Improving reporting standards for polygenic scores in risk prediction studies

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

[#]Shared first authorship
[$]Shared senior authorship

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 Bioinformatics Institute, European Molecular Biology Laboratory, Wellcome Genome Campus, Hinxton, UK
8. National Human Genome Research Institute, Bethesda, MD, USA
9. Mayo Clinic, Rochester, MN, USA
10. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
11. Baylor College of Medicine, Houston, TX, USA
12. Department of Human Genetics, Genentech, South San Francisco, CA, USA
13. Wellcome Centre for Human Genetics, Oxford, UK
14. University of Cambridge, Cambridge, United Kingdom
15. Western University, London, ON, Canada
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. Division of Pharmaceutical Outcomes and Policy, UNC Eshelman School of Pharmacy, Chapel Hill, NC, USA
22. Stanford Center for Biomedical Ethics, Stanford University School of Medicine, Stanford, CA, USA
23. Centers for Disease Control and Prevention, Atlanta, GA, USA
24. Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA
25. Department of Translational and Applied Genomics, Kaiser Permanente Northwest, Portland, OR, USA
26. Center for Health Research, Kaiser Permanente Northwest, Portland, OR, USA
27. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
28. National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge and Cambridge University Hospitals, Cambridge, UK
29. The Alan Turing Institute, London, UK
30. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

*Corresponding author: Dr. Genevieve L. Wojcik, gwojcik1@jhu.edu

Summary

In recent years, polygenic risk scores (PRS) have become an increasingly studied tool to capture the genome-wide liability underlying many human traits and diseases, hoping to better inform an individual's genetic risk. However, a lack of adherence to previous reporting standards has hindered the translation of this important tool into clinical and public health practice with the heterogeneous underreporting of details necessary for benchmarking and reproducibility. To address this gap, the ClinGen Complex Disease Working Group and Polygenic Score (PGS) Catalog have collaborated to develop the 33-item Polygenic Risk Score Reporting Statement (PRS-RS). This framework provides the minimal information expected of authors to promote the internal validity, transparency, and reproducibility of PRS by requiring authors to detail the study population, statistical methods, and clinical utility of a published score. The widespread adoption of this framework will encourage rigorous methodological consideration and facilitate benchmarking to ensure high quality scores are translated into the clinic.

Introduction

The predisposition to common diseases and traits arises from a complex interaction between genetic and nongenetic factors. During the past decade, international collaborations involving cataloged human genetic variation, and large cohorts of well-phenotyped individuals with matched genotype information have enabled the discovery of disease-associated genetic variants.[1–4] In particular, genome-wide association studies (GWAS) have emerged as a powerful approach to identify disease- or trait-associated genetic variants, typically yielding summary statistics describing the magnitude (effect size) and statistical significance of association between an allele and the trait of interest.[4,5] GWAS have been applied to a wide range of complex human traits and diseases, including height, blood pressure, cardiovascular disease, cancer, obesity, and Alzheimer's disease.

The associations identified via GWAS can quantify genetic predisposition to a heritable trait, which can be used to conduct disease risk stratification or predict prognostic outcomes and response to therapy.[6,7] Typically, information across many variants is used to form a weighted sum of allele counts across variants, where the weights reflect the magnitude of association between variant alleles and the trait or disease. These weighted sums can include up to 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 disease risk; or, more generally, **polygenic score(s) (PGS)** when referring to any phenotype (*see **Box 1** for further discussion of nomenclature*). While there is active development of algorithms to decide how many and which variants to include and how much to weigh them so as to maximize the proportion variance explained or the disease discrimination, there is an emerging consensus that the inclusion of variants beyond those meeting stringent GWAS significance levels can boost predictive performance. [8,9]

Box 1. Definitions of relevant genetic risk prediction terms

Polygenic Score(s) (PGS): a single value that estimates the genetic contribution to inter-individual variation in 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. Sometimes referred to as a genetic score.

Polygenic Risk Score(s) (PRS): a 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 of PRS below.

Integrated Risk Model: a risk model combining PGS/PRS with other established risk factors, such as demographics (often age and sex), anthropometrics, biomarkers, and clinical measurements to estimate a specific disease risk.

Categories of use for PRS and/or integrated risk models

The addition of PRS to existing risk models has several potential applications, summarized below. In each, the aim of PRS integration is to improve individual or subgroup classification to the extent that there is clinical benefit.

Disease Risk Prediction – used to estimate an individual's risk of developing a disease, based on the presence of certain genetic and/or clinical variables.

Disease Diagnosis – used to classify whether an individual has a disease, or a disease subtype, linked to a certain etiology based on the presence of certain genetic and/or clinical variables.[8,10]

Disease Prognosis – used to estimate the risk of further adverse outcome(s) *subsequent to* diagnosis of disease.[11]

Therapeutic – used to predict a patient or subgroup's response to a particular treatment.[12]

Frameworks have been developed to establish standards around the transparent, standardized, accurate, complete and meaningful reporting of scientific studies. Those relevant for development and validation of risk prediction models include PICOT[13], TRIPOD[14], STROBE[15], STREGA[16], and, notably, the Genetic Risk Prediction Studies (GRIPS) Statement[17] which specifically address reporting of *genetic* risk prediction studies. However, no framework adequately addresses the emerging use of PRS in clinical care and disease prevention.

It is time to update the GRIPS statement as PRS have become ubiquitous in genetic research while reporting remains heterogeneous, particularly in terms of transparency and enabling reproducibility. The methods utilised for PRS construction and risk-model development have become more diverse and sophisticated.[18–20] Biobanks and large-scale consortia have become dominant, yet frequently have limited access to individual-level data. Standards for reporting genetic ancestry information have been developed[1], and there is a push towards open data sharing as outlined in the FAIR (Findable, Accessible, Interoperable and Reusable) Data Principles.[3,21] Finally, the rapid rise of direct-to-consumer assays and companies

(including 23andMe, Color, MyHeritage, etc.) providing PGS/PRS results to customers has vastly increased the scope and complexity of genetic risk information. The readiness of PRS for implementation varies among phenotypes, with only a few diseases like breast cancer [22–24] and coronary heart disease (CHD) having mature PRS with potential clinical utility *(see Box 2 for additional discussion of CHD)*. These concomitant advances have resulted in healthcare systems developing new infrastructures to deliver genetic risk information, and the field now needs to develop standards for clinical applications of PRS.

Box 2: Current CHD PRS and their potential uses

While many PRS have been developed to predict CHD, they vary greatly in the computational methods used to develop them, the number of variants included (50–6,000,000), and the GWAS and cohorts used for PRS training. For example, the latest and currently most predictive CHD PRS use GWAS summary statistics from the CardiogramPlusC4D study [25], and mainly differ by the computational methods used to select the included variants (including LDpred[26,27], lassosum[28], and meta-scoring approaches[29]), and how they are combined into risk models. These PRS, however, may provide useful information for predicting risk of CHD being largely orthogonal to conventional risk factors (age, sex, blood pressure, cholesterol, BMI, smoking) as well as family history. Clinical applications may include:

- Improved risk prediction for future adverse cardiovascular events when added to traditional risk models (including Framingham Risk Score[30], Pooled Cohort Equations[28,29], QRISK[28]).
- Reclassification of risk categories often leading to recommendations related to risk-reducing treatments like statins.[30–32]

While the data strongly suggest CHD PRS, by refining risk estimates, may improve patient outcomes, clinical utility through randomized clinical trials has yet to be conclusively established. We anticipate this is the future direction of PRS studies, and a number of clinical trials are underway.[33]

At present, there are no uniformly agreed best practices for developing PRS nor any regulation or standards for reporting or assessing their clinical readiness. 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. Here, the Clinical Genome Resource (ClinGen) Complex Disease Working Group and the Polygenic Score (PGS) Catalog jointly present reporting standards that address current PRS research and highlight a complementary centralized repository for PRS with the ClinGen-PGS Catalog joint Polygenic Risk Score Reporting Standards (PRS-RS). We outline a foundation for transparent, standardized, accurate, and clinically meaningful reporting on PRS development and validation in the literature to overcome these barriers.

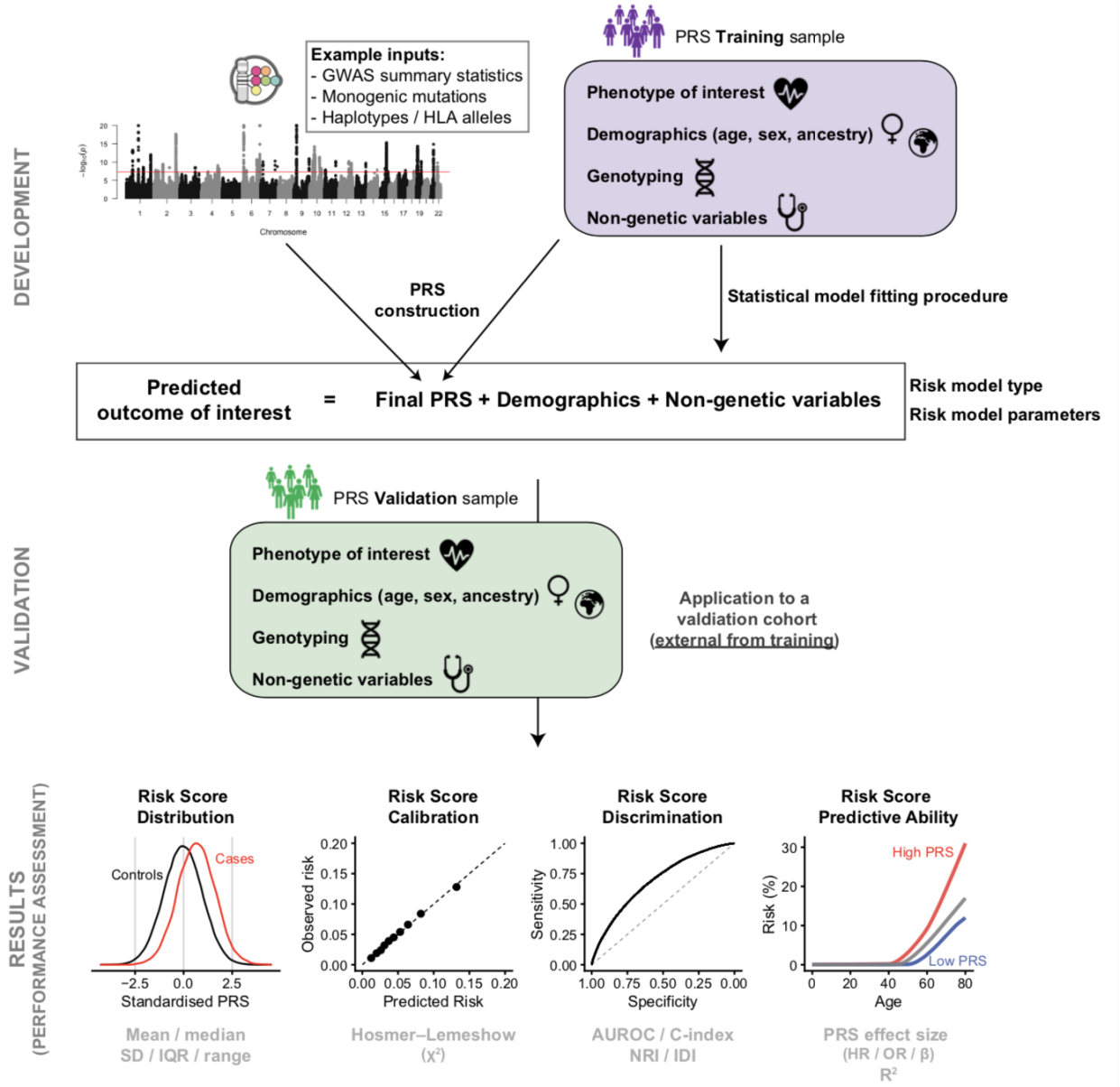

Figure 1: Prototype of PRS development and validation process. Figure 1 displays prototypical steps for PRS construction, development, validation, and performance, with select aspects of the ClinGen PRS reporting guideline highlighted throughout. In PRS development, variants associated with a phenotype of interest, typically identified from a GWAS, are combined as a weighted sum of allele counts across variants. Methods for optimizing variant selection for a PRS (PRS Construction) are not shown. The PRS is tested in a risk model predicting the phenotype of interest and may be combined with other non-clinical variables. Collectively, all variables included in the risk model are referred to as the risk model parameters. After fitting procedures to select the best risk model, this model is validated in an independent sample. The performance of a model is demonstrated though risk score distribution, 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 phenotype of interest criteria, demographics, genotyping, and non-genetic variables should be reported (**Table 1**).

Results

Updated PRS Reporting Guidelines

This guideline aims to specify the minimal criteria needed to accurately interpret a PRS and reproduce results throughout the PRS development process, briefly illustrated in Figure 1. It applies to the development and validation studies for PRS that aim to predict disease, prognosis and response to therapies. **Table 1** presents the full PRS-RS.

Reporting on risk score background: As the PRS-RS are focused on future utility and implementation, authors must outline the study and target population and appropriate outcomes to understand what risk is measured. Authors should use the appropriate data needed to address the intended clinical use, with adequate documentation of dataset characteristics to inform understanding of the nuance in measured risk. For example, when developing a risk score, there can be disparities between the *risk-score predicted clinical end-outcome* (e.g., cardiovascular risk) and the measured *phenotype of interest* (e.g., LDL cholesterol) used in the analysis. If a surrogate outcome is used, this should be clearly stated with an explanation including the limitations of the *study design* or *study recruitment* method, in which the clinical end outcome was not measured.

Reporting on populations: The "who, where, and when" of risk depend on the study population used to derive the risk model. Therefore, authors need to define and characterize the demographics of their study population, especially the *age*, *sex*, and *ancestry* composition. 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. [1,8,34] It is essential for authors to define participants based on their self-identified ancestry, with a standardized framework developed by the NHGRI-EBI GWAS Catalog to enable comparability between studies.[1] While these three demographic variables are the most universally relevant, authors should provide sufficient level of detail for all relevant factors for the outcome of interest, especially if they are included in the final risk model under *risk model parameter specifications*.

Reporting on Methods: There are currently several methods that are commonly used to select variants and fine-tune weights. [7,18–20,35] As the performance and limitations of the risk score are dependent on these methods, authors must provide complete details including the source of genetic information (*risk model genetic data acquisition)* and inclusion/exclusion criteria (*risk model parameter specifications*). Once these parameters are defined, authors should describe the methods used to transform the raw data, often derived from GWAS summary statistics, to a sum of variants for their *polygenic risk score estimation*. Often authors will iterate through numerous models to find the optimal fit. Therefore, in addition to the estimation methods, it is important to detail the *statistical model fitting procedure*, including the measures used for the final model selection.

Reporting on Risk Estimation: Translating the continuous PRS distribution to a risk estimate, whether absolute or relative, is highly dependent on assumptions about and limitations with the specific data set utilised. 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. The *risk score calibration* and *discrimination* should be described for all analyses, although their estimation and interpretation are most relevant for validation, preferably with an external validation set. Any differences in variable definitions between the training and validation sets should be described.

Reporting on Model Parameters: Reporting actual estimates, not only the methods behind decision-making, enables readers to gauge the relative value of an increase in performance against other trade-offs. Making the underlying PRS (variant alleles and derived weights) publicly available and submitting them to the PGS Catalog allows others to reuse existing models (with known validity) and enables direct benchmarking between different PRS for the same trait. The current mathematical form of most PRS—a linear combination of allele counts—facilitates model description and reproducibility. Future genomic risk models may have more complicated forms, e.g. allowing for non-linear epistatic and gene-environment

interactions. It will be important to describe these models in sufficient detail to allow their implementation by other researchers and clinical groups; this might entail sharing open-source code.

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-score intended purpose,* such as how the performance of their PRS compares to other commonly used risk metrics, either from previously published PRS or conventional risk calculators, such as the pooled cohort equations for estimating atherosclerotic cardiovascular disease risk.[36] This is important — what indicates a "good" prediction can differ between outcomes and *intended purposes*.

Lastly, we would like to reiterate the need for both methodological and data transparency. Deposition in a resource such as the PGS Catalog provides an invaluable resource for widespread adoption and improvement of a published PRS. *Supplemental Table 1* provides additional reporting considerations on top of 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 Table 1*, especially those related to discussing the interpretation, limitations, and generalizability of results.

Compatibility with the PGS Catalog submission template

The PGS Catalog (www.PGSCatalog.org; Lambert et al., 2019) provides access to PGS scores and related metadata to support the FAIR principles of data stewardship [21], enabling subsequent applications and assessments of PGS performance and best practices. The goals of the PGS Catalog align with ClinGen with slight differences in how the data is represented in the Catalog (link). Overall, there is a good agreement between the PRS-RS and PGS Catalog representation schema (field by field mapping outlined in *Supplemental Table 2A*), particularly with respect to how study participants are described. Reporting items in the PRS-RS that are not present in the PGS Catalog (*Supplemental Table 2B*) include descriptions, goals, limitations and intended uses of PRS predictions and implementation that are not essential to the

Catalog's goal of indexing available published PGS with the metadata essential for interpretation and reproducibility. PRS described using the PRS-RS items contain sufficient detail for their addition to the PGS Catalog, as such we recommend that authors describe scores using the PRS-RS and submit them to the PGS Catalog upon publication.

Box 3. Many papers lack sufficient detail for interpretation

We carried out a iterature review (representative across a variety of diseases, risk score categories, and populations) to revise the PRS-RS (N=30). It revealed multiple reporting items with insufficient and/or missing details as well as variable details provided on methods, results, and discussion (***Figure 2***). Papers had insufficient detail for items related to study design and variables, particularly phenotype of interest (inclusion/exclusion criteria and control definitions) and ancestry (definition, distribution of participants). There was also insufficient detail and absent reporting for items needed to reproduce or critically assess the analytic validity of a PRS, including statistical validation model, risk score calibration, and data transparency and availability. We observed variable, often absent, discussion about the intended clinical purpose or utility of the score, and on how the PRS compared to standard of care. Papers also varied in their discussion of study limitations, particularly with regard to how the ancestry of the training and validation sets affected PRS generalizability.

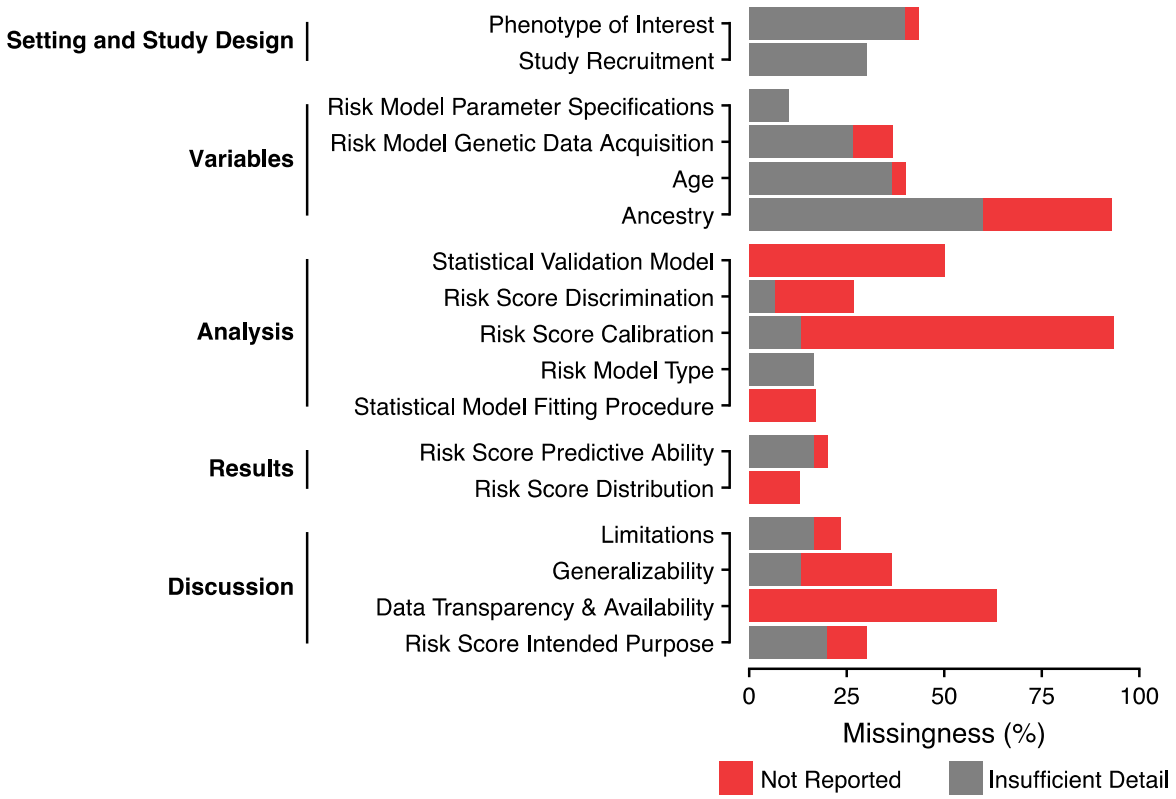

Figure 2. PRS-RS items with missing and insufficient detail. A total of 30 papers were reviewed.

Discussion

Polygenic risk scores have transformed human genetic research and emerged as potentially powerful tools for the translation of genomic discoveries into clinical and public health benefits. However, standardized and robust methods and reporting criteria are urgently needed in this area order to realize its potential. The heterogeneity in the reporting of PRS to date—including what to report and how to report it—highlights the challenges for accurate interpretation and confidence in PRS, especially with respect to assessing clinical readiness. **(Box 3)** Critical aspects of PRS studies, including ancestry, predictive ability, and transparency/availability of information needed to reproduce PRS, are frequently inadequately reported. Without these aspects, PRS cannot be rigorously assessed and compared, even within the same disease or recruitment cohorts. This underscores the need for a reporting standard with clear, specific definitions that conveys the meaningful aspects of PRS development and testing, which are critical to understanding PRS predictive ability and specificity for the intended target population and phenotype of interest.

The ClinGen Complex Disease Working Group utilized an iterative review process incorporating previous standards, expert opinion and practical considerations to create an updated PRS-RS of 33 items, spanning PRS derivation, testing and validation steps ***(Table 1)***. The reporting guideline is complemented by the PGS Catalog (www.PGSCatalog.org), an online database that freely provides the underlying information necessary to calculate the polygenic score (e.g., variants and weights) and curates important meta-data on polygenic scores in structured, standardized formats. The PGS Catalog provides an open platform for implementing reporting standards and lays the foundation for assessing best practices in polygenic score research.

ClinGen has incorporated multiple sources to create a guide that is flexible, pragmatic, and informed. Researchers using this guideline may identify fringe cases that are inadequately covered, and the guide may become dated as PRS research continues to mature. However, by updating previous standards, involving current leaders in the field, and adapting the framework pragmatically to the barriers observed in recent literature, we aimed to provide a comprehensive perspective on the topic. In comparison to the original GRIPS statement, PRS-RS has expanded on elements related to understanding the clinical validity

of PRS and consequent risk models. Items such as risk score predicted clinical end outcome (introduction) and risk score intended purpose (discussion) bookend our guideline with the intended clinical framing of PRS reporting. Most other items in PRS-RS are consistent with the original GRIPS items but are presented in greater detail for more accurately assessing the clinical validity of PRS when needed. For example, the GRIPS item called “study design and setting” has been split into “study design” and “study recruitment,” in acknowledgment of the importance of both distinct pieces of information in understanding study biases or limitations.

While the scope of our work encompasses clinical validity, it does not address the additional requirements needed for clinical or public health utility, such as randomized trials with clinically meaningful outcomes, health economic evaluations, or feasibility studies. In addition, the translation of structured data elements into useful clinical parameters may not be direct. Two relevant examples are (i) the disease case definitions utilized in training or validation in any particular PRS study may deviate (sometimes substantially) from those utilized in any specific health system, and (ii) the definitions used for race/ancestry as outlined in the PGS and GWAS Catalog[1] may also deviate from structured terms used to document ancestry information in clinical care. Such translation issues potentially limit generalizability to target populations and warrant further discussion. Nevertheless, we have emphasized the need for authors to be mindful of their intended purpose and target audience when discussing their findings. Additionally, 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 polygenic scores, or the underlying data thereof. This work can 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 PRS information. The PGS Catalog 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.[37] In addition, it provides a structured database of scores (variants and effect sizes) and metadata requested in 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.

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. In addition, future reporting guidelines should address important questions about clinical readiness, specifically about intended use and target populations to help ease translation into practice. While the working group has begun to address how changes in PRS practices should be accounted for when reporting a PRS, future research should attempt to create a reporting guideline that anticipates the consequences of new methods, such as deep learning. We encourage readers to visit the ClinGen complex disease website (https://clinicalgenome.org/working-groups/complex-disease/) for any future changes or amendments to the reporting guideline.

Methods

ClinGen Complex Disease Working Group

The working group, founded by ClinGen in November 2018, comprised more than thirty experts with epidemiological, statistical, disease-domain specific, implementation science, actionability, and ELSI interests in polygenic risk score application. Members met twice a month to discuss current research, best practices, and limitations within their respective areas of expertise. As a result of these meetings, the workgroup decided to update previous genetic risk-score reporting standards[17] to current PRS practices. This aim was finalized at the NHGRI Genomic Medicine XII: Genomics and Risk Prediction meeting in May 2019 with input from the external scientific community in terms of mission, scope, and long-term objectives of the working group. Current descriptions of workgroup members and goals are available at: https://clinicalgenome.org/working-groups/complex-disease/

The Polygenic Score (PGS) Catalog

The PGS Catalog was founded in 2019 by researchers at the University of Cambridge UK, European Bioinformatics Institute (EMBL-EBI) and Baker Institute, and developed as a sister resource to the NHGRI-

EBI GWAS Catalog[37]. Its goal is to provide an open database of PGS and relevant metadata, so that published PRS/PGS can be distributed, applied, and evaluated in a rigorous and replicable manner in both research and clinical settings. It reports key information about how a PGS has been developed (e.g., variant selection and computational methods), information about the specific datasets used for PGS development and evaluation (e.g., sample size, ancestry, phenotype description), as well as the performance metrics reported during PGS evaluation (e.g., effect sizes, covariates, and/or classification metrics). These data are represented in a schema that links the Scores, Samples, and Performance Metrics presented in each PGS publication. The PGS Catalog is available at www.PGSCatalog.org; additional descriptions of the project, development, methods, full descriptions of the representation schema, along with links for PGS submission can be found in the documentation (www.pgscatalog.org/about/) and will be described in a future publication. (Lambert *et al., in preparation*)

PRS Reporting Framework: Expert Guidelines Approach

PRS-RS was developed in iterative phases utilizing previous standards, expert opinion, and a pragmatic literature review process. First, the entire expert working group created the initial framework draft by adapting previous genetic risk-score reporting standards to current PRS methodologies. This was followed by a second round of revisions using a literature review. These steps led to a series of proposed revisions that were finalized with the entire working group. Finally, PRS-RS and PGS Catalog fields were mapped onto one another, and definitions were modified to reflect shared language, when possible.

Draft guidelines from previous guidelines

The draft PRS-RS largely expanded on the GRIPS guidelines for genetic risk-prediction studies published in 2011.[17] To create the preliminary framework, we relied on expert opinion, and were guided by the PICOT framework used to compare heterogeneous clinical trial outcomes.[13] Our revisions focused on eliciting the individual components from previous standards that experts deemed independently important for transparent interpretation and reproducibility of a risk score (Supplemental Figure 1). We expanded the original GRIPS checklist of 25 items to 44 unique items, of which 33 items were needed for both training and validation cohorts. The majority of these additions were added to explicitly list discrete elements within

an individual GRIPS checklist item if those elements were determined by the work group to have significant impact in the interpretation of a PRS in terms of either analytic validity, clinical validity, or clinical utility. The PICOT framework did not add items to the reporting guidelines, but we did confirm that PICOT concepts were represented in the reporting guideline to facilitate downstream applications of comparing heterogeneous outcomes.

PRS-RS revisions using literature review

We used the PRS-RS checklist to curate original research articles on polygenic risk-score development or validation as a measure of pragmatism and clarity. Thirty-five papers were initially collected via the snowball sampling search based on their use of the term “polygenic risk score” and their research in human populations in preparation for the NHGRI Genomic Medicine XII meeting. Five papers were excluded from the review because they were not original articles, did not develop or validate a PRS, duplicated a previous study, or did not use genetic loci to construct their risk scores. Included articles spanned a variety of disease domains including Alzheimer’s disease, asthma, breast cancer, cerebrovascular event, colon cancer, coronary heart disease, depression, fracture risk, Parkinson’s disease, prostate cancer, and schizophrenia. In addition, articles were selected for variety in the risk score category (development vs. external validation; diagnostic vs. prognostic). Article references are available in the supplement.

Two independent reviewers assessed each article using the draft PRS reporting framework. A 10-person volunteer subgroup of the larger working group met bi-weekly to resolve inter-reviewer discrepancies. If the subgroup was unable to reach a consensus, one of four expert reviewers from the working group was assigned to resolve discrepancies in a third review. This pilot of the reporting guideline on published PRS revealed pragmatic areas for revision. Similar items were combined if they did not individually contribute meaningful concepts for PRS interpretation. Items were removed if they did not contribute to overall interpretation of the risk-score performance or target application. Definitions were expanded and revised to address inconsistencies in inter-reviewer interpretation due to heterogeneous and vague reporting in the literature. Items were kept as discrete items if we observed substantially missing or insufficiently detailed reports on these items in the literature for transparency. When applicable, updated methodology was also

included in definitions. Finally, supplemental considerations were created to address fringe cases (Supplementary Table 1). Proposed reporting guideline revisions were ratified in monthly calls with the entire workgroup. This final 33-item PRS-RS is presented in ***Table 1***.

Papers were re-curated using the final reporting guideline. The majority of papers (25/30) were predicting risk of developing disease with a few characterizing prognostic outcomes. Nearly half of the papers (13/30) developed a novel risk score, while the other half either externally validated a previously published risk score (9/30) or both developed and externally validated the risk score (6/30). Two manuscripts modified a previously published score. The composition of the final published risk scores were limited to genetic variables for the majority of papers (25/30), with only five producing an integrated risk score.

PRS-RS harmonization with PGS Catalog

Two curators mapped reporting fields from the PGS Catalog onto the final PRS-RS guidelines. When possible, similar terminology was adopted between the two resources. A subset of fields in the PGS Catalog differ from PRS-RS due to restrictions in preserving integrity of the data infrastructure. The analogous ClinGen reporting item is listed in the PGS Catalog as a footnote to aid researchers, and this field mapping is available in the supplement (Supplementary Table 2).

Acknowledgements

We would like to acknowledge the input of the ClinGen Complex Disease Working Group members, including Carlos D. Bustamante, Michelle Meyer, Frank Harrell, David Kent, Peter Visscher, Tim Assimes, Sharon Plon, and Jonathan Berg. We would also like to thank Diane Durham for her editorial support and Angela Paolucci for her administrative support in the preparation and submission of this manuscript.

ClinGen is primarily funded by the National Human Genome Research Institute (NHGRI), through the following three grants: U41HG006834, U41HG009649, U41HG009650. ClinGen also receives support for content curation from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), through the following three grants: U24HD093483, U24HD093486, U24HD093487.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Additionally, the views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. Research reported in this publication was supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number U41HG007823 (EBI-NHGRI GWAS Catalog, PGS Catalog). In addition, we acknowledge funding from the European Molecular Biology Laboratory. Individuals were funded from the following sources: MIM was a Wellcome Investigator and an NIHR Senior Investigator with funding from NIDDK (U01-DK105535); Wellcome (090532, 098381, 106130, 203141, 212259). MI and SAL were supported by core funding from: the UK Medical Research Council (MR/L003120/1), the British Heart Foundation (RG/13/13/30194; RG/18/13/33946) and the National Institute for Health Research (Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust). This work was also supported by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome.

Author Contributions

HW, CT, MAI, IJK, CS, RR, KV, KK, KABG, PK, and GLW conducted the literature review of reporting practices. HW, SAL, CT, MIA, JWO, IJK, RR, DB, EV, MIM, ACA, DFE, RAH, AVK, NC, CK, KE, EMR, MCR, KEO, MJK, ACJWJ, KABG, PK, JALM, MI, and GLW provided feedback on the reporting framework. HW, SAL, CT, MAI, JWO, IJK, AVK, JND, HP, EMR, MR, ACJWJ, KABG, PK, JALM, MI, and GLW wrote and edited the manuscript text and items.

Competing Interests

MIM is on the advisory panels Pfizer, Novo Nordisk, and Zoe Global; Honoraria: Merck, Pfizer, Novo Nordisk, and Eli Lilly; Research funding: Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen,

Merck, Novo Nordisk, Pfizer, Roche, Sanofi Aventis, Servier & Takeda. As of June 2019, he is an employee of Genentech with stock and stock options in Roche. No other authors have competing interests to declare.

Bibliography

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).
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).
3. Claussnitzer, M. *et al.* A brief history of human disease genetics. *Nature* (2020).
4. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90,** 7–24 (2012).
5. Pasaniuc, B. & Price, A. L. Dissecting the genetics of complex traits using summary association statistics. *Nat. Rev. Genet.* **18,** 117–127 (2017).
6. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19,** 581–590 (2018).
7. Lambert, S. A., Abraham, G. & Inouye, M. Towards clinical utility of polygenic risk scores. *Hum. Mol. Genet.* **28,** R133–R142 (2019).
8. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51,** 584–591 (2019).
9. 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).
10. 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).
11. Choi, J. *et al.* The associations between immunity-related genes and breast cancer prognosis in Korean women. *PLoS One* **9,** e103593 (2014).
12. Marston, N. A. *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).

13. Rios, L. P., Ye, C. & Thabane, L. Association between framing of the research question using the PICOT format and reporting quality of randomized controlled trials. *BMC Med. Res. Methodol.* **10,** 11 (2010).
14. Moons, K. G. M. *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).
15. 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).
16. Little, J. *et al.* STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. *PLoS Med.* **6,** e22 (2009).
17. Janssens, A. C. J. W. *et al.* Strengthening the reporting of Genetic RIsk Prediction Studies: the GRIPS Statement. *PLoS Med.* **8,** e1000420 (2011).
18. Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X. & Sham, P. C. Polygenic scores via penalized regression on summary statistics. *Genet. Epidemiol.* **41,** 469–480 (2017).
19. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8,** (2019).
20. Vilhjálmsson, B. J. *et al.* Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97,** 576–592 (2015).
21. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3,** 160018 (2016).
22. 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).
23. Kuchenbaecker, K. B. *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).
24. Mavaddat, N. *et al.* Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am. J. Hum. Genet.* **104,** 21–34 (2019).
25. 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).

26. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50,** 1219–1224 (2018).
27. FinnGen *et al.* Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat. Med.* (2020). doi:10.1038/s41591-020-0800-0
28. 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).
29. 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).
30. Abraham, G. *et al.* Genomic prediction of coronary heart disease. *Eur. Heart J.* **37,** 3267–3278 (2016).
31. Ganna, A. *et al.* Multilocus genetic risk scores for coronary heart disease prediction. *Arterioscler. Thromb. Vasc. Biol.* **33,** 2267–2272 (2013).
32. Kullo, I. J. *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).
33. U.S. National Library of Medicine. Home - ClinicalTrials.gov. at <https://clinicaltrials.gov/>
34. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570,** 514–518 (2019).
35. Choi, S. W., Mak, T. S. H. & O'Reilly, P. A guide to performing Polygenic Risk Score analyses. *BioRxiv* (2018). doi:10.1101/416545
36. Yadlowsky, S. *et al.* Clinical implications of revised pooled cohort equations for estimating atherosclerotic cardiovascular disease risk. *Ann. Intern. Med.* **169,** 20–29 (2018).
37. 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).

Table 1. ClinGen PRS Reporting Guideline.

Bold= training and validation, when applicable

Introduction	
Article Type	Specify whether authors are developing a risk score and/or externally validating a previously published PRS. When externally validating or combining previously published PRS, include identifier(s) of original PRS (PMID, PGS catalog ID).
Risk Score Category	Specify the risk score's purpose as risk prediction, diagnostic, prognostic, or therapeutic (or combination of these).
Risk Score Predicted Clinical End Outcome	When describing the risk score purpose, include the clinical end outcome predicted by the final PRS model. If the predicted outcome is a clinical feature or endpoint within a specific disease, state the disease. Inclusion/exclusion criteria are specified in the methods.
Methods: Study Design	
Study Design	Include study design details including study type (e.g., cohort, case control, cross sectional), and whether the predicted clinical end outcome is defined by incidence or prevalence. State whether the data are primary or secondary data. If secondary analysis, include a reference to the original study. For PRS that combine samples from multiple studies, include this information for each study.
Study Recruitment	Stated or reference (if secondary analysis) recruitment details, such as method and years.
Phenotype of Interest	Provide the inclusion and exclusion criteria used to define the predicted clinical end outcome stated in the introduction. If the predicted outcome is a clinical feature or endpoint within a specific disease, provide the inclusion and exclusion criteria used to define that disease. Include details on how inclusion and exclusion information was ascertained (e.g., ICD codes, e-phenotyping algorithms, chart review, self-report). Authors should explicitly state the number of cases and controls included. For dichotomous data, also include inclusion and exclusion criteria for defining the cases and controls. Transformation of continuous data into binary outcomes should be detailed for reproducibility. Authors should explicitly state the number of controls included.
Methods: Variable Definitions	
Ancestry	Include the distribution of ancestral background for all samples used to generate or evaluate a PRS, 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 and ideally include detailed information beyond this when available. When combining samples from multiple studies, aggregate ancestral distribution information is sufficient. If genetic ancestry was considered, estimation methods should be detailed as well as discussing how this was utilized in analysis.

	Geographic location should not be used as a proxy to infer ancestry information.
Age	Include the age distribution of the total data set used to generate a single PRS (whether a single sample set, or the summary of combined samples) using the mean, standard deviation and range. Provide the age distribution by case/control status, if applicable.
Sex	Include the sex distribution of the total data set used to generate a single PRS (whether a single sample set, or the summary of combined samples) using the counts and percentages of total sample. State if sex was inferred from self-report or genetic information. Provide the sex distribution by case/control status, if applicable.
Risk Model Genetic Data Acquisition	Provide method for acquiring genetic information (sequencing vs. genotyping) in the PRS sample, including information about genome build and technical details of the assay. If imputing, specify the populations on the imputation panel, and provide the imputation quality for SNPs included in PRS. Report any imputation quality filters to exclude low quality imputation SNPs.
Risk Model Parameter Specifications	Explicitly state all terms used in the final risk model, including SNPs and any non-genetic variables. Authors should detail inclusion/exclusion criteria for all SNPs and other variables in the final model. Statistical procedures for selecting SNPs from a GWAS for inclusion in the final PRS model are provided in the "Statistical Model Fitting Procedure." Provide a reference for discovery GWAS and whether any adjustments were performed in the GWAS. If parameters were selected from another study, include reference (PMID, PGS catalog ID).
Clinical variable definition(s)	For any non-genetic variables included in the risk model, provide inclusion and exclusion criteria to define each variable, along with data source for that information (e.g., ICD codes, e-phenotyping algorithms, chart review, self-report). Indicate whether the variable is dichotomous or continuous.
Missing Data	Authors should explicitly state how missing data were handled for all variables included in the model, genetic and non-genetic.
Sub-Analyses	For any sub-analyses performed, provide inclusion and exclusion criteria used to stratify or subset the sample, any cut-offs used with justification provided, and the data source for this information (e.g., ICD codes, e-phenotyping algorithms, chart review, self-report). Explicitly state the number of cases included in the sub-analyses.
Methods: Analysis	
Polygenic Risk Score Estimation	Describe the statistical methods used to calculate risk score, primarily how the genetic data are transformed into a single score from individual risk variants. This includes information on how variants were weighted, and how the weights were derived.
Statistical Model Fitting Procedure	State the fitting procedure utilized to select the final version of the model. Details should include criteria for inclusion of SNPs (such as effect size or P-value threshold), if model was selected for optimal performance, and if so, the measures used to assess performance.

Risk Model Type	Detail statistical methods used to estimate risk, either relative or absolute, from the continuous risk score distribution. Authors should detail if risk is cumulative or cross-sectional, as well as the appropriate comparison groups if relative risk presented. Report time *until* predicted risk (eg. 5-year, 10-year, lifetime). In a relative hazard model, the study period or follow up time may be used. In an absolute risk model, state the time until predicted event. Authors should be careful not to simply report total length of study.
Risk Score Calibration	Describe measures used to assess calibration of the risk score and whether any variables were included beyond the risk score in this analysis.
Risk Score Discrimination	Describe measures used to assess discrimination of the risk score and whether any variables were included beyond the risk score in this analysis.
Statistical Validation Model	Outline procedures utilized to validate the risk score, including whether validation was performed, and the specific statistical model used during validation, especially if different than those used in the training set. Other details should include whether validation was internal or with an external validation set. Authors should include how missing data were handled.
Statistical Subgroup Analyses	For any sub-analyses performed, include details about statistical procedures specific to that subgroup that differ from the main analysis.
Results	
Risk Score Distribution	Include a general description of the distribution of the risk score, as well as model fit measures. This details the continuous distribution output directly from the risk model.
Risk Score Predictive Ability	State if risk model output is in terms of absolute or relative risk. If relative risk, include information about the reference population in respect to all relevant variables. It should be explicitly stated if the risk measure has been adjusted for other variables as previously defined in methods. If relative risk, include information about the reference population and risk score distribution in the general population. For absolute risk, include the prevalence/incidence of the predicted outcome in the general population. Enough detail should be included to enable readers to compute measures of risk score predictive ability such as AUC, sensitivity, specificity, PPV, NPV.
Risk Score Discrimination	Include metrics assessing discrimination of the risk score and whether other variables are included beyond the risk score in this analysis.
Risk Score Calibration	Include metrics assessing calibration of the risk score and whether other variables are included beyond the risk score in this analysis.
Risk Score Validation	Report all measures (predictive ability, distribution, discrimination, calibration) conducted in a validation set (either internal or external).

Subgroup Analyses	For any subgroup analyses, report all measures (predictive ability, distribution, discrimination, calibration).
Discussion	
Risk Model Interpretation	Summarize the risk score in terms of what it predicts, how well, and in whom. The predicted outcome predicted should be consistent with the introduction. Explicitly mention the performance of the risk model beyond non-genetic risk factors if non-genetic factors were included in the model. Target population should reflect the sample used in risk model development and validation (eg. ancestry, age, sex, other characteristics).
Limitations	Outline limitations in interpreting results. This includes, but is not limited to, study design restrictions, ascertainment biases, the distribution participant-level traits (ancestry, age, comorbidities), accuracy/specificity of phenotype data, and any statistical considerations. In addition, make note of any unknown reporting items from previous sections. Authors should consider and discuss the impact of these limitations on the interpretation of the risk score and any downstream replication needed.
Generalizability	Discuss which populations this score may be applied to and explicitly address any issues with generalizability beyond the included populations. Discuss whether the risk score has been externally validated, or if the sample is limited with respect to ancestry, age, or other variables.
Risk Score Intended Purpose	Discuss whether there is an intended clinical purpose or utility to the score. If so, discuss the “clinic readiness” and next steps with respect to the interpretation, limitations, and generalizability of the model. Discuss how the predictive ability of the model is benchmarked against current standard of care or other published work (such as existing PRS) on predicting the outcome of interest. If not, discuss why the PRS should not be used for clinical purposes.
Data Transparency and Availability	Information sufficient to calculate the PRS and/or risk model on external samples should be made 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 can be published 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 in the same way.

494