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], Jacqueline S. Dron[10,11,12], Deanna Brockman[10,11], Eric Venner[13], Mark I. McCarthy[14,15], Antonis C. Antoniou[16], Douglas F. Easton[16], Robert A. Hegele[12], Amit V. Khera[10,11], Nilanjan Chatterjee[17,18], Charles Kooperberg[19], Karen Edwards[20], Katherine Vlessis[21], Kim Kinnear[21], John N. Danesh[5,6], Helen Parkinson[6,7], Erin M. Ramos[8], Megan C. Roberts[22], Kelly E. Ormond[21,23], 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[$3-6,29,30], Genevieve Wojcik[$31]

[#]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 Molecular Biology Laboratory, European Bioinformatics Institute, 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. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
12. Western University, London, ON, Canada
13. Baylor College of Medicine, Houston, TX, USA
14. Department of Human Genetics, Genentech, South San Francisco, CA, USA
15. Wellcome Centre for Human Genetics, Oxford, UK
16. University of Cambridge, Cambridge, United Kingdom
17. Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
18. Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD, USA
19. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
20. Department of Epidemiology, University of California, Irvine, CA, USA
21. Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA
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, Center for Health Research, 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. National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge and Cambridge University Hospitals, Cambridge, UK
30. The Alan Turing Institute, London, UK
31. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

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

Abstract

Polygenic risk scores (PRS), often aggregating the results from genome-wide association studies, can bridge the gap between the initial variant discovery efforts and disease risk estimation for clinical applications. However, there is remarkable heterogeneity in the reporting of these risk scores due to a lack of adherence to reporting standards and no accepted standards suited for the current state of PRS development and application. This lack of adherence and best practices hinders the translation of PRS into clinical care. The ClinGen Complex Disease Working Group, in a collaboration with the Polygenic Score (PGS) Catalog, have developed a novel PRS Reporting Statement (PRS-RS), updating previous standards to the current state of the field and to enable downstream utility. Drawing upon experts in epidemiology, statistics, disease-specific applications, implementation, and policy, this 23-item reporting framework defines the minimal information needed to interpret and evaluate a PRS, especially with respect to any downstream clinical applications. Items span detailed descriptions of the study population (recruitment method, key demographics, inclusion/exclusion criteria, and phenotype definition), statistical methods for both PRS development and validation, and considerations for potential limitations of the published risk score and downstream clinical utility. Additionally, emphasis has been placed on data availability and transparency to facilitate reproducibility and benchmarking against other PRS, such as deposition in the publicly available PGS Catalog (www.PGScatalog.org). 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 of PRS into clinical care and progress towards defining best practices.

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 tailored 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 23-item Polygenic Risk Score Reporting Statement (PRS-RS). This framework provides the minimal information expected of authors to promote the validity, transparency, and reproducibility of PRS by requiring authors to detail the study population, statistical methods, and potential 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.

Main Text

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 the landscape of disease-associated genetic variants as a result of the many collaborative consortia and large cohorts of well-phenotyped individuals and matched genetic information.[1–5] In particular, genome-wide association studies (GWAS) have emerged as a powerful approach to identify disease- or trait-associated genetic variants and yielded summary statistics that describe the magnitude (effect size) and statistical significance of association between an allele and the phenotype of interest.[4,6] GWAS have been applied to a wide range of complex human traits and diseases, including height, blood pressure, body mass index/obesity, cardiovascular diseases, cancers, 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 used to form 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 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 risk estimates of disease outcomes; 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 to weigh them so as to maximize the proportion of variance explained or disease discrimination, there is an emerging consensus that the 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, and the utility of these predictions will vary between traits/diseases.[11,12]

In the last decade, the landscape of genetic prediction studies has transformed. There have been over 1,000 publications mentioning PGS/PRS with significant developments in how PGS are constructed and evaluated, as well as myriad new proposed use cases.[13] For instance, 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 isn't available, 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.[14–16] 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 [3,17], with an emphasis on ensuring that research is reproducible by all.

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 of PRS below.

Integrated Risk Model: a risk model combining PRS with other established risk factors for the trait of interest, 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 meaningful 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.[9,18]

Disease Prognosis – used to estimate the risk of further adverse outcome(s) *subsequent to* diagnosis of disease.[19]

Therapeutic – used to predict a patient or subgroup's response to a particular treatment.[20]

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). The readiness of PRS for implementation varies among phenotypes, 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 (including 23andMe, Color, MyHeritage, etc.) providing PGS/PRS results to customers outside of the traditional patient-provider framework. These concomitant advances have resulted in healthcare systems developing new infrastructures to deliver genetic risk information.[21] These advances, individually and combined, have raised significant challenges for PRS reporting standards; from the very basic (e.g. making the raw variant and weight information for a PRS available) to the complicated (e.g. proper study designs reporting performance metrics on an external validation dataset). The field now needs to develop modern standards for the reporting of PRS, conveying their increased scope and complexity for clinical applications.

Box 2: Current CHD PRS and their potential uses

Many PRS have been developed for CHD, varying in the computational methods used, 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 [22], and mainly differ by the computational methods used to select and weight individual variants (including LDpred [23,24], lassosum [25], and meta-scoring approaches [26]), and how they are combined into risk models with conventional risk factors. These PRS, however, may provide useful information for predicting risk of CHD that is largely orthogonal to conventional risk factors (age, sex, blood pressure, cholesterol, BMI, diabetes, smoking) as well as family history. Clinical applications may include:

- Improved prediction of risk for future adverse cardiovascular events when added to conventional risk models (including Framingham Risk Score [27], Pooled Cohort Equations [25,26], QRISK [25]).
- Reclassification of risk categories often leading to recommendations related to risk-reducing treatments like statins. [27–29]

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. [30]

132

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 in the 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. Most of the latest PRS have been developed using GWAS summary statistics and data from the Breast Cancer Association Consortium, using variants passing genome-wide significance (lead SNPs) or methods including stepwise or penalized regressions on individual-level genotypes [31], Pruning + Thresholding [23], and LDpred [24]. In contrast to CHD, genetics is commonly used to measure and understand breast cancer risk vis-a-vis *BRCA1/2* mutation testing; however, routine screening for breast cancer is often performed in older women using mammography and non-genetic risk prediction tools already exist. As such there has been active research into the value of capturing the polygenic risk of common variation on breast cancer, with multiple potential clinical uses and considerations:

- Multiple scores exist to predict risk for specific subtypes of breast cancer (e.g. ER positive/negative, luminal, and triple negative [31,32]), which could be used to stratify patients for more beneficial treatments, and may have prognostic implications.
- PRS can be combined with existing non-genetic risk prediction models (risk factors such as age, family history, mammographic density, hormone replacement therapy) and improve the prediction of incident breast cancer risk. [33–38]
- Breast cancer has been linked to several rare high-or moderate-risk penetrant genes (e.g. *BRCA1, BRCA2, PALB2, CHEK2, ATM)* and screening for pathogenic variants in these genes is part of clinical practice. It has been shown that PRS can provide important stratification of risk among carriers of such pathogenic variants and thus could be further useful for clinical decision making. [24,38–43]

Indeed, the BOADICEA breast cancer risk prediction model includes the effects of common variants (PRS313 [31]) as well as of other rare pathogenic genetic variants [38] and has been implemented in the CanRisk Tool (www.canrisk.org) that has been approved for use by healthcare professionals in the European Economic Area. The utility of PRS information for risk-based breast cancer screening is being tested worldwide, with trials in the US [44] and Europe (MyPeBS; https://mypebs.eu).

133

Some have noted that studies have been poorly designed or described in a way that challenges the validity of PRS to predict its target outcome [45,46]. At present, there are no uniformly agreed best practices for developing PRS, nor updated standards or regulations for reporting that are sufficiently tailored to assess the eventual clinical readiness of a PRS, such as benchmarking, and increased expectation of transparency and reproducibility. There are rapidly 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. [47–49] This in itself is not an issue, but without proper documentation standards to benchmark results, the rapid evolution in both the 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. 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. Those relevant for development and validation of risk prediction models include standards for reporting genetic ancestry information[1], assessing therapeutics (PICOT[50]), multivariate prediction models (TRIPOD [51]), observational epidemiological studies (STROBE[52]), genome-wide association studies (STREGA[53]), and, notably, the Genetic Risk Prediction Studies (GRIPS) Statement[54] which specifically address reporting of genetic risk prediction studies. Despite the strong foundational work of GRIPS, there is a need to update this framework. The field of genetic risk assessment and prediction has expanded and diversified, with new statistical models, the predominance of large-scale biobanks and consortia, and emerging direct uses in clinical care and disease prevention. Additionally, many current studies follow a prototypical process to PRS development evaluation (**Figure 1**) that can be used as a template for standardizing reporting and benchmarking in the field. There is also a need for standards to move beyond reporting guidelines to take into account data sharing principles as outlined in the FAIR Data Principles. [3,17]

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), a synthesis of 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**.

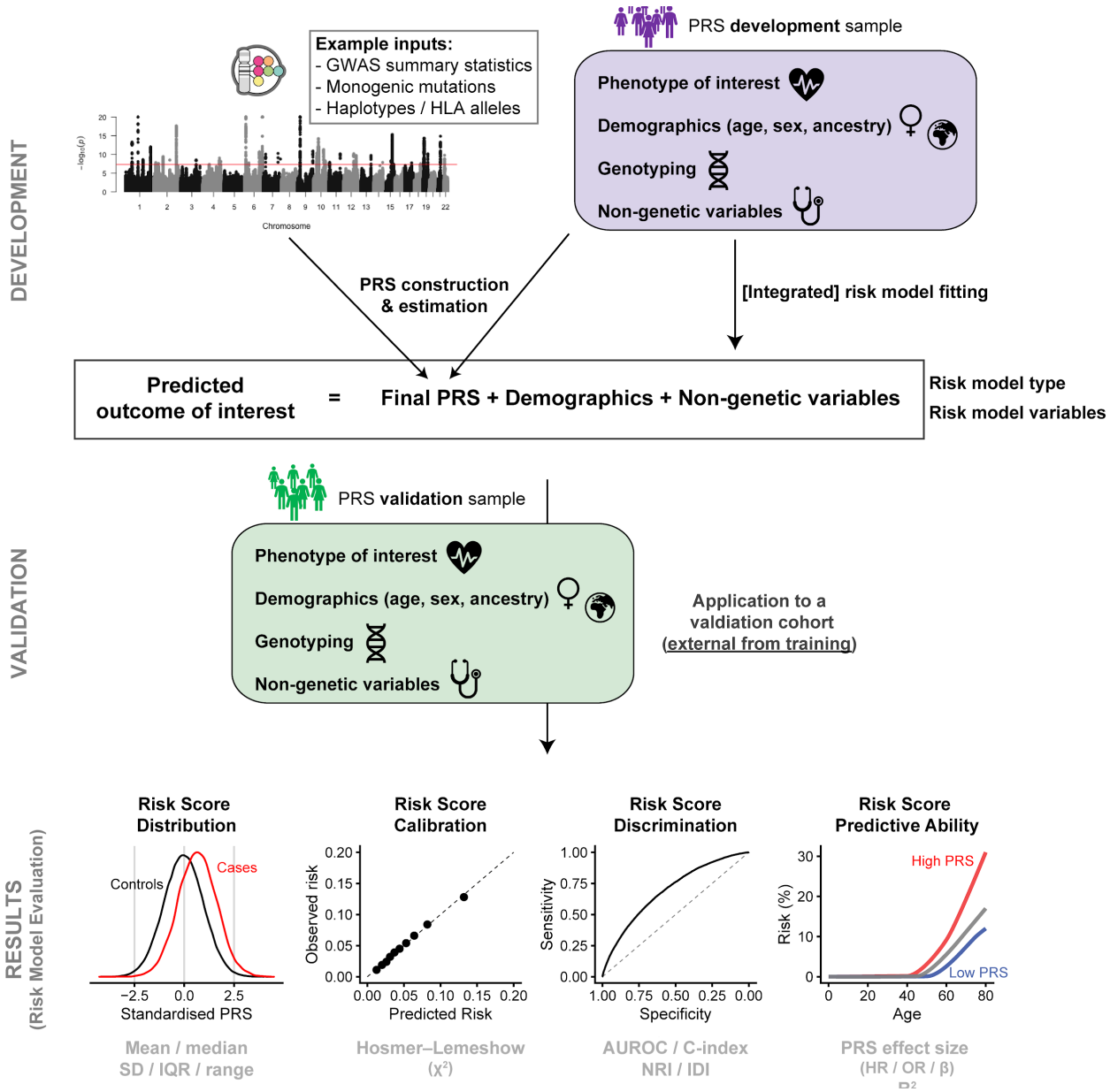

Figure 1: Prototype of PRS development and validation process. Figure 1 displays prototypical steps for PRS construction, risk model development, and validation of performance with select aspects of the PRS-RS guideline labeled in bold text throughout. In PRS development, variants associated with a phenotype 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 & estimation) are not shown. The PRS is tested in a risk model predicting the phenotype of interest and may be combined with other non-genetic variables (e.g. age, sex, ancestry, clinical variables); collectively these are referred to as risk model variables. 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**).

The ClinGen-PGS Catalog 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**. [55] It applies to both PRS development and validation studies that aim to predict disease onset 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, in which criteria are organized into key components along the developmental pipeline of PRS for clear interpretation.

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 *a priori* define the intent (e.g. risk prediction vs. prognosis) and purpose (e.g. to identify individuals at risk for CAD) for developing a PRS and/or risk model in enough detail to understand why the study population and risk model selected are of relevance to addressing this question (e.g. the value for CAD risk stratification and primary prevention is highest in younger individuals compared to those over 80 who have accumulated risk over a lifetime). As the PRS-RS is focused on clinical validity and implementation, authors must outline the study design and measurable predicted outcomes to understand what risk is measured, what the purpose of measuring risk would be, and why this purpose is of clinical relevance. In order to establish the internal validity of a study, authors should use the appropriate data needed to address the intended clinical use (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 their similarity to the original study populations used to derive the risk model. Therefore, authors need to define and characterize the details of the study populations (e.g. *study design* and *recruitment*), and describe their demographic composition in terms of *age*, *sex*, and *ancestry* (if possible, both genetic and self-reported 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. [1,9,56] It is therefore essential for authors to define participants’ genetic ancestry, ideally based on genome-wide data and taking into account any admixture that may exist, following the 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 a sufficient level of description all the relevant factors used to develop the risk model (*clinical variable definition(s)*) and define *phenotype 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 acquisition*).

Reporting on risk model development: There are currently several methods that are commonly used to select variants and fine-tune weights that constitute the PRS. [7,14–16,55] For methods using individual-level genomic data, the original source study should be cited or the assays and quality control described in detail. Methods using GWAS summary statistics should clearly cite the relevant GWAS(s) (preferably using unique and persistent study identifiers from the GWAS Catalog (GCSTs; [57]). 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 selected and combined into a single PRS (*PRS construction & estimation*). Apart from the genetic data in the model, authors should also describe the inclusion/exclusion criteria for other demographic and non-genetic predictors (*clinical variable definitions*) 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 model parameters: Authors should report estimates for all evaluated models, not only the methods behind decision-making, equipping readers with the information necessary to gauge the relative value of an increase in performance against other trade-offs. 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. 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.

Reporting on risk score evaluation: We recommend that authors provide summary information of the *risk score distribution* to aid in model interpretation. The *risk score's predictive ability*, *calibration*, and *discrimination* should also be assessed and described with common descriptions including the risk score effect size, reclassification indices, and metrics like sensitivity, specificity, positive predictive value (PPV),

251 and negative predictive value (NPV). The *risk score calibration* and *discrimination* should be described for
252 all analyses, although their estimation and interpretation are most relevant for the validation sample. It is
253 imperative for the PRS and expanded risk models to be evaluated on a population that is external (e.g.
254 independent, non-overlapping) to the individuals in the study population. The ability of the risk score to
255 classify individuals of interest (*risk score discrimination*) can commonly described and presented in terms
256 of the AUROC, AUPRC or C-index. Any differences in variable definitions or performance discrepancies
257 between the training and validation sets should be described.

258

259 **Reporting on interpretation**: By explicitly describing the *risk model's interpretation* and outlining potential
260 *limitations* to the *generalizability* of their model, authors will empower readers and the wider community to
261 better understand the risk score and its relative merits. Authors should justify the clinical relevance and *risk*
262 *score intended uses,* such as how the performance of their PRS compares to other commonly used risk
263 predictors, or previously published PRS. This may also include comparisons to other genetic predictors of
264 disease (e.g. mutations in high/moderate risk genes associated with Mendelian forms of the disease), family
265 history, simple demographic models, or conventional risk calculators (see Box 2 and 3 for disease-specific
266 examples). What indicates a "good" prediction can differ between outcomes and *intended purposes*, but
267 should be reported with similar metrics to what is described in the evaluation section.

268

269 **Supplemental Note 3** provides reporting considerations in addition to the minimal reporting
270 framework in **Table 1**. Authors intending downstream clinical implementation should aim for the level of
271 transparent and comprehensive reporting covered in both **Table 1** and **Supplemental Notes**, especially
272 those related to discussing the interpretation, limitations, and generalizability of results. The proper
273 reporting of PRS development and performance can also have implications for seeking regulatory approval
274 of the PRS as a clinical test. Though not a comprehensive list of regulatory requirements, we highlight
275 aspects of PRS-RS that would be considered evidence of analytical and clinical validity from the College of
276 American Pathologists (CAP) and the Clinical Laboratory Improvement Amendments (CLIA) perspective
277 (**Supplemental Table 1**). CAP and CLIA approvals are additional incentives for reporting adherence of
278 researchers wishing to translate their work, as well as a caution for researchers wishing to avoid unintended
279 uses of their findings. Lastly, we reiterate the need for both methodological and data transparency and
280 encourage deposition of PRS (variant-level information necessary to recalculate the genetic portion of the
281 score) in the PGS Catalog (www.PGSCatalog.org; [58]), which provides an invaluable resource for
282 widespread adoption and distribution of a published PRS. The PGS Catalog provides access to PGS and
283 related metadata to support the FAIR principles of data stewardship [17], enabling subsequent applications
284 and assessments of PGS performance and best practices (see **Supplemental Table 2** for a description of
285 the metadata captured in the Catalog and it's overlap with the PRS-RS).

Using the PRS-RS and PGS Catalog to improve PRS research and translation: transparency and open data

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 field definitions. For 10 of these publications, we provide detailed annotations using the final field definitions (**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 [59] 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 present an expanded 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**) for flexibility. 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 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* 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 reporting.

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. [60] In addition, the translation of structured data elements into useful clinical parameters may not be direct. A relevant example is that 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, such as coronary heart disease

321 symptoms commonly include angina (chest pain) whereas PRS are frequently trained on stricter definitions
322 excluding angina. Another example is the definitions used for race/ancestry as outlined in the PGS and
323 GWAS Catalogs[1] may also differ from structured terms used to document ancestry information in clinical
324 care, in which case consistent mappings, and potentially parallel analyses, may be necessary between
325 genetically determined ancestries and those routinely used in clinical care. Such translation issues
326 potentially limit generalizability to target populations and warrant further discussion, and we reiterate the
327 need for authors to be mindful of their intended purpose and target audience when discussing their findings.
328 Nevertheless, we have aided authors' understanding of potential translational barriers by considering
329 current CAP/CLIA analytical and clinical validity evidence requirements of peer-reviewed literature to
330 ensure PRS-RS has value in informing later steps of the clinical translation spectrum, including clinical utility
331 (**Supplemental Table 1**). Finally, while the principles of this work are clear, its scope does not include the
332 complex commercial restrictions, such as intellectual property, that may be placed on published studies
333 regarding the reporting or distribution of PGS, or the underlying data thereof. We hope this work will inform
334 downstream regulation and transparency standards for PRS as a commercial clinical tool.

335 The coordinated efforts of the ClinGen Complex Disease Working Group and PGS Catalog provide
336 a set of compatible resources for researchers to deposit PRS information. The PGS Catalog
337 (www.PGSCatalog.org) provides an informatics platform with data integration and harmonization to other
338 PGS as well as the source GWAS study through its sister platform, the GWAS Catalog. [58] In addition, it
339 provides a structured database of scores (variants and effect sizes) and metadata requested in PRS-RS.
340 With these tools, PRS-RS can be mandated by leading peer-reviewed journals and, consequently, the
341 quality and rigor of PRS research will be elevated to a level which facilitates clinical implementation. We
342 encourage readers to visit the ClinGen complex disease website (https://clinicalgenome.org/working-
343 groups/complex-disease/) for any future changes or amendments to the reporting guideline.

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

351

352

353

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). MI was supported by the Baker Institutute's Munz Chair of Cardiovascular Prediction and Prevention. SAL is supported by a Canadian Institutes of Health Research postdoctoral fellowship (MFE-171279). 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, JSD, EMR, MR, ACJWJ, KABG, PK, JALM, MI, and GLW wrote and edited the manuscript text and items.

Competing Interests

MIM has served on the advisory panels for Pfizer, Novo Nordisk, and Zoe Global; has received honoraria from Merck, Pfizer, Novo Nordisk, and Eli Lilly; and research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, Novo Nordisk, Pfizer, Roche, Sanofi Aventis, Servier, and Takeda. As of June 2019, MIM is an employee of Genentech and a holder of Roche stock. 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. Visscher, P. M. *et al.* 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101,** 5–22 (2017).
6. Pasaniuc, B. & Price, A. L. Dissecting the genetics of complex traits using summary association statistics. *Nat. Rev. Genet.* **18,** 117–127 (2017).
7. Lambert, S. A., Abraham, G. & Inouye, M. Towards clinical utility of polygenic risk scores. *Hum. Mol. Genet.* **28,** R133–R142 (2019).
8. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19,** 581–590 (2018).
9. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51,** 584–591 (2019).
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).
11. Wray, N. R., Yang, J., Goddard, M. E. & Visscher, P. M. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet.* **6,** e1000864 (2010).
12. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9,** e1003348 (2013).
13. PubMed Search Results for “Polygenic Score.” at <https://pubmed.ncbi.nlm.nih.gov/?term=%28%22polygenic+score%22%5BTitle%2FAbstract%5D%29+OR+%28%22polygenic+risk+score%22%5BTitle%2FAbstract%5D%29&sort=pubdate>
14. 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).
15. Choi, S. W. & O’Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8,** (2019).
16. Vilhjálmsson, B. J. *et al.* Modeling linkage disequilibrium increases accuracy of polygenic risk scores.

Am. J. Hum. Genet. **97,** 576–592 (2015).

17. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3,** 160018 (2016).
18. 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).
19. Choi, J. *et al.* The associations between immunity-related genes and breast cancer prognosis in Korean women. *PLoS One* **9,** e103593 (2014).
20. 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).
21. 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>
22. 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).
23. 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).
24. 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).
25. 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).
26. 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).
27. Abraham, G. *et al.* Genomic prediction of coronary heart disease. *Eur. Heart J.* **37,** 3267–3278 (2016).
28. Ganna, A. *et al.* Multilocus genetic risk scores for coronary heart disease prediction. *Arterioscler. Thromb. Vasc. Biol.* **33,** 2267–2272 (2013).
29. 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).
30. U.S. National Library of Medicine. Home - ClinicalTrials.gov. at <https://clinicaltrials.gov/>
31. Mavaddat, N. *et al.* Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am. J. Hum. Genet.* **104,** 21–34 (2019).
32. 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).
33. 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).
34. Lakeman, I. M. M. *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).
35. Wilcox, A. N. *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

36. 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

37. 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).

38. Lee, A. *et al.* BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet. Med.* **21,** 1708–1718 (2019).

39. Muranen, T. A. *et al.* Genetic modifiers of CHEK2*1100delC-associated breast cancer risk. *Genet. Med.* **19,** 599–603 (2017).

40. 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).

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

42. Fahed, A. C. *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

43. Barnes, D. R. *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

44. Esserman, L. J. & WISDOM Study and Athena Investigators. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer* **3,** 34 (2017).

45. Martens, F. K., Tonk, E. C. M. & Janssens, A. C. J. W. Evaluation of polygenic risk models using multiple performance measures: a critical assessment of discordant results. *Genet. Med.* **21,** 391–397 (2019).

46. Janssens, A. C. J. W. Validity of polygenic risk scores: are we measuring what we think we are? *Hum. Mol. Genet.* **28,** R143–R150 (2019).

47. 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

48. Matoba, N., Love, M. I. & Stein, J. L. Evaluating brain structure traits as endophenotypes using polygenicity and discoverability. *BioRxiv* (2020). doi:10.1101/2020.07.17.208843

49. Coleman, J. R. I. *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).

50. 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).

51. 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).

52. 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).

53. Little, J. *et al.* STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. *PLoS Med.* **6,** e22 (2009).

54. Janssens, A. C. J. W. *et al.* Strengthening the reporting of Genetic RIsk Prediction Studies: the

511 GRIPS Statement. *PLoS Med.* **8,** e1000420 (2011).

512 55. Choi, S. W., Mak, T. S.-H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score
513 analyses. *Nat. Protoc.* **15,** 2759–2772 (2020).

514 56. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits.
515 *Nature* **570,** 514–518 (2019).

516 57. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies,
517 targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47,** D1005–D1012 (2019).

518 58. Lambert, S. A. *et al.* The Polygenic Score Catalog: an open database for reproducibility and
519 systematic evaluation. *medRxiv* (2020). doi:10.1101/2020.05.20.20108217

520 59. Thelwall, M. *et al.* Is useful research data usually shared? An investigation of genome-wide
521 association study summary statistics. *PLoS One* **15,** e0229578 (2020).

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

524

Table 1. Polygenic Risk Score Reporting Statement (PRS-RS)

Reporting Standard		Description
Risk Score Background	**Type of study**	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 Purpose & Predicted Clinical End Outcome	Specify **what the risk score is intended to predict** and the **purpose** for developing the score: • Will it be used in risk prediction, diagnostic, prognostic, or therapeutic modalities (or combination of these)? • What is 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.
Study Populations *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 populations used in the current study describe the type of sample (e.g., cohort, case-control, cross-sectional), recruitment details (e.g. method and years), 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 the full reference to the original study.
	Ancestry	Include the ancestral background distribution of the sample sets used during PRS development and validation, 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 includes more detailed information when available. When combining samples from multiple studies, aggregate ancestral distribution information is sufficient. The method of ancestry inference should be provided - genomic methods to determine ancestry are preferred, (e.g. principal component analysis).
	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 and evaluate the PRS and/or risk model (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.
	Genetic Data Acquisition	Provide method for acquiring genetic information (e.g. sequencing, genotyping) in each 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. If parameters were selected from another study, include reference (PMID, GWAS Catalog ID).
	Clinical Variable Definition(s)	Define any non-genetic variables that were used to train/fit 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 the scale of each variable, e.g. dichotomous, continuous, categorical, or ordinal. Explicitly state which variables are included in the final model in the **integrated risk model** fitting sections.
	Defining the 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). Transformation of continuous data into binary, ordinal, or categorical outcomes should be detailed for reproducibility. For dichotomous data, also include inclusion and exclusion criteria for defining the cases and controls. Authors should explicitly state the number of cases and controls included.
	Missing Data	Authors should explicitly state how missing data were handled for all variables included in the model—genetic and non-genetic.
Risk Score Development & Application *Describe the relevant methods used to form the*	**Polygenic Risk Score Construction & Estimation**	Describe how genetic data were included in the risk model. Authors should detail inclusion/exclusion criteria for all variants. If individual risk variants are combined into a PRS that is subsequently included in the model as a single variable, 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).

final PRS and/or risk model. Samples used in this stage of the analysis should be referred to as "Score Development" or "Training" samples, and be described according to the items in the Study Populations *section.*	**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.
	Integrated Risk Model Description and Fitting	State the fitting procedure utilized to select the final version of the model (including non-genetic and/or genetic [PRS or variants] variables). If the model was selected for optimal performance, describe measures used to assess performance. Explicitly state all terms used in the final risk model, including PRS/variants and any non-genetic variables.
Risk Score Evaluation *Outline the results and procedures utilized to validate the risk score, specifying whether the validation was performed on internal or external validation samples. Performance results should be described for both the development and validation samples. Specify if the application of the risk score differs between the development and validation samples.*	**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 sensitivity, specificity, PPV, NPV.
	Risk Score Discrimination	Describe and report metrics used to assess the discrimination of the risk score and whether any variables were included beyond the risk score in this analysis. Common metrics include the area under the ROC or Precision-Recall (AUROC/AUPRC) and the Concordance statistic (C-index) for survival models.
	Risk Score Calibration	Describe and report metrics used to assess the calibration of the risk score and whether any variables were included beyond the risk score in this analysis.
	Subgroup Analyses	For any subgroup analyses, report all measures (distribution, predictive ability, discrimination, calibration).
Translation *Discuss the broader context of the study and risk score.*	**Risk Model Interpretation**	Summarize the risk score in terms of what it predicts, how well, and in whom. The predicted outcome should be consistent with the introduction. Explicitly mention the performance of the PRS and/or combined risk model in comparison to conventional risk models. Conventional risk models might include demographic (age, sex), disease-specific risk factors, and/or family history of disease.
	Limitations	Outline limitations in interpreting results, discuss the impact of these limitations on the interpretation of the risk score and any downstream replication needed. Common considerations include: study design restrictions, ascertainment biases, the distribution participant-level traits (ancestry, age, comorbidities), accuracy/specificity of phenotype data, and any statistical considerations. Make note of and discuss the impact of any unknown reporting items from previous sections.
	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 Use Cases	Discuss whether there is an intended clinical use 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 and/or 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.

Further considerations can be found in Supplemental Note 3