM. FOUND., COVERAGE OF BREAST CANCER SCREENING AND PREVENTION SERVICES (Sept. 26, 2019), https://www.kff.org/womens-health-policy/fact-sheet/coverage-of-breast-cancer-screening-and-prevention-services.

415 *See* Babic et al., *supra* note 25, at 1202-03. For more examples, see Boris Babic et al., *When Machine Learning Goes Off the Rails*, HARV. BUS. REV. (Jan.-Feb. 2021), https://hbr.org/2021/01/when-machine-learning-goes-off-the-rails.

416 Babic et al., *supra* note 22, at 1203-04.

417 *See* Babic et al., *supra* note 25, at 1204; I. Glenn Cohen et al., *The European Artificial*

the Sentinel Initiative in response to Congress' mandate in the FDA Amendments Act of 2007[418] to develop novel ways to evaluate the safety of marketed medical products.[419] The FDA also announced in September 2019 that Sentinel will expand to three coordinating centers, one of which, the Sentinel Operations Center, is focusing, among other topics, on AI.[420]

In addition to using a national monitoring system and having an appropriate division of labor,[421] a continuous risk monitoring approach for AI/ML-based SaMD should consist of at least three other elements:

(1) retesting,

(2) simulated checks, and

(3) adversarial stress tests.[422]

First, AI/ML-based SaMD should be continuously retested on all previous cases.[423] Second, AI/ML-based SaMD should be constantly used on "simulated patients" to assess whether their behavior is reliable with regard to an adequate diversity of patient types.[424] For example, previous patient data could be used to create "simulated patients."[425] Third, one could perform algorithmic stress tests throughout the AI/ML-based SaMD's life cycle, borrowing from cybersecurity practices.[426] In particular, AI/ML is vulnerable to adversarial attacks, where a slight change—(almost) undetectable to the human eye—in how inputs are presented to the system alters its output, leading to an incorrect conclusion.[427] This is especially worrisome in cases where the AI/ML-based SaMD is intended to detect, for example, a type of cancer, such as skin cancer, and incorrectly classifies

Intelligence Strategy: Implications and Challenges for Digital Health, 2 LANCET DIGIT. HEALTH e376, e377 (2020); *FDA's Sentinel Initiative*, U.S. FOOD & DRUG ADMIN. (Oct. 18, 2019), https://www.fda.gov/safety/fdas-sentinel-initiative.

418 Food and Drug Administration Amendments Act of 2007, Pub. L. No. 110-85, 121 Stat. 823.

419 *FDA's Sentinel Initiative*, *supra* note 417; *see also* SENTINEL, https://www.sentinelinitiative.org (last visited Mar. 19, 2022) (providing an overview of Sentinel).

420 *FDA's Sentinel Initiative*, *supra* note 417.

421 Monitoring of AI/ML-based SaMD should be carried out by different actors than those developing such devices. *See* Babic et al., *supra* note 25, at 1204.

422 *See id.*

423 *Id.*

424 *Id.*

425 *Id.*

426 *Id.*

427 Samuel G. Finlayson, *Adversarial Attacks on Medical Machine Learning*, 363 SCI. 1287 (2019).

the mole with 100% confidence as malignant instead of benign.[428] Thus, it is essential that AI/ML-based SaMD rigorously undergo algorithmic stress tests throughout their entire life cycle.

As a result, a robust continuous risk monitoring approach, like the one suggested above, can help to ensure that AI/ML-based SaMD remain safe and effective throughout their life cycle. This approach also allows the FDA to quickly recall an AI/ML-based SaMD from the market if necessary.

V. SYSTEM VIEW

It is essential that the FDA broadens its view and considers AI-based medical devices as systems, not *just* devices.[429] The agency should focus more on the environment in which AI-based medical devices are deployed. This system view is crucial to ensure that AI-based medical devices are reasonably safe and effective as well as benefit patients. In this Part, I carve out two components of the system view: (1) considering human-AI interaction and (2) improving patient outcomes.

A. Considering Human-AI Interaction

Generally, when AI-based medical devices enter medical practice, they will interact with humans to varying degrees (from little to collaboratively). Thus, it is essential that regulators like the FDA broaden their view and systematically consider the interaction between the human and the AI. The system view is especially relevant for AI-based medical devices because their performance in the actual practice setting is less predictable than that of traditional medical devices, such as crutches or contact lenses.[430] AI-based medical devices can be biased, opaque, and/or adaptive. Human factors and the interaction of these complex systems with the environment will likely increase variance between such medical devices' performance in simulated testing settings and real life.[431]

For example, imagine an AI-based medical device that is developed and used in a highly specialized clinic and makes sophisticated recommendations to specialist personnel in that clinic. The device shall now be deployed in another hospital in a rural area that is not as specialized as the clinic who developed it and has far fewer medical specialists. It may well be that the recommendations the AI makes are not feasible, useful, safe, and/or cost-effective for less specialized

428 *Id.* at 1287-88.

429 Sara Gerke et al., *The Need for a System View to Regulate Artificial Intelligence/Machine Learning-Based Software as Medical Device*, 3 NPJ DIGIT. MED., no. 53 (2020).

430 *Id.* at 2.

431 *Id.*

personnel in a rural hospital.[432] In other words, as Mildred Cho puts it: "Systems developed in one hospital often flop when deployed in a different facility."[433] Thus, AI bears the risk of "contextual" bias.[434]

Although perhaps desirable, it will likely not be feasible to require licenses at the level of an individual clinic.[435] However, the FDA could at least require rigorous human factors testing for *all* AI-based medical devices that require premarket submission. This would include, for example, a demonstration that users can use the AI-based medical device correctly based merely on reading the labeling and that they can correctly interpret its output and understand that such devices bear the risks of false-positive and false-negative readings. If it is an AI-based home monitoring technology, which is used without (direct) supervision by a health care professional, human factors testing should also include that users do not over-rely on its output and comprehend when to seek medical care.[436] To its credit, the FDA required human factors testing for a few AI-based medical devices that received marketing authorization via the De Novo pathway, such as for IDx-DR and Apple's irregular rhythm notification feature.[437] However, such testing should be standardized and required for all AI-based medical devices that are subject to premarket submission. It is also important that the testing be carried out in actual practice settings since the results will likely vary with the human involvement in decision-making.[438]

Another issue in the human-AI interaction is training and education. A good, although non-AI, example is the da Vinci surgical system. Da Vinci is a robot that helps surgeons to perform minimally invasive surgery. The surgeon uses a console, and the da Vinci system translates the surgeon's hand movements.[439] The FDA first cleared the system in 2000, but since then, unfortunately, many patients have suffered severe complications, some of which even resulted in death.[440] One of the reasons for such complications was a lack of training of the surgeons with the

---

432 Timo Minssen, Sara Gerke, Mateo Aboy, Nicholson Price & Glenn Cohen, *Regulatory Responses to Medical Machine Learning*, 7 J. L. & BIOSCI. l, 17 (2020).

433 Szabo, *supra* note 366.

434 Nicholson Price, *Medical AI and Contextual Bias*, 33 HARV. J.L. & TECH. 66 (2019).

435 Gerke et al., *supra* note 429, at 3.

436 Gerke et al., *supra* note 3, at 1178.

437 U.S. FOOD & DRUG ADMIN., *supra* note 32; Letter from Angela C. Krueger to Donna-Bea Tillman, *supra* note 50.

438 Gerke et al., *supra* note 429, at 4.

439 *About Da Vinci Systems*, INTUITIVE (2022), https://www.davincisurgery.com/da-vinci-systems/about-da-vinci-systems.

440 Kristin Compton, *Da Vinci Surgical System*, DRUGWATCH (2021), https://www.drugwatch.com/davinci-surgery; *see also* Emily R. Siegel et al., *The Da Vinci Surgical Robot: A Medical Breakthrough With Risks for Patients*, NBC NEWS (Dec. 19, 2018), https://www.nbcnews.com/health/health-news/da-vinci-surgical-robot-medical-breakthrough-risks-patients-n949341 (telling the story of Laurie Featherstone, an injured patient).

device.[441]

Training and education, in particular, are crucial for all users of AI-based medical devices since their outcomes can vary considerably the more human involvement there is.[442] For example, in February 2020, the FDA permitted marketing of the first cardiac ultrasound (echocardiography) software, called Caption Guidance, via the De Novo pathway.[443] The software uses AI to help the user capture images of patients' hearts.[444] The peculiarity of the software is that it can be used by non-experts, such as nurses with only a few days of training.[445] Thus, since more AI-based medical devices, similar to IDx-DR and Caption Guidance, that can be used by non-experts are likely to enter the U.S. market in the near future, training and education of the users of such devices at regular intervals will be even more important. Hence, even if the FDA does not regulate the practice of medicine, the agency could more often demand that AI makers set up a training program with instructions on how to use the AI-based medical device, such as the agency did in the case of IDx-DR.[446] Alternatively or additionally, the FDA could more frequently require AI-makers to include a detailed description of the recommended user training in the labeling of the AI/ML-based medical device, as was the case, for example, for Caption Guidance.[447]

A research team at Duke University is also thinking about new ways of labeling health AIs, similar to "nutrition labels" that contain facts on the intended use of the system and how it should be used.[448] More initiatives such as the one at Duke are needed to better understand what content such labeling should include to promote user transparency and comprehension of the benefits, shortcomings, and risks of AI-based medical devices and to mitigate user errors. It is thus to be welcomed that the FDA has recently organized a public workshop on transparency

441 Siegel et al., *supra* note 440.

442 Gerke et al., *supra* note 429, at 2.

443 Letter from Robert Ochs, Deputy Dir. Radiological Health, U.S. Food & Drug Admin. to Sam Surette, RA/QA Manager, Caption Health, Inc. (Feb. 7, 2020), https://www.accessdata.fda.gov/cdrh_docs/pdf19/DEN190040.pdf; *FDA News Release: FDA Authorizes Marketing of First Cardiac Ultrasound Software That Uses Artificial Intelligence to Guide User*, U.S. FOOD & DRUG ADMIN. (Feb. 7, 2020), https://www.fda.gov/news-events/press-announcements/fda-authorizes-marketing-first-cardiac-ultrasound-software-uses-artificial-intelligence-guide-user.

444 *Id.*

445 *Id.*; Casey Ross, *AI Has Arrived in Medical Imaging. Now the FDA Needs to Monitor Its Impact on Patients*, STAT (Febr. 28, 2020), https://www.statnews.com/2020/02/28/ai-medical-imaging-fda-monitor-impact-patients.

446 U.S. FOOD & DRUG ADMIN., *supra* note 32, at 11; Gerke et al., *supra* note 429.

447 Letter from Robert Ochs to Sam Surette., *supra* note 443, at 3.

448 *See* Mark P. Sendak et al., *Presenting Machine Learning Model Information to Clinical End Users With Model Facts Labels*, 3 NPJ DIGIT. MED., no. 41 (2020); Erin Brodwin, *With 'Nutrition Labels' and an Anthropologist's Eye, Duke Pioneers a New Approach to AI in Medicine*, STAT (Oct. 5, 2020), https://www.statnews.com/2020/10/05/duke-artificial-intelligence-hospital-medicine.

of AI/ML-based medical devices, in which the topic of labeling was also discussed, to gather input from stakeholders.[449]

Another example to see the challenges of the interaction between the human and the AI is mortality prediction models. As I have established and argued above,[450] it is highly unclear whether AI-based mortality prediction models are medical devices under current law, but they should be. Imagine that the model predicts the patient will die in the next 12 months. However, the patient's physician did not foresee this. What should the physician do? Should the physician rely on the AI or ignore its prediction? Should the physician start an end-of-life discussion with the patient? Should the physician tell the patient about the AI? Imagine that the physician decides to talk to the patient about the possibility of death in the next 12 months but does not mention the AI. Is this the right choice? What happens if the AI turns out to be wrong and the physician stops (instead of continues) the patient's treatment?

These are tricky questions that have not received enough attention, even though many hospitals are already using these systems on real patients.[451] Suppose a health AI-based product is intended to be used in critical, sensitive situations, such as predicting a patient's death. In that case, it is essential that society starts a discussion about transparency and whether the patient has a right to know that an AI was involved and may have influenced the physician's decision to stop or continue treatment. The interaction between the human and the AI is crucial for a successful outcome. The hospitals that deploy such AIs should develop best practice guidance on how to use these tools. Even if the FDA does not regulate the practice of medicine, there is still something the agency can do. First, as argued above,[452] the FDA could ask Congress to amend the FDCA and clearly classify AI-based mortality prediction models as medical devices and ensure that they are reasonably safe and effective when launched on the U.S. market and used to make such sensitive predictions. Second, once AI-based mortality prediction models are clearly classified as medical devices, the FDA could then demand that AI makers set up a training program with instructions on how to use the device and/or require them to include a detailed description of the recommended user training in the labeling of the device. Third, the FDA may also consider requiring—similar to the

449 *Virtual Public Workshop - Transparency of Artificial Intelligence/Machine Learning-Enabled Medical Devices*, U.S. FOOD & DRUG ADMIN. (Oct. 14, 2021), https://www.fda.gov/medical-devices/workshops-conferences-medical-devices/virtual-public-workshop-transparency-artificial-intelligencemachine-learning-enabled-medical-devices.

450 *See supra* Section II.B.3 and Section II.B.4.

451 *See, e.g.*, Robbins, *supra* note 145; Rebecca Robbins & Erin Brodwin, *An Invisible Hand: Patients Aren't Being Told About the AI Systems Advising Their Care*, STAT (July 15, 2020), https://www.statnews.com/2020/07/15/artificial-intelligence-patient-consent-hospitals.

452 *See supra* Section II.B.4.

case of emergency use authorizations for medical devices[453]—AI makers to develop fact sheets for health professionals and patients (the latter written in plain language) that help them to better understand the device, such as its intended use, its benefits, and its risks. In the fact sheet for health professionals, the manufacturer could also include practical information on how best to handle the situation and predictions by the AI.

In general, a discussion with all stakeholders in the field should begin with the question of whether patients (should) have a right to know about the involvement of an AI-based prediction model. Some hospitals are currently using those systems without telling their patients.[454] Is that morally justifiable? Instead of hiding new AI-based products behind the scenes, is it not better to be frank upfront and promote trust in the doctor-patient relationship? I. Glenn Cohen has recently written about informed consent and medical AI, arguing that "the existing legal doctrine of informed consent does not robustly support an obligation to disclose the use of medical AI/ML," with some exceptions, such as when the patient explicitly asked for the basis of the decision making and is misinformed by the physician.[455] Cohen mentioned in an interview that trust in the health care system and AI could be undercut if patients "were to find out, after the fact, that there's a rash of this being used without anyone ever telling them."[456] Thus, this discussion about the human-AI interaction is crucial and needs to happen now among stakeholders, including patients. As can be seen, many open questions have yet to be answered regarding the human-AI interaction, but the system view can help regulators like the FDA and stakeholders see these issues and address them.

B. Improving Patient Outcomes

The second lesson the system view gives us is that AI-based medical devices do not only need to be safe but should also improve patient outcomes. This is a crucial point, but it has, unfortunately, been neglected so far. As the chess player, Garry Kasparov, correctly pointed out: "Weak human + machine + better process was superior to a strong computer alone and, more remarkably, superior to a strong human + machine + inferior process."[457] Thus, the decisive point is the "process," and if one does not know more about the process of the AI-based medical device,

453 Gerke et al., *supra* note 3, at 1179.

454 Robbins, *supra* note 145. Empathy may also play an important role here. *See* Nicolas Terry, *Appification, AI, and Healthcare's New Iron Triangle*, 20 J. HEALTH CARE L. & POL'Y 117, 159–67 (2018).

455 Cohen, *supra* note 7, at 1467 (2020).

456 Robbins & Brodwin, *supra* note 451.

457 GARRY KASPAROV, DEEP THINKING: WHERE MACHINE INTELLIGENCE ENDS AND HUMAN CREATIVITY BEGINS 214 (2017).

one does not know whether it will improve outcomes.[458] Kasparov teaches us that even if one has an accurate health AI—which is itself challenging to achieve—human factors and the environment in which the product will be deployed need to be considered to ensure that the health AI actually benefits patients.

It seems that so far, however, most AI-based medical devices have not been shown to improve patient outcomes. For example, it is unclear whether IDx-DR, which has already been used in clinical care at over twenty sites across the U.S., improves patient outcomes.[459] To its credit, the company is currently carrying out several studies to examine whether diabetic patients who receive a positive result of more than a mild level of diabetic retinopathy are going to the ophthalmologist and receiving care.[460] The company has also recently launched a care coordination model that will ensure that patients with a positive result receive follow-up care.[461] These are laudable actions, but a rare exception in the field. Thus, the FDA could step in and require, for example, comparative studies for AI-based medical devices where appropriate that demonstrate better outcomes with versus without the device. The FDA could either demand them as a premarket or postmarket requirement, depending on whether the AI-based medical device is urgently needed on the market. Again, the challenge faced by regulators will be to properly balance the different stakeholder interests. The optimal way would be facilitating innovation while simultaneously ensuring that the U.S. market will not be flooded with useless products that do not improve patient outcomes and are also not otherwise valuable, such as products that do not even reduce the labor burden on physicians.

Another example is mobile health apps. There are over 400,000 mobile health apps on the market, but little data on whether or not they actually benefit patients.[462] Most of them, as discussed earlier,[463] are not classified as medical devices and are not FDA reviewed. However, even the ones that are considered to be medical devices have not necessarily been shown to do more good than harm. Take, for example, Apple's irregular rhythm notification feature that is intended to notify the user of possible AFib.[464] Most users of the Apple Watch are young and

458 Gerke et al., *supra* note 429, at 2.

459 Carfagno, *supra* note 35.

460 *Id.*

461 *Id.*

462 *See* Stephan Fihn et al., *Deploying AI in Clinical Settings*, *in* ARTIFICIAL INTELLIGENCE IN HEALTH CARE: THE HOPE, THE HYPE, THE PROMISE, THE PERIL 151, 152 (Michael Matheny et al. eds., 1st ed. 2019); Michael Georgiou, *Developing a Healthcare App in 2022: What do Patients Really Want?*, IMAGINOVATION (Dec. 15, 2021), https://www.imaginovation.net/blog/developing-a-mobile-health-app-what-patients-really-want.

463 *See supra* Section II.B.1.

464 For more information on the app, *see supra* Section I.B.

healthy people who usually are not considered at risk for suffering Afib.[465] Around 70% of individuals with Afib are between 65 and 85 years old.[466] In addition, diagnostic tools can always have false-positive and false-negative results. This may perhaps also be the reason why Apple narrowed the app's indications for use: the app is explicitly "not intended to provide a notification on every episode of irregular rhythm suggestive of Afib" and "is not intended to replace traditional methods of diagnosis or treatment."[467] Still, it is likely that many users do not know that Apple's app is not for diagnosis and therefore the irregular rhythm notification feature gives them a false sense of security. For example, they may think that they are healthy and skip a necessary doctor's appointment because they do not receive alarming notifications from the app. Thus, more user transparency of the indications of use for health apps is needed. Moreover, younger people may also be confronted with a false notification suggestive of Afib and may suffer a shock that can develop further into real psychological or physical harm. In addition, individuals with false notifications may likely sit in the waiting rooms of cardiologists and use unnecessary resources of an already overburdened health care system.[468] In contrast, the ones who would likely benefit most from Apple's app, namely the elderly, are less likely to use the Apple Watch.[469] Thus, it is also essential to make sure that all population groups, particularly the vulnerable ones such as the elderly, benefit from health AI-based products.[470] Furthermore, users who received a notification by using Apple's app and are diagnosed with brief Afib by their cardiologist will likely receive blood-thinning medications as a result. However, one does not know yet whether patients will actually benefit from such medications—or suffer from bleeding risk—and thus whether they would have been better off not to have been diagnosed with brief Afib in the first place.[471] Some people may certainly benefit from Apple's app who would have otherwise

465 Casey Ross, *COVID-19 Apps and Wearables Are Everywhere. Can They Actually Benefit Patients?*, STAT (Aug. 4, 2020), https://www.statnews.com/2020/08/04/covid19-wearables-apps-patient-care.

466 Peter M. Kistler et al., *Electrophysiologic and Electroanatomic Changes in the Human Atrium Associated With Age*, 44 J. AM. COLL. CARDIOLOGY 109, 109 (2004).

467 *See* Letter from Angela C. Krueger to Donna-Bea Tillman, *supra* note 50.

468 Heather Landi, *With Apple's Launch of an ECG Device, Digital Health Leaders, Cardiologists See Possibilities, and Limitations*, HEALTHCARE INNOVATION (Sept. 18, 2018), https://www.hcinnovationgroup.com/interoperability-hie/article/13030721/with-apples-launch-of-an-ecg-device-digital-health-leaders-cardiologists-see-possibilities-and-limitations.

469 *See, e.g.*, Mikey Campbell, *Apple Watch, Other Wearables Increasingly Used to Manage Chronic Health Conditions, Study Says*, APPLEINSIDER (Aug. 30, 2018), https://appleinsider.com/articles/18/08/30/apple-watch-other-wearables-increasingly-used-to-manage-chronic-health-conditions-study-says.

470 For more information on promoting health equity and AI, see Nicolas Terry, *Of Regulating Healthcare AI and Robots*, 21 YALE J.L. & TECH. 133, 186–89 (2019).

471 Landi, *supra* note 468.

perhaps suffered a stroke, but some may not.[472] Thus, regulators like the FDA should apply the system view to not only promote user transparency but also require comparative studies for AI-based medical devices where appropriate to ensure that patients actually benefit from these devices.

CONCLUSION

AI, especially its subset ML, has tremendous potential to improve health care. However, health AI also raises new regulatory challenges. In particular, a new regulatory framework for AI-based medical devices is needed to ensure that such devices are reasonably safe and effective when placed on the market and will remain so throughout their life cycle. Suppose the FDA does not "tame the demon," as Elon Musk would say. In that case, the agency would not have realized the great potential of health AI and patient safety would be jeopardized. Moreover, disparities in health care would likely be exacerbated instead of reduced, presumably to the detriment of vulnerable populations such as racial and ethnic minorities, the economically disadvantaged, the elderly, or people with disabilities.

In this Article, I have especially tried to unpack the complex network of relevant provisions in the FDCA and (draft) guidance documents related to AI-based medical devices. I have shown that the FDA is not yet ready for health AI and that there are significant safety and effectiveness concerns associated with the current regulatory framework. I have advocated for FDA and congressional actions, and I have focused on how the FDA could, with additional statutory authority, regulate AI-based medical devices. What follows are my central claims.

First, the current medical device definition, FDCA section 201(h)(1),[473] is too narrow for health AI. Congress should consider amending the definition to include all CDS, AI-based mortality prediction models, and other models that are intended for use in the prediction or prognosis of disease or other conditions. This suggestion also requires that FDCA section 520(o)(1)(E)[474] is deleted and that FDCA section 520(o)(1)(B)[475] is amended accordingly to reflect the new medical device definition. The FDA should also remain free to exercise its enforcement discretion over lower risk device software functions or lower risk software functions that may meet the medical device definition.

Second, the 510(k) pathway may not be sufficient to identify safety and effectiveness concerns of medical devices. The FDA's reforms to address these issues are welcome. However, the new Safety and Performance Based Pathway

472 *Id.*
473 FDCA § 201(h)(1), 21 U.S.C. § 321(h)(1).
474 FDCA § 520(o)(1)(E), 21 U.S.C. § 360j(o)(1)(E).
475 FDCA § 520(o)(1)(B), 21 U.S.C. § 360j(o)(1)(B).

will likely not be applicable to AI-based medical devices in the near future and is only intended as a voluntary pathway. The Traditional, Special, or Abbreviated 510(k) pathways thus continue to be available to manufacturers. Consequently, I propose a new regulatory framework for premarket review of medical devices, including AI-based medical devices, that would better ensure that medical devices are reasonably safe and effective when placed on the U.S. market. In particular, I argue that the new Safety and Performance Based Pathway—if found to be effective—should replace the Traditional, Special, and Abbreviated 510(k) pathways and become the only available 510(k) pathway. In addition, the De Novo Pathway should be modified to also cover those low to moderate risk medical devices that have a predicate but would not be applicable for the new Safety and Performance Based Pathway. Further, the FDA's envisioned Software Pre-Cert Program raises its own regulatory challenges. If the FDA establishes the Software Pre-Cert Program's specific details, the Pilot proves to be effective, and the agency has statutory authority, the FDA could either implement the Software Pre-Cert Program similarly to the Pre-Cert Pilot Program or entirely separate from the traditional premarket pathways with its own conditions.

Third, the FDA should demand that AI/ML makers use an interpretable AI/ML model if such a model performs better than or as well as the black-box model for its intended use. If the black-box model performs better, the FDA should generally permit its marketing to facilitate innovation, as long as there is sufficient proof that it is safe and effective. A focus on explainable AI/ML is deceptive because the explanations provided are only *ex post* approximations of the black-box algorithms' decisions instead of the actual reasons for them. The FDA should, where feasible, require clinical trials at least for those AI/ML-based medical devices that have a higher risk level. The FDA could leverage the IMDRF framework for risk categorization of SaMD to determine cases where clinical trials are needed. However, in cases where AI/ML-based medical devices are intended to be used to allocate scarce resources, such as ventilators or organs, the FDA should insist on the use of interpretable AI/ML systems.

Fourth, AI/ML-based medical devices can only fully realize their potential if they continuously learn and adapt to novel situations. To address the update problem, the FDA needs to focus on continuous risk monitoring and implement a monitoring system, such as Sentinel, to continuously monitor AI/ML-based SaMD.

Fifth, the FDA should broaden its view and consider AI-based medical devices not just as devices but as systems. In particular, the FDA could require rigorous human factors testing for all AI-based medical devices that require premarket submission to demonstrate that users can read the labeling and use them correctly. The agency could also more often require the AI maker to set up a training program with instructions on how to use the AI-based medical device

and/or to include a detailed description of the recommended user training in the device labeling. In addition, more emphasis should be placed on the AI-based medical devices' ability to improve patient outcomes, not only be safe. This could be demonstrated by comparative studies that the agency could demand, where appropriate, either as a premarket or postmarket requirement, depending on whether the AI-based medical device in question is urgently needed on the U.S. market.

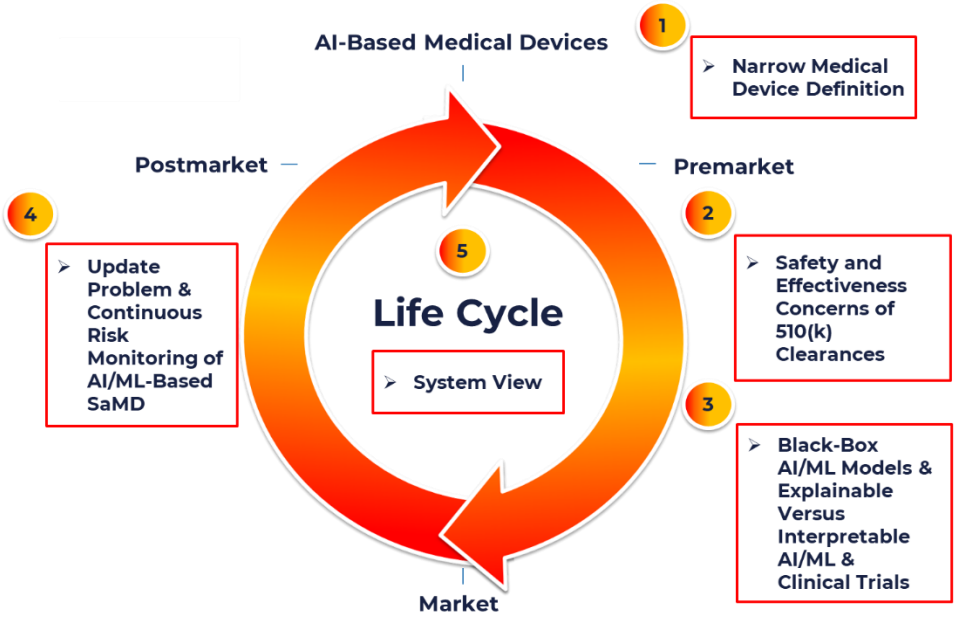

Figure 6: Overview of the Central Claims
Points 1-5 show the central claims. They are arranged in the life cycle of AI-based medical devices — i.e., from premarket to postmarket.

Finally, I conclude that much more work and thinking is required to deliver the full potential of health AI and ensure that such products are reasonably safe and effective. Since the law often lags behind technological advances, it is likewise important that manufacturers design their health AI-based products ethically—irrespective of whether they are classified as medical devices and are subject to FDA regulation. This would, among other things, require AI companies to diversify training data to mitigate biases and ensure that AI-based products are reasonably safe and effective across various subpopulations and remain so throughout their life cycle. Lastly, national, and even international, ethical guidelines for health AI-based products should be developed to establish minimum ethical standards for the design process of such products.