

Managing differential performance of polygenic risk scores across groups: Real-world experience of the eMERGE Network

Anna C.F. Lewis,^{1,2,8,10,*} Rex L. Chisholm,³ John J. Connolly,⁴ Edward D. Esplin,⁵ Joe Glessner,⁴ Adam Gordon,^{3,6} Robert C. Green,^{7,8,9,10} Hakon Hakonarson,^{4,11,12,13} Margaret Harr,⁴ Ingrid A. Holm,^{14,15} Gail P. Jarvik,¹⁶ Beth Karlson,^{7,17} Eimear E. Kenny,^{18,19,20,21} Leah Kottyan,²² Niall Lennon,⁸ Jodell E. Linder,²³ Yuan Luo,²⁴ Lisa J. Martin,^{25,26} Emma Perez,¹⁷ Megan J. Puckelwartz,^{3,6} Laura J. Rasmussen-Torvik,^{3,24} Maya Sabatello,^{27,28} Richard R. Sharp,²⁹ Jordan W. Smoller,^{30,31,32} Rene Sterling,³³ Shannon Terek,⁴ Wei-Qi Wei,³⁴ and Stephanie M. Fullerton³⁵

Summary

The differential performance of polygenic risk scores (PRSs) by group is one of the major ethical barriers to their clinical use. It is also one of the main practical challenges for any implementation effort. The social repercussions of how people are grouped in PRS research must be considered in communications with research participants, including return of results. Here, we outline the decisions faced and choices made by a large multi-site clinical implementation study returning PRSs to diverse participants in handling this issue of differential performance. Our approach to managing the complexities associated with the differential performance of PRSs serves as a case study that can help future implementers of PRSs to plot an anticipatory course in response to this issue.

Introduction

The differential performance of polygenic risk scores (PRSs) by group is often described as one of the major ethical barriers to clinical implementation, as Martin and colleagues clearly articulate in an article of the same name, “Clinical use of current polygenic risk scores may exacerbate health disparities.”¹ In addition to being a major ethical concern, the differential performance of PRSs is also one of the main practical challenges for the clinical use of PRSs. Before PRSs can be returned to patients, careful decisions are needed on how to validate the scores, whether and how to communicate the significance of any differences in performance to patients and healthcare providers, and the terms used to label the groups the scores were validated in.

Decisions about how to validate and what to communicate highlight a related ethical challenge when using PRSs that more reliably predict risk in certain groups. The danger is that differential performance is interpreted as reflecting significantly different genetic risks between socially defined groups. This danger is particularly acute when PRS validation efforts utilize continental ancestry categories that are readily (and frequently) conflated with socially constructed categories of race and ethnicity.² The field of genetics has a long history of entanglement with racialized thinking,³ an entanglement that will continue without carefully considering how—and by whose decision—people are grouped, labeled, and analyzed in the expanding fields of genomic research and genomic medicine.⁴

¹Edmond and Lily Safra Center for Ethics, Harvard University, Cambridge, MA, USA; ²Department of Genetics, Brigham and Women’s Hospital, Boston, MA, USA; ³Center for Genetic Medicine, Northwestern University, Evanston, IL, USA; ⁴Center for Applied Genomics, Children’s Hospital of Philadelphia, Philadelphia, PA, USA; ⁵Invitae, San Francisco, CA, USA; ⁶Department of Pharmacology, Northwestern University, Evanston, IL, USA; ⁷Department of Medicine, Brigham and Women’s Hospital, Boston, MA, USA; ⁸Broad Institute of MIT and Harvard, Cambridge, MA, USA; ⁹Ariadne Labs, Boston, MA, USA; ¹⁰Harvard Medical School, Boston, MA, USA; ¹¹Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; ¹²Division of Human Genetics, Children’s Hospital of Philadelphia, Philadelphia, PA, USA; ¹³Division of Pulmonary Medicine, Children’s Hospital of Philadelphia, Philadelphia, PA, USA; ¹⁴Division of Genetics and Genomics, Boston Children’s Hospital, Boston, MA, USA; ¹⁵Department of Pediatrics, Harvard Medical School, Boston, MA, USA; ¹⁶Division of Medical Genetics, Department of Medicine and Department of Genome Science, University of Washington Medical Center, Seattle, WA, USA; ¹⁷Mass General Brigham Personalized Medicine, Boston, MA, USA; ¹⁸Institute for Genomic Health, Icahn School of Medicine, New York City, NY, USA; ¹⁹Center for Clinical Translational Genomics, Icahn School of Medicine, New York City, NY, USA; ²⁰Division of Genomic Medicine, Department of Medicine, Icahn School of Medicine, New York City, NY, USA; ²¹Department of Genetics and Genomic Sciences, Icahn School of Medicine, New York City, NY, USA; ²²Center for Autoimmune Genomics and Etiology, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA; ²³Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA; ²⁴Department of Preventive Medicine, Northwestern University, Evanston, IL, USA; ²⁵Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA; ²⁶University of Cincinnati College of Medicine, Cincinnati, OH, USA; ²⁷Center for Precision Medicine and Genomics, Department of Medicine, Columbia University Irving Medical Center, New York City, NY, USA; ²⁸Division of Ethics, Department of Medical Humanities and Ethics, Columbia University Irving Medical Center, New York City, NY, USA; ²⁹Biomedical Ethics Program, Mayo Clinic, Rochester, MN, USA; ³⁰Center for Precision Psychiatry, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA; ³¹Psychiatric & Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; ³²Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA; ³³Division of Genomics and Society, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; ³⁴Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA; ³⁵Department of Bioethics & Humanities, University of Washington School of Medicine, Seattle, WA, USA

*Correspondence: aclewis@bwh.harvard.edu
<https://doi.org/10.1016/j.ajhg.2024.04.005>

© 2024 American Society of Human Genetics.

Here, we describe how a large multi-site clinical implementation research study considered, and acted upon, these challenges during the study design phase. The Electronic Medical Records and Genomics (eMERGE) Network is returning genomically informed risk reports that incorporate PRSs for ten conditions to 25,000 participants (Clinical Trial ID NCT05277116).⁵ Each of the 10 study recruitment sites included investigators who conducted linked ethical, legal, and social implications (ELSI) research designed to inform implementation decisions, all of which involved seeking input from various stakeholders.^{6–11}

The main report returned to patient participants in the study is known as the Genomic Informed Risk Assessment (GIRA). The GIRA displays PRS information, monogenic risk, family history, and clinical risk factors (available in the supplementary materials of Linder et al.⁵). For two conditions, breast cancer and coronary heart disease, absolute risk scores that integrate PRSs, personal factors, family history, clinical risk factors, and in the case of breast cancer, monogenic results, are given in the GIRA. A stand-alone PRS report is included as an appendix to the GIRA along with a monogenic risk report. The study is explicitly designed to recruit individuals from groups who are historically medically underserved and/or under-represented in research and seeks to address the ELSI issues associated with PRS implementation.

eMERGE Network: Study decisions relevant to the differential performance of PRSs

Before discussing the practical challenges faced by the eMERGE Network, we outline the current understanding of the performance of PRSs, as defined by the strength of the association of a PRS with the outcome it is designed to predict. The performance of a score typically drops off in a continuous fashion with genetic distance between the data on which a PRS was trained and the test data.¹² This drop-off in performance is believed to mostly be due to differences in how the genetic variants associated with a phenotype in a typical genome-wide association study are correlated with the variants that have some kind of causal impact on that phenotype.¹³ Although these causal variants are mostly the same across our species, correlation patterns, related to the phenomenon of linkage disequilibrium, are variable across human groups, as are allele frequencies.¹⁴ Moreover, because PRSs rely on aggregating the effect sizes of genotype-phenotype associations, their performance is also sensitive to anything that can have an impact on these effect sizes.¹⁵ PRS performance is, for example, known to differ by demographic factors including socioeconomic status, sex, and age.¹⁶ Moreover, many different PRSs have been produced for the same condition, and their performance can differ substantially.¹⁷

Decision 1: How should the groups used for validating PRS performance be defined?

For a PRS to be considered for inclusion on a clinical report, there must be evidence of its clinical validity for the corre-

sponding condition. This involves assessing how predictive the PRS is for the incidence or prevalence of that condition. A choice must be made regarding which groups to use to validate a PRS's performance. In the case of eMERGE, validation cohorts were drawn from many sources (which could differ by condition). These sources relied on different ways of measuring group membership and used different types of population labels. For example, the group label assigned to individuals in the validation cohorts could be based on self-report, extraction from electronic health records, and/or analysis of genetic ancestry. Many of the sources were from legacy datasets, which used pre-determined labels. Given this heterogeneity in labels used, the choice was made to validate the performance of the scores in as many as possible of four groups, designated by the network as European, African, Hispanic, or Asian. This decision—to amalgamate samples that began with different types of labels and to focus on these four ambiguously defined groups—was made in order to maximize the amount of validation data available and to make comparisons across scores possible. The network discussed best practices and methods to genetically define groups but left it up to the leads for each condition to choose the criteria by which individuals would be placed into each of these four groups. This validation process ultimately helped determine which conditions to return as part of the study. Details of the scores selected and the data used for validation are given in Lennon et al.¹⁸

We acknowledge the limitations of this approach to data harmonization. First, the groups chosen aggregated individuals perceived to be similar based on determinations arising from different approaches to measuring (self-identification, genetic derivation) and from different population descriptors (race, ethnicity, ancestry), descriptors that rely on very different concepts of human difference.¹⁹ Combining these different conceptualizations together contributes to the conflation of genetically inferred categories with social identities.² Second, these four groups fail to account for the extensive human genetic variation known to exist, leaving many individuals (e.g., Native American, those with recent ancestors from several of these groups) not represented at all.

Decision 2: Use group-specific scores or the same score for everyone?

For several reasons, the eMERGE Network decided at an early stage not to use group-specific scores but to instead use the same PRS for every individual. The use of group-specific scores would have necessitated a process by which each patient participant was assigned to one of the groups. Input from the ELSI working group emphasized the limitation noted above that there could be many individuals who would not fit into any of the groups. The practical complexities of developing many group-specific reports also weighed in the decision.¹⁸

For each condition, a threshold above which an individual would be identified as “high risk,” regardless of their group membership, was chosen (this too varied by condition).⁵

The odds ratio and associated confidence interval associated with this threshold were determined separately in each of the four groups for which there was sufficient data. For some conditions, the network only had sufficient data for this validation for two or three of the four groups (which groups were available for validation varied by condition).

To ensure a similar distribution of scores for individuals in each validation group, the raw PRSs were calibrated using genetic principal components.¹⁸ It is important to note that it is not possible to “calibrate away” the differential performance issue. The predictive performance of the score could still differ between groups at any chosen threshold. Rather, such calibration helps ensure that a similar fraction of individuals in each group fall above a given percentile threshold. This is helpful because it is not generally believed that the differences in raw score distributions reflect underlying biology.²⁰

Decision 3: Include PRSs for which it was not possible to validate in all the defined validation groups?

The network needed to make a decision on whether to include PRSs for conditions that were not validated in all four groups. Two of the ELSI projects independently conducted interview studies with racially and ethnically diverse participants and ascertained that prospective patients were not concerned about the lower predictive power of PRSs outside European-descent populations and were comfortable receiving scores even if inaccurate for their ancestry.^{6,7} These findings informed the decision to include PRSs that were only validated in two or three of the validation groups. The network wanted to ensure patient participants and their healthcare providers understood which groups a score was validated in, and so both the GIRA and PRS report indicate which of the validation groups were used for a given score, and the explanation that “information is insufficient or not available for populations of other descent.”

Decision 4: What terminology should be used to describe the different validation groups?

Because the established validation groups included different ways of measuring human difference (i.e., self-reported race or ethnicity or genetically inferred ancestry), the network was uncertain about how best to describe these groups to patient participants and their healthcare providers (for example, in the informed consent document and in the GIRA). The term “population” is often used in scientific discourse. The usage of this term is often ambiguous, which may account for its popularity. While this term is highly ambiguous in research settings, in common parlance it is often less ambiguous, referring to people who live in a particular area. The use of the term “population” was hence deemed too confusing for communicating to patient participants and their healthcare providers.

To assist the network in making a decision, the Mass General Brigham site consulted with a group called the Community Coalition for Equity in Research, a Massachusetts-based

group that aims to “build community voice and considerations for health equity into clinical research and strengthen community-academic relationships” (<https://catalyst.harvard.edu/community-engagement/community-coalition/>).²¹ Based in large part on the input from this group, we settled on the term “descent” to characterize the groups. This term was believed to strike a balance between scientific accuracy and lay public understanding. We note that the concept of “descent” is the highest-level concept identified by the recent National Academies report on the use of population descriptors in genetics and genomics research: “The use of such descriptors as race, ethnicity, or ancestry, however, focuses on ‘descent-associated’ groups—sets of individuals whose members are thought to share some characteristic that derives from their common origin.”¹⁹ While we feel that adoption of the “descent” language was a suitable solution for our study, ideally a single concept of human difference would have been used to define the validation groups, and this could have been referred to explicitly.

Decision 5: Communicate the differential performance between validation groups?

The next decision was whether to inform patient participants and their healthcare providers about the differential performance of the PRSs revealed by validation. The ELSI group provided input that each of two strategies considered were ethically defensible. The first strategy to inform about the differential performance was supported by a general argument for the benefits of transparency, as well as the preferences of prospective patient participants, particularly for a study such as eMERGE, where one of the major emphases is on enrolling research participants from historically underrepresented racial and ethnic groups. The second strategy considered was to report an aggregate performance, giving a range encompassing the lowest lower range and the highest upper range of the confidence intervals across the groups, for example, “a high polygenic risk score is associated with 2–4 times increased risk for developing chronic kidney disease relative to a person not in the high-risk category.” This second strategy would be justified by the clinical recommendations being the same regardless of which group(s) a patient might identify with. This second strategy was also supported by the same concerns identified above about the use of the validation groups, as well as the additional concern that the inclusion of all the information could be hard to adequately communicate to patient participants and their healthcare providers and/or could make the reports more unwieldy.

The network opted for a layered approach to communication on the reports to try to secure the advantages of both strategies. On the GIRA, the most prominent information is whether or not an individual was identified as high risk for any of the conditions. The condition-specific page then gives the overall odds ratio range (the second strategy outlined above). The PRS report first indicates in which conditions the individual was determined to be at

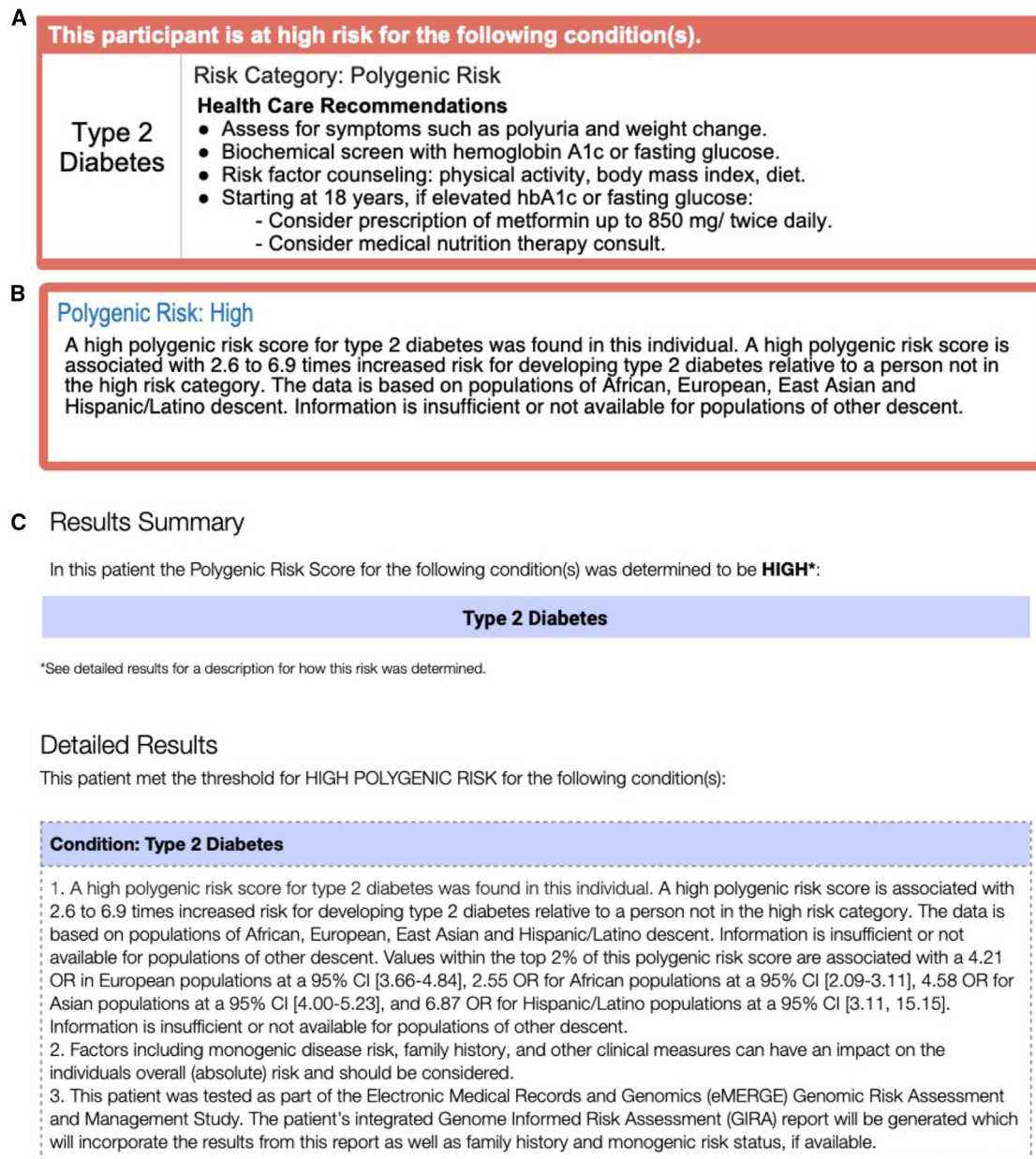


Figure 1. The layered strategy adopted by the eMERGE Network to communicate PRS results, including differential performance
 (A) The summary of findings for type 2 diabetes on the GIRA.
 (B) Detailed results for the high PRS result for type 2 diabetes on the GIRA (see Linder et al.⁵ for complete context for A and B).
 (C) Summary and detailed results on the PRS report for high type 2 diabetes.

high risk, and the detailed results section gives both the aggregate range, odds ratio, and confidence interval for each group included in the validation (thus following the first strategy outlined above). See Figure 1 for mock GIRA and PRS reports illustrating this layered communication strategy.

To aid the conversation between patient participant and healthcare provider for the case where the patient participant identifies with none or multiple of the validation groups, we included the following in a list of frequently asked questions (FAQs) available to prospective patient participants during recruitment: “If you do not identify

with one of the four groups, it is important to discuss this with the study staff and your doctor. Using the overall score, or results from groups you most closely identify with may help you and your doctor make informed decisions about your care.” Likewise, in the FAQs provided alongside the GIRA, the following is included: “You may identify with more than one, none, or all of the listed populations. Across the different results you receive in your GIRA, some may still be meaningful even if you don’t identify with the populations mentioned. You should discuss all of the results you receive with your doctor. Your risk likely falls within the range of risk presented in the report.”

Decision 6: How should the differential performance of the PRSs reported be explained?

Without an explanation as to why there is differential performance of PRSs, inaccurate conclusions could be drawn about, for example, different predispositions of groups to the condition in question. The network developed language to explain the differential performance, which was then used in the informed consent process and on the GIRA. The language in the limitations section of the latter reads, “Genetic research studies need large numbers of participants to understand how human DNA (or genes) contributes to disease risk. When research studies have low representation of some races, ethnicities, or ancestries (populations of descent), there is less genetic information available to understand risks for people in those groups. The GIRA health risk report has been validated (or confirmed) in up to four populations: Asian descent, African descent, European descent, or Hispanic/Latino descent. The report will name the populations included in the validation process. The estimate of risk may not be as accurate for some conditions if the participant is from a population that was not included in the validation process.”

Discussion

An implementation study such as eMERGE encounters many practical challenges, which will be faced in the rollout of PRSs in other contexts. Our study also highlights the ELSI implications of study-design decisions in reaction to some of those challenges. Specifically, we described the repercussions of analytical choices (in this case, at the validation stage) on subsequent decisions regarding patient participant and healthcare provider communication. We note that even if a single concept of difference had informed PRS validation efforts (e.g., by using inferred genetic ancestry to differentiate groups), we would still have faced many of the same questions about how to describe groups as well as how best to communicate and explain differential performance. Use of absolute risk models does not mitigate these challenges, and indeed introduces additional complexities because analytical choices beyond the use of PRSs can potentially affect the accuracy of absolute risk estimates across groups.

The end result of the decisions we made were consent processes, reports, and supporting materials (e.g., FAQs) that included a lot of complexity. This places a burden on investigators who are attempting to communicate the value proposition of the research and on patient participants and healthcare providers to understand the results, particularly among those who might identify with multiple or none of the validation groups. Not all healthcare providers may feel equipped to help navigate these issues with their patients.⁷ This highlights the need for healthcare provider education; the eMERGE Network is making all the materials we have developed to aid in education efforts publicly available.²² The network will be assessing both participant and healthcare provider reactions to results disclosure via surveys and by interviews with healthcare providers at some sites.

The issue of PRS differential performance is most often couched in terms of differing predictive performance across continental genetic ancestry categories.¹ As discussed above, this is an oversimplification. First, because predictive performance falls off in a continuous fashion with measures of genetic similarity. Ideally performance would be assessed in a way that reflected the continuous nature of genetic variation whereby an individual’s location in genetic similarity space—and not their membership in a genetically inferred group—would be used to give risk information. And second, because other, non-genetic dimensions of difference affect predictive performance. A future challenge for the reporting of PRSs will be to identify ways to reflect this more accurate picture. What this would look like remains to be determined.

The field of biomedicine is struggling with questions about how race and ethnicity should be incorporated into clinical tests and decisions.²³ There is a widespread acknowledgment that there is much at stake because of the ways existing practices and policies further health inequities and how they can contribute to racialized thinking. Genetic ancestry is often heralded as part of the solution to the problematic uses of race in biomedicine.^{24,25} But as our case study shows, there are also many challenges if genetic ancestry is framed as a key concept of difference. Not least of these is the empirical observation that both patients and healthcare providers conflate continental ancestry categories with racial categories.⁷

As the experiences of the eMERGE Network suggest, those who choose to report PRSs to research participants or patients face numerous challenges. Currently, there is a paucity of guidance for developers of PRS reports. The recent National Academies report on the use of population descriptors in genomics research explicitly excluded from their scope challenges associated with both data harmonization and clinical implementation.¹⁹ They also gave minimal guidance around choice of language, stating that researchers should “choose wording that transparently reflects the analytical steps taken.” In giving recommendations for which concept of difference to use for respective use cases, the report did not address the use case of validating PRSs. These were some of the issues that we struggled with most, and further guidance in these areas would be helpful. We stress that such guidance would be no substitute for sustained interdisciplinary discussion and community engagement, which should be incorporated at the earliest stages of any new study design.

Finally, we note that the concept of equity drawn upon here to highlight the concern about the differential performance of PRSs is a very narrow one. There are many other factors that will contribute to differences in health outcomes, including who has access to PRS reports and who has access to the recommended clinical actions if identified as at high risk. Attention to these factors will be vital for future work.

ELSI research and expertise were intended to inform the design of our study. Indeed, we believe that different

methodological decisions could have been made in reaction to these complicated implementation problems in the absence of the embedded ELSI approach. Nevertheless, even the opportunity for the network to hear from prospective participants and healthcare providers, as well as to draw upon ELSI reflection and expertise, did not fully address all ethical complexities. This case study demonstrates both the strengths and also some of the limitations of the embedded ELSI approach.²⁶

The researchers involved in the eMERGE Network had the opportunity to recognize and reflect on the numerous decisions needed around the differential performance of PRSs. We hope these reflections can help future projects take an anticipatory approach to these and related issues.

Acknowledgments

The eMERGE Genomic Risk Assessment Network is funded by the National Human Genome Research Institute (NHGRI) through the following grants: U01HG011172 (Cincinnati Children's Hospital Medical Center), U01HG011175 (Children's Hospital of Philadelphia), U01HG008680 (Columbia University), U01HG011176 (Icahn School of Medicine at Mount Sinai), U01HG008685 (Mass General Brigham), U01HG006379 (Mayo Clinic), U01HG011169 (Northwestern University), U01HG011167 (University of Alabama at Birmingham), U01HG008657 (University of Washington), U01HG011181 (Vanderbilt University Medical Center), and U01HG011166 (Vanderbilt University Medical Center serving as the Coordinating Center).

We would like to acknowledge members of the Community Coalition for Equity in Research supported by the National Center for Advancing Translational Sciences through its Clinical and Translational Science Awards Program (UL1TR002541) and Massachusetts Blue Cross Blue Shield for their contributions to this work.

The eMERGE Genomic Risk Assessment study was approved by a central institutional review board at Vanderbilt University Medical Center (#211043) and acknowledged by the local institutional review boards under reliance agreements. Informed consent is required from all enrolled individuals.

Declaration of interests

A.C.F.L. owns some stock in Fabric Genomics; E.D.E. is an employee and stockholder of Invitae, advisor and stockholder of Taproot Health, and advisor and stockholder of Exir; R.C.G. receives compensation for advising the following companies: Allelica, Atria, Fabric, Genomic Life, and Juniper Genomics; and is co-founder of Genome Medical and Nurture Genomics; E.E.K. has received personal fees from Regeneron Pharmaceuticals, 23&Me, Allelica, and Illumina; has received research funding from Allelica; and serves on the advisory boards for Encompass Biosciences, Overtone, and Galatea Bio; N.L. received personal fees from Illumina Inc; E.P. is a paid consultant for Allelica Inc.; M.S. is a member of the Institutional Review Board of the All of Us Research Program; J.W.S. is a member of the Scientific Advisory Board of Sensorium Therapeutics (with equity), has received grant support from Biogen, Inc., and is PI of a collaborative study of the genetics of depression and bipolar disorder sponsored by 23andMe for which 23andMe provides analysis time as in-kind support but no payments.

References

1. Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* *51*, 584–591. <https://doi.org/10.1038/s41588-019-0379-x>.
2. Lewis, A.C.F., Molina, S.J., Appelbaum, P.S., Dauda, B., Di Rienzo, A., Fuentes, A., Fullerton, S.M., Garrison, N.A., Ghosh, N., Hammonds, E.M., et al. (2022). Getting genetic ancestry right for science and society. *Science* *376*, 250–252. <https://doi.org/10.1126/science.abm7530>.
3. Jackson, C.S., Turner, D., June, M., and Miller, M.V. (2023). Facing Our History—Building an Equitable Future. *Am. J. Hum. Genet.* *110*, 377–395. <https://doi.org/10.1016/j.ajhg.2023.02.005>.
4. Roberts, D. (2011). *Fatal Invention: How Science, Politics, and Big Business Re-create Race in the Twenty-First Century* (The New Press).
5. Linder, J.E., Allworth, A., Bland, H.T., Caraballo, P.J., Chisholm, R.L., Clayton, E.W., Crosslin, D.R., Dikilitas, O., DiVietro, A., Esplin, E.D., et al. (2023). Returning integrated genomic risk and clinical recommendations: The eMERGE study. *Genet. Med.* *25*, 100006. <https://doi.org/10.1016/j.gim.2023.100006>.
6. Suckiel, S.A., Braganza, G.T., Aguiñiga, K.L., Odgis, J.A., Bonini, K.E., Kenny, E.E., Hamilton, J.G., and Abul-Husn, N.S. (2022). Perspectives of diverse Spanish- and English-speaking patients on the clinical use of polygenic risk scores. *Genet. Med.* *24*, 1217–1226. <https://doi.org/10.1016/j.gim.2022.03.006>.
7. Lewis, A.C.F., Perez, E.F., Prince, A.E.R., Flaxman, H.R., Gomez, L., Brockman, D.G., Chandler, P.D., Kerman, B.J., Lebo, M.S., Smoller, J.W., et al. (2022). Patient and provider perspectives on polygenic risk scores: implications for clinical reporting and utilization. *Genome Med.* *14*, 114. <https://doi.org/10.1186/s13073-022-01117-8>.
8. May, T., Smith, C.L., Kelley, W., East, K., Orlando, L., Cochran, M., Colletto, S., Moss, I., Nakano-Okuno, M., Korf, B., and Limdi, N. (2023). Does genetic testing offer utility as a supplement to traditional family health history intake for inherited disease risk? *Fam. Pract.* *40*, 760–767. <https://doi.org/10.1093/fampra/cmab017>.
9. Pacyna, J.E., Ennis, J.S., Kullo, I.J., and Sharp, R.R. (2023). Examining the Impact of Polygenic Risk Information in Primary Care. *J. Prim. Care Community Health* *14*, 21501319231151766. <https://doi.org/10.1177/21501319231151766>.
10. Terek, S., Del Rosario, M.C., Hain, H.S., Connolly, J.J., Behr, M.A., Harr, M., Hakonarson, H., and Holm, I.A. (2022). Attitudes among Parents towards Return of Disease-Related Polygenic Risk Scores (PRS) for Their Children. *J. Pers. Med.* *12*, 1945. <https://doi.org/10.3390/jpm12121945>.
11. Sabatello, M., Bakken, S., Chung, W., Cohn, E., Crew, K.D., Kiryluk, K., Kukafka, R., Weng, C., and Appelbaum, P.S. (2024). Return of polygenic risk scores in research: stakeholders' views on the eMERGE-IV study. *Hum. Genet. Genomics Adv.* *5*, 100281.
12. Ding, Y., Hou, K., Xu, Z., Pimplaskar, A., Petter, E., Boulier, K., Privé, F., Vilhjálmsdóttir, B.J., Olde Loohuis, L.M., and Pasiñic, B. (2023). Polygenic scoring accuracy varies across the genetic ancestry continuum. *Nature* *618*, 774–781. <https://doi.org/10.1038/s41586-023-06079-4>.
13. Wang, Y., Guo, J., Ni, G., Yang, J., Visscher, P.M., and Yengo, L. (2020). Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nat. Commun.* *11*, 3865. <https://doi.org/10.1038/s41467-020-17719-y>.

14. Kachuri, L., Chatterjee, N., Hirbo, J., Schaid, D.J., Martin, I., Kullo, I.J., Kenny, E.E., Pasaniuc, B., Polygenic Risk Methods in Diverse Populations PRIMED Consortium Methods Working Group, Witte, J.S., and Ge, T. (2024). Principles and methods for transferring polygenic risk scores across global populations. *Nat. Rev. Genet.* 25, 8–25. <https://doi.org/10.1038/s41576-023-00637-2>.
15. Yair, S., and Coop, G. (2022). Population differentiation of polygenic score predictions under stabilizing selection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 377, 20200416. <https://doi.org/10.1098/rstb.2020.0416>.
16. Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J.K., and Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *Elife* 9, e48376. <https://doi.org/10.7554/eLife.48376>.
17. Clifton, L., Collister, J.A., Liu, X., Littlejohns, T.J., and Hunter, D.J. (2022). Assessing agreement between different polygenic risk scores in the UK Biobank. *Sci. Rep.* 12, 12812. <https://doi.org/10.1038/s41598-022-17012-6>.
18. Lennon, N.J., Kottyan, L.C., Kachulis, C., Abul-Husn, N.S., Arias, J., Belbin, G., Below, J.E., Berndt, S.I., Chung, W.K., Cimino, J.J., et al. (2024). Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations. *Nat. Med.* 30, 480–487. <https://doi.org/10.1038/s41591-024-02796-z>.
19. National Academies of Sciences, E. (2023). *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field* (National Academies Press). <https://doi.org/10.17226/26902>.
20. Martin, A.R., Gignoux, C.R., Walters, R.K., Wojcik, G.L., Neale, B.M., Gravel, S., Daly, M.J., Bustamante, C.D., and Kenny, E.E. (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* 100, 635–649. <https://doi.org/10.1016/j.ajhg.2017.03.004>.
21. Emmons, K.M., Curry, M., Lee, R.M., Pless, A., Ramanadhan, S., and Trujillo, C. (2022). Enabling community input to improve equity in and access to translational research: The Community Coalition for Equity in Research. *J. Clin. Transl. Sci.* 6, e60. <https://doi.org/10.1017/cts.2022.396>.
22. Connolly, J.J., Berner, E.S., Smith, M., Levy, S., Terek, S., Harr, M., Karavite, D., Suckiel, S., Holm, I.A., Dufendach, K., et al. (2023). Education and electronic medical records and genomics network, challenges, and lessons learned from a large-scale clinical trial using polygenic risk scores. *Genet. Med.* 25, 100906. <https://doi.org/10.1016/j.gim.2023.100906>.
23. Vyas, D.A., Eisenstein, L.G., and Jones, D.S. (2020). Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms. *N. Engl. J. Med.* 383, 874–882. <https://doi.org/10.1056/NEJMms2004740>.
24. Borrell, L.N., Elhawary, J.R., Fuentes-Afflick, E., Witonsky, J., Bhakta, N., Wu, A.H.B., Bibbins-Domingo, K., Rodríguez-Santana, J.R., Lenoir, M.A., Gavin, J.R., et al. (2021). Race and Genetic Ancestry in Medicine — A Time for Reckoning with Racism. *N. Engl. J. Med.* 384, 474–480. <https://doi.org/10.1056/NEJMms2029562>.
25. Oni-Orisan, A., Mavura, Y., Banda, Y., Thornton, T.A., and Sebro, R. (2021). Embracing Genetic Diversity to Improve Black Health. *N. Engl. J. Med.* 384, 1163–1167. <https://doi.org/10.1056/NEJMms2031080>.
26. Clayton, E.W., Smith, M.E., Anderson, K.C., Chung, W.K., Connolly, J.J., Fullerton, S.M., McGowan, M.L., Peterson, J.F., Prows, C.A., Sabatello, M., and Holm, I.A. (2023). Studying the impact of translational genomic research: Lessons from eMERGE. *Am. J. Hum. Genet.* 110, 1021–1033. <https://doi.org/10.1016/j.ajhg.2023.05.011>.