*Harvard Journal of Law & Technology*
Volume 38, Number 2 Symposium 2024

# THE REGULATION OF POLYGENIC RISK SCORES

*Jin K. Park* & I. Glenn Cohen***

## ABSTRACT

Polygenic risk scores ("PRSs") provide genome-wide estimates of disease risk by aggregating the effects of thousands of genetic variants across the genome. These scores are the subject of immense scientific interest as research tools and more recently as clinical instruments that may allow for physicians to stratify populations based on underlying genetic predisposition, or to tailor therapeutic interventions based on their needs and likelihood of benefit. While their status as research tools has long-been recognized, these scores are now undergoing clinical trials, increasing the evidence base for their use in clinical settings. These scores have also entered the consumer market, prompting industry experts to call on greater regulatory oversight. However, in part due to the speed of these developments, the legal literature has failed to comprehensively assess the nature of these scores, and whether they differ fundamentally from previous forms of genetic scoring which have been regulated by the complex (yet familiar) regulatory regime for genetic testing. This Article fills this gap in the literature by comparing the state-of-the-art methodological tools used to generate these scores with familiar forms of genetic testing (e.g., IVDs and LDTs). We identify four dimensions that make PRS distinct from previous genetic testing regimes — (1) the underlying method of assessing genetic risk; (2) an evolving evidence base; (3) lack of consensus on methodology; (4) diversity of device functions that PRSs may apply to.

Taking these insights in concert, this Article also offers several principles for regulatory design as it relates to PRSs. These principles include the need for a unified approach across all devices that

---

* Research Fellow, Petrie-Flom Center for Health Law Policy, Biotechnology & Bioethics. MPhil, University of Oxford (2024). JD/MD Candidate, Harvard Medical School and Yale Law School (Expected 2027).

** Deputy Dean, James A. Attwood and Leslie Williams Professor of Law, Harvard Law School. Faculty Director, Petrie-Flom Center for Health Law Policy, Biotechnology & Bioethics.

We thank participants at the 2024 JOLT Symposium, Susannah Baruch, and Anya Prince for comments. We thank Katie Gu for superb research assistance. Cohen and Park's research for this paper was supported by a Novo Nordisk Foundation Grant for a scientifically independent International Collaborative Bioscience Innovation & Law Programme (Inter-CeBIL programme - grant no. NNF23SA0087056). The symposium itself was supported in part by the National Institute of Mental Health and the National Human Genome Research Institute of the National Institutes of Health, grants R01MH128676 and R01HG012402, respectively. Cohen is a member of the ethics advisory board for Illumina and the Bayer Bioethics Council but speaks only in his personal capacity.

incorporate PRSs, the value of taking a risk-based framework, and drawing lessons from AI/ML regulation. Ultimately, while the existing risk-based device framework will serve as a stopgap for the most clinically impactful use cases (and those that pose the most risk to patients and the public), PRSs and other novel technologies may evince the need for updates to the authorities granted to the existing regulatory regime to balance scientific innovation with the public interest.

TABLE OF CONTENTS



## I. INTRODUCTION

Polygenic risk scores are genetic estimates of disease risk that aggregate the contributions of genetic variants across the entire genome. Their use has proliferated over the past decade, with no signs that they will abate in the near future. Orchid is a company that promises to "[m]itigate more risks with the world's most advanced whole genome screening for embryos" and offers "genetic predisposition screening" for a long list of diseases and conditions including Alzheimer's Disease, schizophrenia, bipolar disorder, prostate cancer, celiac disease, and coronary heart disease.[1] Third-party interpretation software such as ADNTRO claims to help consumers understand their "genetic predisposition to alcohol consumption."[2] Other firms offer consumers polygenic risk scores to predict traits outside of traditional medical practice such as educational attainment.[3] But who does and who should regulate

1. *Orchid Home Page*, ORCHID HEALTH, https://www.orchidhealth.com/ [https://perma.cc/CES5-ULQG].

2. Jacob S. Sherkow, Jin K. Park & Christine Y. Lu, *Regulating Direct-to-Consumer Polygenic Risk Score*s, 330 JAMA 691, 691–92 (2023).

3. *See generally* KATHRYN PAIGE HARDEN, THE GENETIC LOTTERY (2021) (insisting upon the importance of understanding genetic factors in driving demonstrable differences in various outcomes including educational attainment and life-expectancy).

polygenic risk scores ("PRSs")?[4] This paper offers a first pass at answering these questions. Part II explains what PRSs are, how they are calculated, and their use cases. Part III looks at the Food and Drug Administration's ("FDA") authority to regulate genetic testing and its suitability to regulate PRSs. Part IV explains the features of PRSs that distinguish them from other genetic tests — namely their novel estimates of genetic risk, lack of expert consensus on their construction, and the diversity of their functionality within a device — and why these features matter for fine-tuning regulatory design. Finally, Part V offers principles for a regulatory framework for PRS-based devices.

## II. WHAT IS A POLYGENIC RISK SCORE?

After publication of the first full sequence of the human genome, the scientific community had high hopes for what was at the time dubbed "the language with which God created life."[5] Although it was known long before the Human Genome Project ("HGP") that many complex traits and diseases are polygenic, HGP has enabled interrogation of *which loci* matter for *which trait* and *in what combination.*[6]

A typical human genome differs from the reference genome at approximately 4.1 million to 5.0 million sites.[7] While many of these changes are individually inconsequential, other changes can mean the difference between life and death. One of the most important forms of genetic variation uncovered by sequencing of the human genome is the single nucleotide polymorphism ("SNP"), which are sites in the genome that differ by just a single base pair.[8] SNPs are the most common form of human genetic variation, and many of them are shared across

4. Hereafter, we use the acronym PRS to indicate the broader category of methods to make genome-wide estimates of disease or trait heritability, though other terms such as "PGS" are also common. Here, we use "PRS" to also encapsulate scores that incorporate rare mutations and private alleles. For discussion, see Iftikhar J. Kullo, Cathryn M. Lewis, Michael Inouye, Alicia R. Martin, Samuli Ripatti & Nilanjan Chatterjee, *Polygenic Scores in Biomedical Research*, 23 NATURE REVS. GENETICS 524, 525–26 (2022).

5. *Text of the White House Statements on the Human Genome Project*, N.Y. TIMES (June 27, 2000), https://archive.nytimes.com/www.nytimes.com/library/national/science/062700sci-genome-text.html [https://perma.cc/B35Q-BY2M].

6. According to some critics, the HGP's ability to improve human health and wellbeing with these discoveries has not been self-executing, and the investment of time and resources it has entailed has redirected the focus of biomedicine to genetic (and not structural or social) causes of ill health. *See generally* JAMES TABERY, TYRANNY OF THE GENE: PERSONALIZED MEDICINE AND ITS THREAT TO PUBLIC HEALTH (2023). *See also* Teneille R. Brown, *The Opposite of Empowering*, 38 HARV. J.L. & TECH. 501, 512–13 (2024) ("Rather than adopting evidence-based public health inventions, in the United States funding agencies have disproportionately financed biomedical market solutions . . . .").

7. The 1000 Genomes Project Consortium, *A Global Reference for Human Genetic Variation*, 526 NATURE 68, 68 (October 2015).

8. *SNP*, NAT. CANCER INSTITUTE, https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/snp [https://perma.cc/B947-A8A8].

various human populations.[9] Due to their ubiquity and unique features, SNPs have been extremely important as research tools, and have been used to elucidate human evolutionary dynamics, the genetic architecture of common diseases, and most relevant for our purposes — the proportion of disease risk that can be attributed to one's genome.[10]

PRSs help to make sense of the significance of SNPs (and other genetic variants) because conceptually, a PRS is simply a weighted sum of multiple disease-related or trait-related genetic variants. In their most familiar construction,[11] these scores are constructed using SNPs and their associated summary statistics drawn from genome-wide association studies ("GWASs"), which are large case-control studies that seek to determine variants associated with a disease or trait.[12] At a general level, the construction of a PRS involves two sets of data.[13] The first (GWAS data) includes all the associations for each SNP and their effect size.[14] These data have become increasingly available in public repositories.[15] The second (target data) consists of genome sequence(s) from the individual(s) for whom the PRS is being calculated.[16] There are important data corrections and quality control measures at that front-end that are standard in the field,[17] but the principle is the same — a PRS allows an investigator to sum the effect sizes of all the variants from an individual's genome by using an index derived from population-level studies. While there are complex nuances underlying these scores, the PRS allows investigators to aggregate the contributions of multiple genomic loci (with varying effect sizes) to the disease/trait of interest.[18]

---

9. Adam Auton, Katarzyna Bryc, Adam R. Boyko, Kirk E. Lohmueller, John Novembre, Andy Reynolds et al., *Global Distribution of Genomic Diversity Underscores Rich Complex History of Continental Human Populations,* 19 GENOME RSCH. 795, 795 (2009).

10. Emil Uffelman, Qin Qin Huang, Nchangwi Syntia Munung, Jantina De Vries, Yukinori Okada, Alicia R. Martin et al., *Genome-Wide Association Studies*, 1 NAT. REVS. METHODS PRIMERS, no. 59, 2021, at 1.

11. There are increasingly efforts to integrate rare variants into PRS generation and reporting. *See, e.g.*, Craig Smail, Nicole M. Ferraro, Qin Hui, Matthew G. Durrant, Matthew Aguirre, Yosuke Tanigawa et al., *Integration of Rare Expression Outlier-Associated Variants Improves Polygenic Risk Prediction*, 109 AM. J. HUM. GENETICS 1055 (2022).

12. *Id.*

13. *See* Shing Wan Choi, Timothy Shin-Heng Mak & Paul F. O'Reilly, *Tutorial: A Guide to Performing Polygenic Risk Score Analyses*, 15 NAT. PROTOCOLS 2759, 2760 (2020).

14. *Id.*

15. *Cf.* Tim Beck, Tom Shorter & Anthony J. Brookes, *GWAS Central is the World's Most Comprehensive Openly Accessible Repository of Summary-Level GWAS Association Information, Providing Over 70 Million P-Values for Over 3800 Studies Investigating Over 1400 Unique Phenotypes*, 48 DATABASE ISSUE NUCLEIC ACIDS RSCH. D933, D933 (2020). Incidentally, this helps to explain in part the rise in so-called third-party interpretation firms that will provide PRS estimates for diseases/traits of interest if provided a raw genome sequence file, without having to do the sequencing themselves.

16. Choi et al., *supra* note 13, at 2760.

17. Choi et al., *supra* note 13, at 2761–62.

18. There are also important questions about how SNPs in GWASs should be aggregated — for example, an additive model would simply sum the small effect sizes of SNPs

The utility of a polygenic score turns on several variables, such as genetic ancestry, age, sex, and the genetic architecture of the disease or trait of interest.[19] The mere fact that a SNP is statistically associated with a given phenotype in a GWAS does not suggest a straightforward causal relationship. A SNP may be correlated with a trait for a variety of reasons, including its relationship to another effect-modifying variant or through its correlation with environmental features that are not randomly distributed in human subpopulations.[20] For a trait such as vertical height, some geneticists believe that as high as forty to fifty percent of the variance in the heritability of height in human populations can be accounted for by aggregated SNPs.[21] In other words, sequencing more genomes won't give us much more information about the *genetic* determinants of height than we already have.[22] For other traits in which the heritability estimate is smaller, more difficult to measure, or is itself contested, the utility of PRS in these cases is also questionable.[23]

These basic facts about PRS are important from a regulatory standpoint because these facts help distinguish PRS from, say, a cancer gene panel or a 23andMe ancestry test. Traditional genetic testing regulated by agencies such as the FDA are evaluated based on demonstrable claims made about the presence (or absence) of a set of alleles and their relationship to human health or disease.[24] But a PRS issues a unitless score for an individual's genetic liability for a given disease or trait based on hundreds and thousands of SNPs whose relative contribution to a complex trait has been inferred through GWAS of extremely large populations.

Consider the use of PRS in the United States to score embryos for preimplantation genetic diagnosis. Although some genetic testing of embryos has been commonplace, the ability to comprehensively score an embryo on several traits of interest, including diabetes, height, and

---

without taking into account epistatic or dominance effects. *But see* Guy Sella & Nicholas Barton, *Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies,* 20 ANN. REVS. GENOMICS & HUM. GENETICS 461, 467 (2019) (discussing why an additive model is often a "sensible starting point" for most discussions).

19. Kangcheng Hou, Ziqi Xu, Yi Ding, Ravi Mandla, Zhuozheng Shi, Kristin Boulier, Arbel Harpak et al., *Calibrated Prediction Intervals for Polygenic Scores Across Diverse Contexts*, 56 NATURE GENETICS 1386, 1386 (2024).

20. K. Paige Harden, *'Reports of My Death Were Greatly Exaggerated': Behavior Genetics in the Postgenomic Era*, 72 ANN. REV. PSYCH. 37, 45 (2021).

21. Loïc Yengo, Sailaja Vedantam, Eirini Marouli, Julia Sidorenko, Eric Martell, Saori Sakaue et al., *A Saturated Map of Common Genetic Variants Associated with Human Height*, 610 NATURE 704, 704 (2022).

22. *Id.* at 704–12.

23. *See* Harden, *supra* note 20, at 47–49; David St Clair & Bing Lang, *Schizophrenia: A Classic Battle Ground of Nature Versus Nurture Debate*, 66 SCIENCE BULL. 1037, 1040–43 (2021). *See generally* Peter M. Visscher, William G. Hill & Naomi R. Wray, *Heritability in the Genomics Era — Concepts and Misconceptions*, 9 NATURE REVS. GENETICS 255 (2008).

24. 21 U.S.C. § 360c(a)(1)(B) (describing the requirements of the FDA).

even educational attainment is new.[25] Companies such as Genomic Prediction and Orchid are offering Polygenic Preimplantation Genetic Testing ("PGT-P") as the next iteration of preimplantation genetic testing, in a context where the demand for in vitro fertilization ("IVF") is expected to swell in the next decade.[26]

Another possible use case will be "population screening to determine individual susceptibility to common disorders such as heart disease, diabetes, and cancer."[27] Studies demonstrate PRS can identify individuals who are at the tail-end of the population distribution for heritable disease risk for common conditions such as diabetes and heart disease.[28] As Nilanjan Chatterjee put it, "PRSs have the unique advantage that they can be applied early in life to simultaneously assess long-term risk of many individual diseases and conditions[.]"[29] These scores can be just as informative as monogenic mutations often found in syndromic conditions.[30] One of the only forms of genetic testing that the US Preventive Services Task Force recommends (as a Grade B recommendation) is BRCA testing for "women with a personal or family history of breast, ovarian, tubal, or peritoneal cancer or an ancestry associated with susceptibility 1 and 2 [BRCA1/2] gene mutations."[31] Polygenic scores are currently being validated for breast cancer, so it may be that population-wide genetic risk stratification (short of whole-

25. *See, e.g.*, Patrick Turley, Michelle N. Meyer, Nancy Wang, David Cesarini, Evelynn Hammonds, Alicia R. Martin et al., *Problems with Using Polygenic Scores to Select Embryos*, 385 NEW ENG. J. MED. 78, 78–79 (2021); Laurent C. A. M. Tellier, Jennifer Eccles, Nathan R. Treff, Louis Lello, Simon Fishel & Stephen Hsu, *Embryo Screening for Polygenic Disease Risk: Recent Advances and Ethical Considerations*, 12 GENES, July 21, 2021, at 6 ("We have described three new technologies that are already making a significant impact on assisted human reproduction (IVF): polygenic risk scores, precision genotyping of embryos, and genomic indices . . .").

26. *Cf.* Max Kozlov, *The Controversial Embryo Tests that Promise a Better Baby*, 609 NATURE 668 (2022).

27. Muin J. Khoury, Linda L. McCabe & Edward R.B. McCabe, *Population Screening in the Age of Genomic Medicine*, 348 NEW ENG. J. MED. 50, 50 (2003).

28. *See, e.g.*, Aniruddh P. Patel, Minxian Wang, Yunfeng Ruan, Satoshi Koyama, Shoa L. Clarke, Xiong Yang et al., *A Multi-Ancestry Polygenic Risk Score Improves Risk Prediction for Coronary Artery Disease*, 29 NATURE MED. 1793, 1801 (2024).

29. Kullo et al., *supra* note 4, at 525.

30. Amit V. Khera, Mark Chaffin, Krishna G. Aragam, Mary E. Haas, Carolina Roselli, Seung Hoan Choi et al., *Genome-Wide Polygenic Scores for Common Diseases Identify Individuals with Risk Equivalent to Monogenic Mutations*, 50 NATURE GENETICS 1219, 1219 (2018).

31. United States Preventive Services Task Force, *Final Recommendation Statement. BRCA-Related Cancer: Risk Assessment, Genetic Counseling, and Genetic Testing*, USPSTF (Aug. 20, 2019), https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/brca-related-cancer-risk-assessment-genetic-counseling-and-genetic-testing [https://perma.cc/3WRN-QZFR].

genome sequencing) in the manner contemplated first by Khoury et al. as early as 2003 may soon be on the horizon.[32]

Many population geneticists insist that PRS as disease prediction or stratification tools are not ready for primetime.[33] To forestall misuse of these scores, some have sought to consolidate and standardize the techniques for generating PRSs.[34] Most notably, Wand et al. have set out standards — called the "Polygenic Risk Score Reporting Standards" ("PRS-RS") — for the generation and reporting of PRSs, to make these scores workable for those "seeking regulatory approval of the PRS as a clinical test."[35]

## III. FDA'S AUTHORITY TO REGULATE GENETIC TESTS AND THE PROSPECTS FOR POLYGENIC RISK SCORE REGULATION

### *A. FDA's Regulation of In Vitro Diagnostics, and Especially Laboratory Developed Tests ("LDTs")*

The FDA's authority to regulate medical devices is traceable primarily to the 1976 Medical Device Amendments to the Food, Drug, and Cosmetic Act ("FDCA") which defines a "medical device" broadly as an instrument "intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or intended to affect the structure or any function of the body of man or other animals."[36] This broad definition also encompasses in vitro devices ("IVDs"), which are devices "intended for use in the collection, preparation, and examination of specimens taken from the human body."[37] Genetic tests are clear examples of IVDs.

The FDA classifies medical devices into three categories based on the description of the device in question as well as its intended use.

---

32. Padma Sheila Rajagopal, Sarah Nielsen & Olufunmilayo I. Olopade, *USPSTF Recommendations for BRCA1 and BRCA2 Testing in the Context of a Transformative National Cancer Control Plan*, JAMA NETWORK OPEN, Aug. 2019, at 1; Khoury et al., *supra* note 27, at 50.

33. *See, e.g.*, Alex Polyakov, David J. Amor, Julian Savulescu, Christopher Gyngell, Ektoras X. Georgiou, Vanessa Ross et al., *Polygenic Risk Score for Embryo Selection — Not Ready for Prime Time*, 37 HUM. REPROD. 2229, 2233 ("[PRSs] are in the early stages of development and clinical applications are currently limited."); Polygenic Risk Score Task Force of the International Common Disease Alliance, *Responsible Use of Polygenic Risk Scores in the Clinic: Potential Benefits, Risks, and Gaps*, 27 NATURE MED. 1876, 1880 (2021) ("Although there is largely a consensus that PRS should be used alongside other informative non-genetic risk factors, gaps remain in determining precisely how this should be done.").

34. *See generally* Hannah Wand, Samuel A. Lambert, Cecelia Tamburro, Michel A. Iacocca, Jack W. O'Sullivan, Catherine Sillari et al., *Improving Reporting Standards for Polygenic Scores in Risk Prediction Studies*, 591 NATURE 211 (2021).

35. *Id.* at 217.

36. 21 U.S.C. § 321(h)(1)(B).

37. 21 C.F.R. § 809.3 (2024).

These classifications correspond to varying degrees of regulatory controls intended to ensure safety and effectiveness of medical devices on the market. Class I devices are low-risk devices that are generally exempt from premarket review.[38] Class II devices are moderate-risk devices that undergo review most commonly through the 510(k) pathway that requires manufacturers to prove that the device in question is "substantially equivalent" to a device already approved for marketing by the FDA — a so-called "predicate device."[39] In 2015, when the FDA first approved 23andMe's direct-to-consumer genetic testing platform, it did so as a Class II medical device, and applied special controls to these devices (i.e., which applies warnings on labeling, imposes accuracy and reproducibility requirements, etc.).[40] Special controls allow the FDA to fine-tune regulatory requirements for certain devices, such as the "promulgation of performance standards, postmarket surveillance, patient registries, development and dissemination of guidelines."[41] Class III devices are high-risk, and generally require a rigorous premarket application ("PMA"), given the vital role these devices often play in sustaining life (e.g., defibrillators, artificial heart valve, etc.).[42]

In practice, the regulation of genetic tests by FDA has been implemented in patchwork fashion, turning on the components of the genetic testing. Some components of genetic tests such as primers and ligands are defined by statute as "analyte specific reagents" which as a category are regulated more stringently than general purpose reagents.[43] These reagents are mostly considered Class I devices, and thus exempt from premarket notification requirements, though there may be general manufacturing and record-keeping requirements.[44] In addition to the regulation of certain components of genetic testing, certain genetic tests with clear clinical validity — for example, a genetic test to detect the presence of a specific variant in a specific locus of a gene that encodes

38. 21 U.S.C. § 360c(a)(1)(A).

39. 21 U.S.C. § 360(o)(1)(A).

40. This was after the FDA sent cease-and-desist letters to 23andMe as well as several other direct-to-consumer testing firms. *See* U.S. FOOD & DRUG ADMIN., WARNING LETTER RE: PERSONAL GENOME SERVICE (2013), https://www.fdanews.com/ext/resources/files/12/12-02-13-23andme.pdf [https://perma.cc/93XX-VZN2]. For discussion, see Kayte Spector-Bagdady & Elizabeth R. Pike, *Consuming Genomics: Regulating Direct-to-Consumer Genetic and Genomic Information*, 92 NEB. L. REV. 677, 704–17 (2014). *See generally* Robert C. Green & Nita A. Farahany, *Regulation: The FDA Is Overcautious on Consumer Genomics*, 505 NATURE 286, 286–87 (2014); George J. Annas & Sherman Elias, *23andMe and the FDA*, 370 NEW ENG. J. MED. 985, 985–88 (2014).

41. 21 U.S.C. § 360c(a)(1)(B).

42. 21 U.S.C. § 360c(a)(1)(C).

43. 21 C.F.R. § 864.4020 (2024); Gail H. Javitt, *In Search of a Coherent Framework: Options for FDA Oversight of Genetic Tests*, 62 FOOD & DRUG L.J. 617, 620–21 (2007).

44. *See* Javitt, *supra* note 43, at 620–21.

a clotting factor — have enjoyed FDA clearance for marketing as a Class II device.[45]

In fact, though, most genetic tests in use in clinical settings have not been approved or cleared by the FDA at all, since they have been treated as so-called "laboratory-developed tests" ("LDTs"). FDA defines an LDT as an "IVD that is intended for clinical use and that is designed, manufactured, and used within a single laboratory that is certified under the Clinical Laboratory Improvement Amendments of 1988 (CLIA)" which set federal standards for facilities that perform clinical testing.[46] Until October 2023, the FDA historically exercised enforcement discretion, choosing not to require testing and approval for LDTs developed in CLIA-certified laboratories and used only locally,[47] relying on CLIA review of those tests to fill in the gaps left by the risk-based framework. The CLIA regime, however, has long-been criticized as problematic for at least two reasons. First, CLIA generally does not assess clinical validly of tests, but only its analytical validity.[48] Thus, while CLIA's mandate clearly enables the Center for Medicare & Medicaid Services ("CMS") to ensure that genetic tests accurately report genomic sequences — literally, the correct sequence of As, Ts, Cs, and Gs in a given sequencing read — as is well-known its mandate does not extend to ensuring the clinical significance of these sequences.[49] Second, while CLIA has allowed CMS to impose additional requirements for laboratories performing tests in "specialty areas" due to their high complexity (e.g., virology, toxicology, cytology, etc.), CLIA did not specify a specialty area for genetic testing (with the exception of some cytogenetics laboratories), in part due to the lack of widely accepted definition of a genetic test and the inability to prove clinical validity.[50]

This risk-based framework has allowed several genetic tests to pass regulatory muster. One of the first genetic test kits approved for marketing by the FDA was for genotyping a single point mutation in the Factor V gene for diagnosis of patients with thrombophilia, with these kits being classified as Class II devices.[51] The FDA has also cleared certain direct-to-consumer ("DTC") tests for direct clinical use —

45. U.S. FOOD & DRUG ADMIN., EVALUATION OF AUTOMATIC CLASS III DESIGNATION FOR DEN160026, at 59 (2017), https://www.accessdata.fda.gov/cdrh_docs/reviews/den160026.pdf [https://perma.cc/H9UG-7HL9].

46. Medical Devices; Laboratory Developed Tests, 88 Fed. Reg. 68006, 68009 (rule in effect July 5, 2024) (codified at 21 C.F.R. § 809.3)).

47. *Id.* at 68006–07.

48. Barbara J. Evans & Ellen Wright Clayton, *Deadly Delay: The FDA's Role in America's COVID-Testing Debacle*, 130 YALE L.J.F. 78, 89–90 (2020).

49. *Id.*

50. Javitt, *supra* note 43, at 624 n.60.

51. *Id.* at 629.

examples include 23andMe's tests for *CYP2C19* and *SLCO1B1*.[52] Even FDA's initial clearance of Illumina's next-generation sequencing machine is based on comparing the performance of the sequencer to a panel of variants identified by expert and professional organizations (e.g., ACMG and ACOG).[53] These genetic tests all are designed around the detection of variants of known clinical significance, enabling an agency to assess their ability to reproducibly return results that may be understood against a backdrop of clear professional and scientific standards.[54] But assessing the clinical validity of genetic tests for variants whose function and disease-significance is unknown is problematic. In medical genetics, these variants are often called "variants of unknown significance" ("VUS") in five variant categorization framework proposed by professional societies such as the American College of Medical Genetics and Genomics ("ACMG").[55] As Barbara Evans has argued, in contrast to assessing the clinical validity of monogenic tests, the FDA's risk-based framework and post-market review is unlikely to be sufficient to assess the clinical validity of whole-genome sequencing concerns because scientific understanding of the significance of variants is far from static.[56]

---

52. U.S. FOOD & DRUG ADMIN., EVALUATION OF AUTOMATIC CLASS III DESIGNATION FOR THE 23ANDME PERSONAL GENOME SERVICE PHARMACOGENETIC REPORTS 1, 16 (2018), https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN180028.pdf [https://perma.cc/3UBG-4UNW].

53. U.S. FOOD & DRUG ADMIN., 510(K) SUBSTANTIAL EQUIVALENCE DETERMINATION DECISION SUMMARY: K124006, at 2 (2013), https://www.accessdata.fda.gov/cdrh_docs/reviews/K124006.pdf [https://perma.cc/7DGB-CKB6] ("The variants include those recommended in 2004 by the American College of Medical Genetics (ACMG) and in 2011 by the American College of Obstetricians and Gynecologists (ACOG).").

54. *See* Barbara J. Evans, Wylie Burke & Gail P. Jarvik, *The FDA and Genomic Tests — Getting Regulation Right*, 372 NEW ENG. J. MED. 2258, 2258 (2015) (discussing the FDA's stance on next-generation genetic tests compared to traditional tests: "[T]he FDA admits that these tests strain its existing regulatory methods and contrasts them with other technologies for detecting genetic variants, such as polymerase- chain-reaction and single-nucleotide-polymorphism arrays, that generally are designed to capture predefined data points that are known in advance of testing").

55. In the ACMG framework, variants are categorized as either pathogenic, likely pathogenic, VUS, likely benign, and benign. *See* Laura M. Amendola, Kathleen Muenzen, Leslie G. Biesecker, Kevin M. Bowling, Greg M. Cooper, Michael O. Dorschner et al., *Variant Classification Concordance Using the ACMG-AMP Variant Interpretation Guidelines Across Nine Genomic Implementation Research Studies*, 107 AM. J. HUM. GENETICS 932, 932–33 (2020).

56. Evans et al. *supra* note 54, at 2259 (noting that "[e]ven in theory, premarket review cannot ensure clinical validity for every variant a genomescale test may detect, because the full range of variants becomes clear only after the test is widely used — presumably after the FDA clears or approves it.").

## IV. HOW IS PRS DIFFERENT AND WHAT SHOULD THAT MEAN FOR THE FDA?

In Part III, we explained FDA's regulatory authority over IVDs and LDTs generally as well as the role of CLIA in the regulatory scheme. We now dive deeper into PRS specifically to illustrate particular features of PRSs that should be relevant to designing a regulatory regime governing them. Not that they are *unique*, but that they have *distinct* features from many other devices including other LDTs. Each feature standing alone may not justify a distinct regulatory analysis, but we believe taken together, they indicate FDA should adjust its existing regulatory pathway in reviewing PRSs.

We consider four dimensions — grounding risk, an evolving evidence base, lack of consensus on methodology, and diversity of device functions and purveyors.

### *A. Estimating Genetic Risk*

PRSs are distinct from other forms of genetic scoring for a variety of reasons, including the underlying theory of genetic risk they presuppose. The FDA's regulatory regime for previous genetic tests was ultimately grounded in the ability of the test to detect known variants of clinical significance. For example, one of 23andMe's first approved DTC tests were "for the detection of the $BLM^{Ash}$ variant in the BLM gene from saliva collected using an FDA cleared collection device."[57] Even high-throughput genetic tests that sequence a stretch of the genome (or a panel of variants), such as Illumina's MiSeq next-generation sequencing test, have been approved on the basis that they correctly call variants across several genomic segments,[58] the clinical significance of which were recommended by professional societies.[59] The FDA placed certain controls on some of these devices in accordance with their status as Class II devices that reported risk that may not be equally accurate or useful for all users.[60]

---

57. U.S. FOOD & DRUG ADMIN., CLASSIFICATION ORDER: DEN 140044, at 2 (2014), https://www.accessdata.fda.gov/cdrh_docs/pdf14/den140044.pdf [https://perma.cc/4WU2-AKQU].

58. Francis S. Collins & Margaret A. Hamburg, *First FDA Authorization for Next-Generation Sequencer*, 369 NEW ENG. J. MED. 2369, 2370 (2013) ("The FDA based its decision to grant marketing authorization for the Illumina instrument platform and reagents on their demonstrated accuracy across numerous genomic segments, spanning 19 human chromosomes.").

59. U.S. FOOD & DRUG ADMIN., 510(K) SUBSTANTIAL EQUIVALENCE DETERMINATION DECISION SUMMARY: K124006, at 2 (2013), https://www.accessdata.fda.gov/cdrh_docs/reviews/k124006.pdf [https://perma.cc/R7FT-QZP6].

60. *Id.*

As we have described in Part II, the model of genetic risk presupposed by PRSs is different from previous genetic tests in two important ways. First, the diseases for which PRSs are likely to be most fruitful (at least in the near-term) are polygenic diseases that are common in the population and account for significant morbidity and mortality (e.g., diabetes, heart disease). For these diseases, polygenic scores are constructed from millions of common variants. Individually each variant is inconsequential, but collectively may stratify individuals at high risk from the population — efforts that could help to direct prevention and screening programs.[61] Unlike previous genetic tests in which an allele or variant of interest can be grounded in a candidate gene or pathophysiological mechanism, the causal network is significantly more complex.[62] To complicate matters, different PRSs vary in their explanatory power depending on the phenotypes/disease. For example, polygenic scores calculated from large GWAS can explain approximately three percent in the phenotypic variance of smoking behavior and depression, but approximately ten percent of the variance in educational attainment.[63]

Secondly, the idea of risk in PRS is best understood as expressing genetic liability of disease risk *in relation to other people.* In other words, it is more akin to a percentile score, rather than an expression of genetic determinism — or, as Raffington et al. put it — a PRS tells us the "propensity toward expressing a phenotype relative to other people, not measures of something 'innate.'"[64] In previous models of genetic testing, the idea of risk was not understood by reference to the broader population distribution. The clinical validity of a cancer gene mutation may involve an assessment of other variants that have a modifying effect on the cancer phenotype, but this does not entail a need for the underlying genetic risk to be understood as a probability distribution.

Instead of a handful of variants, a PRS takes thousands of variants and aggregates their effects through statistical techniques. On the other side, an individual can be told her risk of anything from breast cancer to educational attainment. The question is how this process ought to be regulated. The risk estimate is different from other genetic tests because it is polygenic, and its interpretation is often akin to a population

---

61. Niall J. Lennon, Leah C. Kottyan, Christopher Kachulis, Noura S. Abul-Husn, Josh Arias, Gillian Belbin et al., *Selection, Optimization and Validation of Ten Chronic Disease Polygenic Risk Scores for Clinical Implementation in Diverse US Populations*, 30 NATURE MED. 480, 481 (2024) (describing the selection strategy for conditions based on "condition prevalence and relevance to preventive care . . .").

62. *See generally* Ali Torkamani, Nathan E. Wineinger & Eric J. Topol, *The Personal and Clinical Utility of Polygenic Risk Scores*, 19 NATURE REVS. GENETICS 581 (2018).

63. Laurel Raffington, Travis Mallard & K. Paige Harden, *Polygenic Scores in Developmental Psychology: Invite Genetics In, Leave Biodeterminism Behind*, 2 ANN. REV. DEV. PSYCH. 389, 398 (2020).

64. *Id.* at 390.

distribution rather than rooted in a causal mechanism.[65] The supposedly transformative power of polygenic scoring is that it does not turn on the need to assess the clinical significance of individual variants. The question about regulation of polygenic scores is therefore fundamentally about regulating the *interpretation* of many genetic variants.

What will ultimately be important to a patient is not the probability distribution or the algorithms used to generate polygenic scores, but rather what it means for her life and her future. For these scores to be implemented clinically, the FDA will need to evaluate the risk claims. Geneticists have argued for the importance of tools to translate PRSs into absolute risk, which provide an individual's overall risk of developing a disease within a certain timeframe.[66] Indeed, empirical research suggests that patients may better comprehend genetic risk formulated as absolute risk reduction or relative risk reduction rather than other comparative risk estimates.[67]

The regulatory challenge of polygenic scoring cannot simply be reduced to the familiar problem of inferring clinical validity of rare or novel individual variants, since the practice of polygenic scoring is a practice that *itself* may involve a regulable device function above and beyond the act of simply summing individual risk variants. While a firm-based regulatory approach (e.g., Myriad's BRCAnalysis CDx) is a unique way to promote clinically valid genetic tests within statutory limits in the face of new variants, polygenic scoring has certain "emergent properties" because the relationship between the detection of variants and the reporting of clinical significance is mediated by an intermediate process that *itself may have a novel device function* whose risks cannot be understood by summing the risks of individual variants. In this way, polygenic scoring poses a unique challenge of interpretation and regulation that we believe is novel when compared with previous genetic testing models.

### *B. Evolving Evidence Base*

PRSs may also be difficult to regulate because the base of evidence from which these scores are generated is constantly evolving for three reasons. First, the techniques used to generate these scores are rapidly

---

65. Natalie Ram, *Polygenic Scoring and the Criminal Legal System*, 38 HARV. J.L. & TECH. 577, 584 (2024).

66. *See generally* Nilanjan Chatterjee, Jianxin Shi & Montserrat García-Closas, *Developing and Evaluating Polygenic Risk Prediction Models for Stratified Disease Prevention*, 17 NATURE REVS. GENETICS 392 (2016); Oliver Pain, Alexandra C. Gillett, Jehannine C. Austin, Lasse Folkersen & Cathryn M. Lewis, *A Tool for Translating Polygenic Scores onto the Absolute Scale Using Summary Statistics*, 30 EUR. J. HUM. GENETICS 339 (2022).

67. Stacey L. Sheridan, Michael P. Pignone & Carmen L. Lewis, *A Randomized Comparison of Patients' Understanding of Number Needed to Treat and Other Common Risk Reduction Formats*, 18 J. GEN. INTERNAL MED. 884, 884 (2003).

becoming better and increasingly sophisticated.[68] PRS is an area of intense academic and industrial investment, and each year, significant energy is being invested into techniques that allow for more optimal construction of scores.[69] Second, in response to important criticisms of the lack of portability of PRSs across different ancestral groups, investigators have increasingly turned their attention to ethically sequencing the genomes of diverse populations, as well as testing and improving PRS performance across diverse ancestries.[70] Thus, the accuracy of PRSs across diverse racial and ancestral groups is likely to improve over time, meaning that a PRS regulatory framework will need to readily update consumers, such as through special controls or labeling requirements.[71] Third, as a result of both the evolving evidence base and the updates to reflect improvements in sequencing coverage, the accuracy of these scores will need to be reviewed *individually for each score*, which is likely to significantly increase the burden imposed on regulatory agencies.

Lastly, the broader evidence base is improving over time because we are sequencing more genomes, which may improve PRSs. As Zuk et al. explain, the total heritability determined by biometrical twin- and family-based studies may be skewed upwards due to confounding non-additive genetic (i.e. epistatic or dominant effects) or non-genetic (i.e. environmental) factors.[72] Therefore, if total heritability is defined as the "upper limit" of heritability estimated from twin- and family-based studies assumed to reflect additive genetic effects, then the so-called "chip heritability" lies somewhere below the total heritability and accounts for the proportion of phenotypic variance explained by "all variants assayed by GWAS arrays."[73] Sequencing more genomes allows

---

68. *See, e.g.*, Rikifumi Ohta, Yosuke Tanigawa, Yuta Suzuki, Manolis Kellis & Shinichi Morishita, *A Polygenic Score Method Boosted by Non-Additive Models*, 15 NATURE COMMC'NS, May 29, 2024, at 1 (describing the development of PRS models that incorporate non-additive interactive effects of SNPs).

69. Ying Wang, Kristin Tsuo, Masahiro Kanai, Benjamin M. Neale & Alicia R. Martin, *Challenges and Opportunities for Developing More Generalizable Polygenic Risk Scores*, 5 ANN. REV. BIOMEDICAL DATA SCI. 293, 295 (2022) ("there has been a recent flurry of new PRS construction methods that improve upon methods originally applied in animal breeding to increase accuracy, computational efficiency, and generalizability.").

70. *See generally* Lennon et al., *supra* note 61; Alexander G. Bick, Ginger A. Metcalf, Kelsey R. Mayo, Le Lichtenstein, Shimon Rura, Robert J. Carroll et al., *Genomic Data in the All of Us Research Program*, 627 NATURE 340 (2024); Wang et al., *supra* note 69.

71. *See, e.g.*, U.S. FOOD & DRUG ADMIN., CLASSIFICATION ORDER: DEN180028, at 4 (2019), https://www.accessdata.fda.gov/cdrh_docs/pdf18/DEN180028.pdf [https://perma.cc/4GNV-HR8R] (describing that labelling for 23andMe must include information on the variants for which a given test is most useful).

72. Or Zuk, Eliana Hechter, Shamil R. Sunyaev & Eric S. Lander, *The Mystery of Missing Heritability: Genetic Interactions Create Phantom Heritability*, 109 PROC. NAT'L ACAD. SCIS. 1193, 1196 (2012).

73. John S. Witte, Peter M. Visscher & Naomi R. Wray, *The Contribution of Genetic Variants to Disease Depends on the Ruler*, 15 NATURE REVS. GENETICS 765, 774 (2014).

us to access more SNPs, adding to the pool of variants from which studies can seek to estimate PRSs.

How do these four forms of expanding evidence bear on the regulability of PRS? An evolving device does not necessarily escape regulation. Indeed, this safety and efficacy balance is in many ways the crux of the question in the premarket vs. postmarket evaluations by the FDA. As Gibson and Lemmens argue, an increasingly complex modern drug market means that the FDA and the public at large ought to embrace more expanded postmarket surveillance.[74] In the well-known example of the Risk Evaluation and Mitigation Strategies ("REMS") system, Congress gave the FDA the ability to monitor the effects of a drug or biologic that has been approved by the FDA.[75] The FDA may impose reporting requirements to PRS-based tests on the basis that their safety profile may evolve over time.

Or, compare the problem of PRSs with what Babic et al. call the "update problem" for artificial intelligence and machine learning ("AI/ML") algorithms in medicine.[76] Many AI/ML algorithms are "adaptive" in that they can constantly update and reward or penalize components of the algorithm in response to new inputs, in contrast to a "locked" algorithm that returns the same outputs given identical inputs, and does not change with repeated use.[77] The FDA recently proposed a system of predetermined change control plans to evaluate the safety and efficacy of adaptive AI/ML models.[78] Babic et al. argue that the FDA should adopt a continuous risk-based regulatory approach in which an AI/ML algorithm's safety and efficacy is tested throughout its life-cycle instead of evaluating algorithms as locked at one fixed point in time.[79]

An interesting regulatory framework could have the ability to test and challenge PRS-based genetic tests in the face of an ever-expanding evidence base. A proposal made by Lennerz and Ramamurthy suggests the FDA construct an artificial test data set maintained by the Agency

---

74. *See generally* Shannon Gibson & Trudo Lemmens, *Overcoming "Premarket Syndrome": Promoting Better Postmarket Surveillance in an Evolving Drug Development Context*, *in* FDA IN THE TWENTY-FIRST CENTURY: THE CHALLENGES OF REGULATING DRUGS AND NEW TECHNOLOGIES 268 (Holly Fernandez Lynch & I. Glenn Cohen eds., 2015).

75. *See generally* Food and Drug Administration Amendments Act of 2007, Pub. L. No. 110-85, 121 STAT. 823.

76. *See generally* Boris Babic, Sara Gerke, Theodoros Evgeniou & I. Glenn Cohen, *Algorithms on Regulatory Lockdown in Medicine*, 366 SCI. 1202 (2019).

77. *Id.* at 1203.

78. U.S. FOOD & DRUG ADMIN., PROPOSED REGULATORY FRAMEWORK FOR MODIFICATIONS TO ARTIFICIAL INTELLIGENCE/MACHINE LEARNING (AI/ML)-BASED SOFTWARE AS A MEDICAL DEVICE (2019), https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device [https://perma.cc/G4WG-DHYZ] [hereinafter *Proposed SaMD Framework*].

79. Babic et al., *supra* note 76, at 1204 ("As regulators push forward, their emphasis should be on developing a process to continuously monitor, identify, and manage associated risks due to AI/ML features such as concept drift, covariate shift, and instability.").

"composed of variant-call and phenotype-call files that contain 1000 manually generated artificial clinical interpretations of various mathematical difficulties."[80] If applied to the PRS case, the FDA could maintain a database of artificially-generated genotype files and require PRS-based genetic tests to accurately report polygenic scores within an established confidence interval.[81] However, this assumes the existence of greater consensus on the nature, proper methodology, and value of these scores than currently exists.[82]

### *C. Lack of Consensus on Methodology*

A third concern is the lack of consensus regarding PRS generation, and particularly the construction of scientific consensus about the significance of certain interpretations over others. Indeed, in some ways, the question of analytical validity for genotyping SNPs, which many PRSs are derived from, has been obviated by developments in technology making SNP genotyping arrays more than 99% accurate in most assays on the market.[83] While most PRSs have thus far been calculated from all common SNPs with a certain allele frequency in the population, others have created PRSs from rare variants to create scores that reflect a more comprehensive picture of the genetic liability of disease risk.[84] There are also disagreements about how exactly to generate these scores, and what kinds of underlying assumptions are legitimate to employ with respect to the effects that variants are assumed to have on each other. According to Harden, some of these disagreements may ultimately turn on what function of PRSs investigators seek to highlight.[85] One possible backstop to consider in the face of these profound disagreements, as we discussed in Part II, may be found in the heritability of the trait or disease under discussion. However, heritability

80. Jochen K. Lennerz & Lakshman Ramamurthy, Correspondence to the Editor, *ClinGen and Genetic Testing*, 373 NEW ENG. J. MED. 1376, 1378 (2015).

81. *See* G. Javitt, S. Katsanis, J. Scott & K. Hudson, *Developing the Blueprint for a Genetic Testing Registry*, 13 PUB. HEALTH GENOMICS 95, 99–102 (2009) (discussing the sources of legal authority for the creation of a national genetic testing registry, as well as important features such a registry ought to have).

82. As we described in *supra* Part III, this may be changing. Investigators are increasingly mounting efforts to create standards for PRS explicitly in reference to regulatory understanding. *See generally* Wand et al., *supra* note 34; Lennon et al., *supra* note 61.

83. Thomas LaFramboise, *Single Nucleotide Polymorphism Arrays: A Decade of Biological, Computational and Technological Advances*, 37 NUCLEIC ACIDS RSCH. 4181, 4182–83 (2009).

84. *See generally* Craig Smail Nicole M. Ferraro, Qin Hui, Million Veteran Program, Manuel A. Rivas & Stephen B. Montgomery, *Integration of Rare Expression Outlier-Associated Variants Improves Polygenic Risk Prediction*, 109 AM. J. HUM. GENETICS 1055 (2022); Or Zuk, Stephen Schaffner, Katilin Samocha, Ron Do, Eliana Hechter, et al., *Searching for Missing Heritability: Designing Rare Variant Association Studies*, 111 PROC. NAT'L. ACAD. SCIS. E455 (2014).

85. James W. Madole & K. Paige Harden, *Building Causal Knowledge in Behavior Genetics*, 46 BEHAV. & BRAIN SCIS. 1, 4–7 (2023).

estimates are themselves far from immutable. For example, the heritability of educational attainment increased more than twenty percentage points in post-World War II Norway after the adoption of educational policies to expand opportunity.[86]

How should the lack of consensus on PRS construction impact the prospects for FDA regulation? Although there is significant disagreement about how these scores are generated, what is important from the perspective of the FDA ought to be whether a given entity takes on a device function that falls under the Agency's purview. The mere fact that there is no expert consensus on the methodology used to create a medical device does not generate a prima facie claim against market entry — what matters is what the medical device does in the world, what kinds of claims it makes, and whether it poses substantive risks to consumers and patients.[87]

### *D. Diversity of Device Functions and Purveyors: The Challenge of Integration*

PRSs also may be considered distinct because they have varying degrees of functionality as a device, and also *within* another device. This is because the device function of a PRS is dependent on other contextual factors, such as its integration with other general wellness claims, its relationship to core practice of medicine competencies, and their integration with broader medical systems, such as clinical decision software. We take each in turn.

When the FDA sent cease-and-desist letters to DTC companies in 2013, these firms discontinued health-related genetic reporting, instead offering "interpretation services."[88] As scholars note, the genetic interpretation market has burgeoned over time, with firms offering DTC polygenic scoring in a range of ways, from full-blown clinical interpretation services to "general wellness products."[89] Firms often escape FDA regulatory oversight by claiming in part that their products are "general wellness," and are not meant for the diagnosis or treatment of disease: polygenic scoring is integrated into a broader statistical model — incorporating behavioral and lifestyle factors — to provide a

86. A.C. Heath, K. Berg, L. J. Eaves, M. H. Solaas, L.A. Corey, J. Sundet et al., *Education Policy and the Heritability of Educational Attainment*, 314 NATURE 734, 736 (1985), *cited in* K. Paige Harden, *Genetic Determinism, Essentialism and Reductionism: Semantic Clarity for Contested Science*, 24 NATURE REVS. GENETICS 197, 199 (2023) (synthesizing the literature that demonstrates shifts in heritability for educational attainment after changes in social policy).

87. See *infra* Section V.B.

88. Spector-Bagdady & Pike, *supra* note 40, at 704, 728; *see also* Christi J. Guerrini, Jennifer K. Wagner, Sarah C. Nelson, Gail H. Javitt, & Amy L. McGuire, *Who's on Third? Regulation of Third-Party Genetic Interpretation Services*, 22 GENETICS MED. 4, 5 (2020).

89. *See* Guerrini et al., *supra* note 88, at 5; Sherkow et al., *supra* note 2.

comprehensive assessment of disease risk.[90] However, the line between general wellness products and bona fide devices has been difficult to draw, and FDA has not yet issued a public stance on third-party interpretation software other than to indicate that if the data do not make any claims about the treatment of disease, it does not come into the FDA's gambit.[91] While some scholars have argued that the FDA may establish a regulatory framework for genetic information based broadly on the risk imposed by the data to consumers, time will tell whether there is a clear device function that can be used as the basis for FDA review.[92]

To its credit, the FDA is already responsive to the fact that the use of PRSs in the DTC context raises important regulatory issues. In its October 2023 draft regulation, the FDA contemplated the potentially disparate impact that PRS-based tests may have on individuals with diverse genetic ancestries, citing several studies that demonstrate that the widespread use of PRSs in their current form may exacerbate health disparities.[93] FDA bases part of the reasoning for this 2023 LDT regulation for previous forms of LDTs on the fact that novel technologies that aggregate risk can have disparate impacts on minority populations, obviating the need for closer regulatory scrutiny.[94]

Apart from the DTC context, PRSs may also be integrated into health systems. Polygenic scoring might be integrated into health systems as clinical decision support software. Many commentators have argued that the use of PRSs as a stratification tool is where the use of PRSs is likely to be most relevant in the near-term.[95] In this way, a PRS that reports an individual's risk of diabetes may have a clear device function as a software that enables the diagnosis and treatment of disease. The FDA's recent Software as a Medical Device ("SaMD") guidance document indicates that if software is part of the diagnosis or

---

90. *See* Sarah C. Nelson & Stephanie M. Fullerton, *'Bridge to the Literature?' Third-Party Genetic Interpretation Tools and the Views of Tool Developers*, 27 J. GENETIC COUNSELING 770, 777 (2018). Importantly, these firms may not be seeking to evade FDA regulation *purposefully*, but rather have no other choice in the absence of bright-line guidance regarding the regulatory status of the services they provide. *See id.* ("They [FDA] don't have a policy and they're not in a hurry to develop a policy. And this is like the worst type of regulation, because if there is a policy at least I know what I can and cannot do. When there is no policy . . . you have to guess . . . And you don't get clear answers.").

91. *See* Elizabeth R. Pike & Kayte Spector-Bagdady, *Device-ive Maneuvers: FDA's Risk Assessment of Bifurcated Direct-to-Consumer Genetic Testing*, *in* FDA IN THE TWENTY-FIRST CENTURY: THE CHALLENGES OF REGULATING DRUGS AND NEW TECHNOLOGIES 470, 471 (Holly Fernandez Lynch & I. Glenn Cohen eds., 2015).

92. *See* Guerrini et al., *supra* note 88, at 7.

93. Medical Devices; Laboratory Developed Tests, 89 Fed. Reg. 37286, 37326 (May 6, 2024) (codified at 21 C.F.R. pt. 809).

94. *Id.*

95. *Cf.* Cathryn M. Lewis & Evangelos Vassos, *Polygenic Risk Scores: From Research Tools to Clinical Instruments*, 12 GENOME MED., May 18, 2020, at 1, 8.

treatment of disease, it is regulable by the FDA.[96] What would it mean to think about PRS as a SaMD? We can consider the way that FDA has explained algorithm change protocols in its regulation of devices that include AI/ML algorithms, shifting from a "locked" approach to a "total product lifecycle" approach.[97]

## V. WHAT SHOULD BE THE FUTURE OF PRS REGULATION?

In Part IV, we discussed the features of PRSs that distinguish it from previous genetic tests. In this Part, we synthesize these features, and lay out some general principles for regulatory design as it relates to polygenic scores. These are *general* principles that may be helpful for designing a regulatory framework that is responsive to genetic tests and other devices that incorporate polygenic scores; however, we do not think there are universal prescriptions that can be applied in every case.

### *A. Unified Approach*

First, the regulatory framework should be *unified* with respect to devices that incorporate PRSs. As we discussed in Part IV, an important feature of PRSs is versatility and the diversity of purveyors that may utilize polygenic scoring. Thus, the FDA may need to evaluate devices that incorporate polygenic scores in a diverse array of contexts. PRS scoring explored in Part II includes DTC genetic tests that predict genetic propensity of intelligence as well as complex clinical assessment and decision support tools. There are also use cases that may not have an explicit device function, at least as defined by statute. Whatever the substantive regulatory policy utilized to assess the clinical validity of these scores, the FDA should ensure regulatory policy is unified across its various regulatory portfolios to ensure that these scores are evaluated on a consistent (even if not necessarily identical) basis regardless of the purveyor. In some instances, this may require Congressional alteration to statutory authority.

### *B. Risk-Based Framework: Pre- and Post-Approval Tools*

There are several tools currently available to the FDA. Under existing authority, and as a matter of regulatory policy, the FDA may choose to impose regulatory requirements either during the approval stage or at the post-approval stage. Under the existing statutory authority, any device seeking marketing approval by the FDA as a medical device would either need to go through premarketing approval or seek

96. *See generally Proposed SaMD Framework*, *supra* note 78.
97. *See infra* Section V.C.

exemption through a predicate device.[98] Because there is currently no FDA-approved PRS-based device on the market today, the FDA would have to give a risk classification to this device, just as it did for 23andMe's DTC-GTs.[99] We can make some forecasts how this might go based on how the FDA has previously used its authority.

First, the PRS-based device in question must fit within the FDA's jurisdiction as a medical device. As discussed in Part III, a medical device must make claims about its intended use in the diagnosis or treatment of disease.[100] Interestingly, PRS as a hypothesis-free approach to modeling genetic risk naturally enables claims about disease and illness, which may bring third-party interpretation software closer to the FDA's regulatory jurisdiction. Thus, FDA's current posture with respect to third-party interpretation software might shift, but only if these firms make claims about disease risk, treatment, or propensity — which some companies such as ADNTRO may already be doing.[101] Some scholars have proposed an alternative regulatory framework that phases out this bifurcated strategy in place of a unified data-driven risk-based framework.[102]

Next, FDA's mandate with respect to any medical device — genetic tests or otherwise — is to comprehensively understand the risk that the device poses to patients and consumers. Furthermore, the Agency must also assess its clinical and analytic validity — as we described, this is one way in which the law sets out FDA authority apart from CLIA. Therefore, a primary question will be: what risks does the PRS-based device pose to consumers and patients? While dependent on the specific claims made by the device, the risk-based classification system is not blind to these questions about PRS-based risk. A PRS-based DTC-GT that makes a claim about genetic propensity for autism will pose a different kind of risk profile than a provider-facing electronic health record system that allows the addition of a module that incorporates PRSs in the patient's electronic health record. Even if the FDA is able to differentiate PRSs based on the risk they pose to patients, however, there is still the question of patient comprehension of these risk scores, and regulatory decisions will need to be made about how these firms are required to report genetic risk to patients.[103]

Here, we might consider previous models in which the FDA has involved outside experts to support its functions. For an ever-

98. *Supra* Section III.A.

99. 21 U.S.C. § 360(o)(1)(A); Javitt, *supra* note 43, at 620–21; Spector-Bagdady & Pike, *supra* note 40, at 717.

100. 21 U.S.C. § 321(h).

101. ADNTRO reports a PRS that can explain one's "genetic predisposition to alcohol consumption." *See* Sherkow et al., *supra* note 2, at 691.

102. *See generally* Pike & Spector-Bagdady, *supra* note 91.

103. *See supra* Section IV.A; *see also* Valerie Gutmann Koch, *Previvorship and Medical Uncertainty*, 38 HARV. J.L. & TECH. 401, 407 (2024).

proliferating set of complex medical devices that incorporate machine learning algorithms, the FDA has considered the adoption of "assurance lab networks" to bridge any gap in FDA's existing expertise and capacity.[104] These assurance laboratories would serve as public-private partnerships to help evaluate AI applications in accordance to the life-cycle-based approach that the Agency has proposed.[105]

Then, there are tools within the Agency's disposal to mitigate the risks inherent to PRS-based genetic tests. The FDA could impose annual evidentiary requirements for PRS-based genetic tests, just as it has done for other genetic tests on the market such as 23andMe and Myriad's BRCA CDx.[106] In these cases, the FDA is utilizing its statutory authority to place special controls for certain devices for which FDA has experience regulating. However, special controls are difficult to determine before approval and testing.[107] Second, FDA could carry over labeling requirements that it imposed for other approved genetic tests, including for many of 23andMe's DTC tests, where the firm was required to include in its labeling that some of the findings may be most relevant to certain ethnic groups — for example, 23andMe's reported indications for the G2019S variant of the *LRRK2* gene states that the "test is most relevant for people of European, Ashkenazi Jewish, and North African Berber descent."[108] In the PRS-based context, we might imagine the PRS-based test to have to report the inaccuracy of certain scores for those who belong to certain ancestral groups.

To be sure, these steps may prove costly because, as discussed in Part III, the FDA will likely need an individual review process for every single trait and disease. We may overstate the regulatory burden or it might be mitigated by various mechanisms to automate and streamline components of FDA's review of these scores, but it will require significant investment to build up Agency capacity for evaluating these new tools.[109]

Importantly, the FDA has a uniquely important role in generating high-quality health-related information. PRS-based genetic tests are already on the market and are likely to proliferate.[110] By requiring manufacturers to support certain claims of safety and efficacy over others,

---

104. Robert M. Califf, Comm'r, Food & Drug Admin., U.S. Dep't of Health & Hum. Servs., Remarks to the Coalition for Health AI 3 (Mar. 5, 2024).

105. Nigam H. Shah, John D. Halamka, Suchi Saria, Michael Pencina, Troy Tazbaz, Micky Tripathi et al., *A Nationwide Network of Health AI Assurance Laboratories*, 331 JAMA 245, 245–46 (2024).

106. Annas & Elias, *supra* note 40, at 985.

107. Jacob S. Sherkow & Mateo Aboy, *The FDA De Novo Medical Device Pathway, Patents and Anticompetition*, 38 NATURE BIOTECHNOLOGY 1028, 1028 (2020).

108. U.S. FOOD & DRUG ADMIN., EVALUATION OF AUTOMATIC CLASS III DESIGNATION FOR DEN160026, at 4 (2017), https://www.accessdata.fda.gov/cdrh_docs/reviews/den160026.pdf [https://perma.cc/H9UG-7HL9].

109. *See* Mason Marks, *Automating FDA Regulation*, 71 DUKE L.J. 1207, 1280 (2022).

110. Sherkow *supra* note 2, at 691.

the FDA can shape the flow of research and development in what is largely an unregulated market.[111]

The FDA could also consider putting more emphasis on the post-approval stage and rely on mechanisms such as post-approval surveillance, patient registries, and non-FDA mechanisms to ensure safety and efficacy. Under its REMS authority pursuant to Title IX of the Food and Drug Administration Amendments Act of 2007, the FDA may impose mandatory post-approval reporting, testing, and labeling requirements on drugs.[112] While post-approval monitoring will certainly be useful for keeping track of the many changing inputs that go into these scores,[113] it is unclear whether the challenges inherent to regulating these scores can be mitigated through post-market monitoring alone.

### *C. Additional Lessons from AI/ML*

Finally, it is important to keep in mind, as we described in Part III, that polygenic scores are often hypothesis-free interpretations of GWAS data, and thus must ultimately be generated through algorithmic models. In this way, we may envision a paradigm for the regulation of polygenic scores by conceptualizing them fundamentally as AI/ML software. If so, the FDA's recent guidance on adaptive AI/ML models serves as an important starting point in thinking about how to regulate PRSs, even acknowledging that the analogy is a loose one.

In the AI/ML regulatory space it is common to distinguish "locked" from "adaptive" algorithms. FDA defines a "locked algorithm" as "an algorithm that provides the same result each time the same input is applied to it and does not change with use"[114] as opposed to adaptive algorithms. For AI/ML algorithms that are adaptive, the FDA distinguishes three broad categories of possible alterations to an algorithm — 1) the algorithm's clinical and analytical performance, 2) the inputs incorporated into the algorithm, and 3) the intended use of the algorithm.[115] The first two axes are sure to change for PRSs given their method of construction. FDA's approach to medical devices that incorporates AI/ML algorithms is to take a lifecycle approach, in which safety and efficacy claims are evaluated before market entry and constantly updated after approval to ensure a reasonable level of safety and

111. Larisa Svirsky, Dana Howard & Micah L. Berman, *E-Cigarettes and the Multiple Responsibilities of the FDA*, 22 AM. J. BIOETHICS 5, 7 (2022). Though authors make this claim in relation to e-cigarettes, we believe it is even more relevant to the PRS-based genetic test market.

112. Food and Drug Administration Amendments Act of 2007, Pub. L. No. 110-85, 121 Stat. 823, 852–63.

113. *See* discussion *supra* Part IV.

114. *Proposed SaMD Framework*, *supra* note 78, at 3 n.7.

115. *Id.* at 6.

ongoing efficacy.[116] Part of this approach is to allow for a period of re-review in which manufacturers can tweak and update algorithms without requiring separate premarket review.[117] Indeed, several scholars have urged the importance of focusing regulatory resources on solving this so-called "update problem" for AI/ML algorithms.[118]

PRS models are unlikely to be locked in this way due to the inherent need to constantly update them so that they incorporate the most up-to-date data and modeling schemes. PRS models thus illustrate the profound complexity of solving the update problem for certain types of algorithms and demonstrate the importance of considering algorithms as a class to establish predicates, industry standards, and comparators for regulation. Some of these concerns, however, may be obviated by recent patterns in FDA approval of AI/ML tools. As Muehlematter et al. show, a significant number of devices cleared by the FDA were evaluated based on a predicate (usually a previous generation of the same device) that was not AI/ML-based.[119] Time will tell whether a lifecycle approach will be apt for PRS-based algorithms, though it is a reality that regulatory agencies will have to face in the near future, especially in the increasingly likely scenario where PRS models incorporate other factors such as lifestyle, medical history, etc. to eventually make claims about disease risk or propensity, ultimately collapsing the sharp line between PRSs and other general AI/ML tools.

116. *Id.* at 7–9.

117. Thomas J. Hwang, Aaron S. Kesselheim & Kerstin N. Vokinger, *Lifecycle Regulation of Artificial Intelligence- and Machine Learning-Based Software Devices in Medicine*, 322 JAMA 2285, 2285 (2019).

118. Babic et al., *supra* note 76, at 3; John W. Ayers, Nimit Desai & Davey M. Smith, *Regulate Artificial Intelligence in Health Care by Prioritizing Patient Outcomes*, 331 JAMA 639, 640 (2024).

119. Urs J. Muehlematter, Christian Bluethgen & Kerstin N. Vokinger, *FDA-Cleared Artificial Intelligence and Machine Learning-Based Medical Devices and Their 510(k) Predicate Networks*, 5 LANCET DIGIT. HEALTH e618, e625 (2023).