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], Deanna Brockman[10], Eric Venner[12], Mark I. McCarthy[13,14], Antonis C. Antoniou[15], Douglas F. Easton[15], Robert A. Hegele[11], Amit V. Khera[10], Nilanjan Chatterjee[16,17], Charles Kooperberg[18], Karen Edwards[19], Katherine Vlessis[20], Kim Kinnear[20], John N. Danesh[5,6,21], Helen Parkinson[6,7], Erin M. Ramos[8], Megan C. Roberts[22], Kelly E. Ormond[20,23], Muin J. Khoury[24], A. Cecile J.W. Janssens[25], Katrina A.B. Goddard[26,27], Peter Kraft[28], Jaqueline A. L. MacArthur[7], Michael Inouye[$3-6,21,29], 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 Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, UK
8. National Human Genome Research Institute, Bethesda, MD, USA
9. Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, USA
10. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
11. Western University, London, ON, Canada
12. Baylor College of Medicine, Houston, TX, USA
13. Department of Human Genetics, Genentech, South San Francisco, CA, USA
14. Wellcome Centre for Human Genetics, Oxford, UK
15. University of Cambridge, Cambridge, United Kingdom
16. Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
17. Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD, USA
18. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
19. Department of Epidemiology, University of California, Irvine, CA, USA
20. Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA
21. National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge and Cambridge University Hospitals, Cambridge, UK
22. Division of Pharmaceutical Outcomes and Policy, UNC Eshelman School of Pharmacy, Chapel Hill, NC, USA
23. Stanford Center for Biomedical Ethics, Stanford University School of Medicine, Stanford, CA, USA
24. Centers for Disease Control and Prevention, Atlanta, GA, USA
25. Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA
26. Department of Translational and Applied Genomics, Kaiser Permanente Northwest, Portland, OR, USA
27. Center for Health Research, Kaiser Permanente Northwest, Portland, OR, USA
28. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
29. The Alan Turing Institute, London, UK
30. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

*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 discovery efforts and clinical applications for disease risk estimation. However, there is remarkable heterogeneity in the reporting of these risk scores. This lack of adherence to reporting standards hinders the translation of PRS into clinical care. The ClinGen Complex Disease Working Group, in a collaboration with the Polygenic Score (PGS) Catalog, have updated the Genetic Risk Prediction (GRIPS) Reporting Statement to the current state of the field and to enable downstream utility. Drawing upon experts in epidemiology, statistics, disease-specific applications, implementation, and policy, this 22-item reporting framework defines the minimal information needed to interpret and evaluate a PRS, especially with respect to any downstream clinical applications. Items span detailed descriptions of the study population (recruitment method, key demographic and clinical characteristics, inclusion/exclusion criteria, and outcome 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. 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 increasingly been used 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 existing reporting standards has hindered the translation of this important tool into clinical and public health practice; in particular, details necessary for benchmarking and reproducibility are underreported. To address this gap, the ClinGen Complex Disease Working Group and Polygenic Score (PGS) Catalog have updated the Genetic Risk Prediction (GRIPS) Reporting Statement into the 22-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 encouraging 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 outcome of interest.[4,6] 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 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 outcome (*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.[11–15]

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

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 outcome 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.[9,20]

Disease Prognosis – used to estimate the risk of further adverse outcome(s) *subsequent to* diagnosis of disease.[21]

Therapeutic – used to predict a patient or subgroup's response to a particular treatment.[22]

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 outcomes, with only a few diseases, like coronary heart disease (CHD) and breast cancer, having mature PRS with potential clinical utility (**Boxes 2** and **3,** respectively). There has also been a rapid rise of direct-to-consumer assays and for-profit companies (23andMe, Color, MyHeritage, etc.) providing PGS/PRS results to customers outside of the traditional patient-provider framework. These concomitant advances have resulted in healthcare systems developing new infrastructures to deliver genetic risk information.[23] These advances, individually and combined, have raised significant challenges for PRS reporting standards; from the very basic (e.g. reporting performance metrics on an external validation dataset) to the complicated (e.g. making the raw variant and weight information for a PRS available), necessitating the updating of existing standard for reporting genetic risk prediction studies to convey the increased scope of PRS and complexity for their clinical applications.

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 [24], and mainly differ by the computational methods used to select and weight individual variants (including LDpred [25,26], lassosum [27], and meta-scoring approaches [28], and how they are combined into risk models with conventional risk factors. These PRS may provide useful information for predicting risk of CHD that is largely orthogonal to conventional risk factors (age, sex, hypertension, cholesterol, BMI, diabetes, smoking) as well as family history. Clinical applications may include:

- Improved prediction of risk for future adverse cardiovascular events when added to conventional risk models (such as the Framingham Risk Score [29], Pooled Cohort Equations [27,28], QRISK [27]).
- Reclassification of risk categories often leading to recommendations related to risk-reducing treatments like statins. [29–31]

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. [32]

127

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 (BCAC), using variants passing genome-wide significance (lead SNPs) or methods including stepwise or penalized regressions on individual-level genotypes [33], Pruning + Thresholding [25], and LDpred [26]. 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 PRS exist to predict risk for specific subtypes of breast cancer (e.g. ER positive/negative, luminal, and triple negative[33,34]), 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 (combining risk factors such as age, family history, mammographic density, hormone replacement therapy) and improve the prediction of incident breast cancer risk. [35–42]
- 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.[43] 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. [26,40,44–48]

Indeed, the BOADICEA breast cancer risk prediction model includes the effects of common variants (PRS313 [33]) as well as of other rare pathogenic genetic variants [40] 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. PRS information has been studied via simulation[49] and is being tested in risk-based breast cancer screening trials in the US [50] and Europe (MyPeBS; https://mypebs.eu).

128

Poorly designed or described studies call into question the validity of some PRS to predict their target outcome [51,52], and relatively few studies have benchmarked multiple scores performance externally. At present, there are no uniformly agreed best practices for developing PRS, nor widely adopted standards or regulations sufficiently tailored to assess the eventual clinical readiness of a PRS. There are 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. [53–56] This in itself is not an issue, but 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 when studies are not properly documented. These deficiencies are barriers to PRS being interpreted, compared, and reproduced, and must be addressed to enable the application of PRS to improve clinical practice and public health.

Frameworks have been developed to establish standards around the transparent, standardized, accurate, complete, and meaningful reporting of scientific studies. In 2011, an international working group published the Genetic Risk Prediction Studies (GRIPS) Statement, a reporting guideline for the study of risk prediction models that include genetic variants, from genetic mutations to gene scores.[57] This guideline is analogous to guidelines developed for observational epidemiological studies (STROBE[58]) and genome-wide association studies (STREGA[59]), and is in line with the reporting guideline for multivariate prediction models (TRIPOD [60]). Adherence to reporting statements has been low, and the same holds for GRIPS. One of the reasons might be that researchers feel that GRIPS inadequately addresses PRS, where there has been rapid evolution in score development methodology in recent years. Researchers are frequently uncertain as to what precisely should be reported for a PRS study to be assessed as rigorous, reproducible, and ultimately translatable, especially with the increased push for data availability and transparency. Most PRS studies follow a prototypical process to PRS development evaluation (**Figure 1**) that can be used as a template for standardizing reporting and benchmarking in the field.

Here, the Clinical Genome Resource (ClinGen) Complex Disease Working Group and the Polygenic Score (PGS) Catalog (**Supplemental Note 1**) jointly present the Polygenic Risk Score Reporting Standards (PRS-RS), 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**.

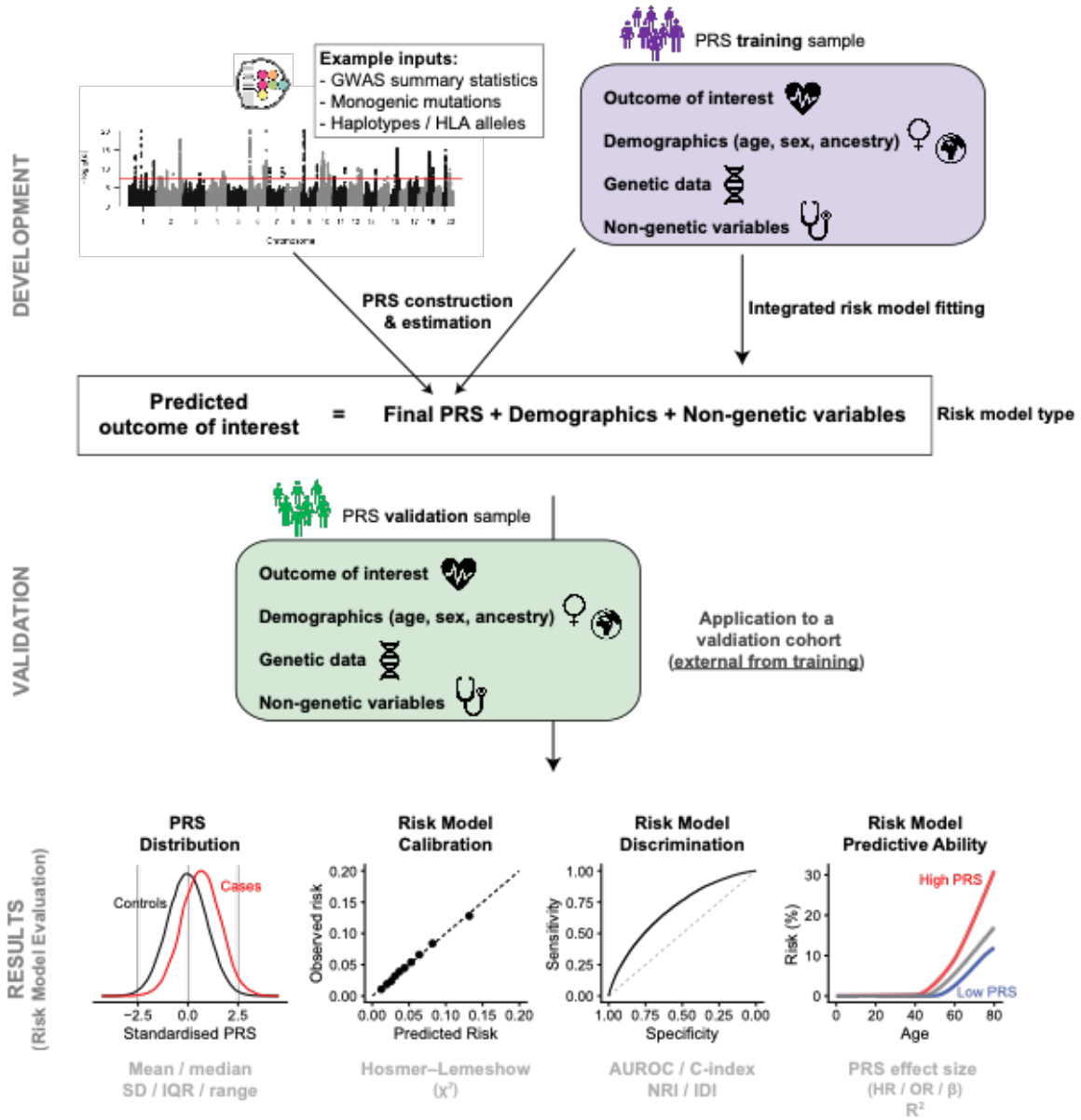

Figure 1: Prototype of PRS development and validation process. Figure 1 displays prototypical steps for PRS construction, risk model development, and validation of performance with select aspects of the PRS-RS guideline labeled in bold text throughout. In PRS development, variants associated with an outcome 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 outcome 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 PRS distribution should be described, and the performance of the risk model demonstrated in terms of its discrimination, predictive ability, and calibration. Though not displayed in the figure, these same results should also be reported for the training sample for comparison to the validation sample. In both training and validation cohorts, the outcome of interest criteria, demographics, genotyping, and non-genetic variables should be reported (**Table 1**).

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**. [61] 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 and to encourage their documentation from the inception of the study well before publication.

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

Reporting on study populations: The applicability of any risk prediction to an external target population (the "who, where, and when") depends on its similarity to the original study populations used to derive the risk model. Therefore, authors need to define and characterize the details of their study population (e.g. *study design and recruitment*), and describe *participant demographics* for key variables (such as age and sex) and *ancestry*. Importantly, there are often inconsistent definitions and levels of detail associated with ancestry, and the transferability of genetic findings between different racial/ethnic groups can be limited. [1,9,62] It is therefore essential for authors to provide a detailed description of participants' genetic ancestry - including how ancestry was determined – using a common controlled vocabulary where possible (e.g. tables 1 and 2 of the standardized framework developed by the NHGRI-EBI GWAS Catalog[1]). While age, sex and ancestry are the most universally relevant characteristics, authors should provide a sufficient level of detailed criteria for defining all the relevant factors used in the risk model (*non-genetic variable(s)*) and the *outcome of interest*. This is particularly important if they are included in the final risk model and should accompany information about how the population was genotyped (*genetic data*).

Reporting on risk model development: There are currently several methods that are commonly used to select

210 variants and fine-tune weights that constitute the PRS. [7,16–18,61] For methods using individual-level genomic
211 data, the original source study should be cited or the assays and quality control described in detail. Methods
212 using GWAS summary statistics should clearly cite the relevant GWAS(s) (preferably using unique and
213 persistent study identifiers from the GWAS Catalog (GCSTs; [63]). As the performance and limitations of the
214 combined risk model are dependent on methodological considerations, authors must provide complete details
215 including the method used and how variants are selected and combined into a single PRS (*PRS construction*
216 *& estimation*). Apart from the genetic data in the model, authors should also describe the defining criteria for
217 other demographic and non-genetic predictors (*non-genetic variables*) included in the model. Often authors will
218 iterate through numerous models to find the optimal fit. In addition to the estimation methods, it is important to
219 detail the *integrated risk model fitting* procedure, including the measures used for final model selection.
220 Translating the continuous PRS distribution to a risk estimate, whether absolute or relative, is highly dependent
221 on assumptions and limitations inherent to the specific data set utilized. When describing the *risk model type*,
222 authors should detail the time scale employed for prediction or the study period/follow-up time for a relative
223 hazard model. Additionally, if relative risk is estimated, the reference group should be well described. These
224 details should be described for the training set, as well as validation and sub-group analyses.
225
226 **Reporting on model parameters:** Authors should report estimates for all evaluated models, not only the
227 methods behind decision-making, equipping readers with the information necessary to gauge the relative value
228 of an increase in performance against other trade-offs (*data transparency and availability*). The underlying PRS
229 (variant alleles and derived weights) should be made publicly available, preferably through direct submission to
230 an indexed repository such as the PGS Catalog, to enable others to reuse existing models (with known validity)
231 and facilitate direct benchmarking between different PRS for the same trait. The current mathematical form of
232 most PRS—a linear combination of allele counts—facilitates clear model description and reproducibility. Future
233 genomic risk models may have more complex forms, e.g. allowing for explicit non-linear epistatic and gene-
234 environment interactions, or deep neural networks of lesser clarity. It will be important to describe these models
235 in sufficient detail to allow their implementation and evaluation by other researchers and clinical groups.
236
237 **Reporting on risk model evaluation:** We recommend that authors provide summary information of the *risk*
238 *score distribution* to aid in model interpretation. The *risk model's predictive ability*, *calibration*, and *discrimination*
239 should also be assessed and described with common descriptions including the risk score effect size, variance
240 explained (R^2), reclassification indices, and metrics like sensitivity, specificity, positive predictive value (PPV),
241 and negative predictive value (NPV). The *risk model calibration* and *discrimination* should be described for all
242 analyses, although their estimation and interpretation are most relevant for the PRS validation sample. It is
243 imperative for the PRS and expanded risk models to be evaluated on a population that is external (e.g.
244 independent, non-overlapping) to the individuals in the study population. The ability of the risk model to classify
245 individuals of interest (*risk model discrimination*) can commonly described and presented in terms of the
246 AUROC, AUPRC or C-index. Any differences in variable definitions or performance discrepancies between the

training and validation sets should be described.

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

Supplemental Note 3 provides reporting considerations in addition to the minimal reporting framework in **Table 1**. Authors intending downstream clinical implementation should aim for the level of transparent and comprehensive reporting covered in both **Table 1** and **Supplemental Notes**, especially those related to discussing the interpretation, limitations, and generalizability of results. The proper reporting of PRS development and performance can also have implications for seeking regulatory approval of the PRS as a clinical test. Though not a comprehensive list of regulatory requirements, we highlight aspects of PRS-RS that would be considered evidence of analytical and clinical validity from the College of American Pathologists (CAP) and the Clinical Laboratory Improvement Amendments (CLIA) perspective (**Supplemental Table 1**). CAP and CLIA approvals are additional incentives for reporting adherence of researchers wishing to translate their work, as well as a caution for researchers wishing to avoid unintended uses of their findings. Lastly, we reiterate the need for both methodological and data transparency and encourage deposition of PRS (variant-level information necessary to recalculate the genetic portion of the score) in the PGS Catalog (www.PGSCatalog.org; [64]), which provides an invaluable resource for widespread adoption and distribution of a published PRS. The PGS Catalog provides access to PGS and related metadata to support the FAIR principles of data stewardship [19], enabling subsequent applications and assessments of PGS performance and best practices (see **Supplemental Table 2** for a description of the metadata captured in the Catalog and it's overlap with the PRS-RS).

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

283 detail necessary for each PRS-RS item (further described in **Supplemental Note 3**). The heterogeneity in the
284 PRS reporting we observed in this pilot highlights a series of challenges**.** Critical aspects of PRS studies,
285 including ancestry, predictive ability, and transparency/availability of information needed to reproduce PRS,
286 were frequently absent or reported in insufficient detail. This underscores the need for PRS-RS to clearly and
287 specifically define meaningful aspects of PRS development, testing, and intended clinical use. However, these
288 deficits in reporting are not unique to PRS; previous reports of underreporting have found that 77% of GWAS
289 publications in 2017 did not share summary statistics [65] and 4% of GWAS do not report any relevant ancestry
290 information [1]. In line with the push towards a culture of reproducibility and open data in genomics, we as the
291 ClinGen Complex Disease Working Group and PGS Catalog joined to create a set of reporting standards (**Table**
292 **1**) specifically tailored to PRS research based on multidisciplinary and international expert opinion for tailoring
293 previous standards.

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

308 While the scope of our work encompasses clinical validity, it does not address the additional
309 requirements needed to establish the clinical or public health utility of a PRS, such as randomized trials with
310 clinically meaningful outcomes, health economic evaluations, or feasibility studies. [66] In addition, the translation
311 of structured data elements into useful clinical parameters may not be direct. A relevant example is that the
312 disease case definitions utilized in training or validation in any particular PRS study may deviate (sometimes
313 substantially) from those utilized in any specific health system, for example CHD symptoms commonly include
314 angina (chest pain), whereas PRS are frequently trained on stricter definitions excluding angina. In addition,
315 the definitions used for race/ancestry as outlined in the PGS and GWAS Catalog[1] may also differ from structured
316 terms used to document ancestry information in clinical care, in which case consistent mappings, and potentially
317 parallel analyses, may be necessary to translate from genetically determined ancestries to those routinely used
318 in clinical care. Such translation issues potentially limit generalizability to target populations and warrant further
319 discussion, and we reiterate the need for authors to be mindful of their intended purpose and target audience

320 when discussing their findings. Nevertheless, we have aided authors' understanding of potential translational
321 barriers by considering current CAP/CLIA analytical and clinical validity evidence requirements of peer-
322 reviewed literature to ensure PRS-RS has value in informing later steps of the clinical translation spectrum,
323 including clinical utility (**Supplemental Table 1**). Finally, while the principles of this work are clear, its scope
324 does not include the complex commercial restrictions, such as intellectual property, that may be placed on
325 published studies regarding the reporting or distribution of PGS, or the underlying data thereof. We hope this
326 work will inform downstream regulation and transparency standards for PRS as a commercial clinical tool.

327 The coordinated efforts of the ClinGen Complex Disease Working Group and PGS Catalog provide a
328 set of compatible resources for researchers to deposit PGS/PRS information. The PGS Catalog
329 (www.PGSCatalog.org) provides an informatics platform with data integration and harmonization to other PGS
330 as well as the source GWAS study through its sister platform, the GWAS Catalog. [64] In addition, it provides a
331 structured database of scores (variants and effect weights) that can be reused, along with metadata requested
332 in the PRS-RS. With these tools, PRS-RS can be mandated by leading peer-reviewed journals and,
333 consequently, the quality and rigor of PRS research will be elevated to a level which facilitates clinical
334 implementation. We encourage readers to visit the ClinGen complex disease website
335 (https://clinicalgenome.org/working-groups/complex-disease/) for any future changes or amendments to the
336 reporting guideline.

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

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, SAL, and JD 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). SAL is supported by a Canadian Institutes of Health Research postdoctoral fellowship (MFE-171279). JD holds a British Heart Foundation Personal Chair and a National Institute for Health Research Senior Investigator Award. 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. 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. Pharoah, P. D. P. *et al.* Polygenic susceptibility to breast cancer and implications for prevention. *Nat. Genet.* **31,** 33–36 (2002).
14. Zhang, Y., Qi, G., Park, J.-H. & Chatterjee, N. Estimation of complex effect-size distributions using

410 summary-level statistics from genome-wide association studies across 32 complex traits. *Nat. Genet.* **50,**
411 1318–1326 (2018).
412 15. Chatterjee, N. *et al.* Projecting the performance of risk prediction based on polygenic analyses of genome-
413 wide association studies. *Nat. Genet.* **45,** 400–5, 405e1 (2013).
414 16. Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X. & Sham, P. C. Polygenic scores via penalized
415 regression on summary statistics. *Genet. Epidemiol.* **41,** 469–480 (2017).
416 17. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience*
417 **8,** (2019).
418 18. Vilhjálmsson, B. J. *et al.* Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am.*
419 *J. Hum. Genet.* **97,** 576–592 (2015).
420 19. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci.*
421 *Data* **3,** 160018 (2016).
422 20. Huo, D. *et al.* Comparison of breast cancer molecular features and survival by african and european
423 ancestry in the cancer genome atlas. *JAMA Oncol.* **3,** 1654–1662 (2017).
424 21. Choi, J. *et al.* The associations between immunity-related genes and breast cancer prognosis in Korean
425 women. *PLoS One* **9,** e103593 (2014).
426 22. Marston, N. A. *et al.* Predicting benefit from evolocumab therapy in patients with atherosclerotic disease
427 using a genetic risk score: results from the FOURIER trial. *Circulation* **141,** 616–623 (2020).
428 23. Ganguly, P. & National Human Genome Research Institute. NIH funds centers to improve the role of
429 genomics in assessing and managing disease risk. (2020). at <https://www.genome.gov/news/news-
430 release/NIH-funds-centers-to-improve-role-of-genomics-in-assessing-and-managing-disease-risk>
431 24. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of
432 coronary artery disease. *Nat. Genet.* **47,** 1121–1130 (2015).
433 25. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk
434 equivalent to monogenic mutations. *Nat. Genet.* **50,** 1219–1224 (2018).
435 26. Mars, N. *et al.* Polygenic and clinical risk scores and their impact on age at onset and prediction of
436 cardiometabolic diseases and common cancers. *Nat. Med.* **26,** 549–557 (2020).
437 27. Elliott, J. *et al.* Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical
438 Risk Score for Coronary Artery Disease. *JAMA* **323,** 636–645 (2020).
439 28. Inouye, M. *et al.* Genomic risk prediction of coronary artery disease in 480,000 adults: implications for
440 primary prevention. *J. Am. Coll. Cardiol.* **72,** 1883–1893 (2018).
441 29. Abraham, G. *et al.* Genomic prediction of coronary heart disease. *Eur. Heart J.* **37,** 3267–3278 (2016).
442 30. Ganna, A. *et al.* Multilocus genetic risk scores for coronary heart disease prediction. *Arterioscler. Thromb.*
443 *Vasc. Biol.* **33,** 2267–2272 (2013).
444 31. Kullo, I. J. *et al.* Incorporating a Genetic Risk Score Into Coronary Heart Disease Risk Estimates: Effect
445 on Low-Density Lipoprotein Cholesterol Levels (the MI-GENES Clinical Trial). *Circulation* **133,** 1181–1188
446 (2016).

447 32. U.S. National Library of Medicine. Home - ClinicalTrials.gov. at <https://clinicaltrials.gov/>
448 33. Mavaddat, N. *et al.* Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am.*
449 *J. Hum. Genet.* **104,** 21–34 (2019).
450 34. Zhang, H. *et al.* Genome-wide association study identifies 32 novel breast cancer susceptibility loci from
451 overall and subtype-specific analyses. *Nat. Genet.* **52,** 572–581 (2020).
452 35. Zhang, X. *et al.* Addition of a polygenic risk score, mammographic density, and endogenous hormones to
453 existing breast cancer risk prediction models: A nested case-control study. *PLoS Med.* **15,** e1002644
454 (2018).
455 36. Lakeman, I. M. M. *et al.* Addition of a 161-SNP polygenic risk score to family history-based risk prediction:
456 impact on clinical management in non-BRCA1/2 breast cancer families. *J. Med. Genet.* **56,** 581–589
457 (2019).
458 37. Wilcox, A. N. *et al.* Prospective Evaluation of a Breast Cancer Risk Model Integrating Classical Risk
459 Factors and Polygenic Risk in 15 Cohorts from Six Countries. *medRxiv* (2019). doi:10.1101/19011171
460 38. Pal Choudhury, P. *et al.* Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk
461 models incorporating classical risk factors and polygenic risk in a population-based prospective cohort.
462 *medRxiv* (2020). doi:10.1101/2020.04.27.20081265
463 39. Maas, P. *et al.* Breast cancer risk from modifiable and nonmodifiable risk factors among white women in
464 the united states. *JAMA Oncol.* **2,** 1295–1302 (2016).
465 40. Lee, A. *et al.* BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and
466 nongenetic risk factors. *Genet. Med.* **21,** 1708–1718 (2019).
467 41. Garcia-Closas, M., Gunsoy, N. B. & Chatterjee, N. Combined associations of genetic and environmental
468 risk factors: implications for prevention of breast cancer. *J. Natl. Cancer Inst.* **106,** (2014).
469 42. Kapoor, P. M. *et al.* Combined associations of a polygenic risk score and classical risk factors with breast
470 cancer risk. *J. Natl. Cancer Inst.* (2020). doi:10.1093/jnci/djaa056
471 43. Easton, D. F. *et al.* Gene-panel sequencing and the prediction of breast-cancer risk. *N. Engl. J. Med.* **372,**
472 2243–2257 (2015).
473 44. Muranen, T. A. *et al.* Genetic modifiers of CHEK2*1100delC-associated breast cancer risk. *Genet. Med.*
474 **19,** 599–603 (2017).
475 45. Kuchenbaecker, K. B. *et al.* Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction
476 in BRCA1 and BRCA2 mutation carriers. *J. Natl. Cancer Inst.* **109,** (2017).
477 46. Mars, N. *et al.* Polygenic risk, susceptibility genes, and breast cancer over the life course. *medRxiv* (2020).
478 doi:10.1101/2020.04.17.20069229
479 47. Fahed, A. C. *et al.* Polygenic background modifies penetrance of monogenic variants conferring risk for
480 coronary artery disease, breast cancer, or colorectal cancer. *medRxiv* (2019). doi:10.1101/19013086
481 48. Barnes, D. R. *et al.* Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of
482 BRCA1 and BRCA2 pathogenic variants. *Genet. Med.* (2020). doi:10.1038/s41436-020-0862-x
483 49. van den Broek, J. J. *et al.* Personalizing breast cancer screening based on polygenic risk and family history.

484 *J. Natl. Cancer Inst.* (2020). doi:10.1093/jnci/djaa127
485 50. Esserman, L. J. & WISDOM Study and Athena Investigators. The WISDOM Study: breaking the deadlock
486 in the breast cancer screening debate. *NPJ Breast Cancer* **3,** 34 (2017).
487 51. Martens, F. K., Tonk, E. C. M. & Janssens, A. C. J. W. Evaluation of polygenic risk models using multiple
488 performance measures: a critical assessment of discordant results. *Genet. Med.* **21,** 391–397 (2019).
489 52. Janssens, A. C. J. W. Validity of polygenic risk scores: are we measuring what we think we are? *Hum.*
490 *Mol. Genet.* **28,** R143–R150 (2019).
491 53. Warrier, V. *et al.* Polygenic scores for intelligence, educational attainment and schizophrenia are
492 differentially associated with core autism features, IQ, and adaptive behaviour in autistic individuals.
493 *medRxiv* (2020). doi:10.1101/2020.07.21.20159228
494 54. Matoba, N., Love, M. I. & Stein, J. L. Evaluating brain structure traits as endophenotypes using polygenicity
495 and discoverability. *BioRxiv* (2020). doi:10.1101/2020.07.17.208843
496 55. Coleman, J. R. I. *et al.* Genome-wide gene-environment analyses of major depressive disorder and
497 reported lifetime traumatic experiences in UK Biobank. *Mol. Psychiatry* **25,** 1430–1446 (2020).
498 56. Kraft, P. & Aschard, H. Finding the missing gene-environment interactions. *Eur J Epidemiol* **30,** 353–355
499 (2015).
500 57. Janssens, A. C. J. W. *et al.* Strengthening the reporting of genetic risk prediction studies (GRIPS):
501 explanation and elaboration. *Eur. J. Hum. Genet.* **19,** 18 p preceding 494 (2011).
502 58. von Elm, E. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)
503 statement: guidelines for reporting observational studies. *BMJ* **335,** 806–808 (2007).
504 59. Little, J. *et al.* STrengthening the REporting of Genetic Association Studies (STREGA): an extension of
505 the STROBE statement. *PLoS Med.* **6,** e22 (2009).
506 60. Moons, K. G. M. *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis
507 or Diagnosis (TRIPOD): explanation and elaboration. *Ann. Intern. Med.* **162,** W1-73 (2015).
508 61. Choi, S. W., Mak, T. S.-H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses.
509 *Nat. Protoc.* **15,** 2759–2772 (2020).
510 62. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature*
511 **570,** 514–518 (2019).
512 63. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted
513 arrays and summary statistics 2019. *Nucleic Acids Res.* **47,** D1005–D1012 (2019).
514 64. Lambert, S. A. *et al.* The Polygenic Score Catalog: an open database for reproducibility and systematic
515 evaluation. *medRxiv* (2020). doi:10.1101/2020.05.20.20108217
516 65. Thelwall, M. *et al.* Is useful research data usually shared? An investigation of genome-wide association
517 study summary statistics. *PLoS One* **15,** e0229578 (2020).
518 66. Burke, W. & Zimmern, R. *Moving Beyond ACCE: An Expanded Framework for Genetic Test Evaluation*.
519 (PHG Foundation, 2007).

520

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

Reporting Standard		Description
Background	**Study Type**	Specify whether the study aims to develop and/or validate a PRS. When externally validating or combining previously published PRS or integrated risk model, include identifier(s) of original PRS (PMID, PGS Catalog ID).
	Risk Model Purpose & Predicted End Outcome	Specify what the risk model is intended to predict and the purpose: • Will it be used in risk prediction, diagnostic, prognostic, or therapeutic modalities? • What is the end outcome predicted by the risk model? If the predicted outcome is a clinical feature or endpoint within a specific disease, state the disease. • What current models for risk prediction are available, if pertinent?
Study Population and Data *Many risk score studies involve multiple populations and cohorts that can be used in different stages of PRS and risk score development and evaluation. Each of the populations used (e.g. training, validation, subgroup analysis samples) in the manuscript should be defined using this common set of descriptors.*	**Study Design & Recruitment**	For each of the data sets describe the study design (e.g., cohort, case-control, cross-sectional), eligbility criteria, recruitment period and setting (e.g. method and years), and follow-up. State whether the data are primary or secondary data. If secondary analysis, include the full reference to the original study.
	Participant Demographic and Clinical Characteristics	Include the distribution of demographic information in each data set (and the combined total if relevant) used to generate a single risk model (whether a single sample set, or the summary of combined samples) including the mean, standard deviation and range. This should at minimum include age, sex and any other characteristics relevant to describe the study population or the performance of the model. Provide demographics stratified by case/control status, if applicable.
	Ancestry	Include the ancestral background distribution of each sample population used during PRS development and validation, 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. The method of ancestry inference should be provided. Genomic methods are preferred, such as principal component analysis.
	Genetic Data	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 imputation panel and give ancestry information if not a standard imputation panel. Report any relevant quanlity control, including imputation quality filters to exclude low quality imputed SNPs. If parameters were selected from another study, include reference (PMID, GWAS Catalog ID).
	Non-Genetic Variables	Define any non-genetic variables that were included in the risk model, provide variable definitions and measurement (e.g., assay, ICD codes, e-phenotyping algorithms, chart review, self-report). Indicate the scale of each variable, e.g. dichotomous, continuous, categorical, or ordinal. Explicitly state which variables are included in the final model.
	Outcome of Interest	Define the predicted outcome of interest. If the predicted outcome is a clinical feature or endpoint within a specific disease, provide the criteria used to define that disease membership. Include details on how information was ascertained (e.g., ICD codes, e-phenotyping algorithms, chart review, self-report). Transformation of continuous data into binary, ordinal, or categorical outcomes should be detailed with justification. Authors should explicitly state the number of cases and controls.
	Missing Data	State explicitly how missing data were handled for all variables included in the model. If imputation was utilized, include detailed of the approach used and any subsequent filtering or post-processing.
Risk Model Development & Application *Describe the relevant methods used to form the final PRS and/or risk model. Samples used in*	**Polygenic Risk Score Construction & Estimation**	Describe how genetic data were included in the PRS. Authors should detail criteria used to determine inclusion in the model for all variants. Define how the variants were selected, weighted and combined into a single score. If the PRS was derived from another study include the reference (PMID, PGS Catalog Score ID).
	Risk Model Type	Detail statistical methods used to estimate risk, either relative or absolute, from the continuous risk score distribution. Detail if risk is cumulative or cross-sectional, as well as the appropriate comparison groups if relative risk presented. Report time until predicted risk (e.g. 5-year, 10-year, lifetime). In a relative hazard model, the study period or follow up

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.*		time may be used. In an absolute risk model, state the time until predicted event and the prevalence/incidence of the predicted outcome in the general population.
	Integrated Risk Model(s) Description and Fitting	State the procedure utilized to develop the risk models that includes non-genetic and/or genetic variables other than the PRS. If the model(s) was selected for optimal performance, describe measures used to assess performance. Explicitly state all variables used in each risk model.
Risk Model Evaluation *Outline the results and procedures utilized to evaluate the risk model, specifying internal or external validation. Performance results should be described for both the development and validation samples. Specify if the application of the risk model differs between the development and validation samples.*	**PRS Distribution**	Include a general description of the distribution of the PRS. This details the continuous distribution output directly from the risk score calculation.
	Risk Model Predictive Ability	Describe and report any metrics of overall performance (proportion of variance explained; R^2) and estimates of risk (such as odds or hazards ratios from regression models) used to evaluate the PRS and/or risk models. Describe the set of genetic/non-genetic variables included in the analysis.
	Risk Model Discrimination	Describe and report metrics used to assess the discrimination of evaluated risk models and whether any non-genetic variables were included beyond a PRS 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. Evaluation of potential clinical utility of models requires also evaluating tail-based measures, such as proportions of populations and cases exceeding specified clinically relevant risk thresholds and measures of reclassfications (e.g. NRI) at such thresholds for comparison of models.
	Risk Model Calibration	Describe and report metrics used to assess the calibration of evaluated risk scores and models. Describe the set of genetic/non-genetic variables included in the analysis.
	Subgroup Analyses	Subgroup size, demographics and clinical characteristics should be given. Relevant evaluation methods and measures (distribution, predictive ability, discrimination, calibration) should be described for each subgroup analysis.
Limitations and Clinical Implications *Discuss the broader context of the study and risk model.*	**Risk Model Interpretation**	Summarize the risk models in terms of what they predict, how well, and in whom. Explicitly mention the 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 of the study with relevance to the results, discuss the impact of these limitations on the interpretation of the risk model and any downstream replication efforts needed. Common considerations include: study design restrictions, use of a surrogate outcome, ascertainment biases, the distribution of participant-level traits (ancestry, age, comorbidities), accuracy/specificity of outcome data, and any statistical considerations. Make note of and discuss the impact of any unknown reporting items from previous sections.
	Generalizability	Discuss the intended target groups or populations this score may be applied to and explicitly address any issues with generalizability beyond the included populations. Discuss whether the study externally validates the score and/or model, or if the sample is limited with respect to ancestry, age, or other variables.
	Risk Model Intended Uses	Discuss whether there is an intended clinical use or utility to the risk model. If so, discuss the “clinic readiness” and next steps with respect to the interpretation, limitations, and generalizability of the model. Discuss how the predictive ability of the model is compared against current standard of care or other published work (such as existing PRS) on predicting the outcome of interest.
Data Transparency and Availability		Information sufficient to calculate the PRS and the risk model(s) on external samples should be made freely available. For genetic variables this would include information about the variants (e.g., rsID, chromosomal location, effect allele, and the effect weight) that comprise the score; PRS with this information should be deposited in the PGS Catalog for findability and to promote re-use and comparison with other established scores. Weights for non-genetic variables should also be provided to make the risk model calculable.

Further reporting considerations can be found in ***Supplemental Note 3****. A reference of relevant manuscript sections for each item is provided in* ***Supplemental Table 4****.*