UC Irvine
UC Irvine Previously Published Works

Title

Improving reporting standards for polygenic scores in risk prediction studies

Permalink

Journal

ISSN

Authors

Wand, Hannah
Lambert, Samuel A
Tamburro, Cecelia
et al.

Publication Date

DOI

Peer reviewed

Improving reporting standards for polygenic scores in risk prediction studies

Hannah Wand[#,1,2], **Samuel A. Lambert**[#,3,4,5,6,7], **Cecelia Tamburro**[8], **Michael A. Iacocca**[1], **Jack W. O'Sullivan**[1,2], **Catherine Sillari**[8], **Iftikhar J. Kullo**[9], **Robb Rowley**[8], **Jacqueline S. Dron**[10,11], **Deanna Brockman**[10], **Eric Venner**[12], **Mark I. McCarthy**[13,14], **Antonis C. Antoniou**[15], **Douglas F. Easton**[15], **Robert A. Hegele**[11], **Amit V. Khera**[10], **Nilanjan Chatterjee**[16,17], **Charles Kooperberg**[18], **Karen Edwards**[19], **Katherine Vlessis**[20], **Kim Kinnear**[20], **John N. Danesh**[5,6,21], **Helen Parkinson**[6,7], **Erin M. Ramos**[8], **Megan C. Roberts**[22], **Kelly E. Ormond**[20,23], **Muin J. Khoury**[24], **A. Cecile J.W. Janssens**[25], **Katrina A.B. Goddard**[26,27], **Peter Kraft**[28], **Jaqueline A. L. MacArthur**[7], **Michael Inouye**[$,3,4,5,6,21,29], **Genevieve L. Wojcik**[$,30]

[1.]Stanford University School of Medicine, Stanford, CA, USA

[2.]Stanford Center for Inherited Cardiovascular Disease, Stanford, CA, USA

[3.]Cambridge Baker Systems Genomic Initiative, University of Cambridge, Cambridge, UK

[4.]Cambridge Baker Systems Genomic Initiative, Baker Heart and Diabetes Institute, Melbourne, AUS

[5.]BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[6.]Health Data Research UK-Cambridge, Wellcome Genome Campus, Hinxton, UK

[7.]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, UK

[8.]National Human Genome Research Institute, Bethesda, MD, USA

[9.]Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, USA

[10.]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

[11.]Western University, London, ON, Canada

12.Baylor College of Medicine, Houston, TX, USA

13.Department of Human Genetics, Genentech, South San Francisco, CA, USA

14.Wellcome Centre for Human Genetics, Oxford, UK

15.University of Cambridge, Cambridge, United Kingdom

16.Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

17.Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD, USA

18.Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

19.Department of Epidemiology, University of California, Irvine, CA, USA

20.Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

21.National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge and Cambridge University Hospitals, Cambridge, UK

22.Division of Pharmaceutical Outcomes and Policy, UNC Eshelman School of Pharmacy, Chapel Hill, NC, USA

23.Stanford Center for Biomedical Ethics, Stanford University School of Medicine, Stanford, CA, USA

24.Centers for Disease Control and Prevention, Atlanta, GA, USA

25.Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA

26.Department of Translational and Applied Genomics, Kaiser Permanente Northwest, Portland, OR, USA

27.Center for Health Research, Kaiser Permanente Northwest, Portland, OR, USA

28.Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

29.The Alan Turing Institute, London, UK

30.Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Preface

Polygenic risk scores (PRS), often aggregating results from genome-wide association studies, can bridge the gap between the initial discovery efforts and clinical applications for disease risk estimation using genetics. However, there is remarkable heterogeneity in the application and reporting of these risk scores, hindering the translation of PRS into clinical care. The ClinGen Complex Disease Working Group, in a collaboration with the Polygenic Score (PGS) Catalog, have updated the Genetic Risk Prediction (GRIPS) Reporting Statement to reflect the current state of the field. Drawing upon experts in epidemiology, statistics, disease-specific applications, implementation, and policy, this 22-item reporting framework defines the minimal information

needed to interpret and evaluate PRS, especially with respect to downstream clinical applications. Items span detailed descriptions of study populations, statistical methods for PRS development and validation, and considerations for potential limitations of the PRS. Additionally, we emphasize the need for data availability and transparency, encouraging researchers to deposit and share PRS via the PGS Catalog to facilitate reproducibility and comparative benchmarking. By providing these criteria in a structured format that builds upon existing standards and ontologies, the use of this framework in publishing PRS will facilitate translation into clinical care and progress towards defining best practices.

The predisposition to common diseases and traits arises from a complex interaction between genetic and nongenetic factors. In the past decade there has been enormous success at discovering disease-associated genetic variants, facilitated by many collaborative consortia and large cohorts of well-phenotyped individuals with matched genetic information.[1-5] In particular, genome-wide association studies (GWAS) have yielded summary statistics that describe the magnitude (effect size) and statistical significance of association between an allele and the outcome of interest.[4,6] GWAS have been applied to many complex human traits and diseases, including height, blood pressure, cardiovascular disease, cancer, obesity, and Alzheimer's disease.

The associations identified via GWAS can be combined to quantify genetic predisposition to a heritable trait, and this information can be used to conduct disease risk stratification or predict prognostic outcomes and response to therapy.[7,8] Typically, information across many variants is combined via a weighted sum of allele counts, where the weights reflect the relative magnitude of association between variant alleles and the trait or disease. These weighted sums can include millions of variants, and are frequently referred to as **polygenic risk score(s) (PRS),** or **genetic or genomic risk score(s) (GRS)**, if they refer to risk estimates of disease outcomes; or, more generally, **polygenic score(s) (PGS)** when referring to any outcome (see Box 1). While there is active development of algorithms to decide how many and which variants to include and how to weigh them so as to maximize the proportion of variance explained or disease discrimination, there is emerging consensus that inclusion of variants beyond those meeting stringent GWAS significance levels can boost predictive performance.[9,10] Methodological research has also established theoretical limits of PGS/PRS performance based on the trait's genetic architecture and heritability.[11-15]

In the last decade, the landscape of genetic prediction studies has transformed. Over 900 publications mention PGS/PRS with significant developments in how PRS are constructed and evaluated, as well as many new proposed uses. The data available in the current era of biomedical research is larger and more consolidated than ever before. Biobanks and large-scale consortia have become dominant, yet frequently researchers have limited access to individual-level data. Since individual data is unavailable, most PRS risk predictions are developed from summary-level data (e.g. GWAS summary statistics) in secondary datasets, each of which come with their own specific methodological considerations.[16-18] At the same time, there has been a push towards open data sharing as outlined in the FAIR (Findable, Accessible, Interoperable and Reusable) Data Principles [3,19], with an emphasis on ensuring that research is reproducible by all.

The capacity of PRS to quantify genetic predisposition for many clinically relevant traits and diseases has begun to be established, with multiple potential clinical uses in settings related to disease risk stratification as well as proposed prognostic uses (e.g., predicting response to intervention/treatment). Readiness for implementation varies by outcome, with only a few diseases, like coronary heart disease (CHD) and breast cancer, having mature PRS with potential clinical utility (Boxes 2 and 3, respectively). There has also been a rapid rise of direct-to-consumer assays and for-profit companies (23andMe, Color, MyHeritage, etc.) providing PGS/PRS results to customers outside of the traditional patient-provider framework. These concomitant developments have resulted in healthcare systems developing new infrastructures to deliver genetic risk information.[20] Individually and combined, these advances have raised significant challenges for PRS reporting standards; from the very basic (e.g. reporting performance metrics on an external validation dataset) to the complicated (e.g. making raw variant and weight information for a PRS available), necessitating the updating of existing standard for reporting genetic risk prediction studies to convey the increased scope of PRS and complexity for their clinical applications.

Poorly designed and/or described studies call into question the validity of some PRS to predict their target outcome [21,22], and relatively few studies have externally benchmarked multiple scores' performance. At present, there are no uniformly agreed best practices for developing PRS, nor widely adopted standards or regulations sufficiently tailored to assess the eventual clinical readiness of a PRS. There are emerging applications of PRS that further compound the heterogeneity in reporting, e.g. using PRS as tools for testing gene x environment interactions or shared etiology between diseases. [23-26] The rapid evolution in both methodological development and applications of PRS make it challenging to compare or reproduce claims about the predictive performance of a PRS for a specific outcome when studies are not properly documented. These deficiencies are barriers to PRS being interpreted, compared, and reproduced, and must be addressed to enable the application of PRS to improve clinical practice and public health.

Frameworks have been developed to establish standards around the transparent, standardized, accurate, complete, and meaningful reporting of scientific studies. In 2011, an international working group published the Genetic Risk Prediction Studies (GRIPS) Statement, a reporting guideline for risk prediction models that include genetic variants, from genetic mutations to gene scores.[27] This guideline is analogous to those developed for observational epidemiological studies (STROBE[28]) and genome-wide association studies (STREGA[29]), and is in line with the reporting guideline for multivariate prediction models (TRIPOD[30]). Adherence to reporting statements has been low, and the same holds for GRIPS. One potential reason might be that researchers feel that GRIPS inadequately addresses PRS. Researchers are frequently uncertain as to what precisely should be reported for a PRS study to be assessed as rigorous, reproducible, and ultimately translatable, especially with the increased push for data availability and transparency. Most PRS studies follow a prototypical process (Figure 1) that can be used as a template for standardizing reporting and benchmarking in the field.

Here, the Clinical Genome Resource (ClinGen) Complex Disease Working Group and the Polygenic Score (PGS) Catalog (Supplemental Note 1) jointly present the Polygenic

Risk Score Reporting Standards (PRS-RS), an expanded reporting standard for PRS that addresses current research environments with advanced methodological developments to inform clinically meaningful reporting on PRS development and validation in the literature with an emphasis on reproducibility and transparency throughout the development process. Additional methods are detailed in Supplemental Note 2.

The Polygenic Risk Score Reporting Standards (PRS-RS)

The PRS-RS is a set of standard items specifying the minimal criteria that need to be described in a manuscript in order to accurately interpret a PRS and reproduce results throughout the PRS development process, briefly illustrated in Figure 1. [31] It applies to both PRS development and validation studies that aim to predict disease onset, diagnosis, and prognosis, as well as response to therapies; however, other research uses of PRS have overlapping steps that should be reported similarly. Table 1 presents the full PRS-RS, with reporting items organized into key components along the developmental pipeline of PRS for clear interpretation and to encourage their documentation from the inception of the study well before publication.

Reporting on risk score background:

The development and validation of a PRS tests a specific hypothesis with a defined outcome and study population. Therefore, authors should define *a priori* the *study type* (e.g. development and/or validation), *purpose* (e.g. risk prediction vs. prognosis) and *predicted outcome* (e.g. CHD) in enough detail to understand why the study population and risk model selected are relevant (e.g. the value for CHD risk stratification and primary prevention is highest in younger individuals compared to those over 80 with lifetime accumulated risk). As the PRS-RS is focused on clinical validity and implementation, authors must outline the study and appropriate outcomes to understand what risk is measured, what the purpose of measuring risk would be, and why this purpose may be of clinical relevance. To establish the internal validity of a study, authors should use the appropriate data for the intended purpose (e.g. prediction of incident disease vs. prognosis), with adequate documentation of dataset characteristics to understand nuances in measured risk.

Reporting on study populations:

The applicability of any risk prediction to an external target population (the "who, where, and when") depends on its similarity to the original study populations used to derive the risk model. Therefore, authors need to define and characterize the details of their study population (*study design and recruitment*), and describe *participant demographics* for key variables (most often *age* and *sex*) and *ancestry*. Importantly, there are often inconsistent definitions and levels of detail associated with ancestry, and the transferability of genetic findings between different racial/ethnic groups can be limited.[1,9,32] It is therefore essential for authors to provide a detailed description of participants' genetic ancestry - including how ancestry was determined – using a common controlled vocabulary where possible (e.g. the standardized framework developed by the NHGRI-EBI GWAS Catalog[1]). Authors should provide a sufficient level of detailed criteria for defining all the factors used in the risk model (*non-genetic variable(s)*) and relevant to the *outcome of interest*. This is particularly

important if they are included in the final risk model and should accompany information about how the population was genotyped (*genetic data*), including assays and all quality control measures.

Reporting on risk model development:

There are currently several commonly used methods to select variants that constitute the PRS and fine-tune their weights.[7,16-18,31] Methods using GWAS summary statistics should clearly cite the relevant GWAS (preferably using unique and persistent study identifiers from the GWAS Catalog [GCSTs] [33]). As the performance and limitations of the combined risk model are dependent on methodological considerations, authors must provide complete details including the method used and how variants are combined into a single PRS (*PRS construction & estimation*). Apart from genetic data, authors should also describe the defining criteria for other demographic and non-genetic predictors (*non-genetic variables*) included in the model. Often authors will iterate through numerous models to find the optimal fit. In addition to the estimation methods, it is important to detail the *integrated risk model fitting* procedure, including the measures used for final model selection. Translating the continuous PRS distribution to a risk estimate, whether absolute or relative, is highly dependent on assumptions and limitations inherent to the specific data set utilized. When describing the *risk model type*, authors should detail the time scale employed for prediction or the study period/follow-up time for a relative hazard model. Additionally, if relative risk is estimated, the reference group should be well described. These details should be described for the training set, as well as validation and sub-group analyses.

Reporting on risk model evaluation:

Authors should report estimates for all evaluated models (including the methods used to derive them) to equip readers with the information necessary to evaluate the relative value of an increase in performance against other trade-offs. We recommend that authors provide summary information of the *risk score distribution* to aid in model interpretation. The *risk model's predictive ability*, *calibration*, and *discrimination* should also be assessed and described with common descriptions including the risk score effect size, variance explained (R^2), reclassification indices, and metrics like sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The *risk model calibration* and *discrimination* should be described for all analyses, although their estimation and interpretation are most relevant for the PRS validation sample. It is imperative for the PRS and integrated risk models to be evaluated on a population that is external (e.g. independent, non-overlapping) to the individuals in the study population. The ability of the risk model to classify individuals of interest (*risk model discrimination*) can commonly described and presented in terms of the area under the receiver operating characteristic (AUROC) or precision-recall curve (AUPRC), or the concordance statistic (C-index). Any differences in variable definitions or performance discrepancies between the training and validation sets should be described.

Reporting on interpretation:

By explicitly describing the *risk model's interpretation* and outlining potential *limitations* to the *generalizability* of their model, authors will empower readers and the wider community

to better understand the risk score and its relative merits. Authors should justify the clinical relevance and *risk model intended uses*, such as how the performance of their PRS compares to other commonly used risk models, or previously published PRS. This may also include comparisons to other genetic predictors of disease (e.g. mutations in high/moderate risk genes associated with Mendelian forms of the disease), family history, simple demographic models, or conventional risk calculators (see Boxes 2 and 3 for disease-specific examples). What indicates a "good" prediction can differ between outcomes and *intended uses*, but should be reported with similar metrics to those described in the evaluation section.

Reporting on model parameters:

The underlying PRS (variant alleles and derived weights) should be made publicly available, preferably through direct submission to an indexed repository such as the PGS Catalog, to enable others to reuse existing models (with known validity) and facilitate direct benchmarking between different PRS for the same trait (*data transparency and availability*). The current mathematical form of most PRS—a linear combination of allele counts— facilitates clear model description and reproducibility. Future genomic risk models may have more complex forms, e.g. allowing for explicit non-linear epistatic and gene-environment interactions, or deep neural networks of lesser clarity. It will be important to describe these models in sufficient detail to allow their implementation and evaluation by other researchers and clinical groups.

Supplemental Note 3 provides explanations of reporting considerations in addition to the minimal reporting framework in Table 1. Authors intending downstream clinical implementation should aim for the level of transparent and comprehensive reporting covered in both Table 1 and Supplemental Notes, especially those related to discussing the interpretation, limitations, and generalizability of results. The proper reporting of PRS development and performance can also have implications for seeking regulatory approval of the PRS as a clinical test. Though not a comprehensive list of regulatory requirements, we highlight aspects of PRS-RS that would be considered evidence of analytical and clinical validity from the College of American Pathologists (CAP) and the Clinical Laboratory Improvement Amendments (CLIA) perspective (Supplemental Table 1). CAP and CLIA approvals are additional incentives for reporting adherence of researchers wishing to translate their work, as well as a caution for researchers wishing to avoid unintended uses of their findings. Lastly, we reiterate the need for both methodological and data transparency and encourage deposition of PRS (variant-level information necessary to recalculate the genetic portion of the score) in the PGS Catalog (www.PGSCatalog.org; [34]), which provides an invaluable resource for widespread adoption and distribution of a published PRS. The PGS Catalog provides access to PGS and related metadata to support the FAIR principles of data stewardship [19], enabling subsequent applications and assessments of PGS performance and best practices (see Supplemental Table 2 for a description of the metadata captured in the Catalog and it's overlap with the PRS-RS).

Improving PRS research and translation

We surveyed 30 publications (selecting for a diversity of disease domains, risk score categories, and populations) to understand how the information in the PRS-RS is presented and displayed as part of the larger iterative process to clarify and improve minimal reporting item descriptions. For 10 of these publications, we provide detailed annotations using the final minimal reporting requirements (Supplemental Table 3) and use these annotations to illustrate the detail necessary for each PRS-RS item (further described in Supplemental Note 3). The heterogeneity in the PRS reporting we observed in this pilot highlights a series of challenges. Critical aspects of PRS studies, including ancestry, predictive ability, and transparency/availability of information needed to reproduce PRS, were frequently absent or reported in insufficient detail. This underscores the need for PRS-RS to clearly and specifically define meaningful aspects of PRS development, testing, and intended clinical use. However, these deficits in reporting are not unique to PRS; previous reports of underreporting have found that 77% of GWAS publications in 2017 did not share summary statistics[35] and 4% of GWAS do not report any relevant ancestry information[1]. In line with the push towards a culture of reproducibility and open data in genomics, we as the ClinGen Complex Disease Working Group and PGS Catalog joined to create this set of reporting standards (Table 1) specifically tailored to PRS research based on multidisciplinary and international expert opinion for tailoring previous standards.

Researchers using PRS-RS may identify fringe cases that are inadequately captured by these reporting items, as we have modeled our guideline on prototypical steps for PRS development (Figure 1). While we anticipate the field may further change as novel methods and technologies are generated, the PRS-RS items can be further expanded and adapted to encompass novel considerations. By updating previous standards, drawing upon current leaders in the field, and tailoring the framework to common barriers observed in recent literature, we aim to provide a comprehensive and pragmatic perspective on the topic. In line with previous standards, PRS-RS includes elements related to understanding the clinical validity of PRS and consequent risk models. Items such as *predicted outcome* and *intended use* bookend our guideline with the intended clinical framing of PRS reporting. In addition, we have modeled the guideline by steps in experimental design, from hypothesis to interpretation, to more clearly emphasize the significance of the intended use case in defining what needs reported and inform documentation throughout the process. As a reference, we have included a guide to where PRS-RS items should be reported in a manuscript in Supplemental Table 4. These expansions will further facilitate the curation and expert annotation of published PRS as we move towards widespread clinical use.

While the scope of our work encompasses clinical validity, it does not address the additional requirements needed to establish the clinical or public health utility of a PRS, such as randomized trials with clinically meaningful outcomes, health economic evaluations, or feasibility studies. [36] In addition, the translation of structured data elements into useful clinical parameters may not be direct. One example is that the case definitions utilized in training or validation in any particular PRS study may deviate (sometimes substantially) from those utilized in any specific health system. CHD symptoms commonly include angina (chest pain), whereas PRS are frequently trained on stricter definitions excluding angina.

Another example is that the definitions used for race/ancestry as outlined in the PGS and GWAS Catalog[1] may differ from structured terms used to document ancestry information in the clinic. Consistent mappings and potentially parallel analyses may be necessary to translate from genetically-determined ancestries to those routinely used in clinical care. Such translation issues potentially limit generalizability to target populations and warrant further discussion, and we reiterate the need for authors to be mindful of their intended purpose and target audience when discussing their findings. Authors' understanding of potential translational barriers can be aided by considering current CAP/CLIA analytical and clinical validity evidence requirements of peer-reviewed literature to ensure PRS-RS has value in informing later steps of the clinical translation spectrum, including clinical utility (Supplemental Table 1). Finally, while the principles of this work are clear, its scope does not include the complex commercial restrictions, such as intellectual property, that may be placed on published studies regarding the reporting or distribution of PGS, or the underlying data thereof. We hope this work will inform downstream regulation and transparency standards for PRS as a commercial clinical tool.

The coordinated efforts of the ClinGen Complex Disease Working Group and PGS Catalog provide a set of compatible resources for researchers to deposit PGS/PRS information. The PGS Catalog (www.PGSCatalog.org) provides an informatics platform with data integration and harmonization to other PGS as well as the source GWAS study through its sister platform, the GWAS Catalog.[34] In addition, it provides a structured database of scores (variants and effect weights) that can be reused, along with metadata requested in the PRS-RS. With these tools, PRS-RS can be mandated by leading peer-reviewed journals and, consequently, the quality and rigor of PRS research will be elevated to a level which facilitates clinical implementation. We encourage readers to visit the ClinGen website (https://clinicalgenome.org/working-groups/complex-disease/) for any future changes or amendments to our reporting standards.

While we have provided explicit recommendations on how to acknowledge study design limitations and their impact on the interpretation and generalizability of a PRS, future research should attempt to establish best practices to guide the field. Moving forward, supplemental frameworks should be developed for the reporting of new methods, such as deep learning, as well as requirements for clinical utility and readiness. Taken together, PRS-RS facilitates the rapid emergence of polygenic risk scores as potentially powerful tools for the translation of genomic discoveries into clinical and public health benefits, and provides a framework for PRS to transform multiple areas of human genetic research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Morales J et al. A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. Genome Biol. 19, 21 (2018). [PubMed: 29448949]

2. MacArthur J et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 45, D896–D901 (2017). [PubMed: 27899670]

3. Claussnitzer M et al. A brief history of human disease genetics. Nature (2020).

4. Visscher PM, Brown MA, McCarthy MI & Yang J Five years of GWAS discovery. Am. J. Hum. Genet 90, 7–24 (2012). [PubMed: 22243964]

5. Visscher PM et al. 10 years of GWAS discovery: biology, function, and translation. Am. J. Hum. Genet 101, 5–22 (2017). [PubMed: 28686856]

6. Pasaniuc B & Price AL Dissecting the genetics of complex traits using summary association statistics. Nat. Rev. Genet 18, 117–127 (2017). [PubMed: 27840428]

7. Lambert SA, Abraham G & Inouye M Towards clinical utility of polygenic risk scores. Hum. Mol. Genet 28, R133–R142 (2019). [PubMed: 31363735]

8. Torkamani A, Wineinger NE & Topol EJ The personal and clinical utility of polygenic risk scores. Nat. Rev. Genet 19, 581–590 (2018). [PubMed: 29789686] Refs 7 & 8 are useful reviews outlining the potential benefits and clinical uses of polygenic risk scores.

9. Martin AR et al. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat. Genet 51, 584–591 (2019). [PubMed: 30926966]

10. Chatterjee N, Shi J & García-Closas M Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat. Rev. Genet 17, 392–406 (2016). [PubMed: 27140283] A comprehensive review of the development of polygenic risk scores with respect to methodological considerations, evaluation frameworks, and downstream utility.

11. Wray NR, Yang J, Goddard ME & Visscher PM The genetic interpretation of area under the ROC curve in genomic profiling. PLoS Genet. 6, e1000864 (2010). [PubMed: 20195508]

12. Dudbridge F Power and predictive accuracy of polygenic risk scores. PLoS Genet. 9, e1003348 (2013). [PubMed: 23555274]

13. Pharoah PDP et al. Polygenic susceptibility to breast cancer and implications for prevention. Nat. Genet 31, 33–36 (2002). [PubMed: 11984562]

14. Zhang Y, Qi G, Park J-H & Chatterjee N Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. Nat. Genet 50, 1318–1326 (2018). [PubMed: 30104760]

15. Chatterjee N et al. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat. Genet 45, 400–5, 405e1 (2013). [PubMed: 23455638]

16. Mak TSH, Porsch RM, Choi SW, Zhou X & Sham PC Polygenic scores via penalized regression on summary statistics. Genet. Epidemiol 41, 469–480 (2017). [PubMed: 28480976]

17. Choi SW & O'Reilly PF PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience 8, (2019).

18. Vilhjálmsson BJ et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am. J. Hum. Genet 97, 576–592 (2015). [PubMed: 26430803]

19. Wilkinson MD et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018 (2016). [PubMed: 26978244]

20. Ganguly P & National Human Genome Research Institute. NIH funds centers to improve the role of genomics in assessing and managing disease risk. (2020). at <https://www.genome.gov/news/news-release/NIH-funds-centers-to-improve-role-of-genomics-in-assessing-and-managing-disease-risk>

21. Martens FK, Tonk ECM & Janssens ACJW Evaluation of polygenic risk models using multiple performance measures: a critical assessment of discordant results. Genet. Med 21, 391–397 (2019). [PubMed: 29895851]

22. Janssens ACJW Validity of polygenic risk scores: are we measuring what we think we are? Hum. Mol. Genet 28, R143–R150 (2019). [PubMed: 31504522]

23. Warrier V et al. Polygenic scores for intelligence, educational attainment and schizophrenia are differentially associated with core autism features, IQ, and adaptive behaviour in autistic individuals. medRxiv (2020). doi:10.1101/2020.07.21.20159228

24. Matoba N, Love MI & Stein JL Evaluating brain structure traits as endophenotypes using polygenicity and discoverability. BioRxiv (2020). doi:10.1101/2020.07.17.208843

25. Coleman JRI et al. Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK Biobank. Mol. Psychiatry 25, 1430–1446 (2020). [PubMed: 31969693]

26. Kraft P & Aschard H Finding the missing gene-environment interactions. Eur J Epidemiol 30, 353–355 (2015). [PubMed: 26026724]

27. Janssens ACJW et al. Strengthening the reporting of genetic risk prediction studies (GRIPS): explanation and elaboration. Eur. J. Hum. Genet 19, 18 p preceding 494 (2011). GRIPS was the first to propose standards for the reporting of risk prediction studies using genetics, and forms the basis of our updated PRS-RS.

28. von Elm E et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 335, 806–808 (2007). [PubMed: 17947786]

29. Little J et al. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. PLoS Med. 6, e22 (2009). [PubMed: 19192942]

30. Moons KGM et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann. Intern. Med 162, W1–73 (2015). [PubMed: 25560730]

31. Choi SW, Mak TS-H & O'Reilly PF Tutorial: a guide to performing polygenic risk score analyses. Nat. Protoc 15, 2759–2772 (2020). [PubMed: 32709988] A step-by-step instruction of key analyis steps in PGS development and evaluation.

32. Wojcik GL et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature 570, 514–518 (2019). [PubMed: 31217584]

33. Buniello A et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 47, D1005–D1012 (2019). [PubMed: 30445434]

34. Lambert SA et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nature Genetics (2021). doi:TBD This paper describes the development and contents of the PGS Catalog (www.PGSCatalog.org), an open database of polygenic scores.

35. Thelwall M et al. Is useful research data usually shared? An investigation of genome-wide association study summary statistics. PLoS One 15, e0229578 (2020). [PubMed: 32084240]

36. Burke W & Zimmern R Moving Beyond ACCE: An Expanded Framework for Genetic Test Evaluation. (PHG Foundation, 2007).

37. Huo D et al. Comparison of breast cancer molecular features and survival by african and european ancestry in the cancer genome atlas. JAMA Oncol. 3, 1654–1662 (2017). [PubMed: 28472234]

38. Choi J et al. The associations between immunity-related genes and breast cancer prognosis in Korean women. PLoS One 9, e103593 (2014). [PubMed: 25075970]

39. Marston NA et al. Predicting benefit from evolocumab therapy in patients with atherosclerotic disease using a genetic risk score: results from the FOURIER trial. Circulation 141, 616–623 (2020). [PubMed: 31707849]

40. Nikpay M et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat. Genet 47, 1121–1130 (2015). [PubMed: 26343387]

41. Khera AV et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat. Genet 50, 1219–1224 (2018). [PubMed: 30104762]

42. Mars N et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. Nat. Med 26, 549–557 (2020). [PubMed: 32273609]

43. Elliott J et al. Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. JAMA 323, 636–645 (2020). [PubMed: 32068818]

44. Inouye M et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. J. Am. Coll. Cardiol 72, 1883–1893 (2018). [PubMed: 30309464]

45. Abraham G et al. Genomic prediction of coronary heart disease. Eur. Heart J 37, 3267–3278 (2016). [PubMed: 27655226]

46. Ganna A et al. Multilocus genetic risk scores for coronary heart disease prediction. Arterioscler. Thromb. Vasc. Biol 33, 2267–2272 (2013). [PubMed: 23685553]

47. Kullo IJ et al. Incorporating a Genetic Risk Score Into Coronary Heart Disease Risk Estimates: Effect on Low-Density Lipoprotein Cholesterol Levels (the MI-GENES Clinical Trial). Circulation 133, 1181–1188 (2016). [PubMed: 26915630]

48. U.S. National Library of Medicine. Home - ClinicalTrials.gov. at <https://clinicaltrials.gov/>

49. Mavaddat N et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am. J. Hum. Genet 104, 21–34 (2019). [PubMed: 30554720]

50. Mars N et al. Polygenic risk, susceptibility genes, and breast cancer over the life course. medRxiv (2020). doi:10.1101/2020.04.17.20069229

51. Zhang H et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. Nat. Genet 52, 572–581 (2020). [PubMed: 32424353]

52. Zhang X et al. Addition of a polygenic risk score, mammographic density, and endogenous hormones to existing breast cancer risk prediction models: A nested case-control study. PLoS Med. 15, e1002644 (2018). [PubMed: 30180161]

53. Lakeman IMM et al. Addition of a 161-SNP polygenic risk score to family history-based risk prediction: impact on clinical management in non-BRCA1/2 breast cancer families. J. Med. Genet 56, 581–589 (2019). [PubMed: 31186341]

54. Wilcox AN et al. Prospective Evaluation of a Breast Cancer Risk Model Integrating Classical Risk Factors and Polygenic Risk in 15 Cohorts from Six Countries. medRxiv (2019). doi:10.1101/19011171

55. Pal Choudhury P et al. Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort. medRxiv (2020). doi:10.1101/2020.04.27.20081265

56. Maas P et al. Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. JAMA Oncol. 2, 1295–1302 (2016). [PubMed: 27228256]

57. Lee A et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet. Med 21, 1708–1718 (2019). [PubMed: 30643217] This paper describes the BOADICEA model, an integrated risk score for breast cancer that includes a PRS. The model is available within the CanRisk Tool (www.canrisk.org) and has been approved for use by healthcare professionals in the European Economic Area.

58. Garcia-Closas M, Gunsoy NB & Chatterjee N Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. J. Natl. Cancer Inst 106, (2014).

59. Kapoor PM et al. Combined associations of a polygenic risk score and classical risk factors with breast cancer risk. J. Natl. Cancer Inst (2020). doi:10.1093/jnci/djaa056

60. Easton DF et al. Gene-panel sequencing and the prediction of breast-cancer risk. N. Engl. J. Med 372, 2243–2257 (2015). [PubMed: 26014596]

61. Muranen TA et al. Genetic modifiers of CHEK2*1100delC-associated breast cancer risk. Genet. Med 19, 599–603 (2017). [PubMed: 27711073]

62. Kuchenbaecker KB et al. Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. J. Natl. Cancer Inst 109, (2017).

63. Fahed AC et al. Polygenic background modifies penetrance of monogenic variants conferring risk for coronary artery disease, breast cancer, or colorectal cancer. medRxiv (2019). doi:10.1101/19013086

64. Barnes DR et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. Genet. Med (2020). doi:10.1038/s41436-020-0862-x

65. van den Broek JJ et al. Personalizing breast cancer screening based on polygenic risk and family history. J. Natl. Cancer Inst (2020). doi:10.1093/jnci/djaa127

66. Esserman LJ & WISDOM Study and Athena Investigators. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. NPJ Breast Cancer 3, 34 (2017). [PubMed: 28944288]

Box 1.

Definitions of relevant genetic risk prediction terms

Polygenic Score(s) (PGS):

a single value that quantifies an individual's genetic predisposition to a trait. Typically calculated by summing the number of trait-associated alleles in an individual weighted by per-allele effect sizes from a discovery GWAS, and normalized using a relevant population distribution. Sometimes referred to as a genetic score.

Polygenic Risk Score(s) (PRS):

a subset of PGS which is used to estimate risk of disease or other clinically relevant outcomes (binary or discrete). Sometimes referred to as a genetic or genomic risk score (GRS). See categories below.

Integrated Risk Model:

a risk model for the outcome of interest which combines PRS with other risk factors, such as demographics (often age and sex), anthropometrics, biomarkers, and clinical measurements.

Categories of use for PRS and/or integrated risk models

The addition of PRS to existing risk models has several potential applications, summarized below. Each aims to improve individual or subgroup classification such that there is clinical benefit.

<u>**Disease Risk Prediction –:**</u> estimate an individual's risk of developing a disease, based on certain genetic and/or clinical variables.

<u>**Disease Diagnosis –:**</u> classify whether an individual has a disease, or a disease subtype, based on certain genetic and/or clinical variables.[9,37]

<u>**Disease Prognosis –:**</u> estimate the risk of further adverse outcome(s) *subsequent to* diagnosis of disease.[38]

<u>**Therapeutic –:**</u> predict a patient or subgroup's response to a particular treatment.[39]

Box 2:

Current CHD PRS and their potential uses

Many PRS have been developed for CHD, varying in the computational methods used, number of variants included (50–6,000,000), and cohorts used for PRS training. For example, many of the latest CHD PRS use GWAS summary statistics from the CardiogramPlusC4D study[40], and differ by the method of selecting and weighting individual variants (including LDpred[41,42], lassosum[43], and meta-scoring approaches[44]) and how they are used in an integrated risk model. These PRS may provide useful information for predicting CHD risk that is largely orthogonal to conventional risk factors (age, sex, hypertension, cholesterol, BMI, diabetes, smoking) as well as family history. Clinical applications may include:

- Improved risk prediction for future adverse cardiovascular events when added to conventional risk models (such as the Framingham Risk Score [45], Pooled Cohort Equations [43,44], QRISK [43]).

- Reclassification of risk categories often leading to recommendations for risk-reducing treatments like statins.[45-47]

While the data for these clinical applications strongly suggest CHD PRS may improve patient outcomes, clinical utility through randomized clinical trials has yet to be established; however, a number of clinical trials are underway.[48]

Box 3:

Current breast cancer PRS and their potential uses

Many of the most recent and most predictive PRS for breast cancer include a smaller number of variants (usually 100-1,000s), possibly owing to a less polygenic architecture and more low-frequency variants having greater effect; however, scores composed of millions of variants also exist. PRS construction typically includes GWAS summary statistics and data from the Breast Cancer Association Consortium (BCAC) then variants passing genome-wide significance (lead SNPs), stepwise regression, penalized regression on individual-level genotypes[49], Clumping + Thresholding[41], or Bayesian methods[42,50]. In contrast to CHD, genetics is commonly used to measure breast cancer risk vis-a-vis *BRCA1/2* mutation testing; however, routine screening for breast cancer is often performed in older women using non-genetic risk prediction tools, such as mammography. Research into PRS for breast cancer include multiple potential clinical uses and considerations:

- Multiple PRS exist to predict risk for subtypes of breast cancer (e.g. ER positive/negative, luminal, and triple negative[49,51]), which could be used to stratify patients according to prognosis or for more beneficial treatments.

- Integrated risk models which combine PRS with non-genetic risk factors (such as age, family history, mammographic density, hormone replacement therapy).[52-59]

- PRS can provide important stratification of risk among carriers of pathogenic variants in genes already screened in clinical practice (e.g. *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM*)[60] and thus could improve clinical decision making.[42,50,57,61-64]

Indeed, the BOADICEA breast cancer risk prediction model includes the effects of common variants (PRS313[49]) as well as other rare pathogenic genetic variants[57] and has been implemented in the CanRisk Tool (www.canrisk.org), which has been approved for use by healthcare professionals in the European Economic Area. PRS utility has been studied via simulation[65] and is being evaluated in risk-based breast cancer screening trials in the US[66] and Europe (MyPeBS; https://mypebs.eu).
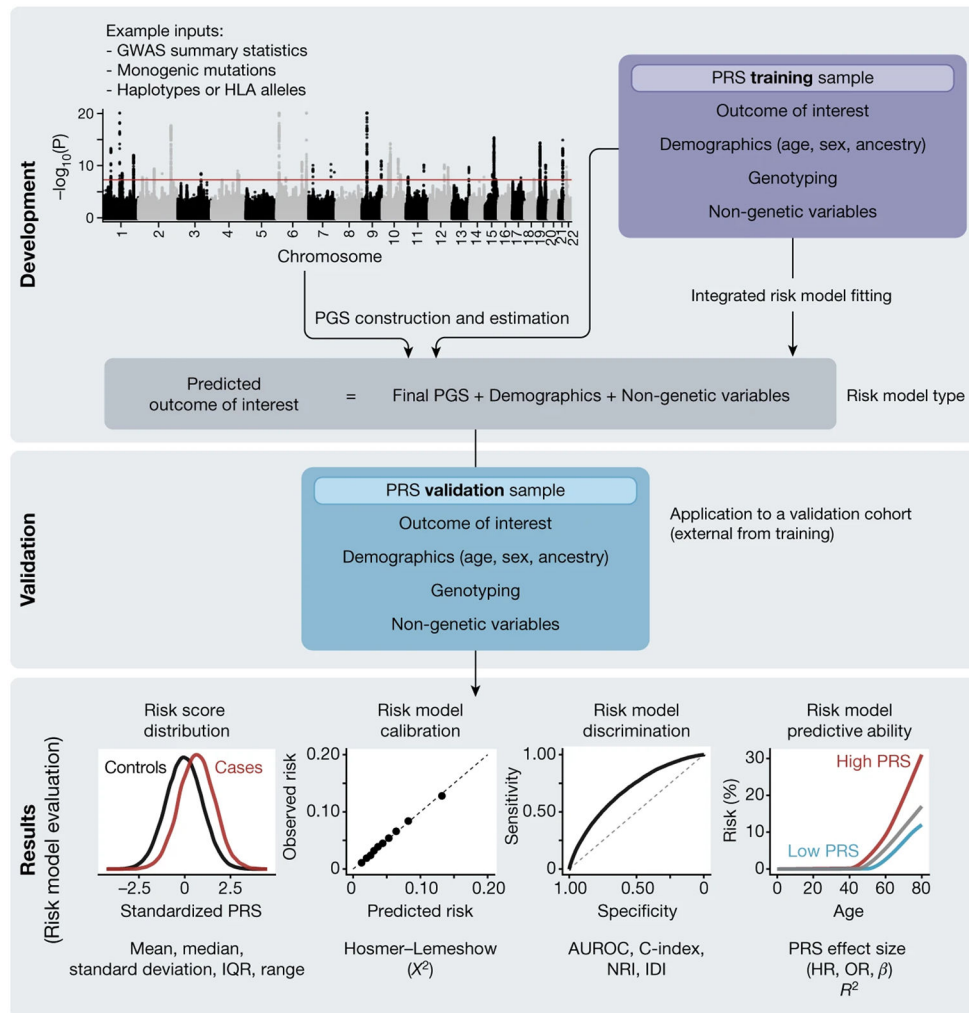
Figure 1: Prototype of PRS development and validation process.
The prototypical steps for PRS construction, risk model development, and validation of performance are displayed with select aspects of the PRS-RS guideline (labeled in **bold**). During PRS development, variants associated with an outcome of interest, typically identified from a GWAS, are combined as a weighted sum of allele counts. Methods for optimizing variant selection (PRS construction & estimation) are not shown. To predict the outcome of interest the PRS is added to a risk model and may be combined with non-genetic variables (e.g. age, sex, ancestry, clinical variables; collectively referred to as risk model variables). After fitting procedures to select the best risk model, this model is validated in an independent sample. The PRS distribution should be described, and the performance of the risk model demonstrated in terms of its discrimination, predictive ability, and calibration. Though not displayed in the figure, these same results should also be reported for the training sample for comparison to the validation sample. In both training and validation cohorts, the outcome of interest criteria, demographics, genotyping, and non-genetic variables should be reported (Table 1).

Table 1.

Polygenic Risk Score Reporting Statement (PRS-RS)

Reporting Standard		Description
Background	**Study Type**	Specify whether the study aims to develop and/or validate a PRS. When externally validating or combining previously published PRS or integrated risk model, include identifier(s) of original PRS (PMID, PGS Catalog ID).
	Risk Model Purpose & Predicted Outcome	Specify what the risk model is intended to predict and the purpose. This includes intended use (risk prediction, diagnostic, prognostic, or therapeutic modalities), predicted outcome (if a clinical feature or endpoint within a specific disease), and the current models available for that outcome.
Study Population and Data *Many risk score studies involve multiple populations and cohorts that can be used in different stages of PRS and risk score development and evaluation. Each of the populations used (e.g. training, validation, subgroup analysis samples) in the manuscript should be defined using this common set of descriptors.*	**Study Design & Recruitment**	For each of the data sets describe the study design (e.g., cohort, case-control, crosssectional), eligbility criteria, recruitment period and setting (e.g. method and years), and follow-up. State whether the data are primary or secondary data. If secondary analysis, include the full reference to the original study.
	Participant Demographic and Clinical Characteristics	Include the distribution of demographic information in each data set (and the combined total if relevant) used to generate a single risk model (whether a single sample set, or the summary of combined samples) including the mean, standard deviation and range. This should at minimum include age, sex and any other characteristics relevant to describe the study population or the performance of the model. Provide demographics stratified by case/control status, if applicable.
	Ancestry	Include the ancestral background distribution of each sample population used during PRS development and validation (including those from any GWAS summary statistics that were included), and the data source of this ancestry information (e.g., self-report, genotyping). Ancestry information should be reported using the standardized framework developed by the NHGRI-EBI GWAS Catalog with detailed information beyond this when available. When combining samples from multiple studies, aggregate ancestral distribution information is sufficient. The method of ancestry inference should be provided.
	Genetic Data	Provide method for acquiring genetic information (e.g. sequencing, genotyping) in each sample, including information about genome build and technical assay details. If imputed, specify the imputation panel and give ancestry information. Report any relevant quality control, including imputation quality filters to exclude low quality imputed SNPs. If parameters were selected from another study, include reference (PMID, GWAS Catalog ID).
	Non-Genetic Variables	Define any non-genetic variables that were included in the risk model, provide variable definitions and measurement (e.g., assay, ICD codes, e-phenotyping algorithms, chart review, self-report). Indicate the scale of each variable, e.g. dichotomous, continuous, categorical, or ordinal. Explicitly state which variables are included in the final model.
	Outcome of Interest	Define the predicted outcome(s) of interest and report distribution. If the predicted outcome is a clinical feature or endpoint within a specific disease, provide the criteria used to define that disease membership. Include details on how information was ascertained (e.g., ICD codes, e-phenotyping algorithms, chart review, self-report). Transformation of continuous data into binary, ordinal, or categorical outcomes should be detailed with justification. State whether the predicted phenotype of the polygenic score is the same or different than the predicted outcome of the risk model. Provide justification for differences, if applicable.
	Missing Data	State explicitly how missing data were handled for all variables included in the model. If imputation was utilized, include detailed of the approach used and any subsequent filtering or post-processing.
Risk Model Development & Application *Describe methods used to form the final*	**Polygenic Risk Score**	Describe how genetic data were included in the PRS. Authors should detail criteria used to determine inclusion in the model for

Reporting Standard		Description
PRS and/or risk model. Samples in this stage of the analysis should be denoted "Score Development" or "Training", and described in <u>Study Populations</u>.	**Construction & Estimation**	all variants. Define how the variants were selected, weighted and combined into a single score. If the PRS was derived from another study include the reference (PMID, PGS Catalog Score ID).
	Risk Model Type	Detail statistical methods used to estimate risk, either relative or absolute, from the continuous risk score distribution. Detail if risk is cumulative or cross-sectional, with appropriate comparison groups if relative risk presented. Report time until predicted risk (e.g. 5-year, 10-year, lifetime). In an absolute risk model, state the time until predicted event and the prevalence/incidence of the predicted outcome in the general population.
	Integrated Risk Model(s) Description and Fitting	State the procedure utilized to develop the risk models that includes non-genetic and/or genetic variables other than the PRS. If the model(s) was selected for optimal performance, describe measures used to assess performance. Explicitly state all variables used in each risk model.
Risk Model Evaluation *Outline results and procedures utilized to evaluate the risk model, specifying internal or external validation. Performance results should be described for both development and validation samples. Specify if the application of the risk model differs between the development and validation samples.*	**PRS**	Include a general description of the distribution of the PRS. This details the continuous
	Distribution	distribution output directly from the risk score calculation.
	Risk Model Predictive Ability	Describe and report metrics of overall performance (proportion of variance explained; R^2) and estimates of risk (such as odds or hazards ratios from regression models) used to evaluate the PRS and/or risk models. Describe the set of genetic/non-genetic variables included in the analysis.
	Risk Model Discrimination	Describe and report metrics (such as ROC or Precision-Recall (AUROC/AUPRC) and the Concordance statistic (C-index) for survival models) used to assess the discrimination of evaluated risk models and whether any non-genetic variables were included beyond a PRS in this analysis. Evaluation of potential clinical utility of models requires evaluating tail-based measures, such as proportions of populations and cases exceeding specified clinically relevant risk thresholds and measures of reclassfications (e.g. NRI) at such thresholds for comparison of models.
	Risk Model Calibration	Describe and report metrics used to assess the calibration of evaluated risk scores and models. Describe the set of genetic/non-genetic variables included in the analysis.
	Subgroup Analyses	Subgroup size, demographics and clinical characteristics should be given. Relevant evaluation methods and measures (distribution, predictive ability, discrimination, calibration) should be described for each subgroup analysis.
Limitations and Clinical Implications *Discuss the broader context of the study and risk model.*	**Risk Model Interpretation**	Summarize the risk models in terms of what they predict, how well, and in whom. Explicitly mention the incremental performance of the PRS and/or combined risk model in comparison to conventional risk models, as well as the performance of the PRS and risk model alone. Conventional risk models might include demographic (age, sex), disease-specific risk factors, and/or family history of disease.
	Limitations	Outline limitations of the study with relevance to the results, discuss the impact of these limitations on the interpretation of the risk model and any downstream replication efforts needed. Common considerations include: study design restrictions, use of a surrogate outcome, ascertainment biases, the distribution of participant-level traits (ancestry, age, comorbidities), accuracy/specificity of outcome data, and any statistical considerations. Note and discuss the impact of any unknown reporting items from previous sections.
	Generalizability	Discuss the intended target groups or populations this score may be applied to and explicitly address any issues with generalizability beyond the included populations. Discuss whether the study externally validates the score and/or model, or if the sample is limited with respect to ancestry, age, or other variables.
	Risk Model Intended Uses	Discuss whether there is an intended clinical use or utility to the risk model. If so, discuss the "clinic readiness" and next steps with respect to the interpretation, limitations, and generalizability of the

Reporting Standard		Description
		model. Discuss how the predictive ability of the model is compared against current standard of care or other published work (such as existing PRS) on predicting the outcome of interest.
Data Transparency and Availability		Information sufficient to calculate the PRS and the risk model(s) on external samples should be made freely available. For genetic variables this would include information about the variants (e.g., rsID, chromosomal location, effect allele, and the effect weight) that comprise the score; PRS with this information should be deposited in the PGS Catalog for findability and to promote re-use and comparison with other established scores. Weights for non-genetic variables should also be provided to make the risk model calculable.

Further reporting considerations beyond the minimal reporting for PRS-RS items can be found in Supplemental Note 3. A reference of relevant manuscript sections for each item is provided in Supplemental Table 4.