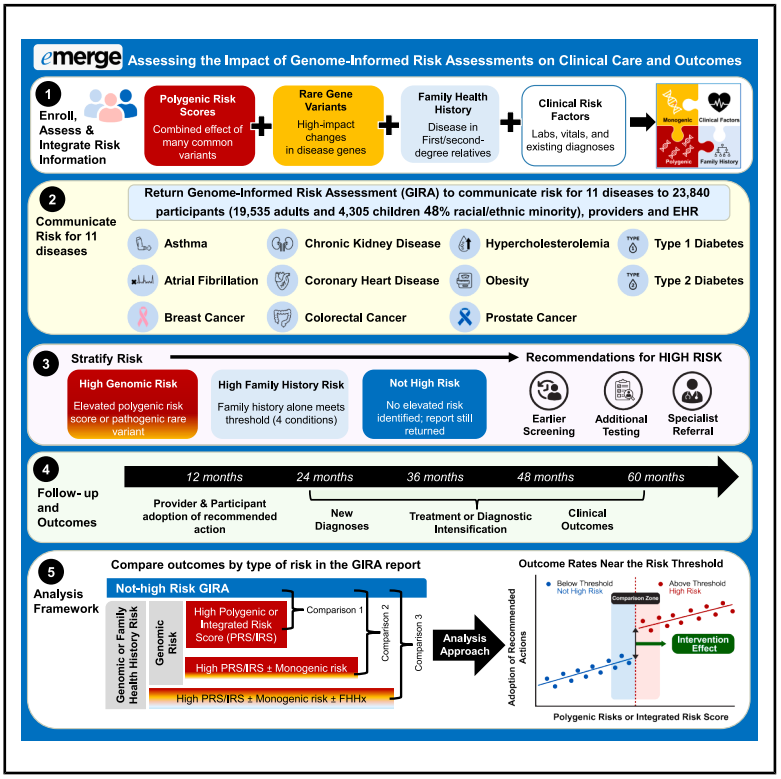
The Electronic Medical Records and Genomics study: Design and analytic framework for assessing the impact of genome-informed risk assessments

Graphical abstract



Authors

Nita Limdi, T. Mark Beasley, Josh Cortopassi, ..., Georgia Wiesner, Noura S. Abul-Husn, David L. Veenstra

Correspondence

nlimdi@uabmc.edu

The eMERGE Network implemented genome-informed risk assessment integrating polygenic, monogenic, clinical, and family history data for 11 diseases, returning results to 23,840 participants. This large prospective study evaluates whether genomic risk disclosure changes care and outcomes. Using quasi-randomized methods, it addresses critical evidence gaps and informs implementation of genomic medicine.

Limdi et al., 2026, The American Journal of Human Genetics *113*, 664–677
April 2, 2026
https://doi.org/10.1016/j.ajhg.2026.02.018

CellPress

ARTICLE

The Electronic Medical Records and Genomics study: Design and analytic framework for assessing the impact of genome-informed risk assessments

Nita Limdi,[1,*] T. Mark Beasley,[2] Josh Cortopassi,[1] Brittney Davis,[1] Hana Bangash,[3] Jingheng Chen,[4] Rex L. Chisholm,[5] Wendy K. Chung,[6] James J. Cimino,[7] John Connolly,[8] David R. Crosslin,[9] Kyle W. Davis,[10] Alanna DiVietro,[11] Edward D. Esplin,[12] Bob Freimuth,[13] Adam Gordon,[5] Hakon Hakonarson,[8] Marwan Hamed,[3,14] Megan He,[11] Christin Hoell,[5] Ingrid Holm,[6] George Hripscak,[15] Margurite R. Irvin,[16] Gail P. Jarvik,[4] Dean Karavite,[17] Elizabeth W. Karlson,[18] Eimear E. Kenny,[19] Atlas Khan,[20] Krzysztof Kiryluk,[20] Sarah Knerr,[21] Bruce Korf,[22] Leah Kottyan,[23] Iftikhar Kullo,[24] Katie Larkin,[25] Niall Lennon,[25] Jodell E. Linder,[11] Teri Manolio,[26] Lisa J. Martin,[23]

(Author list continued on next page)

Summary

The Electronic Medical Records and Genomics (eMERGE) Network developed and implemented a genome-informed risk assessment (GIRA) to communicate genomic (polygenic risk scores [PRSs], integrated risk scores [IRSs], and monogenic results), clinical, and family history-based risk for 11 chronic diseases and provide recommended healthcare recommendations. GIRA reports have now been returned to 23,840 participants and their providers in a large prospective cohort study. We present here the study design and analysis framework for assessing the attributable impact of GIRA return. Pre-specified outcomes include (1) provider/participant adoption of recommended healthcare actions, (2) new diagnosis of disease, (3) treatment initiation/intensification, and (4) clinical outcomes (surrogate markers or clinical events). We assess outcomes in high risk vs. not-high-risk participants, adjusting for covariates. We evaluate the effect of PRS/IRS at pre-established high-risk thresholds using regression discontinuity (RD), a quasi-experimental method that mimics randomization near a cutoff, enabling estimation of causal effects and controlling for unobserved confounders. Monogenic and family history-based risk stratification are analyzed using logistic regression. With 23,840 participants and 12 months of follow-up, the study is powered to detect differences of 2%–11% with 80% power ($\alpha = 0.05$ in the adoption outcome). Longer follow-up will be required to enable assessment of new disease diagnosis, treatment changes, and clinical outcomes. Through innovative RD analyses and defined outcomes and comparison groups, this study will provide new insights into the real-world clinical impact of genomic risk assessment, address critical evidence gaps, advance understanding of genomic medicine outcomes, and inform future research.

Introduction

Chronic diseases such as diabetes, heart disease, kidney disease, and cancer are the leading causes of morbidity and mortality in the United States: 60% of adults and 30% of children have at least one chronic disease, accounting for over 90% of the $4.5 trillion annual healthcare costs.[1–5] Most of this spending is directed toward treating rather than preventing disease, with less than 3% devoted to public health and prevention.[6]

Although the deployment of genomics in the diagnosis and treatment of diseases is increasing, its greatest

[1]Department of Neurology, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA; [2]Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, USA; [3]Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, USA; [4]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA; [5]Center for Genetic Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; [6]Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA; [7]Department of Biomedical Informatics and Data Science, The University of Alabama at Birmingham, Birmingham, AL, USA; [8]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; [9]Division of Biomedical Informatics and Genomics, John W. Deming Department of Medicine, Tulane University School of Medicine, New Orleans, LA, USA; [10]Division of Genetic Medicine and Clinical Pharmacology, Vanderbilt University Medical Center, Nashville, TN, USA; [11]Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA; [12]Stanford University, Stanford, CA, USA; [13]Department of AI and Informatics, Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA; [14]Department of Medicine, University of Colorado School of Medicine, Aurora, CO, USA; [15]Department of Biomedical Informatics, Columbia University, New York, NY, USA; [16]School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA; [17]Department of Biomedical Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; [18]Division of Rheumatology, Inflammation, and Immunity, Department of Medicine, Massachusetts General Brigham, Boston, MA, USA; [19]Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [20]Division of Nephrology, Department of Medicine, Columbia University, New York, NY, USA; [21]Department of Health Services, University of Washington School of Public Health, Seattle, WA, USA; [22]Department of Genetics, Heersink School of Medicine, The University of Alabama at Birmingham, Birmingham, AL, USA; [23]Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH, USA; [24]Division of Cardiovascular Medicine, Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, USA; [25]Broad Institute of MIT and Harvard, Cambridge, MA, USA; [26]Division of Genomic Medicine, NHGRI, NIH, Bethesda, MD, USA; [27]Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; [28]Department of

(Affiliations continued on next page)

Elizabeth M. McNally,[5] Jennifer Morse,[11] Shawn Murphy,[27] Bahram Namjou,[23] Jacqueline A. Odgis,[19] Lori Orlando,[28] Jennifer Pacheco,[29] Josh F. Peterson,[30] Daniel E. Pineda-Alvarez,[31] Cindy Prows,[23] Megan Puckelwartz,[5] Jasmine Purcell,[8] Laura Rasmussen-Torvik,[32] Heidi Rehm,[25] Dan M. Roden,[33] Elisabeth A. Rosenthal,[4] Robb Rowley,[27] Maya Sabatello,[34,35] Daniel Schaid,[36] Richard Sharp,[36] Johanna L. Smith,[36] Jordan W. Smoller,[37] Emily R. Soper,[19] Rene Sterling,[38] Sabrina A. Suckiel,[19] Shannon Terek,[8] Jeritt Thayer,[17] Yi-Lee Ting,[31] Hemant Tiwari,[2] Digna Velez-Edwards,[39] Kavishwar B. Wagholikar,[27] Theresa Walunas,[40] Wei-Qi Wei,[30] Chunua Weng,[15] Georgia Wiesner,[41] Noura S. Abul-Husn,[19,42,44] and David L. Veenstra[43,44]

potential may lie in preventing, delaying, or attenuating disease. Foundational developments in genomics technology and research have propelled our understanding of the genomic basis of chronic diseases. Recently, statistical methodological advances have leveraged genome-wide association studies (GWASs) to create polygenic risk scores (PRSs) that aggregate the effects of many common genetic risk variants to estimate the genetic component of an individual's risk of developing a disease or an extreme disease-associated endophenotype such as LDL cholesterol.[7,8] The confluence of these developments creates an opportunity to integrate the power of genomics with a participant's personal and family health history (FHHx) to identify individuals at the highest risk of disease who could benefit from preventive actions. However, there is limited evidence of the impact of returning genomic risk assessments that include PRS, rare high-effect-size monogenic variants, and FHHx with clinical risk factors in clinical care. Clinical genomics is thus at a crossroads with regard to risk prediction, evidence generation, and implementation policies.

There are three key challenges. First is the tension between methodological rigor and practical feasibility. Although large, prospective randomized controlled trials (RCTs) are considered the gold standard, conducting RCTs of genomic screening for prevention of a disease is inherently difficult due to the relatively low absolute risk of most diseases in the general population. Selecting appropriate controls—such as those not receiving any risk assessment, usual care, or family history-based risk assessment—further complicates study design. Embarking on an assessment for multiple diseases can be even more challenging. In contrast, cohort studies, while more feasible and less expensive, have inherent limitations when used to inform practice guidelines, particularly in a broad population. The key objective is obtaining an unbiased estimate of the effect of providing genomic risk information to participants and their providers, and that requires definition of case and appropriate control groups. Many observational studies lack sufficient clinical information to adjust for potential confounders, and those that do remain vulnerable to confounding from unobserved variables. There is an opportunity to design prospective cohort studies to address these limitations by leveraging key properties of PRSs. One promising quasi-experimental approach taking advantage of the fact that PRS is a continuous variable is known as regression discontinuity (RD), which examines outcomes in individuals above and below a predefined risk threshold, approximating randomization.[9]

The second challenge is that genomic risk prediction is rapidly evolving due to improved sequencing technologies, use of larger datasets, and increasingly robust analytic methods. While these advances enhance predictive performance, they present a challenge for evidence generation for a given intervention. A comprehensive approach in which risk-prediction tools are developed, validated, and implemented within a single study can help ensure that predictions reflect current best practices and are validated in the study population.

The third challenge is the selection of study outcomes in prevention trials. Definitive outcomes such as disease onset may occur many years or decades beyond risk assessment and preventive care. Therefore, it is important to both (1) measure clinical actions that are evidence based and have been shown to improve participant outcomes in other settings, and (2) pre-specify key healthcare actions and outcomes to avoid spurious findings. While it is expected that participants deemed high risk will receive more recommended healthcare actions than those not at high risk, the magnitude of the effect is unknown. Thus, a key goal is to estimate the effect size, specifically to

Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA; [29]Center for Genetic Medicine, Northwestern University, Evanston, IL, USA; [30]Department Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA; [31]Labcorp (formerly Invitae Corp.), San Francisco, CA, USA; [32]Department of Preventative Medicine, Northwestern University, Chicago, IL, USA; [33]Departments of Medicine, Pharmacology, and Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA; [34]Center for Precision Medicine and Genomics, Columbia University Irving Medical Center, Columbia University, New York, NY, USA; [35]Division of Ethics, Department of Medical Humanities and Ethics, Columbia University Irving Medical Center, Columbia University, New York, NY, USA; [36]Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA; [37]Department of Psychiatry and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; [38]Division of Genomes and Society, NHGRI, NIH, Bethesda, MD, USA; [39]Division of Quantitative Science, Department of Obstetrics and Gynecology, Department of Biomedical Sciences, Vanderbilt University Medical Center, Nashville, TN, USA; [40]Department of Medicine and Center for Health Information Partnerships, Northwestern University, Chicago, IL, USA; [41]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; [42]23andMe Research Institute, Palo Alto, CA, USA; [43]School of Pharmacy, University of Washington, Seattle, WA, USA
[44]These authors contributed equally
*Correspondence: nlimdi@uabmc.edu
https://doi.org/10.1016/j.ajhg.2026.02.018.

what degree healthcare providers and participants modify their behaviors to enhance prevention.

The purpose of this manuscript is to describe the study design and statistical analysis framework for evaluating the clinical implementation of a genome-informed risk assessment (GIRA) within the Electronic Medical Records and Genomics (eMERGE) study, with particular emphasis on PRSs. We propose a feasible, rigorous approach to estimate the impact of returning integrated genomic risk information on (1) adoption of pre-specified recommendations, (2) diagnosis of disease, (3) escalation of clinical care, and (4) clinical outcomes.

Subjects and methods

Overall approach

The eMERGE IV study aims to develop, validate, and evaluate the impact of integrating a GIRA into clinical care to support the prevention of chronic diseases (Figure 1).[10,11] Participant recruitment processes, sample collection and genomic risk assessment have been previously described and are summarized below.[10,11]

We developed and validated a PRS for 10 common, chronic diseases relevant to adults and children: asthma,[12] atrial fibrillation (Afib), breast cancer (BCa),[13] chronic kidney disease (CKD),[14] coronary heart disease (CHD),[15] hypercholesterolemia (HC), obesity, prostate cancer (PCa),[16] type 1 diabetes (T1D),[17] and type 2 diabetes (T2D).[18] Monogenic risk (but not PRS) for colorectal cancer (CRC) was also included.[10] For BCa, we developed an integrated risk score (IRS) based on the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) algorithm[19] incorporating PRS, monogenic risk, personal history (including lifestyle and reproductive history), and FHHx.

We then compiled genomic risk (PRS, IRS, and monogenic results) with FHHx and clinical risk factors into a GIRA report to communicate results to participants and their healthcare providers. Importantly, the GIRA report included recommended healthcare actions for participants at high risk based on professional society guidelines and/or expert consensus.[20] These recommendations included additional testing, earlier screening, lifestyle changes, and treatment initiation or intensification. Follow-up commenced at the time of GIRA upload to the electronic health record (EHR).

Analyses will compare outcomes among participants receiving high risk vs. not-high-risk GIRA. Outcomes assessed include pre-specified provider and participant actions, new diagnosis of disease, treatment initiation or intensification, and clinical outcomes such as risk factor control (e.g., LDL, BMI, hemoglobin A1c [HbA1c]) and clinical events (e.g., cardioembolic event, exacerbation).

The eMERGE IV 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 was required from all enrolled participants.

Participants

Eligible participants were able to read or understand English or Spanish, were willing to receive genomic risk results, and had a healthcare provider at their recruitment site able to receive results. Participants who were unable to provide consent (adults only), had a history of transfusion (within 8 weeks) or transplant (solid organ or bone marrow), were unable to identify a healthcare provider at their recruitment site, or were not receiving care at the parent institution were excluded from the study. eMERGE investigators and research staff were also excluded. Each site implemented outreach, engagement, and recruitment approaches to their local clinical care workflows and developed site recruitment goals (Table S1) based on the participant populations served, to ensure broad inclusion and representation of participants.

Participant samples were analyzed at the Broad Institute Clinical Services Laboratory (adult and pediatric participants) for polygenic risk assessment and Invitae (now part of Labcorp) (adults only) for monogenic testing. For adult participants, PRSs were generated for Afib, CKD, CHD, HC, obesity, PCa, and T2D, and IRS was generated for BCa. For pediatric participants, PRSs were generated for asthma, obesity, T1D, and T2D. For adult participants, we sequenced and reported pathogenic or likely pathogenic (P/LP) variants in 16 genes (Table 1) to complement the PRS/IRS assessment for the following conditions: Afib, BCa, CHD, HC, PCa, and CRC.[10]

The GIRA report is a participant-facing summary that presents categorical genomic risk results together with contextual family history and selected clinical risk factors (Table 1). Genomic risk was assessed using PRSs, IRS (for BCa), and monogenic findings. For each condition, PRS or IRS thresholds for a high-risk designation were selected to correspond to a clinically meaningful level of risk, such as a positive family history or an odds ratio ≥ 2. The GIRA report presented categorical risk status (high versus not high).[10,11,19]

Monogenic risk was incorporated for six chronic diseases for which P/LP variants in 16 disease-associated genes were reported. Family health history was displayed on all GIRA reports as a contextual risk factor accompanied by a pedigree when available. A positive FHHx resulted in a high-risk GIRA designation for four conditions: Afib, PCa, CHD, and CKD. Clinical risk factors were included as contextual information for five conditions (asthma, CKD, HC, obesity, and T2D) and were displayed as laboratory values, blood pressure measurements, and International Classification of Diseases (ICD) codes representing known diagnoses. Clinical risk factors alone did not trigger a high-risk GIRA designation. In this issue, Lawson et al. present the prevalence of high-risk triggers in participants receiving GIRA.[21]

Clinical data were used in the calculation of absolute risk for two conditions: the IRS (BOADICEA)[19] for BCa and the pooled cohort equations (PCEs) for CHD.[22] Although these models incorporate clinical variables (and PRS in the case of the IRS for BCa), only the resulting categorical risk designation was displayed on the GIRA report.

Reports for participants identified as high risk by GIRA included healthcare recommendations aimed at risk reduction. Participants who received GIRA reports without a high-risk designation did not receive healthcare recommendations. Recommendations for participants with monogenic or FHHx-based risk were informed by established clinical practice guidelines.[23–25] Recommendations for participants with elevated PRS/IRS were developed by the investigative team through an iterative process with input from clinical experts and have been described previously.[10]

Return of GIRA reports

All participants received their reports. Reports were returned to participants by mail, health portal, or in person and shared with providers via the institutional EHR. Those identified as high risk based on PRS/IRS and/or monogenic findings (but not FHHx) were invited to review results with a genetic counselor,

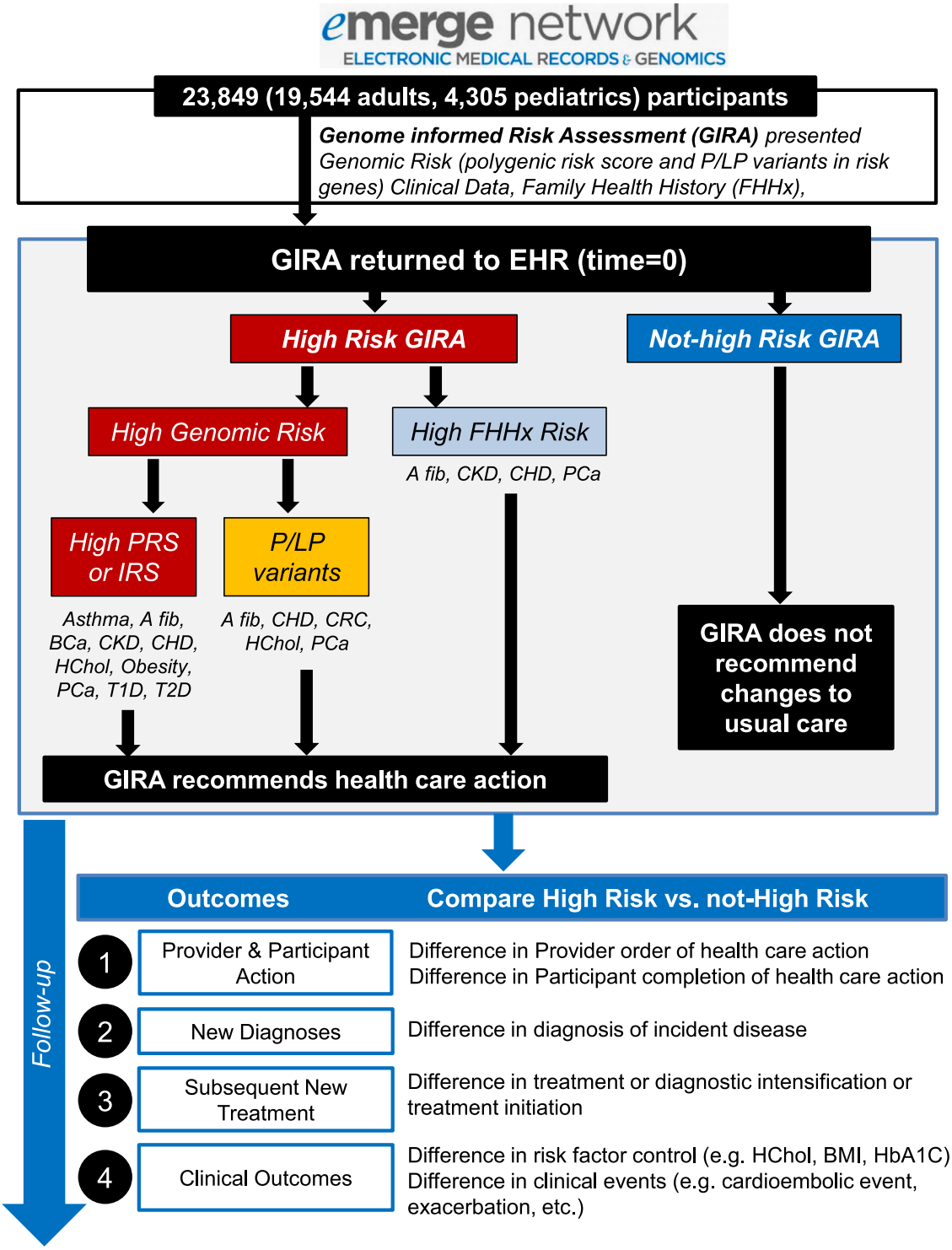


Figure 1. Overall study design of the eMERGE IV cohort, including participant enrollment, return of GIRA reports, follow-up, and cross-condition and condition-specific analyses

Clinical data, family health history (FHHx), survey data, monogenic risk, and polygenic risk are collected and combined to create the GIRA report. Participants identified as having high genomic risk receive in-person return of results. Participants with high risk based on family health history (FHHx) alone, or those with a not-high-risk GIRA, received results through the electronic health record (EHR), patient portal, or mail. The GIRA includes recommended healthcare actions for primary-care providers. Patient and provider adoption of healthcare recommendations will be collected and assessed over a 6- to 12-month follow-up period. New diagnosis of incident disease, treatment initiation or intensification, and clinical outcomes will be collected and assessed over a longer follow-up period (up to 5 years). A fib, atrial fibrillation; BCa, breast cancer; CKD, chronic kidney disease; CRC, colorectal cancer; CHD, coronary heart disease; HChol, hypercholesterolemia; PCa, prostate cancer; T1D, type 1 diabetes; T2D, type 2 diabetes; IS, integrated score used to calculate BCa risk based on the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA); P/LP, pathogenic/likely pathogenic; EHR, electronic medical record; BMI, body mass index; HbA1C, hemoglobin A1C.

Table 1. Triggers designating high-risk status for chronic conditions in the eMERGE IV study

Condition	Expected % with high PRS	Expected % with P/LP variant (genes)	Expected % with FHHx criteria
Asthma	5%	–	–
Atrial fibrillation	3%	0.05% (*LMNA*)	5% (Afib in parents at age <75 years)
CKD	2%	–	9.0% (first degree)
CRC	–	0.54% (*EPCAM*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *STK11*, *PTEN*, *TP53*)	–
CHD	5%	0.89% (*APOB*, *LDLR*, *PCSK9*, *LDLRAP1*)	9.8% (first degree)
HC	3%	0.89% (*APOB*, *LDLR*, *PCSK9*, *LDLRAP1*)	–
Obesity	3%	–	–
PCa	10%	1.69% (*BRCA1*, *BRCA2*, *EPCAM*, *MLH1*, *MSH2*, *MSH6*, *PMS2*)	9.4% (first degree)
T1D	3%	–	–
T2D	2%	–	–
BCa[a]	3% females expected to have an Integrated Score (BOADICEA) ≥25%		

The 10-year risk of CHD based on PCEs does not trigger CHD risk reporting; it is provided as a supplemental piece of information.
P/LP, pathogenic/likely pathogenic; FHHx, family health history.
[a]BCa risk is returned as an integrated score based on the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA), which includes P/LP variants (*BRCA1*, *BRCA2*, *PALB2*, *PTEN*, *TP53*, *STK11*) + PRS + personal history + FHHx + reproductive Hx + others.

pharmacist, study physician, or trained study staff member by phone, in person, or virtually. In this issue, Lawson et al. describe the GIRA return and factors associated with return of results by genetic counselor, pharmacist, study physician, or trained study staff member by phone, in person, or virtually.[21]

Results

Outcomes for all conditions are presented in Table 2, and include (1) pre-specified provider orders and participant adoption of recommended clinical actions, (2) new diagnosis of disease, (3) initiation or intensification of a diagnostic workup or treatment, and (4) differences in clinical outcomes as assessed by surrogate markers (e.g., HbA1c) or clinical events (e.g., exacerbations, stroke, MI).

Timeline for outcome assessment

The impact of GIRA return on adoption of recommended healthcare action by providers and participants will be assessed at 6 and 12 months. The impact of GIRA on new diagnosis, escalation of care, and clinical outcomes will be assessed over a 5-year follow-up period.

Data sources

A key goal of the eMERGE IV study was to streamline outcome assessment using the EHR.[26–28] EHR-based outcomes were measurable for the conditions; the key outcome for PCa was informed decision-making between the provider and participant, which was captured in a participant survey, described below. A pre-specified list of desired laboratory values was created by investigators and extracted from the EHR using Logical Observation Identifiers Names and Codes (LOINC) codes. Drug orders were identified using a pre-specified list of RxNorm codes for key drugs. All ICD and CPT codes were collected for each participant. Visit data for participants included four numerical codes for visit type (inpatient, outpatient, emergency, or other) using the Observational Medical Outcomes Partnership (OMOP) standard concept IDs. Physician orders for laboratory tests and procedures were collected at each site by manually compiling a list of key orders. While all ICD and CPT codes were collected, the data collection for meds, labs, and visits was limited to the period 2017 to the present.

To complement the EHR-based outcomes and collect participant-centered outcomes, we developed a pre-return participant survey and a post-return survey that was deployed 6 months after GIRA return. The pre-return survey captured information across several domains, including demographics, lifestyle and environmental factors, insurance status, and social determinants of health (Table S2). FHHx was self-reported using either MeTree[29] or an abbreviated survey that documented relevant conditions among first- and second-degree relatives. The post-return survey assessed additional domains including the return experience, healthcare actions taken, genomic knowledge, emotional responses, and perceived risk. A specific question about provider-participant informed decision-making for high-risk PCa finding was also included. A 6-month survey was selected to balance participant recall with sufficient follow-up to capture recommended actions and participant perceptions.

Analysis framework

To evaluate the impact of different types of risk information we will conduct stratified analyses based on the type of risk (PRS/IRS, monogenic, or FHHx) identified in the GIRA report. Figure 2 presents the overall framework for each analysis: the composition and estimated proportions of groups classified as high risk and not high risk and

Table 2. Outcomes for all conditions including adoption of pre-specified recommendations, diagnosis of disease, initiation or intensification of a diagnostic or treatment, and clinical outcomes

Condition	Pre-specified recommended healthcare action	Provider (participant) action	New diagnoses	Care intensification or initiation	Clinical outcomes
Asthma	risk assessment visit	ordered (completed) visit	asthma	albuterol/daily inhaler	reduce asthma exacerbations
Afib	ECG	ordered (completed) ECG	AF	proc: cardioversion, ablation, pacemaker Rx: antiarrhythmics, anticoagulants	reduce AF hospitalization, cardioembolic event, stroke
BCa	breast MRI	ordered (completed) breast MRI	BCa	proc: prophylactic mastectomy Rx: tamoxifen, raloxifene, etc.	earlier detection of BCa allowing preemptive prevention
CKD	urine protein or UACR, SCR	ordered (completed) lab test	CKD	Rx: ACEI/ARB, SGLT2 blocker, discontinue NSAID	maintenance of eGFR
CRC	colon screening (e.g., colonoscopy, FIT test)	ordered (completed) screening	CRC	proc: flexible sigmoid, scope, barium enema	earlier detection of CRC allowing preemptive prevention
CHD	lipid profile, imaging (coronary calcium CT, carotid ultrasound)	ordered (completed) lipid profile or imaging	CHD, CHD risk reclassification	Rx: lipid-lowering, anti-platelet or anti-hypertensive drugs	LDL-C levels, angina, MI, stroke
HC	lipid profile	ordered (completed) lipid profile	HC	Rx: lipid-lowering agent	LDL-C levels, angina, MI, stroke
Obesity	referral (weight loss or nutrition consultation)	ordered (completed) referral	obesity	proc: gastric surgery Rx: GLP-1 agonists, other weight-loss meds	reduce BMI, weight loss
PCa	shared decision-making,[a] PSA test	ordered (completed) PSA test	PCa	proc: prostatectomy PSA test frequency	earlier detection of PCa allowing preemptive prevention
T1D	autoantibody test	ordered (completed) autoantibody test	T1D	Rx: insulin, teplizumab	normoglycemia, lower HgA1C
T2D	HgbA1c test	ordered (completed) HgbA1c test	T2D prediabetes	Rx: metformin	normoglycemia, lower HgA1C

Rx, prescription; proc, procedures. ECG, electrocardiogram; UACR, albumin-creatinine ratio; SCR, serum creatinine; GFR, glomerular filtration rate; NSAID, non-steroidal anti-inflammatory drug; CT, computed tomography; PSA, prostate-specific antigen; HbA1c, hemoglobin A1c; AF, atrial fibrillation; BCa, breast cancer; CKD, chronic kidney disease; CRC, colorectal cancer; CHD, coronary heart disease; HC, hypercholesterolemia; MI, myocardial infarction; PCa, prostate cancer; T1D, type1 diabetes; T2D, type 2 diabetes).
[a]Shared decision-making evaluated at 6 months only.

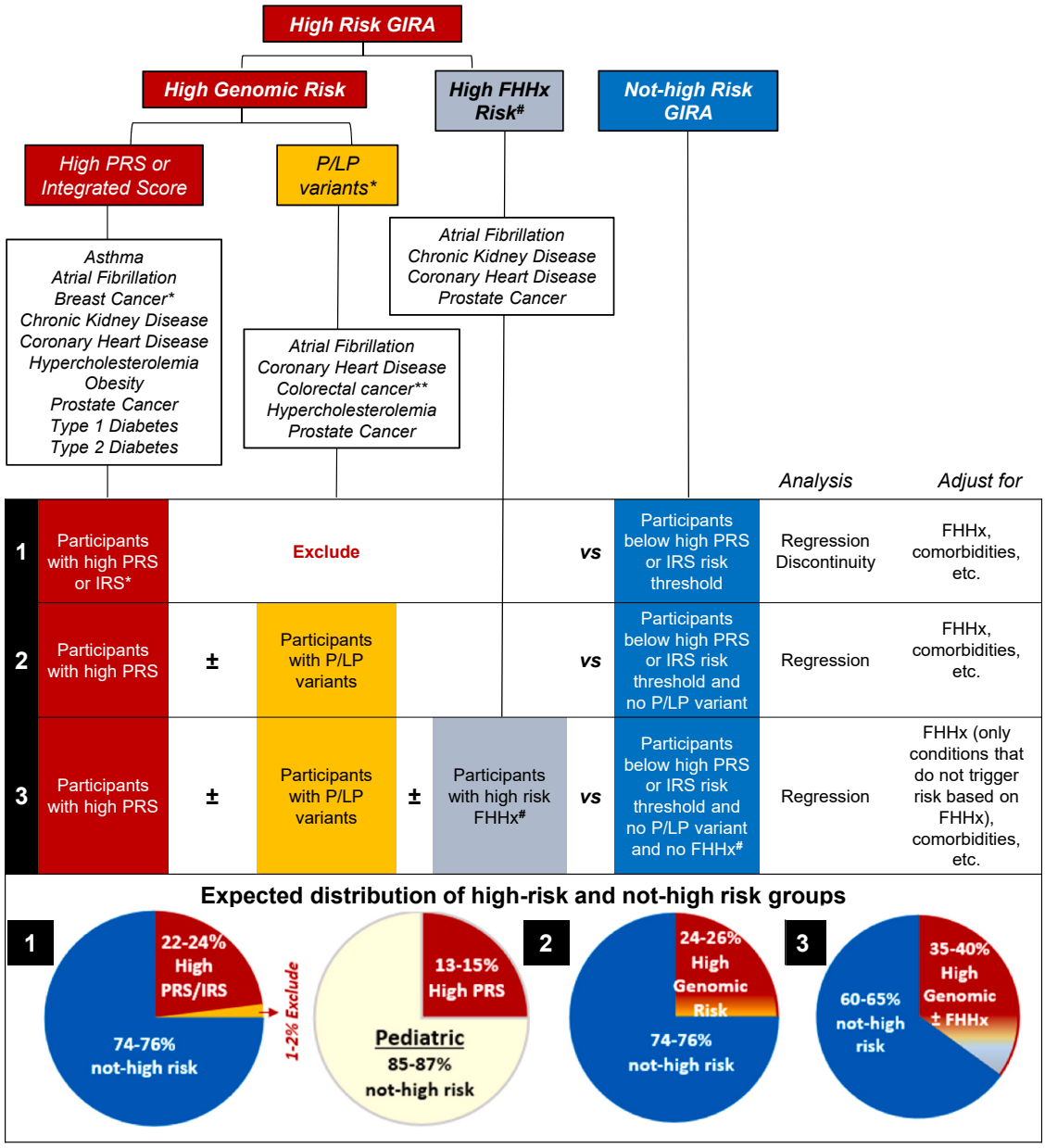
the planned statistical approach. Within each analysis we will access the outcome categories noted above. We provide the definition of high risk for each of the proposed analyses.

1. Proposed analysis 1: PRS or IRS above the pre-established cutoff for 10 conditions (asthma, Afib, BCa, CKD, CHD, HC, obesity, PCa, T1D, T2D). Participants receiving high-risk PRS/IRS form the intervention group and those receiving not high risk below form the comparator group.
2. Proposed analysis 2: PRS or IRS above the pre-established cutoff and/or any monogenic risk for five conditions (Afib, CHD, HC, CRC, PCa). Participants receiving high-risk PRS/IRS or P/LP variants are the intervention group and those receiving not high risk form the comparator group.
3. Proposed analysis 3: PRS or IRS above the pre-established cutoff, and/or any monogenic risk, and/or FHHx risk for four conditions (Afib, CHD, CKD, PCa). Participants receiving high-risk PRS/IRS or P/LP variants or FHHx risk for the four conditions form the intervention group and those receiving not high risk form the comparator group.

Proposed analysis 1: PRS/IRS for 10 conditions

The return of high-risk PRS/IRS information, assigned at clearly defined thresholds using a continuous risk score, enables a pseudo-randomized experiment using a RD analysis to account for unobserved confounders while making it possible to potentially infer causality without randomization.[9] To isolate the attributable impact of risk information in the GIRA, we will utilize control groups selected from the non-high-risk population and the RD as our analytic approach.

We will use RD analysis to evaluate the effect of classifying and communicating high-risk status plus healthcare



*Breast Cancer – P/LP variants and FHHx are incorporated into the integrated risk score (IRS)
FFHx triggers high risk for Afib, CKD, CHD, PCa.
**Colorectal cancer assessment was based on P/LP variants alone

Figure 2. Proposed analysis framework and expected distribution of high-risk and not-high-risk groups
GIRA, genome-informed risk assessment; FHHx, family health history; PRS, polygenic risk score; P/LP, pathogenic/likely pathogenic.

recommendations (i.e., the intervention) on provider and participant adoption of the healthcare recommendations, leveraging the pseudo-randomized design to control unobserved confounders. Participants may have PRS/IRS and recommendation adoption rates for multiple diseases; therefore, each disease will be analyzed separately. For each condition, the comparator group will consist of individuals with PRS/IRS below the risk threshold who did not receive recommendations. The differences in adoption of healthcare recommendations will be assessed using a generalized linear mixed model (GLMM). To apply RD analyses to these GLMMs, the PRS for each disease will be centered at its high-risk (HR) threshold and take the general form of

$$\text{Model 1: } \mu\left(y_j\right) = \beta_0 + \beta_P\left(PRS_j - c_j\right) + \beta_H\left(HR_j\right) + \beta_S(S),$$

where y_j is the primary outcome (adoption) for the j^{th} disease; $\mu(y_j)$ is the mean function of y_j; PRS_j is the PRS for the j^{th} disease; c_j is the PRS (RD) threshold for the j^{th} disease; β_P is the regression coefficient for the partial relationship of PRS to y; HR_j is a dummy-coded variable where 0 represents

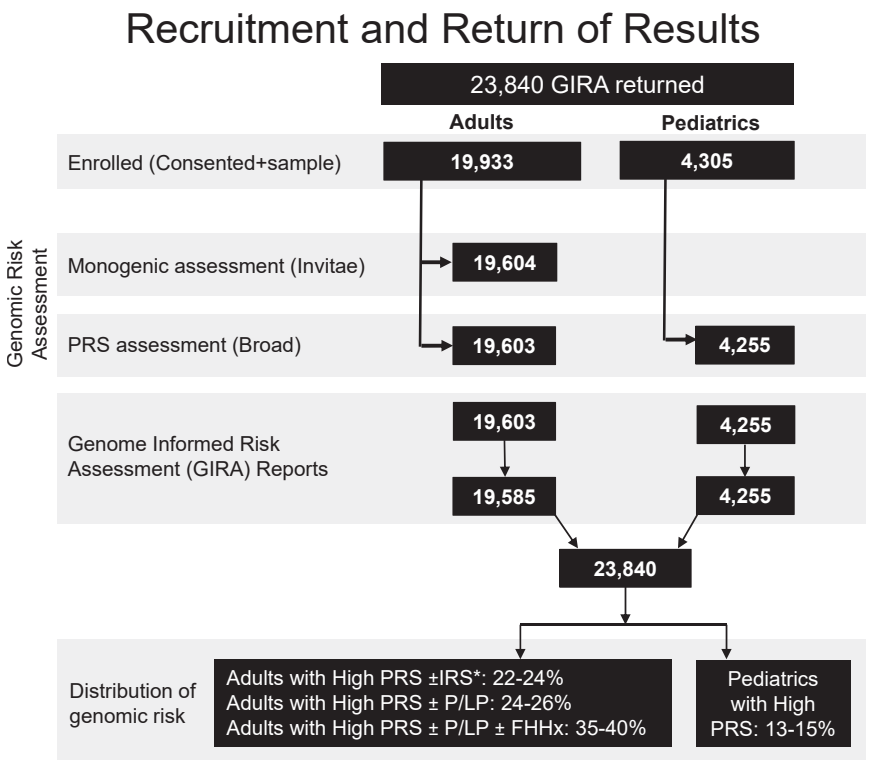


Figure 3. Recruitment and return of results for eMERGE IV participants
PRS, polygenic risk score; IRS, integrated risk score used to calculate BCa risk based on the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA); P/LP, pathogenic/likely pathogenic; FHHx, family health history.

not-high-risk (nHR) classification and 1 represents high risk classification for the j^{th} disease; β_H is the regression coefficient for the partial relationship of risk classification to y and represents the adjusted difference between high risk and not high risk at the PRS threshold for all diseases, which is our primary result of interest; S is a set of coding vectors that represent the site membership (choice of reference site is arbitrary); β_S is a set regression coefficients representing how y differs by network site; and finally, because the PRS is centered, β_0 is the model intercept, which represents the adjusted adoption rate for not high risk at the PRS threshold for the reference site.

Given that the GIRA report only displays risk classification (high or not high) but not the quantitative PRS value itself, we anticipate β_P to be virtually zero for both the high-risk and not-high-risk group and, therefore, assume that the adoption(y)-PRS relationship is linear. However, we will investigate potential non-linearity with graphical methods, and, if necessary, perform models with polynomial or piecewise (spline) terms. We assume that the β_P regression slope will be the same for either risk classification. We will fit HR $\times$ PRS interaction terms to investigate differences in the β_P regression slopes across risk status.

We expect PRS distribution to be similar across the 10 sites from nine states and that site will be independent (uncorrelated) from the risk result (PRS/IRS, monogenic risk, and FHHx) of any disease. We will model sites as a random effect to account for dependency within sites. Because participants will not know their PRS value and are unlikely to communicate health care action adoption status, we anticipate adoption rates to be similar across sites and the dependency due to participants nested in states/sites to be small. We will calculate the intra-class correlation (ICC) to assess the amount of dependency. If the ICC is small, we will consider modeling state/site as a fixed effect, which would reduce model 1 to a logistic regression model.

Proposed analyses 2 and 3: PRS/IRS and/or monogenic risk information for five conditions and combined genomic and/or FHHx risk for four conditions

For the five conditions (Afib, BCa, CRC, CHD, HC, PCa) for which monogenic risk was also assessed and for the four conditions (Afib, CHD, CKD, PCa) for which a positive FHHx also denoted high risk, we will conduct modified RD analyses (covariate adjusted models). Because subjects are reclassified from the not-high-risk (reference) group to a genomic risk group or combined genomic $\pm$ FHHx risk group, the correlation between PRS and the dummy-coded variable representing group status will not be the same as the correlation expected in RD analyses. However, the PRS score will still be correlated with the comparison group status and will be used as a covariate in a regression model. We will also compare the not-high-risk group to the genomic risk group or combined genomic $\pm$ FHHx risk group on a full set of demographic variables to examine whether any of these variables show statistically significant or "substantial" between-group differences and should be included as covariates in the model.

We will examine the distribution of demographic (e.g., age, gender, race/ethnicity) and other background variables (e.g., disease burden, FHHx) across the high-risk and not-high-risk groups to assess potential confounding factors. If substantial group differences are found, these variables will be added as covariates for each of the above analyses.

The EMERGE IV cohort

The eMERGE cohort has recruited and returned GIRA to 23,840 participants (Figure 3; 19,585 adults and 4,255 children), with a robust representation of racial and ethnic minority groups (48%) and medically underserved populations. In this issue, Lawson et al. describe the eMERGE cohort receiving GIRA reports.[21] Participants

to date have accrued approximately 12 months of follow-up. Given the planned interval of analysis, sample size, and heterogeneity of recommendations and outcomes across multiple conditions, the category of "provider action" based on recommendations and "participant receipt of recommended" actions across all conditions were selected as the initial key study outcomes. To avoid multiple hypothesis testing given the multiple potential recommendations for several conditions, one or two key clinical recommendations (Table 2) were identified for each condition.

Expected risk distribution and effect size to achieve 80% power at $\alpha = 0.05$ and 0.01

Effect-size calculations were based on the eMERGE IV cohort sample size, and the proportion of participants expected *a priori* to receive high-risk GIRAs. As genomic risk assessment is not routine in most primary-care clinical practice, the frequency of recommended actions observed in clinical practice informed the proportion of not-high-risk participants expected to receive the recommended action (e.g., laboratory test, imaging order etc.). Based on observed adoption rates in the not-high-risk group (p_{nHR}) in routine clinical care, we present the detectable differences in proportions (p_D) that would need to be observed in the HR group to obtain 80% statistical power at $\alpha = 0.05$ and 0.01.

The RD design creates a collinearity between the PRS covariate and the variables of interest (high risk). The expected correlation between high risk and PRS can be determined from the RD/PRS threshold,[30] and, assuming that the PRS will be normally distributed, we calculated the correlation as a function of truncated Gaussian distributions.[31,32] Because model (1) contains the high-risk variable of interest and the correlated covariate (PRS), calculations were completed using SAS PROC POWER with the LOGISTIC statement and its various options to obtain detectable odds ratios (ORs), which were then transformed into absolute detectable differences in proportions (p_D). As providers and participants, both high risk and not high risk, only know their status as high or not-high-risk categories—and not their exact PRS or percentile risk—we anticipate the regression slope to be virtually zero. We also assume the site is virtually unrelated to the adoption outcome (i.e., low ICC) and to PRS and high risk (i.e., low collinearity). For the genomic risk group or combined genomic ± FHHx risk group vs. those without these risks' analyses, a modification to the RD correlation formula was used to estimate the expected collinearity in the RD analyses.

Figure 4 presents the effect-size differences that the eMERGE IV study is powered to detect. These graphs show detectable differences in the adoption of key recommendations to achieve 80% power at $\alpha = 0.05$ or 0.01. Additionally, effect sizes are presented assuming 100%, 80%, and 60% participant follow-up completion. For asthma, obesity, and T1D and T2D, we present effect sizes based on estimates for both adult and pediatric cohorts to allow for varying adoption rates (Figure 5). We present the effect-size differences the eMERGE IV study is powered to detect for analyses 1, 2, and 3 assuming 100%, 80%, and 60% participant follow-up completion, respectively (Tables S3, S4, and S5). Based on preliminary analyses, we expect to be able to detect 2%–11% differences in adoption with 80% power at $\alpha = 0.05$ and 3%–14% at $\alpha = 0.01$.

Discussion

Genomic medicine is increasingly recognized for its potential to transform disease prevention and management, particularly through improved risk stratification. Screening for relatively rare monogenic variants in disease genes has demonstrated clinical utility in identifying individuals at elevated risk for actionable conditions. However, this approach inherently benefits only a small proportion of the population, typically 1%–5%. While disclosure of such results often leads to meaningful uptake of evidence-based, risk-reducing interventions, a substantial number of at-risk individuals remain undetected through routine clinical care.[28,33–35] In contrast, PRSs offer the ability to stratify risk across a much broader segment of the population, up to 25%–30%, by assessing aggregate genetic risk for common chronic diseases. Unlike monogenic conditions, where variant pathogenicity, penetrance, and clinical actionability are relatively well established, PRS-based risk stratification provides a more probabilistic and nuanced estimate of disease susceptibility. To this end, our prior work defined actionability thresholds and expert-guided recommendations for those at high risk.[10,11,36,37] Central questions that now arise, and will be addressed by the eMERGE IV study, include whether PRS can be assessed and used to stratify participants' risk status for multiple chronic conditions, whether PRS information along with other types of risk information be effectively integrated into clinical care workflows alongside recommended health care actions, whether providers and participants will adopt these recommendations, and whether such adoption will lead to improved clinical outcomes across multiple chronic diseases.

To address these questions, the eMERGE IV study was designed to evaluate the impact of delivering genomic risk information, including PRS, IRS, monogenic risk, clinical risk factors, and FHHx, for common chronic diseases. This paper presents the analytical framework for assessing provider and participant actions and downstream health outcomes following the return of genomic risk information.

We employed a prospective cohort design, allowing participant recruitment and follow-up within routine clinical practice. Genomic risk results were returned to providers and participants using methods consistent with current clinical practice. To enable causal inference in this context, we utilized an RD study design. RD is a quasi-experimental method that leverages predefined

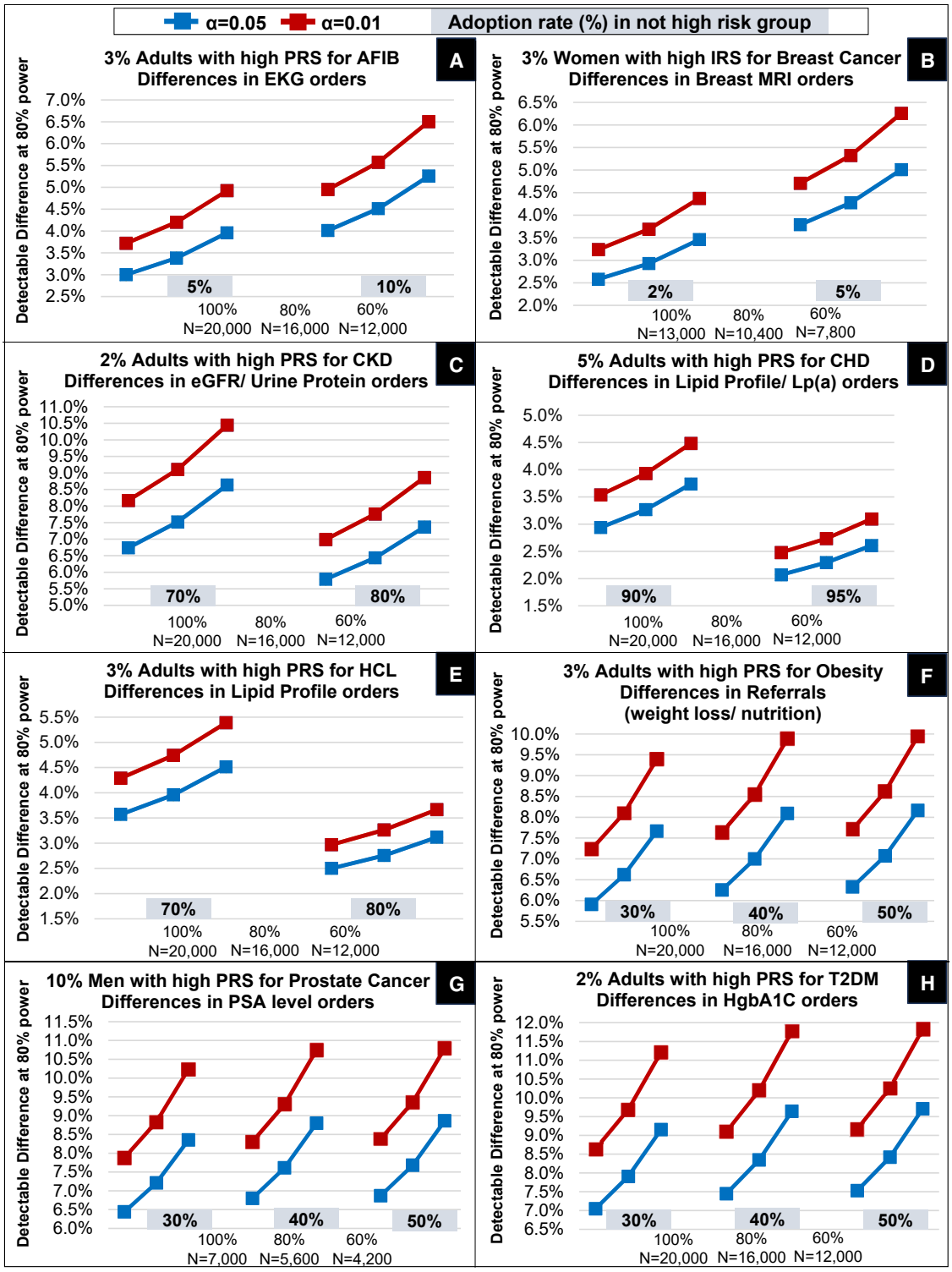


Figure 4. Effect-size differences in the eMERGE IV study

These are powered (80% power at $\alpha = 0.05$ and 0.1) for assessment of high PRS/IRS vs. not-high PRS/IRS for (A) atrial fibrillation, (B) BCa, (C) CKD, (D) CHD, (E) HC, (F) obesity, (G) PCa, and (H) type-2 diabetes assuming 60%, 80% and 100% participant follow-up completion.

thresholds (e.g., the top 3% of a PRS distribution) to approximate random assignment within observational data. When individuals just above and below a given risk threshold are otherwise comparable, differences in downstream outcomes, such as uptake of preventive screening, clinical management, or lifestyle interventions, can be attributed to the effect of being classified as high risk and receiving corresponding care recommendations.

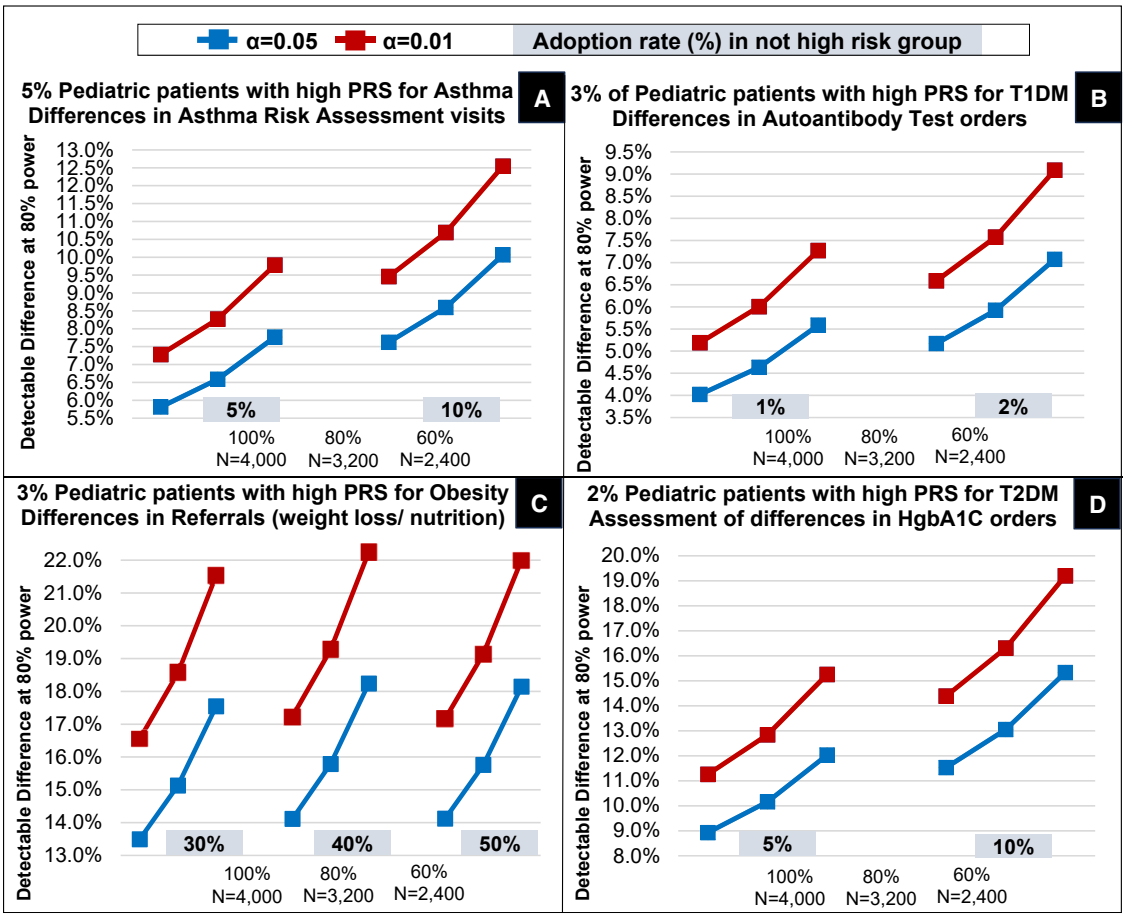


Figure 5. Effect-size differences in the eMERGE IV study

These are powered (80% power at $\alpha = 0.05$, 0.1) for assessment of high PRS/IRS vs. not-high PRS/IRS for (A) asthma, (B) type 1 diabetes, (C) obesity, and (D) type 2 diabetes, assuming 60%, 80%, and 100% participant follow-up completion.

In essence, RD exploits clinical decision rules where participants fall just on either side of a threshold, such as a high PRS or IRS classification. This enables estimation of the causal effect of genomic risk disclosure and risk-guided care. RD has been successfully applied in analogous domains of risk-guided care, such as statin prescribing based on atherosclerotic cardiovascular disease scores and diabetes diagnosis using HbA1c thresholds.[9,38–42]

A selected set of outcomes was identified *a priori* to minimize multiple hypothesis testing and reduce the risk of *post hoc* result fishing. The initial evaluation time frame includes 12 months of EHR follow-up and a 6-month participant survey to capture provider and participant actions. These timelines were informed by the original funding requirements and aligned with recruitment and return-of-result milestones. Within the initial time frame, cross-condition analyses focus on provider and participant actions taken following receipt of the GIRA. These are critical process outcomes that are measurable in the short term and are necessary for longer-term outcomes that may not be observable in the original funding period.[43] Because most outcomes are derived from EHR data, the infrastructure established through this study provides a strong foundation for longer-term follow-up, which is in the planning stages. This extended period will allow for the assessment of the long-term maintenance of risk-reducing interventions and, more importantly, the downstream clinical impact of genomic risk stratification.

There are limitations to this study. An RCT would offer a more rigorous assessment of the impact of genomic risk information; however, RCTs are complex, costly, and may not reflect real-world clinical settings or patient populations. The eMERGE study includes participants recruited from primary-care practices affiliated with academic medical centers; results from the study may not be generalizable to individuals who are not associated with a health system or do not have a primary-care provider. Although the return of GIRA results to participants and their providers reflects real-world workflows, it introduces the possibility that providers and/or participants may not review the GIRA report, follow the provided recommendations, or take other action. These variations present both limitations and opportunities. Specifically, heterogeneity in processes across study sites, such as differences in delivery methods and availability of clinical decision support, creates a natural experiment to evaluate the impact of these factors on the adoption of recommended care in future analyses. This could inform design of

targeted interventions or optimized approaches to deliver results.

The study's primary outcomes are derived from curated and harmonized EHR data across sites using validated phenotyping algorithms. Most of the outcomes include well-documented tasks such as order placement and completion, test results, changes in surrogate measures (e.g., changes in HbA1c), disease diagnoses (based on ICD codes), and medication orders (using RxNorm coding), which can be reliably extracted from the EHR.[44] However, identifying certain outcomes, such as referrals to nutrition, cardiology, or other specialties, remains challenging due to variation in referral mechanisms across health systems. To address these issues, the network is conducting concordance checks between extracted data versus manual chart review to accurately capture these outcomes. For PCa, shared decision-making is a participant-reported outcome and faces the challenges of limited participant recall and survey completion rates. However, prostate-specific antigen (PSA) order placement and completion will be assessed at later time points using EHR data, enabling evaluation of provider and participant actions following shared decision-making.

The eMERGE Network is well positioned to advance understanding of the clinical utility of genomics for the prevention of common chronic diseases. This study represents a real-world application of RD design to evaluate the impact of a genomic medicine intervention. Applying RD within a genomic context offers an innovative and robust framework for estimating the causal impact of genomic risk stratification and result disclosure, generating actionable insights for future implementation. As PRSs are increasingly developed across diverse diseases, this analytical framework provides a feasible and rigorous approach to assessing clinical impact at scale. As health systems increasingly consider integrating genomic data into preventive-care workflows, real-world implementation studies such as eMERGE IV are essential to inform outcomes and evidence-based practice guidelines.

Conducting a study across multiple conditions, incorporating diverse types of risk information, and identifying meaningful outcomes proved challenging but provided a unique opportunity to clarify study design, define critical endpoints, and propose innovative analytic methods to disentangle intervention effects in a real-world setting. Through this work, we have established a scalable implementation framework for integrating genomic information into routine care and generating critical evidence on its effectiveness for population risk stratification for the purpose of preventing common chronic diseases. These methods can serve as a model for future studies and offer a foundation that other researchers can refine and expand. Our approach may also inform clinical guidelines and health policy by providing evidence and highlighting both the challenges and opportunities involved in generating evidence to support recommendations and reimbursement for genomic testing.

Data and code availability

Data utilized in this manuscript are stored in a centralized REDCap database maintained by the eMERGE Consortium. No external datasets were generated or analyzed in this manuscript. We present the design and planned analyses of future work. No datasets or code are available at this time.

Acknowledgments

The network thanks the participant and physician advisory groups that helped shape study design, report content, and education materials. Owing to the structure of the U01 cooperative agreement, the recruitment sites, coordinating center, and funding agency collaborated on the study design as well as collection, management, analysis, and interpretation of the data. This trial was registered with clinicaltrials.gov under identifier NCT05277116. 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).

Author contributions

Writing – original draft, N. Limdi, N.S.A.-H., and D.L.V.; data curation, T.M.B. All authors contributed to conceptualization and writing – review & editing.

Declaration of interests

N.S.A.-H. is an employee of the 23andMe Research Institute and a member of the clinical advisory board for Inflection Medicine. 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 Galateo Bio. E.M.M. consults for PepGen, Tenaya, and Novartis and is a cofounder of Ikaika, Kardigan. D.E.P.-A. and Y.L.T. are employees of Invitae (now part of Labcorp). E.D.E. is a former employee of Labcorp and advisor and stockholder of Exir Bio and ROMTech. B.K. is on advisory boards for SpringWorks, Alexion, Infixion, and Healx and holds stock in Genome Medical. I.H. and J.F.P. are associate editors for *AJHG*. J.F.P. consults for Myome, Inc. T.W. has received research funding from Gilead Sciences.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2026.02.018.

Received: October 9, 2025
Accepted: February 25, 2026
Published: March 23, 2026

References

1. Benavidez, G.A., Zahnd, W.E., Hung, P., and Eberth, J.M. (2024). Chronic Disease Prevalence in the US: Sociodemographic and Geographic Variations by Zip Code Tabulation Area. Prev. Chronic Dis. *21*, E14. https://doi.org/10.5888/pcd21.230267.
2. Hartman, M., Martin, A.B., Whittle, L., Catlin, A.; and National Health Expenditure Accounts Team (2024). National Health Care Spending In 2022: Growth Similar To Prepandemic Rates. Health Aff. *43*, 6–17. https://doi.org/10.1377/hlthaff.2023.01360.
3. Hacker, K. (2024). The Burden of Chronic Disease. Mayo Clin. Proc. Innov. Qual. Outcomes *8*, 112–119. https://doi.org/10.1016/j.mayocpiqo.2023.08.005.
4. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. (2024). Lancet 403, 2100-2132. https://doi.org/10.1016/s0140-6736(24)00367-2.
5. Wisk, L.E., and Sharma, N. (2025). Prevalence and Trends in Pediatric-Onset Chronic Conditions in the United States, 1999-2018. Acad. Pediatr. *25*, 102810. https://doi.org/10.1016/j.acap.2025.102810.
6. Galea, S., and Maani, N. (2020). The cost of preventable disease in the USA. Lancet Public Health *5*, e513–e514. https://doi.org/10.1016/S2468-2667(20)30204-8.
7. Lambert, S.A., Gil, L., Jupp, S., Ritchie, S.C., Xu, Y., Buniello, A., McMahon, A., Abraham, G., Chapman, M., Parkinson, H., et al. (2021). The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nat. Genet. *53*, 420–425. https://doi.org/10.1038/s41588-021-00783-5.
8. Lambert, S.A., Wingfield, B., Gibson, J.T., Gil, L., Ramachandran, S., Yvon, F., Saverimuttu, S., Tinsley, E., Lewis, E., Ritchie, S.C., et al. (2024). Enhancing the Polygenic Score Catalog with tools for score calculation and ancestry normalization. Nat. Genet. *56*, 1989–1994. https://doi.org/10.1038/s41588-024-01937-x.
9. Maciejewski, M.L., and Basu, A. (2020). Regression Discontinuity Design. JAMA *324*, 381–382. https://doi.org/10.1001/jama.2020.3822.
10. 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. Genetics in medicine. Genet. Med. *25*, 100006. https://doi.org/10.1016/j.gim.2023.100006.
11. 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.
12. Namjou, B., Lape, M., Malolepsza, E., DeVore, S.B., Weirauch, M.T., Dikilitas, O., Jarvik, G.P., Kiryluk, K., Kullo, I.J., Liu, C., et al. (2022). Multiancestral polygenic risk score for pediatric asthma. J. Allergy Clin. Immunol. *150*, 1086–1096. https://doi.org/10.1016/j.jaci.2022.03.035.
13. Liu, C., Zeinomar, N., Chung, W.K., Kiryluk, K., Gharavi, A.G., Hripcsak, G., Crew, K.D., Shang, N., Khan, A., Fasel, D., et al. (2021). Generalizability of Polygenic Risk Scores for Breast Cancer Among Women With European, African, and Latinx Ancestry. JAMA Netw. Open *4*, e2119084. https://doi.org/10.1001/jamanetworkopen.2021.19084.
14. Khan, A., Turchin, M.C., Patki, A., Srinivasasainagendra, V., Shang, N., Nadukuru, R., Jones, A.C., Malolepsza, E., Dikilitas, O., Kullo, I.J., et al. (2022). Genome-wide polygenic score to predict chronic kidney disease across ancestries. Nat. Med. *28*, 1412–1420. https://doi.org/10.1038/s41591-022-01869-1.
15. Dikilitas, O., Schaid, D.J., Kosel, M.L., Carroll, R.J., Chute, C.G., Denny, J.A., Fedotov, A., Feng, Q., Hakonarson, H., Jarvik, G.P., et al. (2020). Predictive Utility of Polygenic Risk Scores for Coronary Heart Disease in Three Major Racial and Ethnic Groups. Am. J. Hum. Genet. *106*, 707–716. https://doi.org/10.1016/j.ajhg.2020.04.002.
16. Conti, D.V., Darst, B.F., Moss, L.C., Saunders, E.J., Sheng, X., Chou, A., Schumacher, F.R., Olama, A.A.A., Benlloch, S., Dadaev, T., et al. (2021). Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. Nat. Genet. *53*, 65–75. https://doi.org/10.1038/s41588-020-00748-0.
17. Qu, H.Q., Qu, J., Glessner, J., Liu, Y., Mentch, F., Chang, X., March, M., Li, J., Roizen, J.D., Connolly, J.J., et al. (2022). Improved genetic risk scoring algorithm for type 1 diabetes prediction. Pediatr. Diabetes *23*, 320–323. https://doi.org/10.1111/pedi.13310.
18. Ge, T., Irvin, M.R., Patki, A., Srinivasasainagendra, V., Lin, Y.F., Tiwari, H.K., Armstrong, N.D., Benoit, B., Chen, C.Y., Choi, K.W., et al. (2022). Development and validation of a trans-ancestry polygenic risk score for type 2 diabetes in diverse populations. Genome Med. *14*, 70. https://doi.org/10.1186/s13073-022-01074-2.
19. Lee, A., Mavaddat, N., Wilcox, A.N., Cunningham, A.P., Carver, T., Hartley, S., Babb de Villiers, C., Izquierdo, A., Simard, J., Schmidt, M.K., et al. (2019). BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genetics in medicine. Genet. Med. *21*, 1708–1718. https://doi.org/10.1038/s41436-018-0406-9.
20. Rasmussen-Torvik, L.J., Bonini, K.E., Harr, M.H., Abbass, M.A., Bangash, H., Bland, H.T., Boyd, B.M., Chung, W.K., Clayton, E.W., Cohen, S.J., et al. (2025). Covering medical care costs for participants in the eMERGE Network: Challenges for equity and implementation. Genet. Med. *27*, 101457. https://doi.org/10.1016/j.gim.2025.101457.
21. Lawson, L.P.,P.C., Cortopassi, J., Davis, K.W., Head, M., Martin, L.J., Perez EFm Sobowalw, A., Abul-Husn, N.S., Bangash, H., Bland, H.T., Bonini, K.E., et al. (2026). Return of genome-informed risk assessment results for common conditions to 23,840 adults and children: an eMERGE Network Study. Am. J. Hum. Genet. *113*, 678–691. https://doi.org/10.1016/j.ajhg.2026.02.016.
22. Muntner, P., Colantonio, L.D., Cushman, M., Goff, D.C., Jr., Howard, G., Howard, V.J., Kissela, B., Levitan, E.B., Lloyd-Jones, D.M., and Safford, M.M. (2014). Validation of the atherosclerotic cardiovascular disease Pooled Cohort risk equations. JAMA *311*, 1406–1415. https://doi.org/10.1001/jama.2014.2630.
23. Daly, M.B., Pilarski, R., Yurgelun, M.B., Berry, M.P., Buys, S.S., Dickson, P., Domchek, S.M., Elkhanany, A., Friedman, S., Garber, J.E., et al. (2020). NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic, Version 1.2020. J. Natl. Compr. Canc.

Netw. *18*, 380–391. https://doi.org/10.6004/jnccn.2020.0017.

24. Gupta, S., Provenzale, D., Regenbogen, S.E., Hampel, H., Slavin, T.P., Hall, M.J., Llor, X., Chung, D.C., Ahnen, D.J., Bray, T., et al. (2017). NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Colorectal, Version 3.2017. J. Natl. Compr. Canc. Netw. *15*, 1465–1475. https://doi.org/10.6004/jnccn.2017.0176.
25. Grundy, S.M., Stone, N.J., Bailey, A.L., Beam, C., Birtcher, K.K., Blumenthal, R.S., Braun, L.T., de Ferranti, S., Faiella-Tommasino, J., Forman, D.E., et al. (2019). 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation *139*, e1082–e1143. https://doi.org/10.1161/CIR.0000000000000625.
26. Newton, K.M., Peissig, P.L., Kho, A.N., Bielinski, S.J., Berg, R.L., Choudhary, V., Basford, M., Chute, C.G., Kullo, I.J., Li, R., et al. (2013). Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. J. Am. Med. Inform. Assoc. *20*, e147–e154. https://doi.org/10.1136/amiajnl-2012-000896.
27. Pacheco, J.A., Rasmussen, L.V., Wiley, K., Jr., Person, T.N., Cronkite, D.J., Sohn, S., Murphy, S., Gundelach, J.H., Gainer, V., Castro, V.M., et al. (2023). Evaluation of the portability of computable phenotypes with natural language processing in the eMERGE network. Sci. Rep. *13*, 1971. https://doi.org/10.1038/s41598-023-27481-y.
28. Linder, J.E., Tao, R., Chung, W.K., Kiryluk, K., Liu, C., Weng, C., Connolly, J.J., Hakonarson, H., Harr, M., Leppig, K.A., et al. (2023). Prospective, multi-site study of healthcare utilization after actionable monogenic findings from clinical sequencing. Am. J. Hum. Genet. *110*, 1950–1958. https://doi.org/10.1016/j.ajhg.2023.10.006.
29. Orlando, L.A., Buchanan, A.H., Hahn, S.E., Christianson, C.A., Powell, K.P., Skinner, C.S., Chesnut, B., Blach, C., Due, B., Ginsburg, G.S., and Henrich, V.C. (2013). Development and validation of a primary care-based family health history and decision support program (MeTree). N. C. Med. J. *74*, 287–296.
30. Cappelleri, J.C., and Trochim, W.M. (1994). An illustrative statistical analysis of cutoff-based randomized clinical trials. J. Clin. Epidemiol. *47*, 261–270. https://doi.org/10.1016/0895-4356(94)90007-8.
31. Johnson, W.D., Elston, R.C., and Wickremasinghe, A.R. (1994). Testing specific hypotheses by fitting underlying distributions to categorical data. J. Biopharm. Stat. *4*, 53–64. https://doi.org/10.1080/10543409408835072.
32. Cappelleri, J.C., Darlington, R.B., and Trochim, W.M.K. (1994). Power Analysis of Cutoff-Based Randomized Clinical Trials. Eval. Rev. *18*, 141–152. https://doi.org/10.1177/0193841x9401800202.
33. Buchanan, A.H., Lester Kirchner, H., Schwartz, M.L.B., Kelly, M.A., Schmidlen, T., Jones, L.K., Hallquist, M.L.G., Rocha, H., Betts, M., Schwiter, R., et al. (2020). Clinical outcomes of a genomic screening program for actionable genetic conditions. Genet. Med. *22*, 1874–1882. https://doi.org/10.1038/s41436-020-0876-4.
34. Savatt, J.M., Johns, A., Schwartz, M.L.B., McDonald, W.S., Salvati, Z.M., Oritz, N.M., Masnick, M., Hatchell, K., Hao, J., Buchanan, A.H., and Williams, M.S. (2023). Testing and Management of Iron Overload After Genetic Screening-Identified Hemochromatosis. JAMA Netw. Open *6*, e2338995. https://doi.org/10.1001/jamanetworkopen.2023.38995.
35. Hao, J., Hassen, D., Manickam, K., Murray, M.F., Hartzel, D.N., Hu, Y., Liu, K., Rahm, A.K., Williams, M.S., Lazzeri, A., et al. (2020). Healthcare Utilization and Costs after Receiving a Positive BRCA1/2 Result from a Genomic Screening Program. J. Pers. Med. *10*, 7. https://doi.org/10.3390/jpm10010007.
36. Kerman, B.J., Brunette, C.A., Harris, E.J., Antwi, A.A., Lemke, A.A., and Vassy, J.L. (2023). Primary care physician use of patient race and polygenic risk scores in medical decision-making. Genetics in medicine. Genet. Med. *25*, 100800. https://doi.org/10.1016/j.gim.2023.100800.
37. Muse, E.D., Chen, S.F., Liu, S., Fernandez, B., Schrader, B., Molparia, B., León, A.N., Lee, R., Pubbi, N., Mejia, N., et al. (2022). Impact of polygenic risk communication: an observational mobile application-based coronary artery disease study. npj Digit. Med. *5*, 30. https://doi.org/10.1038/s41746-022-00578-w.
38. Odden, M.C., Zhang, A., Jawadekar, N., Tan, A., Moran, A.E., Glymour, M.M., Brayne, C., Zeki Al Hazzouri, A., and Calonico, S. (2023). Regression discontinuity design to evaluate the effect of statins on myocardial infarction in electronic health records. Eur. J. Epidemiol. *38*, 393–402. https://doi.org/10.1007/s10654-023-00982-w.
39. Petersen, I., Nicolaisen, S.K., Ricciardi, F., Sharma, M., Thomsen, R.W., Baio, G., and Pedersen, L. (2020). Impact of Being Eligible for Type 2 Diabetes Treatment on All-Cause Mortality and Cardiovascular Events: Regression Discontinuity Design Study. Clin. Epidemiol. *12*, 569–577. https://doi.org/10.2147/CLEP.S251704.
40. Adeleke, M.O., O'Keeffe, A.G., and Baio, G. (2023). Approaches to risk ratio estimation in a regression discontinuity design: Application to the prescription of statins for cholesterol reduction in UK primary care. Stat. Methods Med. Res. *32*, 1994–2015. https://doi.org/10.1177/09622802231192958.
41. van Leeuwen, N., Lingsma, H.F., de Craen, A.J.M., Nieboer, D., Mooijaart, S.P., Richard, E., and Steyerberg, E.W. (2016). Regression Discontinuity Design: Simulation and Application in Two Cardiovascular Trials with Continuous Outcomes. Epidemiology *27*, 503–511. https://doi.org/10.1097/EDE.0000000000000486.
42. Thomsen, M.K., Nicolaisen, S.K., Pedersen, L., Lash, T.L., Erichsen, R., Sørensen, H.T., and Mikkelsen, E.M. (2023). Effect of a FIT-Based Colorectal Cancer Screening Program on Mortality Estimated by the Regression Discontinuity Design. Am. J. Epidemiol. *192*, 1475–1484. https://doi.org/10.1093/aje/kwad096.
43. Peterson, J.F., Roden, D.M., Orlando, L.A., Ramirez, A.H., Mensah, G.A., and Williams, M.S. (2019). Building evidence and measuring clinical outcomes for genomic medicine. Lancet *394*, 604–610. https://doi.org/10.1016/s0140-6736(19)31278-4.
44. Wagholikar, K.B., Pacheco, J.A., Gordon, A.S., Khan, A., Khales, B.N., Benoit, B., Kerman, B.J., Weng, C., Ta, C., Prows, C.A., et al. (2025). Current Limitations of Electronic Health Record Systems in Supporting Pragmatic Clinical Trials: Insights from the eMERGE Consortium. medRxiv. https://doi.org/10.1101/2025.04.01.25325049.