AUTOMATING FDA REGULATION

MASON MARKS†

ABSTRACT

In the twentieth century, the Food and Drug Administration ("FDA") rose to prominence as a respected scientific agency. By the middle of the century, it transformed the U.S. medical marketplace from an unregulated haven for dangerous products and false claims to a respected exemplar of public health. More recently, the FDA's objectivity has increasingly been questioned. Critics argue the agency has become overly political and too accommodating to industry while lowering its standards for safety and efficacy. The FDA's accelerated pathways for product testing and approval are partly to blame. They require lower-quality evidence, such as surrogate endpoints, and shift the FDA's focus from premarket clinical trials toward postmarket surveillance, requiring less evidence up front while promising enhanced scrutiny on the back end. To further streamline product testing and approval, the FDA is adopting outputs from computer models, enhanced by artificial intelligence ("AI"), as surrogates for direct evidence of safety and efficacy.

This Article analyzes how the FDA uses computer models and simulations to save resources, reduce costs, infer product safety and efficacy, and make regulatory decisions. To test medical products, the FDA assembles cohorts of virtual humans and conducts digital clinical trials. Using molecular modeling, it simulates how substances interact with cellular targets to predict adverse effects and determine how drugs should be regulated. Though legal scholars have commented on the role of AI as a medical product that is regulated by the FDA, they have largely overlooked the role of AI as a medical product regulator. Modeling and simulation could eventually reduce the exposure of

† Assistant Professor, University of New Hampshire Franklin Pierce School of Law; Senior Fellow and Project Lead of the Project on Psychedelics Law and Regulation (POPLAR), Petrie-From Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School; Affiliated Fellow, Information Society Project at Yale Law School; Doctoral Researcher, Center for Law and Digital Technologies at Leiden Law School. Special thanks to I. Glenn Cohen, Michael B. Abramowicz, Daniel E. Walters, Mark Fenster, and Ryan Calo for their helpful comments. Professor Marks receives compensation from POPLAR, which is funded by the Saisei Foundation, a nonprofit organization. He advises clients on state and federal controlled substance regulation.

volunteers to risks and help protect the public. However, these technologies lower safety and efficacy standards and may erode public trust in the FDA while undermining its transparency, accountability, objectivity, and legitimacy. Bias in computer models and simulations may prioritize efficiency and speed over other values such as maximizing safety, equity, and public health. By analyzing FDA guidance documents and industry and agency simulation standards, this Article offers recommendations for safer and more equitable automation of FDA regulation.

TABLE OF CONTENTS

Introduction 1209
I. From Snake Oil to Surrogates: The Rise and Fall of Clinical Evidence 1214
II. Building an Algorithmic FDA 1219
A. Automating the Administrative State 1219
B. FDA Modeling and Simulation 1221
1. *Molecular Models* 1227
2. *Virtual Humans and Patient-Specific Models.* 1236
3. *Simulated Clinical Trials* 1238
III. Automated FDA Regulation and Administrative Law 1246
A. Black Box Simulations and FDA Transparency and Accountability 1247
B. Modeling, Simulation, and Nondelegation 1260
IV. Good Simulation Practices 1262
A. Industry and Agency Guidance 1262
B. Evaluating the Credibility of a Regulatory Model 1272
C. Recommendations 1276
1. *Form Independent Boards to Assess Model Credibility* 1276
2. *Consider the Full Spectrum of Algorithmic Harms* 1277
3. *Assess Institutional and Algorithmic Opacity.* 1278
4. *Analyze the Pedigree of Training Data and Incentivize Disclosure* 1278
5. *Identify Alternate Methods of Answering Questions of Interest* 1279
Conclusion 1279

INTRODUCTION

In the twentieth century, the U.S. Food and Drug Administration ("FDA") rose to prominence as a respected scientific agency.[1] By the middle of the century, it had transformed the U.S. medical marketplace from an unregulated haven for dangerous products and false claims to an exemplar of public health.[2] In 1962, the Kefauver-Harris Amendment to the Federal Food, Drug, and Cosmetic Act ("FDCA") empowered the agency to require proof of drug efficacy, instead of safety alone, prior to regulatory approval.[3]

As a result, the modern phased system of clinical trials became the norm for generating scientific knowledge regarding new therapeutics.[4] Though these changes protected the public, they made the FDA approval process longer and more expensive.[5] In the 1980s, drug companies and patient advocates pressured the agency to streamline its procedures.[6] In response, the FDA developed pathways to accelerate product testing and approval.[7] Most allow the agency to rely on surrogate endpoints—measurements such as laboratory tests—that substitute for direct evidence of symptomatic improvement.[8]

---

1. *See* DANIEL CARPENTER, REPUTATION AND POWER: ORGANIZATIONAL IMAGE AND PHARMACEUTICAL REGULATION AT THE FDA 73–75 (2014) (describing how the FDA's response to a series of public health disasters, such as the thalidomaide scandal and sulfamilamide tragedy, contributed to the agency's reputation in the twentieth century).

2. *80 Years of the Federal Food, Drug, and Cosmetic Act*, U.S. FOOD & DRUG ADMIN. (July 11, 2018), https://www.fda.gov/about-fda/fda-history-exhibits/80-years-federal-food-drug-and-cosmetic-act [https://perma.cc/V3U8-VKXF] (describing harmful and deceptively marketed products of the early twentieth century displayed in the FDA's Chamber of Horrors Photo Gallery, and noting that the FDA transformed the medical product marketplace following enactment of landmark legislation, such as the Food, Drug, and Cosmetic Act of 1938).

3. Jeremy A. Greene & Scott H. Podolsky, *Reform, Regulation, and Pharmaceuticals—the Kefauver-Harris Amendments at 50*, 367 NEW ENG. J. MED. 1481, 1481 (2012).

4. *Id.*

5. *Id.* at 1482.

6. *See* Jonathan J. Darrow, Jerry Avorn & Aaron S. Kesselheim, *New FDA Breakthrough-Drug Category — Implications for Patients*, 370 NEW ENG. J. MED. 1252, 1257 (2014) [hereinafter Darrow, Avorn & Kesselheim, *New FDA Breakthrough-Drug Category*] (explaining how some patient advocacy groups argue that lengthy FDA approval requirements delay access to treatments and restrict personal autonomy).

7. *Id.* at 1253.

8. *See* Kevin Knopf, Michael Baum, William S. Shimp, Charles L. Bennett, Dinah Faith, Marc L. Fishman & William J. M. Hrushesky, *Interpretation of Surrogate Endpoints in the Era of the 21st Century Cures Act*, 355 BMJ i6286, i6286 (2016); *see also* William S. Weintraub, Thomas F. Lüscher & Stuart Pocock, *The Perils of Surrogate Endpoints*, 36 EUR. HEART J. 2212, 2213 tbl.1

Commentators have critiqued accelerated pathways, surrogate endpoints, or both.[9] They argue that by adopting these innovations, the FDA succumbed to external pressures to hasten approval, abandoning its commitment to scientific rigor.[10] Some fear that accelerated pathways and surrogate endpoints allow the FDA to cut corners, exposing consumers to riskier drugs and burdening the public with paying for ineffective medical treatments.[11]

Concerns regarding industry influence and the FDA's evidentiary standards gained national attention in 2021, following the agency's controversial approval of the Alzheimer's drug aducanumab, a process that relied on surrogate endpoints.[12] Experts claim the FDA ignored a lack of clear evidence for safety and efficacy when it approved aducanumab against the advice of an expert advisory committee.[13] Three respected advisors resigned in protest.[14] Some say the agency reached a new low.[15]

In addition to substituting surrogate endpoints for direct observations of product efficacy, the FDA is augmenting and replacing laboratory and clinical trial data with predictions from computer models and simulations.[16] These technologies increasingly rely on

(2015) (providing a table of potential surrogate endpoints, including biochemical markers measured in body fluids, radiographic data, physical measurements, and other variables).

9. *See, e.g.*, Knopf et al., *supra* note 8 (arguing that surrogate endpoints lower the level of evidence required to rush products to market, which can facilitate opportunistic behavior); *see also* Weintraub et al., *supra* note 8, at 2214 (claiming that surrogate endpoints are often unreliable).

10. *See* Jonathan J. Darrow, Jerry Avorn & Aaron S. Kesselheim, *FDA Approval and Regulation of Pharmaceuticals, 1983-2018*, 323 JAMA 164, 173 (2020) (describing pressure from patients and drug manufacturers to speed FDA approval and to adopt surrogate measures for efficacy).

11. *See, e.g.*, Knopf et al., *supra* note 8, at i6288.

12. Aaron S. Kesselheim & Jerry Avorn, *The F.D.A. Has Reached a New Low*, N.Y. TIMES (June 15, 2021), https://www.nytimes.com/2021/06/15/opinion/alzheimers-drug-aducanumab-fda.html [https://perma.cc/3B28-ZPC8].

13. *Id.*

14. Andrew Joseph, *Third Member of FDA Expert Committee Resigns Over Controversial Alzheimer's Therapy Decision*, STAT (June 10, 2021), https://www.statnews.com/2021/06/10/third-member-of-fda-expert-committee-resigns-over-controversial-alzheimers-therapy-decision [https://perma.cc/9MDE-UV2M].

15. Kesselheim & Avorn, *supra* note 12.

16. *See* Scott Gottlieb, *How FDA Plans To Help Consumers Capitalize on Advances in Science*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/news-events/fda-voices/how-fda-plans-help-consumers-capitalize-advances-science [https://perma.cc/TJ5M-8R4R] (last updated June 28, 2018) (announcing that the FDA uses computer models and simulations "to predict

artificial intelligence ("AI") to infer the effects of drugs and medical devices.[17] In 2018, the FDA completed the Virtual Imaging Trial for Regulatory Evaluation ("VICTRE"), a digital trial that analyzed the images of 2,986 virtual patients to predict the effectiveness of a new form of digital breast imaging.[18]

Modeling and simulation are part of a larger trend in which administrative agencies automate many facets of their operations. In the context of public health specifically, AI is changing how the FDA collects information, performs research, evaluates medical products, makes policy decisions, and monitors the population for adverse events.[19] In each case, AI impacts the quantity and quality of evidence on which the FDA relies to make policy and regulatory decisions, replacing direct evidence with algorithmic predictions.

Existing scholarship concerning AI at the FDA addresses how the agency should regulate medical devices that make algorithmic predictions to diagnose and treat patients.[20] These products include modern versions of traditional medical devices, such as pacemakers and insulin pumps enhanced by AI, as well as new classes of products that are entirely software-based, such as smartphone apps and clinical decision support software. Both types of products pose unique risks to consumers, and the FDA is determining how to best evaluate their safety and efficacy.[21]

---

clinical outcomes, inform clinical trial designs, support evidence of effectiveness, optimize dosing, predict product safety, and evaluate potential adverse event mechanisms").

17. *See* Likhitha Kolla, Fred K. Gruber, Omar Khalid, Colin Hill & Ravi B. Parikh, *The Case for AI-Driven Cancer Clinical Trials – The Efficacy Arm In Silico*, 1876 BBA – REVS. ON CANCER 188572, 188572–74 (2021) (explaining the advantages of adopting AI models for data analysis over biomathematical models to conduct simulated clinical trials).

18. *Id.* at 188574.

19. *See id.*; *see also* Diksha Sharma, Christian G. Graff, Andreu Badal, Rongping Zeng, Purva Sawant, Aunnasha Sengupta, Eshan Dahal & Aldo Badano, *Technical Note: In Silico Imaging Tools from the VICTRE Clinical Trial*, 46 MED. PHYSICS 3924, 3924 (2019) (describing the FDA's use of a computational model to perform a virtual clinical trial to predict the effectiveness of a new form of diagnostic imaging).

20. *See, e.g.*, Andrea M. Matwyshyn, *The Internet of Bodies*, 61 WM. & MARY L. REV. 77, 131 (2019) (recommending that the FDA require extensive premarket disclosures from companies that produce medical products that utilize AI, including the results of third-party code audits).

21. *See generally* U.S. FOOD & DRUG ADMIN., ARTIFICIAL INTELLIGENCE/MACHINE LEARNING (AI/ML)-BASED SOFTWARE AS A MEDICAL DEVICE (SAMD) ACTION PLAN (2021) (describing the FDA's proposed steps for regulating AI-based medical devices in response to stakeholder input).

This Article, however, addresses the other side of AI at the FDA. Instead of serving only as a regulated medical product, AI increasingly regulates other medical products through its role in computer modeling and simulation.[22] To analyze how these tools are changing FDA regulation, this Article introduces three nascent technologies: molecular modeling, virtual humans, and *in silico* ("simulated") clinical trials. These tools use AI to simulate biological phenomena, and like surrogate endpoints, they allow the FDA to reduce its reliance on clinical evidence.[23] Instead of interpreting data from living research subjects, the agency may rely on algorithmic predictions drawn from simulated people and medical products.

Industry and agency representatives praise simulations for reducing the risk and expense of clinical trials while decreasing the time required to obtain FDA approval.[24] They argue that simulations make clinical research more efficient, equitable, and inclusive.[25] Some claim the prevailing model for clinical trials, which contributed to the FDA's strong reputation during the twentieth century, is burdensome, costly, and perhaps even outdated.[26]

Despite their potential benefits, models and simulations have significant limitations and may pose serious public health risks. Like surrogate endpoints and accelerated approval pathways, they could

22. *See* Gottlieb, *supra* note 16; *see also* Kolla et al., *supra* note 17, at 188572 (describing the role of simulations in evaluating new drugs where computer models and AI might conserve resources, replace the control and intervention arms of clinical trials, and optimize the recruitment of research subjects).

23. *See, e.g.*, Biovia, *SAOE 2018: Credible Modeling and Simulation at the FDA*, YOUTUBE (July 10, 2018) [hereinafter Biovia], https://www.youtube.com/watch?v=CRF8JWPaVXU [https://perma.cc/MVQ6-C25H] (describing the benefits of virtual clinical trials, including reduced reliance on human trials).

24. *See, e.g.*, Ansys, *In Silico Healthcare: The Power To Cure Diseases, To Change the World*, YOUTUBE (Aug. 17, 2018), https://www.youtube.com/watch?v=fpPx0D0ZTlM [https://perma.cc/LS4M-JEPW] (arguing that simulation significantly reduces the time and expense of obtaining FDA approval and that companies adopting virtual clinical trials receive a 500 percent return on their investment); *see also* Biovia, *supra* note 23 (claiming that simulations can reduce the cost of research and predict the performance of medical products in conditions that would be too dangerous for human research subjects).

25. *See, e.g.*, Aldo Badano, *In Silico Imaging Clinical Trials: Cheaper, Better, Faster, and More Scalable*, 22 TRIALS 1, 1, 4–5 (2021) (stating that simulated imaging trials can be substituted for clinical trials "at a fraction of the cost," while supporting similar regulatory decisions); *see also* Kolla et al., *supra* note 17, at 188573 (stating that simulated trials lower the cost of research by shortening its duration, eliminating costs associated with recruiting patients, and reducing the physical resources required).

26. *See, e.g.*, Kolla et al., *supra* note 17, at 188572.

erode the FDA's evidentiary standards and expose the public to dangerous and unreliable products.[27] They may also camouflage industry influence over the regulatory process.[28] Yet agency and industry reports usually exclude these potential harms.[29] Moreover, academic scholarship on AI has largely overlooked its role in FDA computer models and simulations.[30]

This Article helps fill the gap in AI research by analyzing the ethical, legal, and social implications of relying on models and simulations in the context of public health regulation. Without careful consideration of the risks and the implementation of appropriate safeguards, automating FDA regulation could harm consumers, exacerbate existing inequality, erode public trust, and undermine agency accountability and legitimacy.

This Article contains four Parts. Part I summarizes FDA history, including how the agency's authority and reputation have changed over time. It describes the FDA's increasing reliance on premarket clinical evidence in the mid-twentieth century, its shift away from clinical data by the turn of the century, and its increasing emphasis on postmarket surveillance in the early twenty-first century. Part I concludes by analyzing how political and industry influence—and a series of questionable decisions—have eroded public trust in the FDA.

Part II evaluates the FDA's existing and future uses for AI by introducing three nascent modeling and simulation technologies. It explains how they streamline agency functions, hasten the shift away

27. *See, e.g.*, Knopf et al., *supra* note 8, at i6288 (describing the risks of relying on surrogate endpoints).

28. *See id.* at i6286 (explaining how stakeholders may exploit the use of surrogate endpoints, and other means of subverting fundamental principles of randomized trials, for personal gain while rushing therapies to market).

29. *See, e.g.*, Christopher R. Ellis, Rebecca Racz, Naomi L. Kruhlak, Marlene T. Kim, Alexey V. Zakharov, Noel Southall, Edward G. Hawkins, Keith Burkhart, David G. Strauss & Lidiya Stavitskaya, *Evaluating Kratom Alkaloids Using PHASE*, 15 PLOS ONE e0229646, e0229646 (2020) (omitting discussion of the public health risks associated with biased or inaccurate predictions).

30. For two examples that have addressed the role of AI in FDA models and simulations, see DAVID FREEMAN ENGSTROM, DANIEL E. HO, CATHERINE M. SHARKEY & MARIANO-FLORENTINO CUÉLLAR, GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES 53 (2020), https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf [https://perma.cc/5S6Q-XC3S] (discussing attempts to use AI to analyze reports in the FDA Adverse Event Reporting System, which the FDA concluded were unsuccessful) and David W. Opderbeck, *Artificial Intelligence in Pharmaceuticals, Biologics, and Medical Devices: Present and Future Regulatory Models*, 88 FORDHAM L. REV. 553, 570–71 (2019) (introducing the FDA's perspective on in silico trials).

from clinical evidence, and transform AI from a regulated medical product to a regulator of other medical products. This Part draws upon the fields of AI ethics and data protection regulation to analyze the risks and benefits of these technologies, including their impact on patient safety, research equity, medical innovation, and public health.

Part III analyzes how administrative law doctrine might apply to FDA modeling and simulation. It evaluates the impact of these technologies on agency transparency, accountability, and legitimacy, while evaluating the limits of administrative law to effectively address the effects. It concludes that in some cases, automating FDA regulation could undermine historical justifications for delegating congressional authority to agencies and erode the FDA's credibility.

Part IV analyzes agency guidance documents and proposals for good simulation practices. It concludes that administrative law and the FDA should adopt principles of AI ethics and offers recommendations for safer and more equitable automation of public health regulation.

I. FROM SNAKE OIL TO SURROGATES: THE RISE AND FALL OF CLINICAL EVIDENCE

To contextualize the FDA's adoption of computer models and simulations, this Part summarizes the agency's history and how its treatment of evidence has changed over time. The FDA is a public health agency housed within the Department of Health and Human Services ("HHS"), and its effects on the economy and public safety are substantial.[31] FDA-regulated products account for approximately 20 percent of U.S. household spending, equaling over $2.5 trillion.[32] However, the agency had far more humble beginnings. In the early twentieth century, its predecessor, the Bureau of Chemistry, provided

31. *See FDA Organization Charts*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/about-fda/fda-organization/fda-organization-charts [https://perma.cc/2GX5-6ZVY] (last updated Mar. 21, 2019) (depicting the organizational hierarchy of the FDA's nine different center-level organizations); *see also* ENGSTROM ET AL., *supra* note 30, at 54 (describing the FDA's impact on the U.S. economy).

32. ENGSTROM ET AL., *supra* note 30, at 54.

only limited protection from harmful products.[33] The Bureau required no testing for product safety or efficacy.[34]

In 1906, Congress enacted the Pure Food and Drugs Act in response to widespread drug contamination and adulteration.[35] The act required accurate labeling to prevent false or misleading claims regarding a drug's identity and composition.[36] However, the law did not prohibit false therapeutic claims.[37]

In 1927, the Bureau split into two new agencies, and regulatory functions were moved to the newly formed Food, Drug, and Insecticide Administration.[38] In 1938, after the agency's name was shortened to the Food and Drug Administration,[39] Congress increased its power by enacting the FDCA, which empowered the agency to require evidence of safety before drugs could be marketed.[40] Despite this advancement, the FDCA had shortcomings. If the FDA did not act on a new drug application within sixty days, the drug was automatically approved.[41]

In 1959, following the high-profile thalidomide scandal, in which a drug sold for treating morning sickness in pregnant women caused birth defects in thousands of babies, Senator Estes Kefauver held hearings to improve the FDCA.[42] The resulting Kefauver-Harris Drug Amendments required manufacturers to provide evidence of safety and efficacy before drugs could be marketed.[43] Evidence was obtained

---

33. *See Milestones in U.S. Food and Drug Law*, U.S. FOOD & DRUG ADMIN. [hereinafter *Milestones*], https://www.fda.gov/about-fda/fda-history/milestones-us-food-and-drug-law [https://perma.cc/CS8T-E4DP] (last updated Jan. 31, 2018).

34. *See id.* (noting that evidence of safety was not required until 1938 and testing for efficacy was not required until 1962).

35. *Id.*

36. United States v. Johnson, 221 U.S. 488, 497 (1911).

37. *See id.*

38. Wallace F. Janssen, *The Story of the Laws Behind the Labels*, FDA CONSUMER MAG., June 1981, https://lessonbank.kyae.ky.gov/wp-content/uploads/2019/02/The-Story-of-the-Laws-Behind-the-Labels.pdf [https://perma.cc/QQS3-ZYMY]; *see also Milestones*, *supra* note 33 (describing how nonregulatory research was shifted to the newly formed Bureau of Chemistry and Soils).

39. Janssen, *supra* note 38.

40. *See* CARPENTER, *supra* note 1.

41. TOM BRODY, CLINICAL TRIALS 782 (2d ed. 2016).

42. *Id.* at 783 (describing enactment of the Kefauver-Harris Amendments); *see also* Frederick Dove, *What's Happened to Thalidomide Babies*, BBC NEWS (Nov. 3, 2011), https://www.bbc.com/news/magazine-15536544 [https://perma.cc/KL4C-KEWY] (describing the harmful effects of thalidomide).

43. Greene & Podolsky, *supra* note 3.

from "adequate and well-controlled investigations."[44] Modern three-phased randomized controlled trials meet this standard.

In clinical research, Phase 1 trials evaluate the safety of new medical products in small numbers of healthy volunteers.[45] Drug manufacturers must submit an Investigational New Drug application before commencing a Phase 1 trial.[46] This stage typically includes twenty to one hundred participants and can last several months.[47] Phase 2 trials add more participants, last longer, and evaluate effectiveness in addition to safety.[48] They include several hundred participants with specific medical conditions and last from several months to two years.[49]

Phase 3 trials are the final stage before evidence is submitted to the FDA for review and approval.[50] Like Phase 2 trials, they evaluate both safety and efficacy in people with specific health conditions.[51] However, Phase 3 trials are much larger, consisting of a few hundred to a few thousand participants, and they last from one to four years.[52] After completing Phase 3 trials, drug makers must submit a New Drug Application before the FDA will consider their products for approval, which allows drugs to be marketed in the United States.[53]

Phase 4 is a more recent innovation that consists of postmarket surveillance in which manufacturers monitor public use of their FDA-approved products.[54] Postmarket surveillance has been criticized because manufacturers often fail to follow through with it, and

---

44. Drug Amendments of 1962 § 102, 21 U.S.C. § 355(d); *see also* Greene & Podolsky, *supra* note 3, at 1482.

45. *Step 3: Clinical Research*, U.S. FOOD & DRUG ADMIN. (Jan. 4, 2018) [hereinafter *Clinical Research*], https://www.fda.gov/patients/drug-development-process/step-3-clinical-research [https://perma.cc/4MGV-T2MG].

46. *Id.*

47. *Id.*

48. *See id.*

49. *Id.*

50. *New Drug Application (NDA)*, U.S. FOOD & DRUG ADMIN. [hereinafter *NDA*], https://www.fda.gov/drugs/types-applications/new-drug-application-nda [https://perma.cc/E4H7-TDDV] (last updated June 10, 2019).

51. *Clinical Research*, *supra* note 45.

52. *Id.*

53. *NDA*, *supra* note 50.

54. *See* Daniel Carpenter, *Reputation, Gatekeeping and the Politics of Post-Marketing Drug Regulation*, 8 ETHICS J. AM. MED. ASS'N 403, 404 (2006).

compliance can be difficult to monitor and enforce.[55] Despite these shortcomings, in recent years, Congress and the FDA have pushed for increasing reliance on postmarket surveillance, which was codified in the 21st Century Cures Act.[56]

In addition to being lengthy and expensive, clinical trials can also expose people to physical and psychological risks.[57] Participants can be harmed by unproven interventions, or, if placed in a control group, they can receive placebos instead of the experimental treatment they had hoped to receive.[58] Due to the length and expense of randomized controlled trials, patient advocacy groups and industry stakeholders pressured the FDA to accelerate the process by creating exceptions to its standard practices.[59]

Calls for faster FDA approval prompted the creation of four accelerated pathways, including priority review, accelerated approval, fast track designation, and breakthrough therapy designation.[60] The first path, priority review, speeds up FDA review of completed trial data.[61] The latter three pathways reduce the mean time to FDA approval by several years, in part by allowing the substitution of surrogate endpoints for clinical evidence.[62]

---

55. *See id.* (reporting that many postmarketing commitments are not honored, in part because the FDA has more power over sponsors during the premarket period).

56. *See, e.g.*, Deborah Mazer & Gregory D. Curfman, *21st Century Cures Act Lowers Confidence in FDA Approved Drugs and Devices*, HEALTH AFFS. BLOG (Feb. 14, 2017), https://www.healthaffairs.org/do/10.1377/hblog20170214.058710/full [https://perma.cc/QY3V-G837] (criticizing the 21st Century Cures Act for expanding the use of postmarket surveillance data).

57. *See, e.g.*, Charles Schmidt, *The Struggle To Do No Harm*, 552 NATURE S74, S74 (2017) (describing unexplained deaths that have occurred during cancer therapy trials).

58. *See id.*; *see also* S. Ito, *Placebo in Clinical Trials*, 90 CLINICAL PHARMACOLOGY & THERAPEUTICS 637, 637 (2011) (defining placebos and describing their use in clinical trials).

59. *See* Darrow, Avorn & Kesselheim, *supra* note 10, at 173; *see also* Christopher H. Foreman, Jr., *The Fast Track: Federal Agencies and the Political Demand for AIDS Drugs*, 9 BROOKINGS REV. 30, 31–32 (1991) (describing the impact of AIDS activism on FDA approval timelines).

60. *See* Darrow, Avorn & Kesselheim, *supra* note 10, at 169.

61. *Priority Review*, U.S. FOOD & DRUG ADMIN. (Jan. 4, 2018), https://www.fda.gov/patients/fast-track-breakthrough-therapy-accelerated-approval-priority-review/priority-review [https://perma.cc/4T2J-XH3L] (explaining that priority review indicates the FDA's intent to review a new drug application within six months, compared to the usual review period of ten months).

62. Darrow, Avorn & Kesselheim, *New FDA Breakthrough-Drug Category*, *supra* note 6, at 1253–55.

Some experts criticize the FDA's increasing reliance on accelerated pathways, arguing it erodes evidentiary standards.[63] While approving the Gardasil vaccine for cervical cancer through its fast track pathway, for example, the FDA accepted surrogate endpoints in lieu of direct evidence.[64] Critics claimed the surrogates were of little value in determining the vaccine's long-term effectiveness for cancer prevention.[65]

Further, some experts believe the shift away from direct evidence of clinical improvement has occurred because the agency grew too close to the industries it regulates.[66] When the FDA approved aducanumab based on surrogate endpoints, it went against the advice of an expert advisory committee.[67] Three members of the committee resigned.[68] One of them, Dr. Aaron Kesselheim, described the decision as "probably the worst drug approval decision in recent US history."[69] Kesselheim said the agency and representatives of drug sponsor Biogen seemed unusually aligned in support of approving aducanumab.[70] Daniel Carpenter, professor of government at Harvard University, concluded, "We're seeing what happens when the FDA loses credibility."[71]

Automating FDA regulation through the substitution of models and simulations for clinical evidence may further erode the FDA's evidentiary standards and undermine agency credibility. Part II describes how these tools currently influence the FDA's research and regulatory decisions.

---

63. *Id.* at 1255; Darrow, Avorn & Kesselheim, *supra* note 10, at 173.

64. *See* Lucija Tomljenovic & Christopher A. Shaw, *Too Fast or Not Too Fast: The FDA's Approval of Merck's HPV Vaccine Gardasil*, 40 J.L. MED. & ETHICS 673, 674–75 (2012).

65. *Id.* at 675.

66. *See* Jeffrey Toobin, *The Road to Aduhelm: What One Ex-FDA Advisor Called 'Probably the Worst Drug Approval Decision in Recent US History' for an Alzheimer's Treatment*, CNN (Sept. 27, 2021, 10:01 AM), https://www.cnn.com/2021/09/26/politics/alzheimers-drug-aduhelm-fda-approval/index.html [https://perma.cc/9P89-DNSV] (describing how the FDA overruled its expert advisory committee).

67. *Id.*

68. *Id.*

69. *Id.*

70. *Id.*

71. Daniel Carpenter, *We're Seeing What Happens When the FDA Loses Credibility*, WASH. POST (July 21, 2021, 6:00 AM), https://www.washingtonpost.com/politics/2021/07/21/were-seeing-what-happens-when-fda-loses-credibility [https://perma.cc/2BA9-9BTJ].

II. BUILDING AN ALGORITHMIC FDA

This Part introduces three technologies used by the FDA to simulate biological systems in the interest of supporting research and regulatory decision-making. Molecular modeling, virtual humans, and simulated clinical trials are part of a larger trend in which sixty-four agencies have adopted, or plan to adopt, AI systems to automate a variety of governance tasks.[72] The following discussion summarizes this phenomenon.

A. Automating the Administrative State

AI has captured the imagination of the media, the public, and industries from Wall Street to Silicon Valley.[73] Federal agencies are also experimenting with AI to automate their operations.[74] According to a 2020 report drafted for the Administrative Conference of the United States ("ACUS"), the maturation of AI technology, and its adoption by federal agencies, may be one of the most important developments for the administrative state in decades.[75]

The ACUS report canvassed 142 federal departments and agencies to determine which had adopted AI tools.[76] It concluded that 45 percent of the organizations studied had experimented with AI.[77] The authors identified 157 use cases across sixty-four agencies.[78] The Office of Justice Programs had the most use cases with a total of 12.[79] The Securities and Exchange Commission had the second-highest

72. ENGSTROM ET AL., *supra* note 30, at 16.

73. *See, e.g.*, FRANCESCO MARCONI, NEWSMAKERS: ARTIFICIAL INTELLIGENCE AND THE FUTURE OF JOURNALISM 20 (2020) (describing journalism of the future in which human reporters collaborate with AI); George Dvorsky, *How an Artificial Superintelligence Might Actually Destroy Humanity*, GIZMODO (May 26, 2021, 10:20 AM), https://gizmodo.com/how-an-artificial-superintelligence-might-actually-dest-1846968207 [https://perma.cc/2XYD-FTXA] (arguing that AI poses an existential threat to humanity); *see also* Steve Taplin, *How AI Is Connecting Employers with Software Engineers*, ENTREPRENEUR (Dec. 10, 2021), https://www.entrepreneur.com/article/396480 [https://perma.cc/LSD6-SHN2] (describing how employers increasingly use AI to recruit top talent). *See generally* Mason Marks, *Artificial Intelligence Based Suicide Prediction*, 21 YALE J.L. & TECH. 98 (2019) (describing how social media companies use AI to predict which users have mental health conditions and who is most likely to attempt suicide).

74. ENGSTROM ET AL., *supra* note 30, at 6.

75. *Id.* at 9.

76. *Id.* at 6.

77. *Id.*

78. *Id.* at 16.

79. *Id.* at 16 tbl.2.

number of cases with 10, and the National Aeronautics and Space Administration had 9.[80] With 8 use cases, the FDA tied with two other agencies, the U.S. Geological Survey and the U.S. Postal Service, for the fourth-highest number of cases.[81] Four other agencies, the Social Security Administration, the U.S. Patent and Trademark Office ("PTO"), the Bureau of Labor Statistics, and U.S. Customs and Border Protection, each had between 4 and 7 use cases.[82]

Agencies are adopting AI across diverse policy areas to assist with many government functions, including adjudication, enforcement, public engagement, internal management, and regulatory research, analysis, and monitoring.[83] For instance, the Bureau of Labor Statistics uses AI to sort injury narratives submitted by workers, the Securities and Exchange Commission uses AI to identify people most at risk of violating securities regulations, and the PTO uses AI to help adjudicate patent applications.[84]

The ACUS report also indicated that law enforcement was the policy area most likely to receive AI support, with over 30 use cases, and education was the least likely with fewer than 5.[85] Health was the second most automated area with nearly 20 use cases, almost half as many as law enforcement.[86] Financial regulation, social welfare, commerce, and environmental regulation followed close behind with between 10 and 15 use cases each.[87]

Regulatory research, analysis, and monitoring were the most frequently automated governance tasks with about 80 uses.[88] This category includes AI tools that help agencies collect and analyze data to aid policymaking.[89] Adjudication had the fewest use cases with approximately 12.[90] Adjudication includes AI tools that assist formal or informal adjudication of rights and benefits.[91]

---

80. *Id.*
81. *Id.*
82. *Id.*
83. *Id.* at 10 tbl.1.
84. *Id.*
85. *Id.* at 17 fig.1.
86. *Id.*
87. *Id.*
88. *Id.* at 17 fig.2.
89. *Id.*
90. *Id.*
91. *Id.*

Only one third of use cases had been fully deployed by agencies.[92] Just over 40 had been partially deployed or were restricted to pilot programs, and the majority of uses, about 60, remained in the planning stages.[93] Over half of the AI systems, representing about 84 cases, had been created in-house by agency staff.[94] Nearly 50 were developed externally by commercial contractors, and about 20 were produced through noncommercial collaborations between agencies and academic labs or through public-facing competitions.[95] The implications of who develops agency AI systems and what kind of data they analyze will be discussed further below.

Most use cases, about 80 of 157, analyzed structured data consisting of numerical information.[96] The remaining cases analyzed unstructured data, which consisted of text in about 70 cases, images in approximately 40 cases, and audio in about 5 cases.[97]

Arguably, some federal agencies can automate many operations without endangering the public.[98] However, because the FDA creates and enforces public health policies and regulates numerous medical products across diverse industries, automating its operations could profoundly impact the health of millions.[99] Moreover, because the agency strongly influences international policy, the impact is potentially global.[100] The following section introduces three nascent AI technologies that the FDA uses to automate a variety of governance tasks.

B. FDA Modeling and Simulation

Computer modeling uses software, mathematics, and physics to create abstract representations of complex systems and study their

92. *Id.* at 18.

93. *Id.* at 18 fig.3.

94. *Id.* at 18 fig.4.

95. *Id.*

96. *Id.* at 19 fig.6.

97. *Id.*

98. *See, e.g.*, Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine Learning Era*, GEO. L.J. 1147, 1169 (2017) (arguing that automated mail sorting by the U.S. Postal Service poses little social or legal risk).

99. *See What Does FDA Regulate?*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/about-fda/fda-basics/what-does-fda-regulate [https://perma.cc/5LC3-Y6YD] (last updated June 24, 2021).

100. *See* CARPENTER, *supra* note 1, at 687 (describing the strong international influence of FDA policies).

behavior.[101] Models are built to represent many complex systems including molecules, aircraft, weather systems, and earthquakes. They are particularly useful for solving problems associated with systems for which there are no easier means of analysis.[102]

Each model contains variables that collectively approximate the corresponding system.[103] Simulation of various conditions can be achieved by adjusting one or more variables and observing the results.[104] Connecting different models allows researchers to simulate more complex phenomena. For instance, a model of the solar system might contain other models that represent the planets, and each planetary model might contain models for an atmosphere and its weather. The term "multiscale modeling" describes the process of combining models representing phenomena at different scales.[105] Combining models at the tissue, organ, and organismal levels, for example, enables simulation of disease processes and the effects of therapeutic intervention.[106]

The FDA ranks computer modeling and simulation among its core strategic priorities.[107] The agency aims to reduce its reliance on clinical trials by replacing them with simulated trials, in which products are tested on populations of virtual patients comprising multiscale models representing different physiologic processes.[108] According to Dr. Tina Morrison, director of the Office of Regulatory Science and Innovation

---

101. *See Computational Modeling*, NAT'L INST. OF BIOMEDICAL IMAGING & BIOENGINEERING, https://www.nibib.nih.gov/science-education/science-topics/computational-modeling [https://perma.cc/JXP7-8GXX] (last updated May 2020); *see also* William A. Menner, *Introduction to Modeling and Simulation*, 16 JOHNS HOPKINS APL TECH. DIG. 6, 6 (1995) (defining a "model" as an abstraction of a system).

102. Menner, *supra* note 101.

103. *Computational Modeling*, *supra* note 101.

104. *Id.*

105. *Id.*

106. *Id.*

107. *See Credibility of Computational Models Program: Research on Computational Models and Simulation Associated with Medical Devices*, U.S. FOOD & DRUG ADMIN. [hereinafter *Credibility of Computational Models Program*], https://www.fda.gov/medical-devices/medical-device-regulatory-science-research-programs-conducted-osel/credibility-computational-models-program-research-computational-models-and-simulation-associated [https://perma.cc/LE65-G9Q8] (last updated Mar. 24, 2021); *see also* Tina M. Morrison, Prasanna Hariharan, Chloe M. Funkhouser, Payman Afshari, Mark Goodin & Marc Horner, *Assessing Computational Model Credibility Using a Risk-Based Framework: Application to Hemolysis in Centrifugal Blood Pumps*, 65 ASAIO J. 349, 349 (2019) ("Computational modeling continues to be a top regulatory science priority for CDRH for medical device evaluation.").

108. *Credibility of Computational Models Program*, *supra* note 107.

at the FDA, "[R]ealistically, modeling and simulation does not play a driving role in regulatory decision making. But we want it to."[109]

Alluding to the FDA's shift away from premarket data, Morrison added, "We want to lessen the burden of evidence from clinical and animal studies. We want to do the work that's necessary, and not the work we think is needed just because that's what we've done for twenty years."[110] However, despite the promise of computer models, regulators acknowledge that their credibility is not well established.[111] Moreover, adopting models hastily or haphazardly can produce erroneous conclusions.[112]

Credibility reflects a model's ability to elicit trust in its predictions for a specific context of use.[113] According to FDA guidance, many existing and proposed algorithmic models have not been rigorously evaluated, and their credibility is unknown.[114] Others have known deficiencies that negatively affect their credibility.[115]

The FDA identifies several major gaps and challenges of establishing model credibility.[116] There is a paucity of quality clinical and experimental data to develop and validate models and simulations, especially information derived from humans under real-world conditions, that includes data from sufficiently diverse populations.[117] There is also a lack of validation tools and metrics, and there are no reliable methods to evaluate the acceptability of virtual patients in a cohort.[118] Moreover, there are no tools to reliably verify relevant computer code and calculations, and decision-making frameworks for evaluating the overall credibility of models and simulations remain rudimentary.[119]

---

109. Biovia, *supra* note 23, at 16:44.

110. *Id.* at 16:50.

111. *Credibility of Computational Models Program*, *supra* note 107.

112. Menner, *supra* note 101.

113. Carl Popelar, Tina Morrison, Andrew Rau & Ryan Crane, *V&V for Computational Modeling for Medical Devices*, ASME, https://www.imagwiki.nibib.nih.gov/sites/default/files/asme_vv_40_model_simulation_standard.pdf [https://perma.cc/Y249-45RE].

114. *Credibility of Computational Models Program*, *supra* note 107.

115. *Id.*

116. *See id.*

117. *Id.*

118. *Id.*

119. *Id.*

Importantly, there are also no established best practices for using models and simulations or relying on them to make regulatory decisions.[120] To address these concerns, the FDA's Center for Devices and Radiologic Health ("CDRH") created an initiative called the Credibility of Computational Models Program, which is one of twenty research initiatives within the FDA's Office of Science and Engineering Laboratories.[121]

When FDA officials and industry stakeholders discuss the shortcomings of computer models and simulations, their descriptions often include vague concepts such as trust and credibility, while omitting specifics regarding the risks of substituting algorithmic models for FDA staff judgment and how to mitigate those risks.[122] One frequently overlooked concern is how algorithmic bias becomes baked into machine learning models.

Machine learning is a form of AI that excels at pattern recognition.[123] It analyzes large data sets and discovers correlations that are imperceptible to humans.[124] Before deploying machine learning models, programmers must train them on historical data.[125] The correlations derived from training data can then be deployed as rules to make predictions from a new data set. In this manner, the outputs of machine learning models are shaped by the examples to which they are exposed during training, and biased sets of training data produce biased models and predictions.[126] For instance, training predictive models on data sets that underrepresent certain populations can produce predictions that are inaccurate for those groups.[127] Moreover, training data that reflects prejudices directed at certain communities

120. *Id.*

121. *See id.*

122. *See generally* Ellis et al., *supra* note 29 (describing the FDA's use of computer models and artificial intelligence to predict how substances affect the body and impact public health).

123. David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 671 (2017).

124. *See* Mason Marks, *Emergent Medical Data: Health Information Inferred by Artificial Intelligence*, 11 U.C. IRVINE L. REV. 995, 1005–06 (2021).

125. *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 678 (2016).

126. *See id.* at 680–81.

127. *See* Effy Vayena, Alessandro Blasimme & I. Glenn Cohen, *Machine Learning in Medicine: Addressing Ethical Challenges*, 15 PLOS MED. e1002689, e1002690 (2018) (describing how an algorithm trained largely on data from a population of older white men would yield inaccurate predictions for younger Black women).

often produces models that retain those prejudices, and their biased outputs may disproportionately impact certain groups.[128] In one example, a machine learning model concluded that Black patients were healthier than they actually were.[129] The algorithm made this error because it used healthcare costs as a proxy for medical need; because less money was spent on Black patients compared to white patients of comparable health, the algorithm incorrectly concluded that the Black patients were healthier than their white counterparts.[130] In this case, an existing bias in healthcare spending was incorporated into the algorithm and caused a false conclusion.[131] If relied on to direct healthcare resources, the algorithm could have incorrectly directed those resources away from Black patients.

Not every computer model uses machine learning. Some use well-known mathematical relationships and physical principles to predict outcomes.[132] Because the models rely on known equations that are not derived from training data, the outputs of these *physics-based* models may be more explainable than *data-driven* machine learning algorithms that create their own rules by analyzing large data sets. Some machine learning algorithms, called white box models, are more explainable than others because they rely on established patterns, rules, or decision trees.[133] Other models, called black boxes, can be difficult to understand and explain, even by experts in the relevant field.[134] In practice the line between white and black box machine learning models is not always clear, and some models called grey boxes are hybrids of

128. Barocas & Selbst, *supra* note 125, at 681.

129. *See* Kerstin N. Vokinger, Stefan Feuerriegel & Aaron S. Kesselheim, *Mitigating Bias in Machine Learning for Medicine*, COMMC'NS MED., Aug. 23, 2021, at 1, 1.

130. *Id.*

131. *Id.*

132. *See* Hector Klie, *A Tale of Two Approaches: Physics-Based vs. Data-Driven Models*, J. PETROL. TECH. (May 3, 2021), https://jpt.spe.org/a-tale-of-two-approaches-physics-based-vs-data-driven-models [https://perma.cc/NZP7-JLCP].

133. *See* Octavio Loyola-González, *Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses from a Practical Point of View*, 7 IEEE ACCESS 154096, 154096 (2019) (defining white box models).

134. *Id.* at 154097.

the two.[135] Combining elements of physics-based and data-driven models can improve predictions to help solve real-world problems.[136]

Biased predictions can occur in the absence of nefarious intent because bias can creep into algorithmic models at many stages.[137] Researchers can introduce bias while selecting training data, when processing and labeling the data to make it more easily readable by computers, while selecting models for a task, or when interpreting algorithmic predictions.[138] A detailed description of these steps is beyond the scope of this Article, and other authors provide comprehensive explanations of this topic.[139] Suffice to say, accounting for bias in computer models and simulations is complex and challenging.

One might assume that as a scientific agency, the FDA's models are less susceptible to bias than algorithms used in other contexts such as employment screening, policing, and criminal sentencing. Unlike predictive models used in those contexts, which directly impact individuals, the FDA's predictions often affect healthcare products, medical research, and the rights of manufacturers. Because algorithmic bias impacts people less directly in the public health context, it may be less obvious. However, because the FDA regulates thousands of products across numerous industries that affect the health and safety of millions, this algorithmic bias is no less harmful.[140]

The following discussions analyze the current and potential roles of three modeling technologies in FDA regulation. Like surrogate endpoints, the outputs of these technologies serve as indirect evidence of product safety and efficacy. Despite acknowledging that they lack credibility, the FDA increasingly relies on them to shape its policies.

---

135. *See* Emmanuel Pintelas, Ioannis E. Livieris & Panagiotis Pintelas, *A Grey-Box Ensemble Model Exploiting Black-Box Accuracy and White-Box Intrinsic Interpretability*, 13 ALGORITHMS 17, 18 (2020).

136. Klie, *supra* note 132.

137. *See id.* at 674, 681–93.

138. *Id.* at 681–93; *see also* Lehr & Ohm, *supra* note 123, at 653–54 (listing the stages of machine learning).

139. *See generally* Lehr & Ohm, *supra* note 123 (describing the steps in the machine learning process).

140. *See Fact Sheet: FDA at a Glance*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/media/154548/download [https://perma.cc/JD75-V36T] (last updated Nov. 2021) (stating that the FDA oversees more than $2.7 trillion in food, medical product, and tobacco consumption, including about 78 percent of the U.S. food supply, over 20,000 prescription drugs, over 6,700 medical devices, and over 100,000 tobacco products).

1. *Molecular Models.* In 2018, FDA Commissioner Scott Gottlieb issued a statement regarding kratom, a botanical product derived from the tree *Mitragyna speciosa* that acts as a pain reliever and mild stimulant.[141] In Southeast Asia, many chew its leaves to remain alert, much like others drink caffeinated beverages to boost concentration.[142] Some use kratom to relieve pain or as a substitute for opium and other opioids.[143] In his statement, Gottlieb claimed kratom's most common active ingredients, including mitragynine and 7-hydroxymitragynine, collectively referred to as the mitragynines, are harmful opioids that pose serious risks to humans.[144]

The FDA had previously discouraged kratom use.[145] However, Gottlieb's announcement was surprising because he unveiled a computational methodology called Public Health Assessment via Structural Evaluation ("PHASE").[146] When the FDA identifies a potentially harmful substance, it now uses PHASE to assess the risk to public safety.[147]

---

141. Press Release, U.S. Food & Drug Administration, Statement from FDA Commissioner Scott Gottlieb, M.D., on the Agency's Scientific Evidence on the Presence of Opioid Compounds in Kratom, Underscoring Its Potential for Abuse (Feb. 6, 2018) [hereinafter Gottlieb 2018], https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-agencys-scientific-evidence-presence-opioid-compounds [https://perma.cc/UG9J-BLNH].

142. *See, e.g.*, Darshan Singh, Suresh Narayanan, Balasingam Vicknasingam, Ornella Corazza, Rita Santacroce & Andres Roman Urrestarazu, *Changing Trends in the Use of Kratom (*Mitragyna Speciosa*) in Southeast Asia*, 32 HUM. PSYCHOPHARMACOLOGY (SPECIAL ISSUE) e2582, e2583 (2017) (stating that in Southeast Asia, many individuals consume kratom in the morning to combat fatigue and increase productivity).

143. *Id.*

144. Gottlieb 2018, *supra* note 141.

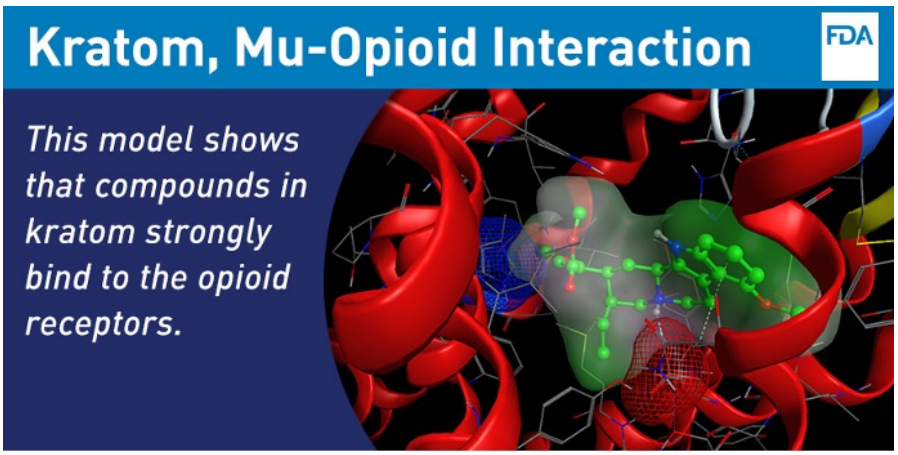
145. Scott Gottlieb, Comm'r, U.S. Food & Drug Admin., Remarks at the FDA Office of Criminal Investigations Meeting (Nov. 14, 2017), https://www.fda.gov/news-events/speeches-fda-officials/remarks-fda-office-criminal-investigations-meeting-11142017 [https://perma.cc/P9RV-N4EG] (describing the FDA's use of its PHASE methodology to predict the physical effects and public health impact of the twenty-five most common compounds in kratom); *see also* Ellis et al., *supra* note 29, at 2–3 (describing mitragynine and 7-hydroxymitagynine as compounds that bind to the μ-opioid receptor and comprise a significant portion of the alkaloids found in kratom).

146. *See* Gottlieb 2018, *supra* note 141; Christopher R. Ellis, Rebecca Racz, Naomi L. Kruhlak, Marlene T. Kim, Edward G. Hawkins, David G. Strauss & Lidiya Stavitskaya, *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs of Abuse to Controlled Substances Using Public Health Assessment via Structural Evaluation*, 106 CLINICAL PHARMACOLOGY & THERAPEUTICS 116, 116 (2019) [hereinafter Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*].

147. *See* Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*, *supra* note 146, at 117.

When announcing PHASE, Gottlieb concluded that kratom posed an imminent risk to public health.[148] He claimed the FDA's models predicted that the mitragynines bind strongly to the μ-opioid receptor, which also binds opioids like morphine and heroin.[149] However, Gottlieb revealed few details on how PHASE works, beyond describing its underlying models as "an advanced, common and reliable tool for understanding the behavior of drugs in the body."[150] FDA scientists later revealed that PHASE relies on "machine learning techniques, such as random forest, support vector machine, and artificial neural networks."[151]

Figure 1: FDA Kratom model illustration



Some experts criticized the FDA's use of PHASE to predict kratom's potential for harm.[152] According to Andrew Kruegel, a research chemist at Columbia University, the FDA's claim that kratom has risks comparable to those of morphine is like saying that all molecules that bind to the μ-opioid receptor have the same effects,

148. *See* Gottlieb 2018, *supra* note 141.

149. *Id.*

150. *Id.*

151. *See* Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*, *supra* note 146, at 118.

152. *See, e.g.*, Nick Wing, *FDA Releases Kratom Death Data, Undermines Its Own Claims About Drug's Deadly Harms*, HUFFPOST (Feb. 7, 2018, 11:20 AM), https://www.huffpost.com/entry/kratom-deaths-fda_n_5a7a3549e4b07af4e81eda8b [https://perma.cc/37EX-QEUW] (reporting that the FDA's analysis of kratom was far less rigorous than previous studies).

"which is not true based on what we've learned about these compounds."[153] Numerous factors determine a drug's effects and potential for harm, including how it is absorbed, how easily it crosses the blood-brain barrier, how its binding to cell surface receptors affects the interior of cells, and how quickly it is metabolized and excreted.[154] Some of these variables are represented by the phrase absorption, distribution, metabolism, and excretion ("ADME").

In particular, through a phenomenon called biased agonism, the mitragynines produce intracellular effects that differ from those triggered by classic opioids.[155] Specifically, they fail to activate an intracellular protein called -arrestin 2, which triggers a cascade of intracellular events that are linked to the potentially fatal effects of opioids, such as slow and shallow breathing.[156] Biased agonism and ADME are more difficult to model than chemical structure and binding affinity due to their complexity.[157] The PHASE methodology did not account for these factors, and Gottlieb omitted its limitations from his public statements.[158] Thus, concerns about accuracy and reliability are warranted.

Police sometimes use smart technologies, such as AI- and internet-enabled cameras and microphones, to predict who will commit crimes based on similarities between one's appearance or behavior and those

153. *Id.*

154. *See* Josh Bloom, *The FDA Concludes That Kratom Is an Opioid, And . . .*, AM. COUNCIL SCI. HEALTH (Feb. 6, 2018), https://www.acsh.org/news/2018/02/06/fda-concludes-kratom-opioid-and-12537 [https://perma.cc/Y9VG-YCE6] (explaining that absorption, metabolism, and ease of traversing the blood-brain barrier influence drug effects and are more difficult to predict than is binding affinity); *see also* Andrew C. Kruegel, Madalee M. Gassaway, Abhijeet Kapoor, Andras Varadi, Susruta Majumdar, Marta Filizola, Jonathan A. Javitch & Dalibor Sames, *Synthetic and Receptor Signaling Explorations of the Mitragyna Alkaloids: Mitragynine as an Atypical Molecular Framework for Opioid Receptor Modulators*, 138 J. AM. CHEM. SOC'Y 6754, 6757 (2016) (explaining how biased agonism allows substances to produce different intracellular effects when binding to the same cell surface receptor).

155. *See* Kruegel et al., *supra* note 154.

156. *See id.*; C. E. Groer, K. Tidgewell, R. A. Moyer, W. W. Harding, R. B. Rothman, T. E. Prisinzano & L. M. Bohn, *An Opioid Receptor Agonist that Does Not Induce μ-Opioid Receptor – Arrestin Interactions or Receptor Internalization*, 71 MOLECULAR PHARMACOLOGY 549, 555 (2007); *see also* Iris Bachmutsky, Xin Paul Wei, Eszter Kish & Kevin Yackle, *Opioids Depress Breathing Through Two Small Brainstem Sites*, 9 ELIFE e52694, e52694 (2020).

157. *See* Bloom, *supra* note 154; *see also* Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*, *supra* note 146, at 121 (stating that PHASE does not consider lipid solubility, blood-brain barrier transmission, and drug clearance).

158. *See* Gottlieb 2018, *supra* note 141.

of people known to have committed crimes in the past.[159] Similarly, the FDA uses molecular modeling to predict the effects of substances based on their chemical structure and predicted binding affinities, and their similarities to substances known to cause harm.[160] However, facial features and behavior are not reliable predictors of future acts,[161] and similarly, due to phenomena like biased agonism, chemical structure and binding affinity are unreliable predictors of drug effects.[162]

Typically, FDA officials and industry stakeholders justify the use of computer models and simulations on the ground that they reduce the burden on drug companies and agency staff when developing and evaluating substances and medical products.[163] In the case of kratom and other unregulated substances, the FDA justifies using PHASE on slightly different grounds. Instead of reducing the burden on manufacturers and agency staff during the approval process, FDA scientists claim PHASE reduces their burden when evaluating substances for potential regulation or prohibition by U.S. Drug Enforcement Administration ("DEA").[164]

Before the DEA can place a drug on the controlled substances list, a process called scheduling, the FDA performs a scientific analysis based on eight factors defined by the Controlled Substances Act ("CSA").[165] FDA scientists claim that using PHASE as a substitute for the eight-factor analysis of unknown substances, such as suspected

159. *See* Ángel Díaz, *Data-Driven Policing's Threat to Our Constitutional Rights*, BROOKINGS (Sept. 13, 2021), https://www.brookings.edu/techstream/data-driven-policings-threat-to-our-constitutional-rights [https://perma.cc/7WJY-4FZS]; *see also* Doug Wyllie, *How AI Software Could Monitor Real-Time Camera Feeds To Detect Criminal Behavior*, POLICE1 (Dec. 7, 2017), https://www.police1.com/police-products/intelligence-led-policing/articles/how-ai-software-could-monitor-real-time-camera-feeds-to-detect-criminal-behavior-NW2h8yWh6CvLxPCP [https://perma.cc/4HPE-Z7AS].

160. *See generally* Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*, *supra* note 146 (noting PHASE uses molecular structure to predict biological function and to compare similarities between controlled substances and newly emerging illicit drugs).

161. *See* Diaz, *supra* note 159 (stating predictive policing technologies are often biased and inaccurate).

162. *See* Bloom, *supra* note 154; Kruegel et al., *supra* note 154.

163. *See, e.g.*, Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*, *supra* note 146, at 117 (stating that the FDA developed molecular modeling to save time when evaluating newly identified substances for potential harm).

164. *Id.*

165. 21 U.S.C. § 811(c).

fentanyl analogs, conserves time and resources.[166] However, PHASE is a poor substitute for the eight-factor analysis, which addresses complex historical, epidemiological, and psychological factors in addition to chemical and physical properties.[167]

Two years after Gottlieb unveiled PHASE, the FDA augmented it with predictive software called Clarity, which is developed and licensed by Chemotargets, a private company.[168] Clarity's algorithms were developed using "an expertly curated training set of 2.6 million compounds," with data "derived from patents, journals and public databases."[169] This training set contains many sources of potential bias that could undermine the accuracy and credibility of PHASE and its predictions.[170] For instance, it includes at least 240,000 U.S. patents from a database called SureChEMBL, which is maintained by the European Bioinformatics Institute.[171]

Patent documents are not peer reviewed, and they can contain inaccurate and misleading information.[172] They are drafted to persuade the PTO that inventions meet the requirements for patentability. Inventors are incentivized to make broad claims to monopolize the

166. Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*, *supra* note 146, at 117 ("The human resource requirements to perform an eight-factor analysis, coupled with the vast number of possible fentanyl analogs, is prohibitively resource intensive," claim FDA scientists).

167. *See* 21 U.S.C. § 811(c) (requiring consideration of the risk of psychological dependence, the history and current pattern of abuse, and the scope, duration, and significance of abuse).

168. *The FDA Licenses Chemotargets CLARITY Platform*, UNIV. BARCELONA (May 5, 2020), https://www.pcb.ub.edu/en/the-fda-licenses-chemotargets-clarity-platform [https://perma.cc/DLC7-4RDF].

169. *Clarity Highlights*, CHEMOTARGETS, https://www.chemotargets.com/PRODUCTS/CLARITY-HIGHLIGHTS [https://perma.cc/UVH7-3YNB].

170. *See* Ellis et al., *supra* note 29, at e0229649.

171. Anna Gaulton, *Identifying Relevant Compounds in Patents*, CHEMBL-OG (May 13, 2021), https://chembl.blogspot.com/2021/05/identifying-relevant-compounds-in.html [https://perma.cc/8CZZ-HQFB]. *See generally* George Papadatos, Mark Davies, Nathan Dedman, Jon Chambers, Anna Gaulton, James Siddle, Richard Koks, Sean A. Irvine, Joe Pettersson, Nicko Goncharoff, Anne Hersey & John P. Overington, *SureChEMBL: A Large-Scale, Chemically Annotated Patent Document Database*, 44 NUCLEIC ACID RES. D1220 (2015) (describing SureChEMBL, a publicly available database containing information on seventeen million compounds extracted from patent documents).

172. *See* Lisa Larrimore Ouellette, *Who Reads Patents?*, 35 NATURE BIOTECHNOLOGY 421, 423 (2017) (reporting the results of a survey in which scientist respondents stated that patents do not contain useful information).

widest swaths of technology possible.[173] Furthermore, patent documents often describe inventions that do not exist and experiments that have not been conducted.[174] These imaginary examples are potentially harmful because they can mislead and spread misinformation.[175] If incorporated into predictive models, they can confuse and mislead regulators.

The predictions of Clarity and PHASE may also be biased by the inclusion of journal articles and public databases in the training set. Until at least early 2020, this information was drawn from a database called ChEMBL version 24.[176] Because the FDA's predictive models rely on Clarity's predictions, one must know more about the contents of ChEMBL to evaluate the models' credibility. Journals and public databases were reportedly included to "ensure comprehensive coverage of the chemical space pharmacology," which suggests that journals from other fields may have been omitted.[177] ChEMBL version 26 was released on March 3, 2020.[178] It contains 76,076 documents drawn from approximately twenty chemistry journals.[179] A query of the database using the search term "kratom" yielded only two articles from the Journal of Medicine Chemistry, and a search using the term "mitragynine" returned those documents and two additional articles from the same journal.[180]

---

173. *See* Alan C. Marco, Joshua D. Sarnoff & Charles A.W. DeGrazia, *Patent Claims and Patent Scope*, 48 RES. POL'Y 1, 2 (2019) (stating that overly broad claims can be exploited for rent seeking).

174. Janet Freilich, *Prophetic Patents*, 53 U.C. DAVIS L. REV. 663, 668 (2019) (reporting that in over two million chemistry and biology patents issued between 1976 and 2017, 17 percent of examples were prophetic, and of the patents containing examples, 24 percent had at least some prophetic examples); *see also* Jorge L. Contreras, *Patent Fakes: How Fraudulent Inventions Threaten Public Health, Innovation, and the Economy*, HARV. L. SCH. BILL HEALTH (July 1, 2020), https://blog.petrieflom.law.harvard.edu/2020/07/01/patent-fakes-fraud-inventions-covid [https://perma.cc/URT9-39AF] (explaining how a company that never developed its technology could nevertheless have obtained hundreds of patents protecting that nonexistent technology).

175. Freilich, *supra* note 174, at 712.

176. Ellis et al., *supra* note 29, at e0229649.

177. *CLARITY Highlights*, *supra* note 169.

178. *ChEMBL 26 Released*, CHEMBL-OG (Mar. 3, 2020), http://chembl.blogspot.com/2020/03/chembl-26-released.html [https://perma.cc/P6Y8-HYRT].

179. *Id.*

180. Samuel Obeng, Shyam H. Kamble, Morgan E. Reeves, Luis F. Restrepo, Avi Patel, Mira Behnke, Nelson J.-Y. Chear, Surash Ramanathan, Abhisheak Sharma, Francisco León, Takato Hiranita, Bonnie A. Avery, Lance R. McMahon & Christopher R. McCurdy, *Investigation of the Adrenergic and Opioid Binding Affinities, Metabolic Stability, Plasma Protein Binding Properties, and Functional Effects of Selected Indole-Based Kratom Alkaloids*, 63 J. MED. CHEMISTRY 433

Training predictive models using data from only certain fields while excluding others could bias their predictions. Chemistry articles may emphasize chemical structure and reactivity, pharmacology articles may focus on toxicity and mechanisms of action, and drug policy articles may emphasize the social impact of the availability or prohibition of substances. Including or excluding each source might bias predictions in one direction or another. The geographic origin of articles could also affect predictions. Authors who have spent time in Southeast Asia, where kratom is commonly grown and consumed, may have greater familiarity with its effects and more favorable views regarding its risks and benefits.[181] Excluding their viewpoints could negatively bias model predictions. Similarly, including or excluding various types of "grey literature," such as newsletters, internal reports, blog posts, and corporate promotional materials, could further bias results.[182]

Six months after Gottlieb unveiled PHASE, HHS Assistant Secretary for Health Dr. Brett Giroir sent a letter to the DEA, instructing its administrator not to schedule kratom.[183] His order rescinded an earlier HHS opinion that had recommended prohibition.[184] "This decision is based on many factors," said Giroir, including "the relative lack of evidence, combined with an unknown and potentially substantial risk to public health if these chemicals were scheduled at this time."[185] Giroir's statement regarding a "relative lack of evidence" for harm suggests that HHS was not impressed by the FDA's PHASE predictions.[186]

(2020); András Váradi et al., *Mitragynine/Corynantheidine Pseudoindoxyls as Opioid Analgesics with Mu Agonism and Delta Antagonism, Which Do Not Recruit -Arrestin-2*, 59 J. MED. CHEMISTRY 8381 (2016).

181. *See*, *e.g.*, Singh et al., *supra* note 142, at e2586 (stating kratom does not appear to adversely affect the health or social functioning of people who consume it in Southeast Asia, and cases of kratom toxicity are rare).

182. *See, e.g.*, Jean Adams, Frances C. Hillier-Brown, Helen J. Moore, Amelia A. Lake, Vera Araujo-Soares, Martin White & Carolyn Summerbell, *Searching and Synthesizing 'Grey Literature' and 'Grey Information' in Public Health: Critical Reflections on Three Case Studies*, 5 SYSTEMATIC REVS. 164, 172 (2016).

183. Letter from Brett P. Giroir to Uttam Dhillon 1 (Aug. 16, 2018) (on file with author) [hereinafter Giroir 2018].

184. *Id.*

185. *Id.*

186. *See id.*

In particular, Giroir explained that prohibiting kratom could have "immediate adverse public health consequences for potentially millions of users," including "intractable pain" and "switching to highly lethal opioids, including potent and deadly prescription opioids," which could cause "thousands of deaths from overdoses and infectious diseases associated with IV drug use."[187] He expressed concern that scheduling kratom would further stigmatize the substance, inhibiting people from discussing it with doctors and stifling important research on its "potentially useful chemistry."[188] Moreover, he cited a peer-reviewed animal study, which found that "mitragynine does not have abuse potential and actually reduced morphine intake."[189]

It is unclear why the FDA continues to use a methodology discredited by its parent agency.[190] In his letter to the DEA, Giroir listed several details that the FDA had failed to provide such as "[a] scientific assessment of the *actual* scale and degree of dependence and/or addiction of Americans utilizing *kratom*," and "[a] scientific determination based on data whether *kratom actually* serves as a gateway drug that promotes further use of more dangerous opioids."[191] Giroir may have been responding to points that Gottlieb raised in the 2018 public statement in which he introduced PHASE, said molecular models predicted that kratom affects the body "just like opioids," and claimed the FDA's concerns were "rooted in sound science."[192] Further, Giroir's use of "actual" and "actually" may have been intended to distinguish real evidence of harm from predictions based on models and simulations.[193]

Could the FDA's persistence be motivated by something other than public health? Long before adopting PHASE, Gottlieb took a negative stance on kratom.[194] In the years that followed the HHS ruling, FDA scientists published two articles on PHASE and its

187. *Id.* at 3–4.

188. *Id.*

189. *Id.* at 3.

190. *Id.* ("The level of scientific data and analysis presented by the FDA and available in the literature do not meet the criteria for inclusion of *kratom* or its chemical components in Schedule I.").

191. *Id.* (emphasis added).

192. Gottlieb 2018, *supra* note 141.

193. *See* Giroir 2018, *supra* note 183, at 3.

194. *See* Gottlieb, *supra* note 145 ("[Kratom] carries risks of abuse, addiction, and death.").

utility.[195] A 2019 article explains its benefits as a substitute for the CSA's eight-factor analysis while screening fentanyl analogs.[196] The article acknowledged that PHASE ignores several drug characteristics including absorption, transmission across the blood-brain barrier, and elimination.[197] However, the authors appear to shrug off these limitations, claiming that whenever predictive models for these variables are developed, they will incorporate them into PHASE.[198] In 2020, the same authors published a second article that praises PHASE as a means of analyzing kratom's harm potential.[199] The article ignores the public health risks of prohibition and the lack of actual scientific data, which were both described by Giroir two years earlier.[200]

In 2021, Gottlieb, who had by then resigned from the FDA to serve on the board of drugmaker Pfizer, took to Twitter to express his dissatisfaction with HHS: "We were prevented by HHS from moving forward with the scheduling of Kratom, and I'm convinced it's fueling the opioid addiction crisis."[201] Thirty minutes later, Giroir responded, claiming that he had rejected Gottlieb's recommendation due to "embarrassingly poor evidence & data, and a failure to consider overall public health."[202] Giroir added, "If #Kratom is fueling opioid addiction, prove it . . . ."[203]

Having been overruled, and failing to prohibit kratom domestically, the FDA threatened to go over the heads of HHS

195. *See generally* Ellis et al., *supra* note 29 (stating PHASE can identify potential safety signals, providing a tool for prioritizing evaluation of high-risk compounds); Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*, *supra* note 146 (claiming PHASE can evaluate the similarity of a newly identified drug of abuse to known controlled substances and inform the public health response).

196. *See* Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*, *supra* note 146, at 117.

197. *Id.* at 121.

198. *See id.*

199. *See* Ellis et al., *supra* note 29, at e0229648 ("The PHASE binding profiles of the kratom alkaloids indicated several biological targets associated with potential adverse effects.").

200. *See* Giroir 2018, *supra* note 183, at 1, 3.

201. Scott Gottlieb (@ScottGottliebMD), TWITTER (May 21, 2021, 6:17 PM), https://twitter.com/ScottGottliebMD/status/1395866461776891908 [https://perma.cc/5N3E-26K2]; *see also* Sheila Kaplan, *Elizabeth Warren Calls on Former F.D.A. Chief To Quit Pfizer Board*, N.Y. TIMES (July 2, 2019), https://www.nytimes.com/2019/07/02/health/elizabeth-warren-scott-gottlieb-pfizer.html [https://perma.cc/3FCD-8M2Y] (reporting Gottlieb's move from the FDA to Pfizer).

202. Brett Giroir (@DrGiroir), TWITTER (May 21, 2021, 6:49 PM), https://twitter.com/drgiroir/status/1395874443726102533 [https://perma.cc/FGV3-6FJK].

203. *Id.*

officials by requesting a global ban from the World Health Organization.[204] If it had been successful, the FDA could have effectively banned kratom domestically and internationally based on discredited evidence from computer models. However, the World Health Organization concluded there was "insufficient evidence to support a critical review of kratom" and declined to pursue a ban.[205]

2. *Virtual Humans and Patient-Specific Models.* Despite the limitations of molecular modeling, which simulates the interactions between small molecules and their biological targets, the FDA is pursuing more complex multiscale modeling to simulate human organs and entire human bodies. The following discussion explains how the agency uses multiscale models called virtual humans to evaluate new medical products.

One of the earliest efforts to simulate human physiology was the Virtual Physiological Human Project, which was founded in 2007.[206] Supported by the European Commission, the project developed along two paths. One approach, developed through a project called Discipulus, supported the creation of digital patients to predict health outcomes and create personalized therapies for individuals.[207] A second path, developed through a project called the Avicenna Support Action, guided the development of cohorts of virtual humans on which drugs and medical devices could be tested.[208]

---

204. Mason Marks, *FDA's Kratom Ban Would Harm the Public and Damage the Agency's Credibility*, STAT FIRST OP. (Aug. 23, 2021), https://www.statnews.com/2021/08/23/fdas-kratom-ban-would-harm-the-public-and-damage-the-agencys-credibility [https://perma.cc/5QYY-25DZ].

205. COMM'N ON NARCOTIC DRUGS, SUMMARY OF ASSESSMENTS, FINDINGS AND RECOMMENDATIONS OF THE 44TH WORLD HEALTH ORGANIZATION'S (WHO) EXPERT COMMITTEE ON DRUG DEPENDENCE (ECDD), 11-15 OCTOBER 2021, at 9 (2021), https://www.unodc.org/documents/commissions/CND/CND_Sessions/CND_64Reconvened/ECN72021_CRP12_V2108992.pdf [https://perma.cc/NHJ2-BSBY].

206. Peter J. Hunter & Marco Viceconti, *VPH-Physiome Project: Standards and Tools for Multiscale Modeling in Clinical Applications*, 2 IEEE REV. BIOMEDICAL ENG'G, 2009, at 40, 40.

207. *See* DISCIPULUS, VPH INST., ROADMAP FOR THE DIGITAL PATIENT 9 (2013), https://www.vph-institute.org/upload/discipulus-digital-patient-research-roadmap_5270f44c03856.pdf [https://perma.cc/PVT7-DKVQ].

208. *See generally* Marco Viceconti, Adriano Henney & Edwin Morley-Fletcher, *In Silico Clinical Trials: How Computer Simulation Will Transform the Biomedical Industry*, 3 INT'L J. CLINICAL TRIALS 37 (2016) (report review) [hereinafter Viceconti et al., *In Silico Clinical Trials*] (describing the Avicenna Support Action, which ran from 2013 to 2015); MARCO VICECONTI, MARCO VICECONTI, JAMES KENNEDY, ADRIANO HENNEY, MARKUS REITERER, SEBASTIAN POLAK, DIRK COLAERT, JEAN-PIERRE BOISSEL, MARTINA CONTIN, CLAUDIA MAZZÀ, ANNAMARIA CARUSI, ENRICO DALL'ARA, MATTHEW BURNETT, IWONA ZWIERZAK, KAREN

Significant progress has been made in creating digital patients to personalize the care of individuals. Simulating patient physiology helps surgeons plan procedures and reduce related risk. Prior to complex heart surgery, such as aortic valve replacement, a patient-specific model can be built to simulate deployment of an implantable valve and help predict the outcome.[209] One product, called HeartFlow, generates patient-specific models of coronary blood vessels from radiographic images of patients' hearts.[210] It predicts the values that a measurement called fractional flow reserve would produce without subjecting patients to the risks of cardiac catheterization, an invasive procedure that is required to obtain the measurement.[211]

If used incorrectly, or if their predictions are misinterpreted, patient-specific models can expose people to unnecessary interventions or cause them to forego lifesaving procedures. HeartFlow belongs to a class of models that predict the results of risky procedures that would otherwise be performed if they were less costly, complex, or dangerous.[212] In this context, the lack of safe alternatives likely justifies their use.

Because patient-specific models are used to diagnose and treat individuals, the FDA regulates them as medical devices.[213] However, they can be adapted to evaluate the safety and efficacy of other medical devices in populations instead of individuals.[214] For instance, the SIMULIA Living Heart model has been used to evaluate the effectiveness of an implantable device designed to treat mitral valve

---

EL-ARIFI, MASSIMO CELLA, GIUSEPPE ASSOGNA, ROBERT HESTER & FILIPE HELDER MOTA, IN SILICO CLINICAL TRIALS: HOW COMPUTER SIMULATION WILL TRANSFORM THE BIOMEDICAL INDUSTRY (2016), https://avicenna-alliance.com/files/user_upload/PDF/Avicenna_Roadmap.pdf [https://perma.cc/PCD2-2TLY] (reporting the Avicenna Action's findings and recommendations); Marco Viceconti & Peter Hunter, *The Virtual Physiologic Human: Ten Years After*, 18 ANN. REV. BIOMED. ENG'G 103, 107 (2016).

209. Viceconti & Hunter, *supra* note 208, at 111.

210. *Id.*

211. *Id.*

212. *Id.*

213. *See* Federal Food, Drug, and Cosmetic Act § 201(h), 21 U.S.C. § 321(h) (defining medical device); U.S. FOOD & DRUG ADMIN., HOW TO DETERMINE IF YOUR PRODUCT IS A MEDICAL DEVICE (2019), https://www.fda.gov/medical-devices/classify-your-medical-device/how-determine-if-your-product-medical-device [https://perma.cc/Q8DP-7JVC].

214. *E.g.*, Karl D'Souza, *Technology To Transform Lives: The SIMULIA Living Heart Model*, BENCHMARK MAG. (July 2015), https://www.3ds.com/fileadmin/Industries/life-sciences/pdf/NAFEMS-Benchmark-Technology-to-Save-Lives-LHP-07-01-15.pdf [https://perma.cc/M7AT-Y257] (describing the SIMULIA Living Heart).

regurgitation, where blood leaks back across the mitral valve instead of into the aorta.[215]

Since 2007, the FDA has made four anatomically correct whole-body models available.[216] Called the Virtual Family, these models simulate the effects of heat, ultrasound, electromagnetic fields, and other variables.[217] The latest versions represent about three hundred organs and tissues.[218] They can be downloaded for free, and as of 2017, their predictions had accompanied over 160 FDA medical device submissions.[219] A larger collection, called the Computable Virtual Population, contains seventeen virtual humans ranging in age from eight weeks to eighty-four years, collectively representing over 120 anatomical features and more than 300 tissues.[220]

The future of FDA regulation involves assembling large cohorts of virtual humans to conduct simulated clinical trials that augment or replace randomized controlled trials.[221] However, when used to approve novel drugs and medical devices, biased algorithmic predictions could promote FDA approval of medical products that lack adequate testing or prevent the FDA from approving safe and effective products that are incorrectly predicted to be harmful or ineffective. The following discussion describes ongoing efforts to augment or replace clinical trials with simulations.

3. *Simulated Clinical Trials.* The FDA envisions a future where simulated trials replace a significant portion of its actionable health-

215. *See id.*

216. U.S. FOOD & DRUG ADMIN., VIRTUAL FAMILY (2017), https://www.fda.gov/about-fda/cdrh-offices/virtual-family [https://perma.cc/425F-LPE9] (describing the FDA's virtual family that was created using magnetic resonance image data from volunteers and contains models representing a thirty-four-year-old adult male, a twenty-six-year-old female, an eleven-year-old female, and a six-year-old male).

217. *Id.*

218. *Id.*

219. U.S. FOOD & DRUG ADMIN., U.S. FOOD AND DRUG ADMINISTRATION CENTER FOR DEVICES AND RADIOLOGICAL HEALTH: PROGRESS IN ACHIEVING OUR VISION OF PATIENTS FIRST 4 (2016), https://www.fda.gov/media/104262/download [https://perma.cc/N5TM-5P8D].

220. *Computable Virtual Population: Resolution at Its Limit*, IT'IS FOUND., https://itis.swiss/virtual-population/virtual-population/vip3 [https://perma.cc/LRD2-UVZP].

221. *See* Francesco Pappalardo, Giulia Russo, Flora Musuamba Tshinanu & Marco Viceconti, *In Silico Clinical Trials: Concepts and Early Adoptions*, 20 BRIEFINGS BIOINFORMATICS 1699, 1704 (2019) (describing the incorporation of multiple virtual patients into cohorts to conduct simulated clinical trials).

care research.[222] Agency officials and industry stakeholders extol the benefits of simulated clinical trials over traditional *in vivo* experiments, which are conducted with humans and animals.[223] They argue that simulations make research less expensive, lower the regulatory burden on product manufacturers, and shorten the time required to obtain FDA approval, allowing lifesaving therapies to reach the market sooner.[224]

Incorporating AI into simulated trials could be advantageous compared to the use of purely biomathematical models.[225] Whereas linear statistical models are useful only for analyzing structured data derived from clinical studies, AI models can potentially mine and incorporate data from a wide variety of sources, including unstructured data from electronic medical records, radiographic images, and genomic analysis.[226] Proponents also claim that simulated trials pose fewer risks to human research subjects and will make trials more inclusive because people with diverse traits can be modeled and included in virtual cohorts.[227]

Additionally, proponents contend that simulated trials can potentially substitute for the control arms of randomized controlled trials, eliminating the need for people to receive placebos.[228] Simulating control arms could allow every human research participant to join the intervention arm of a trial and receive the experimental therapy.[229]

---

222. *See* Tina Morrison, Chair, FDA Modeling & Simulation Working Grp., FDA Grand Rounds: Advancing Regulatory Science with Modeling and Simulation at FDA (Aug. 9, 2018), https://collaboration.fda.gov/p4r7q3qweuv [https://perma.cc/XTQ2-F75S] (presenting "Future Opportunities for *In Silico* Clinical Trials").

223. *See id.* (presenting "[s]uccess stories with modeling and simulation at FDA"); *see also* Viceconti et al., *In Silico Clinical Trials*, *supra* note 208, at 41–42 (discussing how and why simulated clinical trials could replace human and animal clinical trials).

224. *See* Morrison, *supra* note 222, at 25:00 ("[In silico clinical trials . . . enable] safe and effective new therapeutics to advance more efficiently through the different stages of clinical trials." (quoting Gottlieb, *supra* note 16)); Kolla et al., *supra* note 17, at 188572 ("[In silico trials] can salvage the resources devoted to failed pharmacological studies by enabling better powered trials, simulating control and efficacy arms, and optimizing patient recruitment and drug protocols in efficacy arms.").

225. Kolla et al., *supra* note 17, at 188573; *see also* Pappalardo et al., *supra* note 221, at 1701 ("[A]rtificial intelligence technology . . . improve[s] our ability to investigate drug mechanism of action and drug effects that would be missed entirely by conventional statistical tests.").

226. Kolla et al., *supra* note 17, at 188573–74.

227. *Id.* at 188573.

228. *See id.*

229. *Id.*

Advocates also contend that simulated control arms could eliminate concerns regarding unblinding, which occurs when research subjects discover that they were assigned to a control arm and received placebos.[230] Some proponents even claim that simulated trials could replace the treatment arms of clinical trials instead of the control arm, or simulate both, potentially eliminating the need for human research subjects altogether.[231]

Unlike traditional trials, which are difficult to repeat due to the associated expense and risks to participants, simulated trials can be run repeatedly. Through an iterative process, each consecutive cycle can integrate new patients and data and learn from errors in previous cycles to improve the model's predictions.[232] Algorithms can potentially add unlimited numbers of virtual patients to simulated trials and split the populations into groups for parallel processing, which further increases speed and reduces the computational resources required.[233] If they are safer, faster, and less expensive than human trials, simulated trials could reduce the regulatory burden on medical product manufacturers.[234]

Simulated trials could also help achieve racial equity. Historically, racial minorities have been abused and exploited in the name of scientific research.[235] Minorities have also been underrepresented in clinical trials, and significant disparities persist in contemporary

230. *Id.*

231. *See* Pappalardo et al., *supra* note 221 (stating that when used in conjunction with postmarket surveillance to confirm simulation predictions, in silico trials could be used as "conclusive evidence" of safety and efficacy, replacing Phase III clinical trials); *see also* Kolla et al., *supra* note 17, at 188573 ("Given the early success of synthetic control arms, companies have started to extend this approach for the simulation of intervention arm drug effects —*in silico* efficacy arms.").

232. Kolla et al., *supra* note 17, at 188574.

233. *Id.*

234. *See* Tina M. Morrison, Pras Pathmanathan, Mariam Adwan & Edward Margerrison, *Advancing Regulatory Science with Computational Modeling for Medical Devices at the FDA's Office of Science and Engineering Laboratories*, 5 FRONTIERS MED., Sept. 2018, at 1, 7 [hereinafter Morrison et al., *Advancing Regulatory Science*] (describing the FDA's desire to use modeling to reduce burdensome data collection associated with animal and human studies).

235. *E.g.*, Allan M. Brandt, *Racism and Research: The Case of the Tuskegee Syphilis Study*, 8 HASTINGS CTR. REP. 21, 21–24 (1978) (describing the infamous case in which the U.S. Public Health Service unethically studied untreated syphilis in four hundred Black individuals in the United States under the guise of providing treatment and withheld treatment when it subsequently became available).

healthcare and biomedical research.[236] In addition to being unjust, a lack of diversity can negatively impact trial results, decrease generalizability, and place patients at risk.[237] During the COVID-19 pandemic, researchers discovered that spirometers, devices that measure lung function, systematically treat Black and Asian patients differently than their white counterparts, potentially producing dangerous misdiagnoses.[238] In addition to making research more equitable, advocates claim simulated trials could supplement human participants with comparable virtual humans, yielding more data on rare diseases because individuals with those conditions can be challenging to recruit.[239]

Though augmenting clinical trials with simulations *may* promote equity and increase the generalizability of research results, it remains unclear whether simulations *will* achieve these goals. Substituting algorithmic predictions for clinical data could introduce new sources of bias and create unknown risks. Moreover, in many cases, instead of simulating people from underrepresented communities, it may be preferable to address the social obstacles to their equal participation in clinical research. Participating in clinical trials can have benefits other than establishing the safety and efficacy of medical products. Some patients rely on clinical trials to access experimental treatments.[240]

---

236. Luther T. Clark, Laurence Watkins, Ileana L. Piña, Mary Elmer, Ola Akinboboye, Millicent Gorham, Brenda Jamerson, Cassandra McCullough, Christine Pierre, Adam B. Polis, Gary Puckrein & Jeanne M. Regnante, *Increasing Diversity in Clinical Trials: Overcoming Critical Barriers*, 44 CURRENT PROBS. CARDIOLOGY, May 2019, at 148, 149. *See generally* Merlin Chowkwanyun & Adolph L. Reed, Jr., *Racial Health Disparities and COVID-19 — Caution and Context*, 383 NEW. ENG. J. MED. 201 (2020) (discussing COVID-19's disparate impact on racial minority populations and the importance of studying and contextualizing the disparity).

237. *See, e.g.*, Michael W. Sjoding, Robert P. Dickson, Theodore J. Iwashyna, Steven E. Gay & Thomas S. Valley, *Racial Bias in Pulse Oximetry Measurement*, 383 NEW ENG. J. MED. 2477, 2477–78 (2020) (reporting racial bias in pulse oximetry measurements, which may have resulted from testing oximeters in populations lacking racial diversity).

238. *See* Meredith A. Anderson, Atul Malhotra & Amy L. Non, *Could Routine Race-Adjustment of Spirometers Exacerbate Racial Disparities in COVID-19 Recovery?*, 9 LANCET RESPIRATORY MED. 124, 124 (2021) (reporting that spirometers systematically apply a race-based correction or "ethnic adjustment" that assumes Black and Asian patients have lower lung capacities than their Caucasian counterparts).

239. A. Carlier, A. Vasilevich, M. Maréchal, J. de Boer & L. Geris, In Silico *Clinical Trials for Pediatric Orphan Diseases*, 8 SCI. REPS., Feb. 6, 2018, at 1, 1–2, 6, https://www.nature.com/articles/s41598-018-20737-y [https://perma.cc/KF4K-RLC8].

240. *See* Arthur L. Caplan & Alison Bateman-House, *Should Patients In Need Be Given Access to Experimental Drugs?*, 16 EXPERT OP. PHARMACOTHERAPY 1275, 1276 (2015) ("Ideally, patients who wish to try an experimental drug should do so in a clinical trial.").

Unless alternate means are expanded, some communities may lose access to experimental therapies if human medical research is reduced or replaced with simulated trials.

For over a decade, simulations have been used to assess the safety and efficacy of medical devices.[241] In 2011, the FDA deemed a Medtronic pacemaker safe for use with magnetic resonance imaging ("MRI") machines based solely on simulation data.[242] Similarly, the Medical Device Innovation Consortium ("MDIC"), a public-private partnership between FDA and industry stakeholders, built a model to simulate a cardiac pacemaker lead, the wire that delivers electrical impulses from pacemakers to heart tissue.[243]

The MDIC told the FDA it could simulate a thousand cohorts of 1,000 patients to determine the device's safety, replacing a 500-person clinical study that would otherwise be necessary.[244] Unlike the Medtronic pacemaker submission, the MDIC simulation was a mock FDA submission.[245] In 2018, the FDA completed the VICTRE simulated trial.[246] A computer model analyzed the breast images of 2,986 virtual patients to reach a conclusion the FDA claims was comparable to that of a 600-person human clinical trial.[247]

Simulated trials have also been used to predict the effects of pharmaceuticals by replacing control arms with cohorts of virtual humans.[248] These simulated research subjects are exposed to virtual placebos, and the results are compared to the effects of interventions administered to real humans in corresponding experimental arms.[249] This process has been used to accelerate the approval of several drugs.

In 2015, a pharmaceutical company utilized a simulated control arm with 68 virtual patients to evaluate the safety and efficacy of

241. *Id.*

242. Biovia, *supra* note 23.

243. *Id.*

244. *Id.*

245. *Id.*

246. Sharma et al., *supra* note 19.

247. Aldo Badano, Christian G. Graff, Andreu Badal, Diksha Sharma, Rongping Zeng, Frank W. Samuelson, Stephen J. Glick & Kyle J. Myers, *Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial*, 7 JAMA NETWORK OPEN, Nov. 30, 2018, at e185474, e185479.

248. Kolla et al., *supra* note 17, at 188573.

249. *Id.*

alectinib, a lung cancer therapy.[250] The simulation reduced the time to FDA approval and increased the duration of European marketing exclusivity by eighteen months.[251] Another team used a similar process with 694 virtual patients to hasten approval of blinatumomab, which treats a rare form of leukemia.[252] Due to their success with simulated control arms, companies have expanded this approach to simulating experimental intervention arms.[253]

In 2007, a medical device manufacturer simulated joints with rheumatoid arthritis to compare the efficacy of two interventions for preventing bone erosion in severe cases.[254] The model predicted that one treatment, rituximab, would be superior to the other intervention, anti-tumor-necrosis-factor therapies.[255] Years later, the prediction was confirmed through traditional randomized controlled trials.[256] More recently, cancer researchers reported success using simulated trials to predict which patients with blood cancers were unlikely to respond to standard treatments.[257]

Though the technologies described above are impressive, the FDA acknowledges their limitations. Surprisingly, there has been little or no discussion of the potential for overreliance on models and simulations to harm consumers and negatively impact public health. Though proponents argue that they will democratize medical research, there is little reason to believe that simulated trials are any less susceptible to bias than algorithmic models used in other contexts. Because the FDA's reliance on them substitutes algorithmic predictions for direct observations of safety and efficacy, simulated trials share many risks with the agency's reliance on surrogate endpoints. In fact, algorithmic predictions should be framed as a type of surrogate endpoint.[258] However, instead of being surrogates for direct observation of clinical effects, models and simulations are

250. *Id.*
251. *Id.*
252. *Id.*
253. *Id.*
254. *Id.*
255. *Id.*
256. *Id.*
257. *Id.*
258. *See* Weintraub et al., *supra* note 8, at 2212 (stating that all endpoints other than improvements in health status, survival, or cost are surrogate endpoints).

surrogates for evidence of causation. Consequently, they may be even less reliable than traditional surrogates.

The best surrogate endpoints lie within the causal chain leading to a true endpoint, a clinical outcome of interest such as increased longevity or quality of life.[259] Surrogates can be biomarkers produced by pathologic processes leading to a medical condition of interest.[260] In contrast, the least reliable surrogates are merely associated with a true endpoint and do not lie within its causal path.[261] Even if simulated trials can infer the likelihood of certain outcomes, these virtual endpoints remain surrogates for their real-life counterparts because they are algorithmic predictions and may be of limited epistemological value.

Though imperfect, randomized controlled trials attempt to establish causality by eliminating, to the extent possible, the influence of potential causes other than the therapeutic intervention being investigated.[262] Simulated trials cannot ensure that all possible causes have been considered or eliminated.[263] For that reason, some scholars question their epistemic value.[264] Barbara Osimani and colleagues argue that simulations can supplement, but not replace, randomized controlled trials because simulations "cannot be equated to experiments" and "cannot establish new causal laws per se."[265] Nonetheless, some researchers describe simulations as acceptable surrogates for human trials and other data sources.

Even real-world surrogate endpoints can be problematic substitutes for clinical outcomes. Though trials using surrogates are often faster and less expensive than trials yielding clinical outcomes, the results of the former are less reliable because the relationship between surrogates and clinical endpoints is often uncertain.[266]

---

259. *See id.* at 2212, 2214 (explaining that a surrogate is most useful when it "consistently predicts events in the future," and thus a causal path is stronger than a mere association).

260. *See id.* at 2214.

261. *See id.* (explaining that mere association confounds the relationship between a surrogate and an outcome event).

262. *See* Barbara Osimani, Marta Bertolaso, Roland Poellinger & Emanuele Frontoni, *Real and Virtual Clinical Trials: A Formal Analysis*, 38 TOPOI 411, 412 (2019).

263. *Id.*

264. *Id.* at 416.

265. *Id.* at 420.

266. Weintraub et al., *supra* note 8, at 2216.

In particular, surrogates are rarely so reliable that they should substitute for clinical endpoints in the context of FDA approval.[267] A 2015 metanalysis of oncology trials found that in 52 percent of studies, the correlation between surrogate endpoints and overall patient survival was low.[268] Twenty-five percent of studies showed a moderate correlation, and only 23 percent showed a high correlation.[269] According to some experts, "it seems strange that a clinical trial using surrogate endpoints and comparing regimens with different timings could be deemed credible by governmental and private payers, and least of all by thoughtful physician scientists."[270]

Another barrier to trusting simulated trials is the fact that there are no industry standards or best practices. Though industry stakeholders often act as though simulations are comparable to human trials, simulated trials constitute a patchwork of computer models that were often designed for other contexts. The cobbling together of different computer models contributes to a lack of uniform practices. The fact that most computer models are developed by industry stakeholders to gather evidence on the safety and efficacy of their own products creates conflicts of interest that must be addressed. Specifically, medical product manufacturers typically produce the data on which FDA officials base regulatory approval.[271] To the extent that computer models are proprietary black boxes, and manufacturers conceal their code and training data, their opacity may conceal bias that favors industry stakeholders.

Considering the aforementioned technologies and associated risks, the following Part explains how administrative law doctrine might apply to FDA models and simulations. It concludes that existing doctrine is likely too inflexible and unpredictable to address the transparency and accountability concerns raised by automated FDA regulation.

---

267. *Id.*

268. Vinay Prasad, Chul Kim, Mauricio Burotto & Andrae Vandross, *The Strength of Association Between Surrogate End Points and Survival in Oncology: A Systematic Review of Trial-Level Meta-Analysis*, 175 JAMA INTERNAL MED. 1389, 1392 (2015).

269. *Id.*

270. Knopf et al., *supra* note 8, at i6287.

271. *See* Morrison et al., *Advancing Regulatory Science*, supra note 234 (stating that FDA reviewers in the Center for Devices and Radiological Health do not run simulations and instead rely on reports submitted by industry stakeholders).

III. AUTOMATED FDA REGULATION AND ADMINISTRATIVE LAW

A significant portion of the legal scholarship on automation and the administrative state has focused on how agencies should regulate AI and related technologies.[272] An increasing number of authors also analyze the role of AI in regulatory decision-making.[273] However, few have addressed the role of AI in FDA research or regulation.[274] The following discussion builds on previous scholarship to analyze the implications of FDA models and simulations for administrative law and vice versa.

The central doctrine of administrative law coalesced long before federal agencies adopted computer models and simulations. Much of it rests upon principles of accountability expressed in statutes such as the Administrative Procedure Act ("APA"), enacted in 1946, and the Freedom of Information Act ("FOIA"), enacted in 1966.[275] These statutes often require agencies to operate with a certain degree of transparency.[276] In many cases, when agency action affects people's rights, the government must explain its reasoning.[277] However, the

272. *E.g.*, Ira S. Rubinstein, Ronald D. Lee & Paul M. Schwartz, *Data Mining and Internet Profiling: Emerging Regulatory and Technological Approaches*, 75 U. CHI. L. REV. 261, 261–62 (2008) (describing the use of data mining by government agencies and private companies to predict behavior and how this practice should be regulated). *See generally* Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 556 (2015) (proposing a unified agency, such as a Federal Robotics Commission, to regulate and provide technical expertise on robotics and related technologies); Terrel McSweeny, *Psychographics, Predictive Analytics, Artificial Intelligence, & Bots: Is the FTC Keeping Pace?*, 2 GA. L. TECH. REV. 514 (2018) (discussing Federal Trade Commission regulation of AI and related technologies); W. Nicholson Price II, *Regulating Black-Box Medicine*, 116 MICH. L. REV. 421 (2017) (describing how the FDA should regulate medical AI); Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83 (2017) (proposing an FDA-like agency to regulate all applications of AI).

273. *See generally* David Freeman Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State*, 37 YALE J. ON REGUL. 800 (2020) (explaining how various federal agencies use AI); ENGSTROM ET AL., *supra* note 30 (same); Frank Pasquale, *Normative Dimensions of Consensual Application of Black Box Artificial Intelligence in Administrative Adjudication of Benefits Claims*, 84 LAW & CONTEMP. PROBS. 35 (2021) (detailing how governments can use black box AI for administrative adjudication); Coglianese & Lehr, *supra* note 98 (analyzing how government agencies can use AI for administrative decisions); Arti K. Rai, *Machine Learning at the Patent Office: Lessons for Patents and Administrative Law*, 104 IOWA L. REV. 2617 (2019) (explaining how the PTO uses AI); Danielle Keats Citron, *Technological Due Process*, 85 WASH. L. REV. 1249 (2008) (arguing that usage of AI by administrative agencies threatens due process norms).

274. *See* ENGSTROM ET AL., *supra* note 30, at 3, 53; Opderbeck, *supra* note 30.

275. Coglianese & Lehr, *supra* note 98, at 1205.

276. *Id.* at 1205–06.

277. Engstrom & Ho, *supra* note 273, at 823–24.

automation of administrative processes complicates matters because the inner workings of algorithmic models are often opaque and inscrutable to human observers, which inhibits meaningful explanation.[278] The following discussion explains how the opacity of the FDA's computer models intersects with current theories of administrative law. Part III.B then analyzes how the adoption of AI by FDA staff effectively delegates legislative authority to software developers, potentially undermining agency legitimacy.

A. Black Box Simulations and FDA Transparency and Accountability

Machine learning algorithms are useful because they analyze large datasets to find and leverage correlations that are imperceptible to humans.[279] These impressive abilities add complexity, and consequently, many algorithmic models are subject to what experts call the black box problem.[280] Because its reasoning tends to be inscrutable, algorithmic decision-making can be difficult to comprehend or explain. The black box problem is part of what Frank Pasquale describes as "the triple barriers" to public accountability; other factors include trade secrecy, nondisclosure agreements, and technical complexity.[281]

Scholars disagree on the social and legal significance of the black box problem. Some see it as an unavoidable feature of algorithmic predictions rather than a cause for concern.[282] David Freeman Engstrom and Daniel E. Ho acknowledge that algorithmic opacity could erode administrative accountability by making agency decisions less accessible and understandable.[283] However, they suggest that in some cases, algorithms might improve agency transparency.[284] By adopting algorithmic rules as substitutes for unpredictable human

278. *See id.* at 824.

279. *See* Marks, *supra* note 124.

280. *See* Frank Pasquale, *Licensure as Data Governance*, KNIGHT FIRST AMENDMENT INST. (Sept. 28, 2021), https://knightcolumbia.org/content/licensure-as-data-governance [https://perma.cc/CP7W-5LW2].

281. *Id.*

282. *See* Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1024–25 (2017) (arguing that having accessible inputs and outputs is more important than algorithmic transparency for preventing algorithmic discrimination); *see also* Coglianese & Lehr, *supra* note 98, at 1206–07 (arguing that, despite the inherent black box nature of algorithms, algorithms can be examined by humans to promote transparency).

283. Engstrom & Ho, *supra* note 273, at 821.

284. *Id.*

judgment, agencies could standardize their procedures, to some extent, making their decision-making processes more transparent and predictable.[285]

Other scholars question whether transparency in administrative decision-making is always beneficial or desirable.[286] David Pozen argues that government transparency should not be sought as an end in itself because, in some cases, greater transparency promotes negative outcomes.[287] Pozen calls for scholars to avoid making abstract generalizations about transparency policies in favor of drilling down "into the specific legal, institutional, historical, political, and cultural contexts in which these policies are crafted and implemented."[288] This sociological approach to transparency calls for utilizing mixed methods and "reckoning with transparency in its full complexity as a social phenomenon."[289] In other words, one must define what one means by transparency and explain why it is beneficial in different contexts.

Margaret Kwoka and Bridget DuPey distinguish transparency from the related concept of disclosure.[290] They define the former as "across-the-board requirements for government openness," and the latter as "targeted requirements of private organizations to release otherwise closely held information."[291] However, not everyone makes this distinction, and people often use the terms interchangeably.[292]

Some scholars argue that black box algorithms apply a veneer of objectivity to agency decisions, making them appear more logical, impartial, and scientific—even when they are not.[293] This façade can promote undeserved trust in agency models and regulatory decisions.

---

285. *Id.*

286. *See, e.g.*, David E. Pozen, *Seeing Transparency More Clearly*, 80 PUB. ADMIN. REV. 326, 326–28 (2019).

287. *Id.* at 327–28 (arguing that bad actors can exploit the information revealed through government transparency, and in some cases, transparency ties the hands of agencies, creates a culture of suspicion that reduces trust, and places citizens and the state in an adversarial relationship).

288. *Id.* at 326.

289. *Id.* at 330.

290. Margaret Kwoka & Bridget DuPey, *Targeted Transparency as Regulation*, 48 FLA. ST. L. REV. 389, 390 (2021).

291. *Id.*

292. *Id.* at 406.

293. *See* Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 797, 805 (2021).

The FDA's evangelism for molecular modeling illustrates this point.[294] Superficially, the PHASE methodology sounds futuristic and impressive; yet when one digs deeper, there is little substance to be found.[295] Misplaced trust in algorithmic decisions can have harmful societal and public health consequences.[296] Giroir's 2018 letter to the DEA describes how lives can be endangered by poor regulatory decisions.[297] More disclosure and greater transparency surrounding how the FDA and DEA rely on algorithms might reduce the risk by increasing agency accountability. Because the rules encoded by predictive algorithms are often hidden from view, they may violate the letter and spirit of open government laws intended to make agency conduct accessible. [298] According to Danielle Citron, algorithmic predictions "endanger the basic right to be given notice of an agency's intended actions," which threatens constitutional and statutory guarantees of due process.[299]

Concerns regarding due process guarantees may appear less salient in the public health context. Courts have failed to find that FDA decisions threaten fundamental rights or trigger constitutional protections.[300] Nevertheless, FDA research and regulatory decisions are central to public health, and there are ample opportunities for outside influence to negatively impact agency decision-making, elevating the importance of disclosure and transparency to promote

294. *See* Gottlieb 2018, *supra* note 141.

295. *See* Giroir 2018, *supra* note 183, at 3 (dismissing the FDA's evidence because it failed to meet the criteria required for scheduling kratom).

296. *Id.* at 3–4.

297. *Id.* (describing the harms that may follow the unnecessary scheduling of kratom under the CSA).

298. *See* Citron, *supra* note 273, at 1281–82 (arguing that automation jeopardizes transparency and public participation in administrative decisions).

299. *Id.*

300. *See, e.g.*, Abigail All. v. Eschenbach, 495 F.3d 695, 711 (D.C. Cir. 2007) (finding no fundamental right for terminally ill patients to access experimental therapies); *see also* Mitchell v. Clayton, 995 F.2d 772, 775 (7th Cir. 1993) ("[M]ost federal courts have held that a patient does not have a constitutional right to obtain a particular type of treatment or to obtain treatment from a particular provider if the government has reasonably prohibited that type of treatment or provider.").

agency accountability.[301] Moreover, reliance on algorithmic models in public health may compound past injustice.[302]

Deborah Hellman promotes an anticompounding injustice principle,[303] which is particularly relevant to public health regulation given the government's history of exploiting marginalized communities in the name of research and the disparate impact of drug laws on those communities.[304] When taking Pozen's sociological approach to transparency, one must consider these historical, political, and institutional factors.[305]

In the context of controlled substance regulation, evidence suggests that the drug scheduling system has racist origins.[306] Moreover, the war on drugs, which was bolstered by passage of the CSA, has devastated communities of color.[307] Even if one accepts that government transparency is not always desirable, these historical concerns, numerous examples of algorithmic bias, and existing social

---

301. *See, e.g.*, Press Release, U.S. Food & Drug Administration, FDA and DoD Launch Program To Expedite Availability of Medical Products for the Emergency Care of American Military Personnel (Jan. 16, 2018), https://www.fda.gov/news-events/press-announcements/fda-and-dod-launch-program-expedite-availability-medical-products-emergency-care-american-military [https://perma.cc/9P74-FG6B] (announcing the start of an FDA program prioritizing the development of safe and effective medical products to save the lives of military personnel); *see also* Press Release, U.S. Food & Drug Administration, FDA and DHS Increase Coordination of Responses to Medical Device Cybersecurity Threats Under New Partnership; A Part of the Two Agencies' Broader Effort to Protect Patient Safety (Oct. 16, 2018), https://www.fda.gov/news-events/press-announcements/fda-and-dhs-increase-coordination-responses-medical-device-cybersecurity-threats-under-new [https://perma.cc/E8N3-C3HG] (emphasizing the FDA's role in ensuring patient protection).

302. *See* Deborah Hellman, *Big Data and Compounding Injustice*, U. VA. PUB. L. & LEGAL THEORY RES. PAPER SERIES, May 2021, at 1, 1 (defining the anticompounding injustice principle, which states "the fact that an action will compound a prior injustice counts as a reason against doing that action").

303. *Id.*

304. *E.g.*, Brandt, *supra* note 235, at 21–24; *see* Doris Marie Provine, *Race and Inequality in the War on Drugs*, 7 ANN. REV. L. SOC. SCI. 41, 46, 49–51 (2011) (arguing the war on drugs led to a vast increase in imprisonment of Black individuals in the United States).

305. *See* Pozen, *supra* note 286, at 326 (describing the sociological approach to transparency).

306. See Erik Sherman, *Nixon's Drug War, an Excuse To Lock Up Blacks and Protesters, Continues*, FORBES (Mar. 23, 2016), https://www.forbes.com/sites/eriksherman/2016/03/23/nixons-drug-war-an-excuse-to-lock-up-blacks-and-protesters-continues [https://perma.cc/QNR9-CPF2] (discussing a 1994 interview in which former Richard Nixon aide John Ehrlichman stated, "The Nixon campaign in 1968, and the Nixon White House after that, had two enemies: the antiwar left and black people." (quoting Dan Baum, *Legalize It All*, HARPER'S MAG. (Apr. 2016), https://harpers.org/archive/2016/04/legalize-it-all [https://perma.cc/VG3B-GTXL])).

307. *See generally* Provine, *supra* note 304 (examining the role of social racism in the war on drugs).

and healthcare inequities highlight the importance of promoting algorithmic transparency in the public health context.

Outsourcing the design and control of algorithms exacerbates transparency and accountability concerns.[308] When agencies adopt proprietary models, third-party developers may treat their algorithms and training data as trade secrets.[309] Courts have required agencies to reveal some information about algorithmic decisions, such as the data inputted into the systems.[310] Yet, there is little consensus regarding the level of detail required.[311] FOIA can potentially shed some light on algorithmic agency decisions.[312] However, the statute has constraints that diminish its utility.[313] Under one FOIA exemption, "trade secrets and commercial or financial information" need not be revealed, which may limit the public's ability to compel disclosure of proprietary algorithms.[314] Hannah Bloch-Wehba emphasizes that FOIA's trade secret and confidentiality exemption intends to promote information sharing between the private sector and the government.[315] The exception was not created to obscure agency decisions from public view.[316] However, it could be invoked for that purpose.[317]

Another FOIA exception creates a "deliberative process privilege" that shields aspects of government deliberations, such as interagency or intra-agency memoranda, from outside scrutiny and interference.[318] In theory, it exempts the formulation of agency policy from public disclosure while leaving accessible the facts on which those policies are based.[319] In 2006, the Environmental Protection Agency

---

308. *See* Citron, *supra* note 273, at 1290.

309. *Id.* at 1293.

310. *See* Hanna Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1291 (2020).

311. *Id.*

312. *See id.* at 1298 (stating that Congress enacted FOIA to open agency actions to public scrutiny by creating a process for requesting administrative records and by giving federal courts jurisdiction to prevent agencies from withholding them—and to order their production when improperly withheld).

313. *See id.* at 1299 (explaining that because FOIA only requires the disclosure of records that an agency controls, some algorithmic agency decisions may still remain secret).

314. *Id.* at 1300.

315. *Id.*

316. *Id.*

317. *Id.*

318. *Id.* at 1302; *see* Reilly v. EPA, 429 F. Supp. 2d 335, 341 (D. Mass. 2006).

319. Bloch-Wehba, *supra* note 310, at 1302–03.

("EPA") invoked the deliberative process privilege to avoid disclosing details of a computer model it used to analyze different approaches to pollution control.[320]

The U.S. District Court for the District of Massachusetts rejected the EPA's claim that its model represented the agency's deliberative process, finding instead that the model was a fact-finding tool.[321] Despite acknowledging that the model's operation might reflect some aspects of the EPA's reasoning, the court framed it as an investigative instrument that generates evidence to inform agency rulemaking.[322] In other words, on the continuum separating factual information from deliberative processes, the model's calculations more closely resembled facts.[323] The court reasoned that in agency fact-finding investigations, "knowing what questions are asked or which witnesses are interviewed reveals aspects of what the investigator deemed important or worthy of consideration."[324]

This discussion might suggest that details of the FDA's computer models should be accessible under FOIA. However, the agency's culture typically resists attempts to compel transparency.[325] Besides, as an enforcement agency, the FDA may qualify for further FOIA exceptions, potentially limiting public access to the inner workings of its algorithms.[326]

Some agencies have taken steps to make their algorithmic predictions more transparent.[327] For instance, the PTO has expressed an intolerance for black box algorithms.[328] In 2016, it awarded a contract to a company called AI Patents, which develops patent search algorithms.[329] According to the company, its contract was not renewed

320. *See Reilly*, 429 F. Supp. 2d at 336–37.

321. *Id.* at 352–54.

322. *Id.* at 352.

323. *Id.*

324. *Id.*

325. *See* David Gortler, *How the FDA's Lack of Transparency Undermines Public Trust*, FORBES (Aug. 24, 2021, 12:36 PM), https://www.forbes.com/sites/davidgortler/2021/08/24/how-the-fdas-lack-of-transparency-undermines-public-trust [https://perma.cc/BV2K-NM5C].

326. *See* Daniel C. Taylor, *Taking* Touhy *Too Far: Why It Is Improper for Federal Agencies To Unilaterally Convert Subpoenas into FOIA Requests*, 99 GEO. L. REV. 1227, 1242–48 (2011) (describing FOIA exemptions and their application to the FDA).

327. *See* Rai, *supra* note 273, at 2639 (describing the PTO's emphasis on algorithmic transparency).

328. *Id.*

329. *Id.* at 2638.

due to disagreements with the agency over full transparency.[330] For similar reasons, the EPA prefers adopting nonproprietary computer models when possible, and for circumstances in which they must be used, the agency provides guidance for promoting transparency.[331] Like the FDA, the EPA has its own molecular modeling platform, which it calls ToxCast.[332] However, unlike the FDA, the EPA makes its model available for public use and provides an "owner's manual" to help people understand and use the system.[333]

As algorithmic models increasingly influence government decisions, agencies could employ more staff capable of designing, applying, and interpreting them.[334] Requiring agencies to build models internally would keep them up to date on algorithmic technologies and promote independence instead of reliance on the private sector. In addition to transparency concerns, close ties with industry create conflicts of interest, which may be obscured by trade secrecy, the black box problem, and the veil of objectivity it creates. Moreover, according to the 2018 ACUS report, designing algorithms in-house makes for more effective models that are more likely to be designed and deployed lawfully.[335]

Aside from developing computer models in-house, there are legal solutions for promoting transparency. Congress could require third-party vendors to make their source code and training data publicly available.[336] The Office of Management and Budget could condition funding for technology acquisitions on the adoption of nonproprietary models and open-source code.[337] Some experts contend that certain AI

330. *Id.*

331. U.S. ENV'T PROT. AGENCY, GUIDANCE ON THE DEVELOPMENT, EVALUATION, AND APPLICATION OF ENVIRONMENTAL MODELS 31–32 (2009), https://www.epa.gov/sites/default/files/2015-04/documents/cred_guidance_0309.pdf [https://perma.cc/NP6R-SPD9].

332. *Toxicity Forecasting: Advancing the Next Generation of Chemical Evaluation*, U.S. ENV'T PROT. AGENCY, https://www.epa.gov/chemical-research/toxicity-forecasting [https://perma.cc/68HQ-TGU2].

333. *See ToxCast Owner's Manual – Guidance for Exploring Data*, U.S. ENV'T PROT. AGENCY, https://www.epa.gov/chemical-research/toxcast-owners-manual-guidance-exploring-data [https://perma.cc/6Q2L-TNYB].

334. *See* ENGSTROM ET AL., *supra* note 30, at 7 (stating the importance of in-house AI expertise).

335. *Id.*

336. *See* Citron, *supra* note 273, at 1308 (arguing that releasing source code would increase transparency by revealing how a system works).

337. *See id.* at 1309.

applications should remain confidential due to national security concerns.[338] However, mandating open-source code at the FDA will likely mitigate more public health risks than it creates, and publishing code will allow independent researchers to examine it and run their own simulations. In fact, Citron suggests that agencies should be required to test their software using hypothetical scenarios provided by independent experts, which would help expose unreliable systems.[339]

Further, the APA's notice-and-comment requirements could help improve transparency and accountability when agencies adopt models and simulations.[340] Though the APA requires legislative rules to be published, not all agency actions are legislative.[341] The FDA famously avoids notice-and-comment rulemaking by issuing most recommendations as industry guidance, which is ostensibly nonbinding.[342] This practice allows the agency to issue de facto rules while escaping certain procedural safeguards.[343] However, unlike other agencies, the FDA must submit major guidance documents for public comment prior to their final adoption.[344] This requirement stems from the Food and Drug Modernization Act of 1997, which mandates public participation in the development of significant FDA guidance.[345] However, unlike notice-and-comment rulemaking under the APA, the 1997 Act does not require the FDA to respond to the public comments it receives in response to draft guidance.[346]

Definitions for agency guidance vary. However, the FDA's announcement that it had adopted PHASE would arguably qualify. FDA regulation defines guidance as documents created for applicants, sponsors, agency staff, or the public that describe FDA policy

338. *E.g.*, *id.*

339. *Id.* at 1310.

340. *See* Engstrom & Ho, *supra* note 273, at 836.

341. *See id.* at 836–37.

342. *See* Lars Noah, *Governance by the Backdoor: Administrative Law(lessness) at the FDA*, 93 NEB. L. REV. 89, 97 (2014) (explaining how the FDA may utilize nonbinding industry guidance as de facto rules while avoiding procedural safeguards).

343. *Id.* at 97.

344. *Id.* at 99.

345. 21 U.S.C. § 371(h); *see also* Cary Coglianese, *Illuminating Regulatory Guidance*, 9 MICH. J. ENV'T & ADMIN. L. 243, 257–58 (2020) (distinguishing significant (Level 1) FDA guidance that requires publication prior to implementation from less significant (Level 2) guidance that does not).

346. Noah, *supra* note 342, at 102.

regarding, or interpretation of, a regulatory issue.[347] Guidance can include documents related to the design and testing of regulated products, the evaluation or approval of submissions, enforcement policies, and other matters.[348] Gottlieb's PHASE announcement ostensibly described agency policies regarding kratom regulation and the overdose crisis, and it was directed to the public. However, the FDA might argue that Gottlieb's announcement was not guidance because it constituted press materials, editorial content, or general information provided to consumers, all of which are excluded from the agency's definition of guidance.[349]

One potential solution is to frame the adoption of algorithmic models as legislative rulemaking by default, requiring it to be subjected to notice and comment under the APA.[350] Engstrom and Ho describe the factors courts use to identify agency decisions that rise to the level of legislation.[351] However, distinguishing legislative rules from nonlegislative action may be the most vexing problem of administrative law.[352] One factor courts may consider is the extent to which an agency's pronouncement binds its future behavior.[353] Whether staff are bound turns largely on the degree of human involvement that remains after computer models are adopted, which is open to interpretation.[354]

Regarding the FDA's use of molecular modeling, arguably agency staff could simply disregard model outputs. FDA scientists claim these predictions are subject to expert review before a scheduling recommendation is made.[355] Moreover, they would likely contend that because humans remain in the loop, the FDA is not bound by molecular modeling predictions.[356]

---

347. 21 C.F.R. § 10.115(b)(1) (2021).

348. *Id.* § 10.115(b)(2).

349. *Id.* § 10.115(b)(3).

350. Engstrom & Ho, *supra* note 273, at 836.

351. *Id.* at 837.

352. *See* David L. Franklin, *Legislative Rules, Nonlegislative Rules, and the Perils of the Short Cut*, 120 YALE L.J. 276, 278 (2010).

353. *Id.*

354. *Id.*

355. Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*, *supra* note 146, at 117.

356. *See* Engstrom & Ho, *supra* note 273, at 837.

However, at least some caselaw argues against this interpretation. In *McLouth Steel Products Corp. v. Thomas*,[357] the D.C. Circuit considered whether an EPA computer model was subject to notice-and-comment requirements.[358] According to the court, it was practically irrelevant whether the EPA considered itself bound to the model's predictions.[359] The agency's past statements and prior applications of the model were most determinative.[360] Though the EPA had previously stated that it retained discretion to ignore the model, and that its predictions were only one of several factors considered in regulatory decisions, the court pointed to other statements within the same document that suggested that the agency was bound by the model's predictions.[361] The EPA had framed its model as "the quantitative approach" that "*will* be used" to make predictions regarding environmental toxins.[362] The court said the agency's use of the word "will" reflected "the rigor of a rule, not the pliancy of a policy."[363]

Still, FDA staff argue that what makes molecular modeling so useful is its ability to perform tasks that humans would otherwise perform.[364] They claim the list of substances requiring regulatory review is so long that it would be cost prohibitive to review each case manually by performing the entire eight-factor analysis required by the CSA.[365] Accordingly, to the extent that staff forego performance of the eight-factor analysis, they rely on models to fulfill their statutory obligations. Instead of merely supporting FDA staff in this role, algorithm predictions replace their judgments, making agency staff dependent on them in this context.

Commissioner Gottlieb relied on molecular modeling when he unveiled PHASE in 2018.[366] According to his statement at the time, "[D]ata from the PHASE model shows us that kratom compounds are

357. McLouth Steel Prods. Corp. v. Thomas, 838 F.2d 1317 (D.C. Cir. 1988).

358. *Id.* at 1319.

359. *Id.* at 1320.

360. *Id.*

361. *Id.*

362. *Id.*

363. *Id.* at 1320–21.

364. *See* Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*, *supra* note 146, at 117.

365. *Id.*

366. Gottlieb 2018, *supra* note 141.

predicted to affect the body just like opioids."[367] Further, "this highlights the power of our computational model-based approach to rapidly assess any newly identified natural or synthetic opioids to respond to a public health emergency."[368] Like the EPA's intent to rely on environmental models, Gottlieb's statements reflect the FDA's reliance on algorithms to address the overdose crisis, and they could represent the "will" described in *McLouth Steel*.[369] However, some courts are unmoved by a lack of human involvement and refuse to categorize rulemaking as legislation, even when the promulgating agency is left with little or no discretion.[370]

A second factor that can distinguish legislative from nonlegislative rules is whether a rule has the "force of law," impacting the rights and interests of regulated entities.[371] In other words, legislative rules implement the relevant statute instead of merely interpreting it.[372] One might argue that the FDA's use of models to make scheduling recommendations lacks legal effect. However, the agency performs scheduling analyses on behalf of the secretary of HHS, and the secretary's determinations bind the DEA.[373] Adding a substance to Schedule I, for instance, renders it illegal and halts related scientific research.[374] Institutions hoping to commercialize or study the substance would face higher legal and financial barriers.[375] Consequently, adopting algorithmic models to guide scheduling decisions substantially impacts the interests of regulated parties. In the context

367. *Id.*

368. *Id.*

369. McLouth Steel Prods. Corp. v. Thomas, 838 F.2d 1317, 1320–21 (D.C. Cir. 1988).

370. Engstrom & Ho, *supra* note 273, at 838 (citing *Ass'n of Irritated Residents v. EPA*, 494 F.3d 1027, 1034 (D.C. Cir. 2007), where the D.C. Circuit held agreements made by the EPA were not legislative because ruling otherwise might lower the bar for what constitutes a legislative rule and potentially make nearly all agency consent agreements subject to notice-and-comment requirements).

371. Nat'l Mining Ass'n v. McCarthy, 758 F.3d 243, 250 (D.C. Cir. 2014) ("Legislative rules have the 'force and effect of law'. . . ."); Engstrom & Ho, *supra* note 273, at 837 (stating that whether a rule "substantially alters the rights and interests of regulated parties" affects whether it is legislative).

372. *Nat'l Mining Ass'n*, 758 F.3d at 251–52.

373. 21 U.S.C. § 811(b).

374. *See* Alex Kreit, *Controlled Substances, Uncontrolled Law*, 6 ALB. GOV'T L. REV. 331, 352–53 (2013) (describing the effect on research of categorizing a substance in Schedule I).

375. *See* Mason Marks, *Psychedelic Medicine for Mental Illness and Substance Use Disorders: Overcoming Social and Legal Obstacles*, 21 N.Y.U. J. LEGIS. & PUB. POL'Y 69, 89 (2018) (describing the legal and financial barriers to Schedule I drug manufacturing and research).

of FDA approval, substituting model predictions for clinical data impacts the rights of drug sponsors and medical device manufacturers. When the FDA deemed Medtronic's pacemaker-lead MRI safe using only simulation data, it saved Medtronic from conducting human studies with hundreds of patients.[376] Consequently, the FDA's adoption of models for regulatory decision-making arguably impacts the rights and interests of regulated parties and could constitute a rule having the force of law.

Despite the aforementioned arguments, this area of administrative law is notoriously murky, and establishing that the FDA's adoption of computer models constitutes a legislative rule is far from a clear-cut case.[377] Furthermore, notice-and-comment review itself is an imperfect mechanism for promoting transparency and accountability. A common criticism is that well-funded entities are likely best equipped to understand proposed rules and shape them to promote their interests.[378] The public may also lack the motivation to participate in notice-and-comment rulemaking.[379] However, when citizen groups are impacted by agency decisions, they can mobilize their membership and capitalize on their relevant expertise to provide impactful feedback on proposed legislation.[380] Consequently, the utility and importance of notice and comments should not be underestimated.[381]

Skeptics may also raise concerns regarding rulemaking ossification, the theory that bolstering procedural requirements inordinately increases the time required for agencies to promulgate

376. Owen Faris & Jeffrey Shuren, *An FDA Viewpoint on Unique Considerations for Medical-Device Clinical Trials*, 376 NEW ENG. J. MED. 1350, 1353 (2017).

377. *See* Paralyzed Veterans of Am. v. D.C. Arena L.P., 117 F.3d 579, 587 (D.C. Cir. 1997) ("[I]t is quite difficult to draw a line between substantive and interpretive rules."); *see also* Franklin, *supra* note 352, at 287 (reciting a common refrain of courts and administrative law scholars that describes the distinction between legislative and nonlegislative rules as "'fuzzy,' 'tenuous,' 'blurred,' 'baffling,' and 'enshrouded in considerable smog'" (first quoting Am. Hosp. Ass'n v. Bowen, 834 F.2d 1037, 1046 (D.C. Cir. 1987); then quoting Chisholm v. FCC, 538 F.2d 349, 393 (D.C. Cir. 1976); then quoting Cmty. Nutrition Inst. v. Young, 818 F.2d 943, 946 (D.C. Cir. 1987); and then quoting Noel v. Chapman, 508 F.2d 1023, 1030 (2d Cir. 1975))).

378. Katherine J. Strandburg, *Rulemaking and Inscrutable Automated Decision Tools*, COLUM. L. REV. 1851, 1869 (2019).

379. *Id.* at 1869–70.

380. *Id.*

381. *Id.* at 1870.

regulations.[382] Procedural requirements may include performing cost-benefit analyses, subjecting rules to notice and comment, explaining agency reasoning in anticipation of judicial review, and documenting the quality of the data that informed the rulemaking process.[383] Critics of ossification argue that these burdens cause unnecessary delay, increase the likelihood that regulation will become obsolete, and discourage agencies from using the notice-and-comment rulemaking process.[384] Some scholars worry that agencies may respond to ossification by favoring other means of promulgating rules such as adjudication.[385]

Not everyone agrees that ossification is harmful or undesirable. Some accept it as the cost of ensuring that rules are scientifically grounded.[386] Others contend that ossification can itself be beneficial.[387] Aaron Nielson argues that ossification has benefits aside from ensuring the soundness of promulgated rules.[388] For instance, because it makes changing rules more onerous, ossification helps commit agencies to the rules they make.[389] Relatedly, when changes become more burdensome, regulated parties can be more confident that promulgated rules will endure.[390] This enhanced confidence, Nielson argues, may increase the likelihood that regulated parties act in ways that agencies want them to act.[391] Ossification might also increase the legitimacy of agency actions because delay allows time for public participation, which in turn builds trust.[392]

---

382. *See* Aaron L. Nielson, *Optimal Ossification*, 86 GEO. WASH. L. REV. 1209, 1210–11 (2018) [hereinafter Nielson, *Optimal Ossification*] (defining rulemaking ossification); *see also* Aaron L. Nielson, *Sticky Regulations*, 85 U. CHI. L. REV. 85, 94 (2018) (presenting a common argument against rulemaking ossification).

383. *See* Nielson, *Optimal Ossification*, *supra* note 382, at 1217; *see also* Stuart Shapiro, *Embracing Ossification*, REGULATION, Winter 2018–2019, at 8, 8.

384. Nielson, *Optimal Ossification*, *supra* note 382, at 1215–17.

385. *See* Jason Webb Yackee & Susan Webb Yackee, *Testing the Ossification Thesis: An Empirical Examination of Federal Regulatory Volume and Speed, 1950-1990*, 80 GEO. WASH. L. REV. 1414, 1463 (2012) (describing how agencies may increasingly turn to adjudication as rulemaking becomes more difficult).

386. Nielson, *Optimal Ossification*, *supra* note 382, at 1219.

387. *Id.* at 1219–20.

388. *Id.*

389. *Id.* at 1220.

390. *Id.*

391. *See id.* (explaining how the durability of administrative rules affects the actions of regulated parties).

392. *Id.* at 1227.

In some cases, it may be desirable to promulgate rules quickly. However, substituting algorithmic models for clinical trials, and other well-established practices for collecting and analyzing scientific evidence, should not be one of them. The public health risks of moving hastily are great, and opportunities for exploiting confusing or otherwise poorly conceived rules are plentiful. In a time when the FDA's reputation is under fire and confidence in public health agencies is waning, delay that allows for careful deliberation and public participation seems prudent. To date, the FDA has not sought public comment on its adoption of computer models.

Notably, some agencies take a different approach. In 2015, the EPA announced its intent to use molecular modeling in the Federal Register.[393] The technology it disclosed is analogous to the FDA's molecular models because it uses algorithmic predictions to infer which compounds are potentially harmful.[394] In addition to seeking public input, compared to the FDA, the EPA has been more forthcoming regarding the limitations of computer models used in this context.[395] In the Federal Register, the EPA cautioned the public not to confuse predictions made by its model with agency determinations that a particular substance causes harm.[396] This caveat stands in contrast to Gottlieb's reliance on PHASE to conclude that kratom's ingredients "affect the body just like opioids."[397]

B. Modeling, Simulation, and Nondelegation

Article II of the U.S. Constitution vests all legislative powers in Congress.[398] However, courts have long permitted Congress to delegate legislative authority to agencies led by appointed officers.[399] In theory, the nondelegation doctrine imposes limits on the power delegated to federal agencies.[400] Constitutional delegations must be accompanied by

393. Use of High Throughput Assays and Computational Tools, 80 Fed. Reg. 35,350, 35,350 (June 19, 2015).

394. *See id.* at 35,352 ("To reduce non-specific results, the computational model can use results from multiple assays and technologies to predict whether a chemical is truly bioactive in the pathway being evaluated.").

395. *See, e.g.*, *id.*

396. *Id.* at 35,353.

397. Gottlieb 2018, *supra* note 141.

398. U.S. CONST. art. 1, § 1.

399. Coglianese & Lehr, *supra* note 98, at 1178.

400. *Id.*

an intelligible principle to guide agency officials.[401] However, for nearly a century, the Supreme Court has declined to find delegations of legislative authority unconstitutional, regardless of their breadth.[402]

Many scholars view nondelegation doctrine as an aspirational concept of little or no practical importance.[403] Some say it was doomed from the start because the Founders never contemplated such a principle nor is it reflected in the Constitution.[404] Others believe the doctrine could be salvaged if only it was reimagined.[405] In recent years, a few Supreme Court opinions suggest there may be renewed interest in the doctrine. Meanwhile, some have questioned whether replacing agency judgment with algorithmic predictions could constitute the unlawful delegation of legislative authority.[406]

Cary Coglianese and David Lehr are unconcerned. They argue that all machine learning algorithms have objective functions that can substitute for the intelligible principle required by common law.[407] In contrast, Citron and Ryan Calo question whether an algorithm's objective function "bears the slightest resemblance to an intelligible principle."[408] The traditional justification for delegating legislative authority to agencies assumes that they possess specialized knowledge that Congress lacks.[409] However, algorithms, as currently designed and implemented, are poor replacements for real expertise.[410] According to Citron and Calo, by delegating authority to opaque computer models, agencies transfer power from Congress to software, discarding discretion and expertise, and rendering the principles underlying permissible delegations unintelligible.[411]

Coglianese and Lehr disagree. They frame algorithms as mere measurement tools, similar to other instruments adopted by agencies,

401. *Id.*

402. ERWIN CHEMERINSKY, CONSTITUTIONAL LAW 295 (6th ed. 2019).

403. Ronald A. Cass, *Delegation Reconsidered: A Delegation Doctrine for the Modern Administrative State*, 40 HARV. J.L. & PUB. POL'Y 147, 150 (2016).

404. *See, e.g.*, Julian Davis Mortenson & Nicholas Bagley, *Delegation at the Founding*, 121 COLUM. L. REV. 277, 279–80 (2021).

405. *See, e.g.*, Cass, *supra* note 403, at 151 (arguing that nondelegation doctrine should be reoriented to focus on the nature of the granted authority instead of its scope).

406. *See* Coglianese & Lehr, *supra* note 98, at 1178–80.

407. *Id.* at 1179.

408. Calo & Citron, *supra* note 293, at 817.

409. *Id.* at 816.

410. *Id.* at 833.

411. *Id.*

such as rulers or calculators.[412] Nonetheless, it seems nonsensical to suggest that using rulers and calculators could transfer decision making authority to their manufacturers, while the notion that adopting computer models might transfer such authority has greated intuitive appeal. At the very least, when agencies rely on outside firms to produce computer models because they lack the expertise to create and train the models themselves, agencies may undermine their own credibility and legitimacy.[413] The FDA's continued reliance on molecular models overruled by HHS, the agency to which Congress delegated the authority to conduct the CSA's eight-factor analysis, lends credibility to this argument.

This discussion suggests that administrative law may be unprepared for automated FDA regulation. Though some aspects of the law may be flexible enough to promote algorithmic transparency and accountability, others must be updated, or other areas of law should pick up the slack. The following Part suggests blending principles of AI ethics with existing industry and agency guidance to fill the gaps left by administrative law.

IV. GOOD SIMULATION PRACTICES

There are currently no standards to guide either the automation of agency decision-making generally or the use of AI to guide medical and public health regulatory decisions specifically. To move toward this goal, this Part analyzes industry and FDA guidance on models and simulations. After evaluating the strengths and weaknesses of agency guidance, this Part concludes by incorporating principles of AI ethics into existing guidance to create good simulation practices that promote transparency while acknowledging the full spectrum of potential harms.

A. Industry and Agency Guidance

The FDA has published a series of documents on modeling and simulation.[414] It identifies the adoption of these technologies as agency

412. Coglianese & Lehr, *supra* note 98, at 1181.

413. *See* Calo & Citron, *supra* note 293, at 833, 835.

414. *See* U.S. FOOD & DRUG ADMIN., REPORTING OF COMPUTATIONAL MODELING STUDIES IN MEDICAL DEVICE SUBMISSIONS: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF 2 (2016) (providing guidance to industry stakeholders on how to submit modeling information to the FDA); *see also Credibility of Computational Models Program*, *supra*

priorities.[415] One of the FDA's primary goals is to reduce its reliance on clinical trials by augmenting or replacing them with simulations in which devices are tested on cohorts of virtual patients.[416] However, the FDA acknowledges that the credibility of computer models has not been fully established.[417] Accordingly, it created an initiative called the Credibility of Computational Models Program.[418]

Both the FDA and the American Society of Mechanical Engineers ("ASME") acknowledge that the credibility of models turns on whether people trust their predictive capabilities.[419] The FDA identifies major gaps and challenges associated with establishing trust.[420] According to agency reports, many existing and proposed models have not been rigorously evaluated, and their credibility is unknown.[421] Some have known deficiencies that negatively affect their credibility.[422] There is a paucity of quality clinical or experimental data available to help develop and validate models and simulations, especially data derived from humans under real-world conditions.[423] There are insufficient validation tools and metrics, and reliable methods do not exist to evaluate the acceptability of virtual patients in a cohort.[424] Moreover, there is a lack of tools to verify relevant computer code and calculations and a lack of decision-making frameworks to evaluate the overall credibility of models.[425] Finally, there are no established best practices.[426]

---

note 107 (providing draft guidance to industry on establishing the credibility of computer models); U.S. FOOD & DRUG ADMIN., ASSESSING THE CREDIBILITY OF COMPUTATIONAL MODELING AND SIMULATION IN MEDICAL DEVICE SUBMISSIONS: DRAFT GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF 4–5 (2021) [hereinafter FDA DRAFT GUIDANCE 2021] (providing draft guidance for industry and FDA staff).

415. *Credibility of Computational Models Program*, *supra* note 107; *see also* Morrison et al., *supra* note 107 (stating that computer modeling is a top priority for the FDA's Center for Devices and Radiological Health).

416. *Credibility of Computational Models Program*, *supra* note 107.

417. *Id.*

418. *Id.*

419. *Id.*; Morrison et al., *supra* note 107, at 349.

420. *Credibility of Computational Models Program*, *supra* note 107.

421. *Id.*

422. *Id.*

423. *Id.*

424. *Id.*

425. *Id.*

426. *Id.*

One existing framework for evaluating model credibility was designed by the ASME Verification & Validation 40 ("V&V 40") Subcommittee.[427] According to ASME, verification involves proving that the equations solved by a model are mathematically correct, and validation entails demonstrating that the right equations are being solved to address the questions the model is being used to answer.[428] This definition serves as a reminder that computer models are mathematical abstractions bearing no meaningful resemblance to their real-world counterparts.

To illustrate how the V&V 40 framework is applied, FDA and industry researchers used it to evaluate a hypothetical model intended to simulate a centrifugal blood pump, a medical device that aids circulation, to predict the risk of red blood cell rupture, or hemolysis.[429] The V&V 40 framework assumes that the level of evidence required to establish model credibility should vary with the risk associated with using the model to guide clinical or regulatory decisions.[430] In other words, the framework should adapt to the unique risks of different applications. It describes four steps for determining the level of evidence required: identifying a question the model is intended to address; defining the context in which the model will be used and drafting a detailed statement determining its specific role in addressing the question of interest; assessing the model risk, defined as the likelihood that adopting the model will produce decisions that harm patients or lead to other undesirable outcomes; and identifying credibility factors, goals for various aspects of the verification and validation process, which are influenced by the model risk.[431]

Model risk is affected by the degree to which a model influences decision-making, called the model influence, and the significance of adverse outcomes resulting from reliance on the model, called the decision consequence.[432] Models that play only minor roles in a

427. Morrison et al., *supra* note 107, at 350.

428. Colleen Kuemmel, Yuching Yang, Xinyuan Zhang, Jeffry Florian, Hao Zhu, Million Tegenge, Shiew-Mei Huang, Yaning Wang, Tina Morrison & Issam Zineh, *Consideration of a Credibility Assessment Framework in Model-Informed Drug Development: Potential Application to Physiologically-Based Pharmacokinetic Modeling and Simulation*, 9 CPT PHARMACOMETRICS SYS. PHARMACOLOGY 21, 25 (2020).

429. Morrison et al., *supra* note 107, at 350.

430. *Id.*

431. *Id.*

432. *Id.*

decision have lower model influence than models that are heavily relied upon.[433] Both model influence and decision consequence are determined by the context of use, which requires model risk calculations to be context specific.[434] Moreover, the credibility of a model should match its risk. High-risk models require significantly more evidence of credibility than low- or moderate-risk models.[435] ASME has provided a list of thirteen factors that can help establish credibility.[436] Like model influence and model risk, credibility factors may vary with the context of use.[437] Analyzing the relevance of credibility factors to the context of use, and the relevance of the phenomena being modeled, is called determining model *applicability*.[438]

When using the V&V 40 framework to determine the evidence required to establish the credibility of the blood pump model, researchers first identified a question the model would help them answer. They wanted to know whether hemolysis caused by the pump would fall within acceptable levels.[439] They identified two contexts of use for comparison.[440] In the first, the model would predict how the pump would perform in a benchtop test prior to surgery.[441] This scenario was presumably chosen because it represents a low- to moderate-risk application. In the second context, the model would simulate hemolysis levels of an implanted pump that assists a patient's circulation at home.[442]

Researchers determined that in the context of simulating preoperative testing, the model influence would be low because surgical decision making would be based on the outcome of preoperative benchtop testing instead of being influenced directly by the simulation, which would only determine the testing parameters.[443]

---

433. *See* Kuemmel et al., *supra* note 428, at 24 (providing a table that defines low, medium, and high model influence based on the ratio between the influence of model evidence and clinical evidence).

434. Kuemmel et al., *supra* note 428.

435. *See id.* at 22.

436. *Id.* at 23 tbl.2.

437. *Id.* at 22.

438. *See* Morrison et al., *supra* note 107, at 355.

439. *Id.* at 351.

440. *Id.*

441. *See id.* at 351, 352 tbl.2.

442. *See id.*

443. *Id.* at 352 tbl.2.

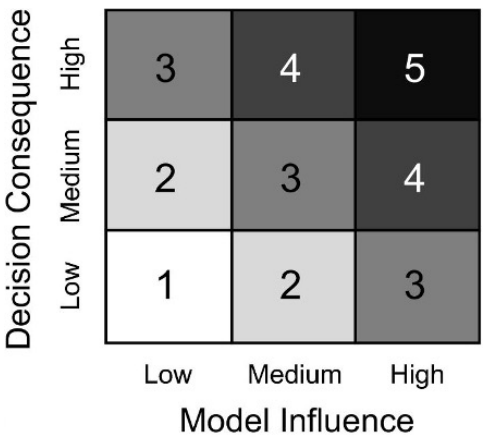
The researchers estimated the decision consequence of the model to be medium because if the pump caused severe hemolysis in the operating room, surgeons could quickly mitigate the harm and the pump could even be replaced.[444]

For the second context of use, simulating at-home use of an implanted pump, the researchers determined the model influence to be high because clinical decisions would be directly influenced by the simulation instead of a preclinical test.[445] Similarly, they found the decision consequence to be high because the pump would be implanted and patients would be far from an operating room, making it more difficult to mitigate harm or to replace a failing pump.[446]

To summarize their results, researchers created a three-by-three matrix called a model risk map, with decision consequence on the x-axis and model influence on the y-axis.[447] The matrix suggests that contexts of use with high model influence and high decision consequence have the greatest model risk—five out of five—requiring the strongest evidence of model credibility to justify adoption.[448] In contrast, contexts of use with low model influence and low decision consequence have the lowest model risk—one out of five—requiring the least evidence of credibility.[449] The researchers concluded that using the simulation to inform benchtop testing prior to surgery carried a model risk of two, and using it to guide decisions outside the operating room carried a model risk of five.[450]

444. *Id.*
445. *Id.*
446. *See id.*
447. *Id.* at 355; Kuemmel et al., *supra* note 428, at 24.
448. Morrison et al., *supra* note 107, at 353, 355.
449. *Id.*
450. *Id.* at 351, 355.

Figure 2: Discrete Risk Map



Though useful for explaining how model influence and decision consequence shape model risk, this application of the V&V 40 framework has shortcomings that limit generalizability. For instance, the researchers only evaluated models for simulating clinical and preclinical scenarios.[451] They did not assess the framework's ability to evaluate models that guide regulatory decisions, which would entail different kinds of risks and credibility factors. In addition, they did not consider AI or its associated risks and harms, such as low-quality training data and algorithmic bias. This omission may not be surprising because the V&V 40 framework was not designed for models that use AI. Instead, it was created to assess the credibility of physics-based models. The framework also assumes that the risks associated with models and each context of use are foreseeable. However, AI-related risks are often difficult to forecast because they involve many unknowns.[452]

Another limitation of this example is the nature of the simulated technology. Centrifugal blood pumps are relatively simple mechanical

451. *Id.* at 352 tbl.2.

452. *See* Wayne Holmes, Kaska Porayska-Pomsta, Ken Holstein, Emma Sutherland, Toby Baker, Simon Buckingham Shum, Olga C. Santos, Mercedes T. Rodrigo, Mutlu Cukurova, Ig Ibert Bittencourt & Kenneth R. Koedinger, *Ethics of AI in Education: Towards a Community-Wide Framework*, INT'L J. A.I. EDUC., Apr. 9, 2021, at 1, 18–19, https://link.springer.com/content/pdf/10.1007/s40593-021-00239-1.pdf [https://perma.cc/4UJ2-B5YF] (discussing the importance of considering unknown unknowns regarding the use of AI in education).

devices. Consequently, they are more easily emulated using well-known equations for modeling fluid dynamics, rather than more complex biological structures such as molecules, organs, and people. "[W]e're really good at simulating medical devices, they're hunks of metal with electronics and all these small parts," said the FDA's Morrison.[453] It is far more challenging to simulate complex biological systems. However, such systems are increasingly modeled by both the FDA and product manufacturers hoping to secure regulatory approval, and such models increasingly rely on AI.[454]

In a 2020 article coauthored with other FDA scientists, Morrison claims there is "no consensus among modeling and simulation approaches or regulatory authorities on how to establish or assess the credibility of a model for regulatory purposes."[455] The lack of standards for assessing model safety and fairness is concerning due to rising pressure from government and industry to model biological systems and substitute simulations for clinical evidence. The authors describe the need for an expanded framework to establishing the credibility of models used in regulatory contexts.[456] Building upon previous discussions of the V&V 40 framework, they introduce more robust ways to determine decision consequence, which could make the framework more useful to regulatory agencies.[457] For instance, they suggest that the decision consequence should increase with the number of people potentially impacted by incorrect decisions, the severity of potential harms, and the likelihood that those harms could occur.[458] Decisions that would not adversely affect patient safety have a low decision consequence.[459] In contrast, decisions that could produce minor or moderate harms have a medium decision consequence, and those that could produce severe harms are assigned a high decision consequence.[460]

---

453. Biovia, *supra* note 23.

454. *Id.*; *see also Artificial Intelligence for Regulatory Science Research*, U.S. FOOD & DRUG ADMIN. (May 14, 2020), https://www.fda.gov/science-research/fda-grand-rounds/artificial-intelligence-regulatory-science-research-05142020-05142020 [https://perma.cc/C46J-CV45] (describing how the FDAs National Center for Toxicological Research utilizes AI in its regulatory work).

455. Kuemmel et al., *supra* note 428, at 21.

456. *Id.*

457. *See id.* at 24.

458. *See id.*

459. *Id.*

460. *Id.*

The authors apply the updated framework to two contexts of use In which clinical trials would evaluate drug safety and efficacy.[461] The first involves using a model to infer how an experimental drug should be dosed when coadministered with another drug with known metabolic effects.[462] The authors determine the decision influence to be high because model predictions would be used in lieu of clinical trials for most of the patient population.[463] The second context of use involves determining the optimal dose of the drug for children.[464] The authors determine the decision influence was low because clinical trials would be conducted for all patient populations.[465] Only the starting dose for the pediatric clinical trial would be determined by algorithmic predictions alone, and the final dose of the labeled product would be influenced by a combination of clinical and algorithmic data.[466]

The authors make several important observations and recommendations. They acknowledge that adopting the ASME credibility framework would require regulators and product manufacturers to change how they assess computer models.[467] Specifically, regulators must understand and accept that model risk determines the selection of credibility factors, and the more rigorous the factors, the higher their applicability to the intended use, and the higher the model credibility.[468] They suggest that a team of experts determine the appropriate level of rigor.[469] They also indicate that different people applying the framework might reach different conclusions, and, to alleviate this concern, they emphasize that this possibility exists whether or not a framework is adopted.[470] Finally, they advocate for public discussions regarding the framework involving multiple stakeholders.[471]

461. *Id.* at 25.
462. *Id.* at 23, 25.
463. *See id.* at 25.
464. *Id.* at 23, 25.
465. *Id.* at 24–25.
466. *Id.* at 25.
467. *See id.* (outlining new processes for verifying the accuracy and reliability of models).
468. *Id.* at 22, 24–25.
469. *Id.* at 25.
470. *Id.* at 27.
471. *Id.*

Other FDA scientists have attempted to clarify the types of information that can help establish model credibility.[472] They provide eight factors, such as the pedigree of input data and the quantification of uncertainty.[473] They also suggest creating a preassessment plan describing the methods for acquiring and analyzing data and including evidence to establish that data sources are of sufficient quality to establish model credibility.[474] The inclusion of these factors and recommendations adds new dimensions to the V&V 40 framework.

In late 2021, the FDA published draft guidance for assessing the credibility of computational models used to evaluate medical devices.[475] The draft incorporates suggestions made by agency scientists in their analysis of the V&V 40 framework.[476] It recommends that device manufacturers provide a credibility assessment plan and seek FDA feedback.[477] Additionally, the agency recommends inclusion of a credibility assessment report when submitting the results of modeling studies or simulated trials.[478] The report should state the evidence and rationale supporting model credibility.[479] However, the document provides no details on how to gather that evidence.[480]

Despite making useful recommendations, the draft guidance omits many prior suggestions from FDA staff. There is no mention of training data or the importance of evaluating its quality and pedigree.[481] The recommendation to adjust decision consequence based on the number of people potentially impacted by a model's adoption was also overlooked. Moreover, some sections of the draft guidance are likely to confuse industry stakeholders and the public. For instance, the FDA states that its proposed guidance "is not intended to apply to statistical or data-driven models such as machine learning or artificial intelligence."[482] However, the document mentions augmenting or replacing human trials with simulated trials consisting of virtual

472. *See* POPELAR ET AL., *supra* note 113.
473. *Id.*
474. *Id.*
475. FDA DRAFT GUIDANCE 2021, *supra* note 414, at 4–5.
476. *Id.*
477. *Id.* at 34.
478. *Id.* at 35.
479. *Id.*
480. *Id.* at 7.
481. *Id.*
482. *Id.* at 5.

patients,[483] which requires AI to be successful according to Morrison and other FDA scientists.[484] Moreover, the draft guidance discusses models used to estimate measurements such as fractional flow reserve.[485] Though not mentioned by name in the document, HeartFlow is a prominent example of such a model,[486] and it uses "an advanced form of AI called deep learning."[487] Consequently, readers may not understand whether the draft guidance applies to simulated trials or to HeartFlow and similar products that infer fractional flow research and other measurements. The document's omission of AI-based models is also surprising given their status among the FDA's strategic priorities.[488]

The draft guidance has other shortcomings. It applies only to models used to address medical devices and does not address models for evaluating drugs and biological products, for which there is only limited guidance that similarly excludes discussion of AI.[489] Moreover, the FDA frames the draft guidance as a risk-based framework for evaluating model credibility.[490] However, it defines risk narrowly in terms of harm to individual patients, trial participants, or healthcare providers, overlooking the broader impact to society and public health.[491] The document only emphasizes factors that support model credibility. It does not encourage manufacturers to analyze and acknowledge the shortcomings of their models. Finally, the draft omits discussion of proprietary models and the benefits of open-source code,

483. *Id.* at 5–6 (referencing in silico trials).

484. Morrison et al., *Advancing Regulatory Science*, *supra* note 234, at 8 (stating that achieving the same results as human trials with simulated trials "relies on statistical models, deep learning and artificial intelligence").

485. FDA DRAFT GUIDANCE 2021, *supra* note 414, at 21.

486. *See supra* notes 209–212 and accompanying text.

487. *Our Technology Core*, HEARTFLOW (Mar. 31, 2018), https://www.heartflow.com/heartflow-ffrct-analysis/article/our-technology-core [https://perma.cc/GN9Q-N2RS].

488. *See* Morrison, *Advancing Regulatory Science*, *supra* note 234, at 3 (emphasizing the role of data mining, machine learning, and deep learning in computational modeling).

489. FDA DRAFT GUIDANCE 2021, *supra* note 414, at 7; *see also* U.S. FOOD & DRUG ADMIN., POPULATION PHARMACOKINETICS 1 (2021) (providing guidance for regulatory submissions that rely on population-based pharmacokinetic models); U.S. FOOD & DRUG ADMIN., PHYSIOLOGICALLY BASED PHARMACOKINETIC ANALYSIS—FORMAT AND CONTENT 1 (2018) (providing guidance for regulatory submissions that rely on physiologically-based pharmacokinetic models).

490. FDA DRAFT GUIDANCE 2021, *supra* note 414, at 4–5.

491. *See id.* at 14–15.

methods, and data sets.[492] Instead of creating clear standards and promoting uniform regulatory submissions, which would increase transparency and efficiency, the draft's lack of specificity and discussion of non-proprietary models may promote further expansion of the existing patchwork of modeling approaches.

Though the draft guidance is intended for FDA staff as well as industry stakeholders, it omits discussion of staff utilization or interpretation of computer models. To help fill this gap, the following discussion applies the V&V 40 framework to agency decision-making. Specifically, it analyzes the FDA's adoption of molecular modeling to make scheduling recommendations. Though the V&V 40 framework was not intended to evaluate the credibility of AI-based models, its principles remain useful when applied to them. The following discussion highlights the need for expansion of existing credibility frameworks and FDA guidance.

B. Evaluating the Credibility of a Regulatory Model

The FDA's draft guidance recommends that medical device manufacturers use the V&V 40 framework to evaluate the credibility of their computer models. However, the agency makes no comparable recommendations to its staff and departments to evaluate the credibility of models they develop or adopt. This Section applies the V&V 40 framework to the FDA's PHASE methodology.

It should be acknowledged that using models to predict the harm potential of substances for scheduling purposes differs from using them to simulate patients and clinical trials to evaluate product safety and effectiveness. However, the principles remain are the same. In both contexts, models draw inferences from data, and biased training data will produce biased algorithms and outputs. Moreover, in both cases, inaccurate and biased predictions can harm people and compound past injustice.

In the context of substance regulation, the model influence would be high because the FDA substitutes model predictions for elements of the CSA's eight-factor analysis.[493] Agency staff might argue that the model influence is at most moderate because they can ignore the model

492. *See generally id.* (lacking discussion of proprietary models or open-source code, methods, and data sets).

493. *See* Gottlieb 2018, *supra* note 141; *see also* Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*, *supra* note 146, at 117.

predictions, and decision consequence is at most moderate because inaccurate prediction do not guarantee that kratom will be scheduled. However, FDA staff rely heavily on molecular models, which they systematically use to automate the evaluation of unregulated substances.[494] Moreover, using the rubric for model influence provided by FDA scientists, if no clinical trial data will be used to inform a regulatory decision, and a computer model provides the bulk of the actionable information, then the model influence should be high.[495]

The decision consequence would also be high because an incorrect decision from the FDA could result in a ban that would indefinitely remove a substance from the marketplace and inhibit scientific research.[496] Scheduling substances is a nuclear option that can cause significant harm.[497] As Giroir describes in his 2018 letter, prohibiting kratom could have caused "immediate adverse public health consequences" for millions of users, including intractable pain, the replacement of kratom with more harmful substances, and fatal overdose associated with consuming those substances.[498] These results represent the downstream public health impact, which should be considered when relying on models to make regulatory decisions. The number of affected individuals would be high because millions of people use kratom in the United States.

As explained by FDA scientists, the number of people affected by an incorrect conclusion affects the decision consequence.[499] Because millions of individuals across the United States could potentially be impacted by a kratom ban, the decision consequence of using models to determine kratom's regulatory status should increase.[500] Finally, because the model influence and decision consequence are high, the model risk should equal five.

494. *See* Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*, *supra* note 146, at 117.

495. *See* Kuemmel et al., *supra* note 428, at 21, 24.

496. *See id.* at 24.

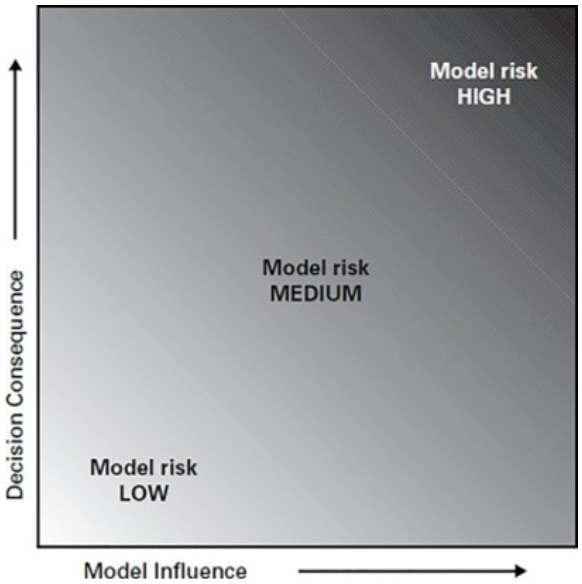
497. *See* Giroir 2018, *supra* note 183, at 3.

498. *Id.* at 3–4.

499. *See* Kuemmel et al., *supra* note 428, at 24.

500. *See* Giroir 2018, *supra* note 183, at 3.

Figure 3: Continuous Risk Map



The FDA might claim that the benefits of using molecular models are so great that their positive impact on public health justifies their use. For instance, adopting the model conserves agency resources. However, the V&V 40 framework does not account for the benefits of adopting computer models. Instead, as a risk-based framework, it focuses on the harms of drawing incorrect conclusions.[501]

Perhaps the most concerning aspect of the FDA's adoption of molecular modeling is the technology's poor applicability for the task of predicting whether substances are harmful.[502] In its PHASE methodology, the FDA uses models that simulate chemical structure and predict binding affinity.[503] As discussed previously, these variables are only two of many factors that determine drug effects.[504] The exclusion of other equally important variables, such as biased agonism, renders the FDA's models a poor choice for predicting harm and determining scheduling status. These shortcomings affect how

501. *Id.*

502. *See, e.g.*, Morrison et al., *supra* note 107, at 355 (finding a molecular model could not predict hemolysis).

503. Gottlieb 2018, *supra* note 141.

504. *See* Bloom, *supra* note 154; *see also* Ellis et al., *Assessing the Structural and Pharmacological Similarity of Newly Identified Drugs*, *supra* note 146, at 121.

applicable the models are to this context of use, which should have prevented their adoption. At the very least, with respect to credibility assessment, poor applicability should significantly increase the quality and quantity of factors required to establish credibility.

Other factors that the FDA should have considered before adopting molecular models include the opacity of the algorithms and whether they are proprietary or open source; the pedigree of the model's training and input data; who developed the models and whether they have conflicts of interest; whether the model's adoption was subjected to notice and comment; whether the question of interest could be answered using other means that are more transparent and produce higher quality data; and whether the models' adoption could compound past injustice.

The logic of the FDA's molecular modeling software is opaque largely because it relies on proprietary technology developed by a private drug company, which also creates conflicts of interest. The models' training data contain numerous sources of potential bias such as patent documents and academic articles drawn from a narrow range of sources. Though the FDA's adoption of molecular modeling arguably constitutes a rule, or at the very least industry guidance, it was not published or subjected to notice and comment requirements. Other credible sources of information, such as animal studies and scientific reports, were available to inform scheduling decisions, yet the FDA's recommendations were largely based on model outputs. Scheduling kratom would potentially impact millions of people who consume kratom, threatening public health, and scientific progress would be prevented by increased restrictions on kratom research. Because many users of kratom belong to historically marginalized communities, adopting molecular modeling in this context could compound past injustice.[505] Considering this analysis, the FDA's molecular models lack credibility, and their use in drug scheduling should be reevaluated.

505. *See* Kirsten E. Smith, Kelly Dunn, Oliver Grundmann, Albert Garcia-Romeu, Jeffrey M. Rogers, Marc T. Swogger & David H. Epstein, *Social, Psychological, and Substance Use Characteristics of U.S. Adults Who Use Kratom: Initial Findings from an Online, Crowdsourced Study*, EXPERIMENTAL & CLINICAL PSYCHOPHARMACOLOGY 2, 6 (Nov. 4, 2021), https://psycnet.apa.org/record/2022-00103-001 [https://perma.cc/G45B-DP7F] (reporting that substance use and mental health conditions are common in people who use kratom, and social, psychological, and health indicators are lower in groups that consume kratom compared to those who do not).

The following Section makes recommendations for improving existing credibility frameworks to increase their utility in all public health contexts.

C. Recommendations

The FDA should update its practices and guidance to account for the full impact of model design, training, and deployment on model credibility and public health. Drawing from the field of AI ethics and previous suggestions from FDA staff, this Section makes recommendations to enhance the utility of credibility frameworks like the V&V 40. These recommendations should be followed whenever the FDA adopts or interprets models to make regulatory decisions. They are equally applicable to research institutions, product manufacturers, and other stakeholders.

1. *Form Independent Boards to Assess Model Credibility.* Research institutions, product manufacturers, and public health agencies should form independent boards to assess model risk and credibility. In some settings, institutional review boards could fill this role.[506] However, they may require modification and training to meet the challenges of assessing models credibility.[507] To avoid conflicts of interest, review boards should be independent, and their members should lack ties to researchers, regulators, and the products being evaluated.[508] To promote equity and consideration of the full spectrum of risks, boards should be diverse and include representatives from communities likely to be impacted by the models being evaluated.[509] To avoid playing only a symbolic role, review boards should have authority to require that changes be made to models, their training data, and their range and mode of application.[510]

Models should neither be utilized outside the range of tasks for which they were designed nor used to analyze information that differs

506. *See generally* Phoebe Friesen, Rachel Douglas-Jones, Mason Marks, Robin Pierce, Katherine Fletcher, Abhishek Mishra, Jessica Lorimer, Carissa Véliz, Nina Hallowell, Mackenzie Graham, Mei Sum Chan, Huw Davies & Taj Sallamuddin, *Governing AI-Driven Health Research: Are IRBs Up to the Task?*, 43 ETHICS & HUMAN RES. 35 (2021) (discussing the AI industry's use of internal review boards and their efficacy).

507. *See id.* at 37–38.

508. *Id.* at 39.

509. *Id.*

510. *Id.*

significantly from their training data. When models are deployed near the limits of their optimal context of use, boards should require stronger evidence of credibility, and they should deem models uncredible if applied outside those bounds.

2. *Consider the Full Spectrum of Algorithmic Harms.* The FDA is not accustomed to evaluating algorithmic harms.[511] Its risk analysis framework is outdated and focuses too narrowly on harms associated with traditional medical products, such as physical injury from malfunctioning devices.[512] Even the FDA's recent guidance on evaluating model credibility, which the agency describes as a "risk-based framework," takes an antiquated approach by defining risk only in terms of harm to individual patients or research participants.[513]

When adopting computer models and interpreting or reporting their outputs, FDA staff and industry stakeholders should consider the full spectrum of potential harms, including the potential for downstream public health effects and compounding prior injustice. For instance, they should assess whether regulatory decisions informed by computer models might impact historically marginalized communities. If the potential to compound prior injustice is high, then model risk should be considered high, and the model should not be adopted. If the potential to compound prior injustice is medium, then stronger evidence should be required to establish model credibility compared to models for which the potential to compound prior injustice is low. Similarly, stronger evidence of credibility should be required when the size or vulnerability of the impacted population grows.

When the FDA adopts computer models to assess product safety, or to support regulatory decision-making, it should complete an algorithmic impact assessment and publish the results for public comment.[514] Moreover, when industry stakeholders submit model data

511. *See* U.S. FOOD & DRUG ADMIN., ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN SOFTWARE AS A MEDICAL DEVICE 1 (2021), https://www.fda.gov/media/145022/download [https://perma.cc/8J34-R5QJ] (describing the FDA's plan for regulating AI, which has not been finalized).

512. *See* Marks, *supra* note 124, at 1044–45 (describing the FDA's outdated risk analysis framework).

513. FDA DRAFT GUIDANCE 2021, *supra* note 414, at 15.

514. DILLON REISMAN, JASON SCHULTZ, KATE CRAWFORD & MEREDITH WHITTAKER, ALGORITHMIC IMPACT ASSESSMENTS: A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY ACCOUNTABILITY 15 (2018) (describing the ability of algorithmic impact assessments to increase an agency's internal capacity to better understand and explain the impacts of a system before implementing it).

to the FDA, they should include the results of an internal algorithmic impact assessment. Product manufacturers should acknowledge and disclose the limitations of their models, and a lack of disclosure should negatively impact model credibility. The FDA should update current guidance to inform stakeholders of these requirements.

3. *Assess Institutional and Algorithmic Opacity.* Like the PTO, the FDA should prioritize in-house development of computer models. Doing so will bolster agency expertise, build public trust, and enhance the legitimacy of agency decisions. Congress should fund the FDA's development of nonproprietary models that, like the virtual family of anatomical models, are made freely available to researchers and industry stakeholders.

When adopting models created by industry is unavoidable, or when the FDA must interpret their outputs, the agency should consider whether the models and their training data are proprietary or open source. If a model's source code or training data is proprietary, significantly stronger evidence should be required to establish its credibility. Manufacturers and the FDA should disclose conflicts of interest and acknowledge how reliance on proprietary models might bias results and mislead regulators and the public.

To promote trust and the legitimacy of FDA decisions, the agency should treat its adoption of computer models as rulemaking subject to notice-and-comment requirements. At the very least, the agency should follow its own good guidance practices and treat its adoption of models as significant (Level 1) guidance requiring publication in the Federal Register. If the FDA's adoption of a computer model has not been published, then far stronger evidence should be required to establish its credibility.

4. *Analyze the Pedigree of Training Data and Incentivize Disclosure.* The FDA and industry stakeholders should assess the pedigree of training data by analyzing the diversity and reliability of data sources, the potential for bias, and the extent to which the appropriateness and accuracy of data sets can be verified. All stakeholders should acknowledge which variables have been omitted from training data and anticipate the impact on model predictions. Low-quality or potentially biased training data should decrease credibility, and high-quality data sets that are well justified and appropriate for the population to be analyzed should enhance credibility. The sources and quality of training data should be disclosed

with regulatory submissions, and a lack of disclosure should negatively impact credibility assessments. The pedigree of training data should be conveyed to healthcare providers and the public through product labeling.

5. *Identify Alternate Methods of Answering Questions of Interest.* Until effective methods for evaluating model credibility become available and are standardized, agency and industry stakeholders should identify alternate approaches to answering questions of interest, and the adoption of computer models should be reserved for situations in which other methods are unacceptably dangerous or ineffective. Safety and public health should be prioritized over cost effectiveness and the desire to innovate, and when alternate methods have not been assessed and disclosed, stronger evidence should be required to establish model credibility.

The FDA should incorporate the above recommendations into future guidance. Taking a more holistic approach to credibility assessment will build trust in computer models and minimize algorithmic risks, which include harms broader than physical injury to patients and research subjects. The FDA defines credibility as a model's ability to elicit trust in its predictions. However, the agency should not overlook its role in building trust. How the FDA adopts and interprets computer models affects public faith in the technology and its own credibility. When respect for the FDA may be waning, the agency should emphasize transparency and increase evidentiary standards instead of letting them lapse.

CONCLUSION

Throughout the twentieth century, Congress enhanced the FDA's powers in response to a series of public health disasters. With each increase in responsibility, the FDA also enhanced its evidentiary standards, which earned public trust and improved the agency's reputation. However, in the past few decades, the FDA's evidentiary standards have decreased, and the agency has made a series of questionable decisions. Its adoption of computer models and simulations may exacerbate this trend. In addition to relying on surrogate endpoints in lieu of direct evidence of symptomatic improvement, the FDA and its industry partners increasingly rely on computer models and simulations to augment or replace data from clinical trials. Though the FDA has drafted guidance to help agency staff and industry partners assess the credibility of computer models, it

does not address how models should inform FDA decision making. Further, it overlooks models that rely on AI, the risks of algorithmic bias, and the potential for computer models to compound past injustice.

When adopting proprietary models to make regulatory decisions, the FDA may effectively delegate legislative power to private parties. This practice conceals the logic of FDA decisions from public view and arguably undermines the legitimacy of agency decisions. Until the Supreme Court revives the nondelegation doctrine and updates it to address the democratic risks of algorithmic governance, administrative law may be of little help. In the meantime, before investing further in computer models, the FDA should enhance existing frameworks for evaluating model credibility with principles of AI ethics to account for algorithmic bias and harms. It should follow the example of agencies that exhibit greater transparency and public participation when adopting computer models, and it should require similar transparency from its industry partners.

AI has the potential to streamline aspects of FDA approval, substitute for dangerous procedures or experiments, and conserve healthcare resources. However, it can also contribute to bad regulatory decisions, harm marginalized communities, and impede scientific progress. In the twentieth century, it took a series of public health disasters to enhance the FDA's evidentiary standards. In the twenty-first, it should not take an algorithmic public health disaster to implement appropriate safeguards.