REVIEW

Understanding polygenic models, their development and the potential application of polygenic scores in healthcare

Chantal Babb de Villiers [ID] , Mark Kroese, Sowmiya Moorthie

PHG Foundation, University of Cambridge, Cambridge, UK

Correspondence to
Dr Chantal Babb de Villiers, PHG Foundation, University of Cambridge, Cambridge CB1 8RN, UK;
chantal.babbdevilliers@phgfoundation.org

Check for updates

ABSTRACT

The use of genomic information to better understand and prevent common complex diseases has been an ongoing goal of genetic research. Over the past few years, research in this area has proliferated with several proposed methods of generating polygenic scores. This has been driven by the availability of larger data sets, primarily from genome-wide association studies and concomitant developments in statistical methodologies. Here we provide an overview of the methodological aspects of polygenic model construction. In addition, we consider the state of the field and implications for potential applications of polygenic scores for risk estimation within healthcare.

INTRODUCTION

There has been considerable interest in elucidating the contribution of genetic factors to the development of common diseases and using this information for better prediction of disease risk. The common disease common variant hypothesis predicts that variants that are common in the population play a role in disease susceptibility.[1] Genome-wide association studies (GWAS) using single nucleotide polymorphism (SNP) arrays were developed as a mechanism by which to investigate these genetic factors and it was hoped this would lead to identification of variants associated with disease risk and subsequent development of predictive tests. Variants identified as associated with particular traits by these studies, for the large part, are SNPs that individually have a minor effect on disease risk and hence, by themselves, cannot be reliably used in disease prediction. Looking at the aggregate impact of these SNPs in the form of a polygenic score (PGS) appeared to be one possible means of using this information to predict disease.[2] It is thought this will be of benefit as our genetic make-up is largely stable from birth and dictates a 'baseline risk' on which external influences act and modulate. Therefore, PGS are a potential mechanism to act as a risk predictor by capturing information on this genetic liability.

The use of PGS as a predictive biomarker is being explored in a number of different disease areas, including cancer,[3 4] psychiatric disorders,[5–7] metabolic disorders (diabetes,[8] obesity[9]) and coronary artery disease (CAD).[10] The proposed applications range from aiding disease diagnosis, informing selection of therapeutic interventions, improvement of risk prediction, informing disease screening and, on a personal level, informing life planning. Therefore, genetic risk information in the form of a PGS is considered to have potential in informing both clinical and individual-level decision-making.

Recent advances in statistical techniques, improved computational power and the availability of large data sets have led to rapid developments in this area over the past few years. This has resulted in a variety of approaches to construction of models for score calculation and the investigation of these scores for prediction of common diseases.[11] Several review articles aimed at researchers with a working knowledge of this field have been produced.[6 11–17] In this article, we provide an overview of the key aspects of PGS construction to assist clinicians and researchers in other areas of academia to gain an understanding of the processes involved in score construction. We also consider the implications of evolving methodologies for the development of applications of PGS in healthcare.

EVOLUTION IN POLYGENIC MODEL CONSTRUCTION METHODOLOGIES

Terminology with respect to PGS has evolved over time, reflecting evolving approaches and methodology. Other terms include PGS, polygenic risk score, polygenic load, genotype score, genetic burden, polygenic hazard score, genetic risk score (GRS), metaGRS and allelic risk score. Throughout this article we use the terms polygenic models to refer to the method used to calculate an output in the form of a PGS. Different polygenic models can be used to calculate a PGS and analysis of these scores can be used to examine associations with particular markers or to predict an individuals risk of diseases.[12]

Usual practice in calculating PGS is as a weighted sum of a number of risk alleles carried by an individual, where the risk alleles and their weights are defined by SNPs and their measured effects (figure 1).[11] Polygenic models have been constructed using a few, hundreds or thousands of SNPs, and more recently SNPs across the whole genome. Consequently, determining which SNPs to include and the disease-associated weighting to assign to SNPs are important aspects of model construction (figure 2).[18] These aspects are influenced by available genotype data and effect size estimates as well as the methodology employed in turning this information into model parameters (ie, weighted SNPs).

$$PGS = \beta_1 k_1 + \beta_2 k_2 + \beta_3 k_3 + \ldots \beta_n k_n$$

β = Weights (example is β as the per allele log odds ratio for the minor allele)

k = allele dosage for the SNP (0, 1, or 2)

n = number of SNPs

Figure 1 Polygenic score calculation. This calculation aggregates the SNPs and their weights selected for a polygenic score. Common diseases are thought to be influenced by many genetic variants with small individual effect sizes, such that meaningful risk prediction necessitates examining the aggregated impact of these multiple variants including their weightings. PGS, polygenic score.

Changes in data availability over time have had an impact on the approach taken in SNP selection and weighting. Early studies to identify variants associated with common diseases took the form of candidate gene studies. The small size of candidate gene studies, the limitation of technologies available for genotyping and stringent significance thresholds meant that these studies investigated fewer variants and those that were identified with disease associations had relatively large effect sizes.[19] Taken together, this meant that a relatively small number of variants were available for consideration for inclusion in a polygenic model.[20 21] Furthermore, weighting parameters for these few variants were often simplistic, such as counts of the number of risk alleles carried, ignoring their individual effect sizes.[16]

The advent of GWAS enabled assessment of SNPs across the genome, leading to the identification of a larger number of disease-associated variants and therefore more variants suitable for inclusion in a polygenic model. In addition, the increasing number of individuals in the association studies meant that the power of these studies increased, allowing for more precise estimates of effect sizes.[19] Furthermore, some theorised that lowering stringent significance thresholds set for SNP–trait associations could also identify SNPs that might play a part in disease risk.[11 16] This resulted in more options with respect to polygenic model parameters of SNPs to include and weights to assign to them. However, the inclusion of more SNPs and direct application of GWAS effect sizes as a weighting parameter does not always equate to better predictive performance.[4 16] This is because GWAS do not provide perfect information with respect to the causal SNP, the effect sizes or the number of SNPs that contribute to the trait. Therefore, different methods have been developed to address these issues and optimise predictive performance of the score. Current common practice is to construct models with different iterations of SNPs and weighting, with assessment of the performance of each to identify the optimum configuration of SNPs and their weights (figure 2).

METHODS USED IN SNP SELECTION AND WEIGHTING ASSIGNMENT

Some methods of model development will initially involve selection of SNPs followed by optimisation of weighting, whereas others may involve optimisation of weightings for all SNPs that have been genotyped using their overall GWAS effect sizes, the linkage disequilibrium (LD) and an estimate of the proportion of SNPs that are expected to contribute to the risk.[22]

LD is the phenomenon where some SNPs are coinherited more frequently with other SNPs due to their close proximity on the genome. Segments with strong LD between SNPs are referred to as haplotype blocks. This phenomenon means that GWAS often identify multiple SNPs in the same haplotype block associated with disease and the true causal SNP is not known. As models have started to assess more SNPs, careful consideration is required to take into account possible correlation between SNPs as a result of this phenomenon. Correlation between SNPs can lead to double counting of SNPs and association redundancy, where multiple SNPs in a region of LD are identified as being associated with the outcome. This can lead to reduction in the predictive performance of the model. Therefore, processes for filtering SNPs and using one SNP (tag SNP) to act as a marker in an area of high LD, through LD thinning, were developed. Through these processes SNPs correlated with other SNPs in a block are removed, by either pruning or clumping. Pruning ignores p value thresholds and 'eliminates' SNPs by a process of iterative comparison between a pair of SNPs to assess if they are correlated, and subsequently could remove SNPs that are deemed to have evidence of association. Clumping (also known as informed pruning) is guided by GWAS p values and chooses the most significant SNP, therefore keeping the most significant SNP within a block.[23] This is all done with the aim of pinpointing relatively small areas of the genome that contribute to risk of the trait. Different significance thresholds may be used to select SNPs from this subgroup for inclusion in models.

Poor performance of a model can result from imperfect tagging with the underlying causal SNP.[16] This is because the causal SNP that is associated with disease might not be in LD with the tag SNP that is in the model but is in LD with another SNP which is not in the model. This particularly occurs where the LD and variant frequency differs between population groups.[24] An
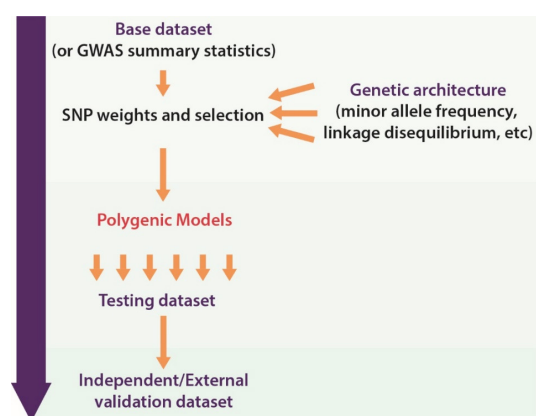

Figure 2 Construction of a polygenic score. In the process of developing a polygenic score, numerous models are tested and then compared. The model that performs best (as determined by one or more measures) is then selected for validation in the external data set. GWAS, genome-wide association studies.

alternate approach to filter SNPs is stepwise regression where SNPs are selected based on how much the SNPs improve the model's performance. This is a statistical approach and does not consider the impact of LD or effect size.

As described above, early studies used simple weighting approaches or directly applied effect sizes from GWAS as weighting parameters for SNPs. However, application of effect sizes as a weighting parameter directly from a GWAS may not be optimal, due to differences in the population that the GWAS was conducted in and the target population. Also as described above, LD and the fact that not all SNPs may contribute to the trait mean that these effect sizes from GWAS are imperfect estimates. Therefore, methods have been developed that adjust effect size estimates from GWAS using statistical techniques which make assumptions about factors such as the number of causal SNPs, level of LD between SNPs or knowledge of their potential function to better reflect their impact on a trait. Numerous statistical methodologies have been developed to improve weighting with a view to enhancing the discriminative power of a PGS.[25 26] Examples of some methodological approaches are LDpred,[22] winner's curse correction,[23] empirical Bayes estimation,[27] shrinkage regression (Lasso),[28] linear mixed models,[29] with more being developed or tested. An additional improvement on the methods is to embed non-genetic information (eg, age-specific ORs).[6] Determination of which methodology or hybrid of methodologies is most appropriate for various settings and conditions is continuously being explored and is evolving with new statistical approaches developing at a rapid pace.

In summary, model development has evolved in an attempt to gain the most from available GWAS data and address some of the issues that arise due to working with data sets which cannot be directly translated into parameters for prediction models. The different approaches taken in SNP selection and weighting, and the impact on the predictive performance of a model are important to consider when assessing different models. This is because different approaches to PGS modelling can achieve the same or a similar level of prediction. From a health system implementation perspective, particular approaches may be preferred following practical considerations and trade-offs between obtaining genotype data, processes for score construction and model performance. In addition, the degree to which these parameters need to be optimised will also be impacted by the input data and validation data set, and the quality control procedures that need to be applied to these data sets.[12]

SOURCES OF INPUT DATA FOR SCORE CONSTRUCTION

Key to the development of a polygenic model is the availability of data sets that can provide input parameters for model construction. Genotype data used in model construction can either be available as raw GWAS data or provided as GWAS summary statistics. Data in the raw format are individual-level data from a SNP array and may not have undergone basic quality control such as assessment of missingness, sex discrepancy checks, deviation from Hardy-Weinberg equilibrium, heterozygosity rate, relatedness or assessment for outliers.[30 31] Availability of raw GWAS data allows for different polygenic models to be developed because of the richness of the data, however computational issues arise because of the size of the data sets. Data based on genome sequencing, as opposed to SNP arrays, could also be used in model construction. There have been limited studies of PGS developed from this form of data due to limitations in data availability, which is mainly due to cost restraints.[15 32] Individual-level

genomic data are also often not available to researchers due to privacy concerns.

Due to these issues, the focus of polygenic model development has therefore been on using well-powered GWAS summary statistics.[33] These are available from open access repositories and contain summary information such as the allele positions, ORs, CIs and allele frequency, without containing confidential information on individuals. These data sets have usually been through the basic quality control measures mentioned above. There are, however, no standards for publicly available files, meaning some further processing steps may be required, in particular when various data sets are combined for a meta-analysis. Quality control on summary statistics is only possible if information such as missing genotype rate, minor allele frequency, Hardy-Weinberg equilibrium failures and non-Mendelian transmission rates is provided.[12]

Processing of GWAS data may include additional quality control steps, imputation and filtering of the SNP information, which can be done at the level of genotype or summary statistics data. SNP arrays used in GWAS only have common SNPs represented on them as they rely on LD between SNPs to cover the entire genome. As described above, one tag SNP on the array can represent many other SNPs. Imputation of SNPs is common in GWAS and describes the process of predicting genotypes that have not been directly genotyped but are statistically inferred (imputed) based on haplotype blocks from a reference sequence.[33–35] Often association tests between the imputed SNPs and trait are repeated. As genotype imputation requires individual-level data, researchers have proposed summary statistics imputation as a mechanism to infer the association between untyped SNPs and a trait. The performance of imputation has been evaluated and shown that, with certain limitations, summary statistics imputation is an efficient and cost-effective methodology to identify loci associated with traits when compared with imputation done on genotypes.[36]

An alternative source of input data for the selection of SNPs and their weightings is through literature or in existing databases, where already known trait-associated SNPs and their effect sizes are used as the input parameters in model development. A number of studies have taken this approach[37 38] and it is possible to use multiple sources when developing various polygenic models and establishing the preferred parameters to use.

Currently, there does not appear to be one methodology that works across all contexts and traits, each trait will need to be assessed to determine which method is the most suitable for the trait being evaluated. For example, four different polygenic model construction strategies were explored for three skin cancer subtypes[4] by using data on SNPs and their effect sizes from different sources, such as the latest GWAS meta-analysis results, the National Human Genome Research Institute (NHGRI) EBI GWAS catalogue, UK Biobank GWAS summary statistics with different thresholds and GWAS summary statistics with LDpred. In this setting for basal cell carcinoma and melanoma, the meta-analysis and catalogue-derived models were found to perform similarly but that the latter was ultimately used as it included more SNPs. For squamous cell carcinoma the meta-analysis-derived model performed better than the catalogue-derived model. This demonstrates how each disease subtype, model construction strategy and data set can have their own limitations and advantages.

Knowledge of the sources of input data and its subsequent use in model development is important in understanding the limitations of available models. Models that are developed using data sets that reflect the population in which prediction is to be

727

carried out will perform better. For example, data collected from a symptomatic or high-risk population may not be suitable as an input data set for the development of a polygenic model that will be used for disease prediction in the general population. Large GWAS studies were previously focused on high-risk individuals, such as patients with breast cancer with a strong family history or known pathogenic variants in BRCA1 or BRCA2. These studies would not be suitable for the development of PGS for use in the general population but can inform risk assessment in high-risk individuals. The source of the data for SNP selection and weighting also has implications for downstream uses and validation. For example, variant frequency and LD patterns can vary between populations and this can translate to poor performance of the polygenic model if the external validation population is different from that of the input data set.[39–41] Furthermore, the power and validity of polygenic analyses are influenced by the input data sources.[12 42]

FROM A MODEL TO A SCORE

PGS can be calculated using one of the methodologies discussed above. The resulting PGS units of measurement depend on which measurement is used for the weighting.[12] For example, the weightings may have been calculated based on logOR for discrete traits or linear regression coefficient (β/beta) in continuous traits from univariate regression tests carried out in the GWAS. The resulting scores are then usually transformed to a standard normal distribution to give scores ranging from −1 to 1, or 0 to 100 for ease of interpretation. This enables further examination of the association between the score and a trait and the predictive ability of different scores generated by different models. Similar to other biomarker analyses, this involves using the PGS as a predictor of a trait with other covariates (eg, age, smoking, and so on) added, if appropriate, in a target sample. Examination of differences in the distribution of scores in cases and controls, or by examining differences in traits between different strata of PGS can enable assessment of predictive ability (figure 3). Common practice is for individual-level PGS values to be used to stratify populations into distinct groups of risk based on percentile cut-off or threshold values (eg, the top 1%).

MODEL VALIDATION

Polygenic model development is reliant on further data sets for model testing and validation and the composition of these data sets is important in ensuring that the models are appropriate for a particular purpose. The development of a model to calculate a PGS involves refinement of the previously discussed input
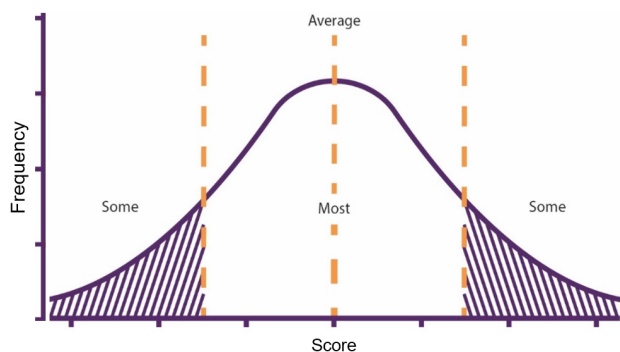

Figure 3 Example distribution of polygenic scores across a population. Thresholds can be set to stratify risk as low (some), average (most) and high (some).

parameters, and selection of the 'best' of several models based on performance (figure 2). Therefore, a testing/training data set is often required to assess the model's ability to accurately predict the trait of interest. This is often a data set that is independent of the base/input/discovery data set. It may comprise a subset of the discovery data set that is only used for testing and was not included in the initial development of the model but should ideally be a separate independent data set.

Genotype and phenotype data are needed in these data sets. Polygenic models are used to calculate PGS for individuals in the training data set and regression analysis is performed with the PGS as a predictor of a trait; other covariates may also be included, if appropriate. This testing phase can be considered a process for identifying models with better overall performance and/or informing refinements needed. Hence, this phase often involves comparison of different models that are developed using the same input data set to identify those models that have optimal performance.

The primary purpose is to determine which model best discriminates between cases and controls. The area under the curve (AUC) or the C-statistic is the most commonly used measure in assessing discriminative ability. It has been criticised as being an insensitive measure that is not able to fully capture all aspects of predictive ability. For instance, in some instances, AUC can remain unchanged between models but the individuals within are categorised into a different risk group.[43] Alternative metrics that have been used to evaluate model performance include increase in risk difference, integrated discrimination improvement, R2 (estimate of variance explained by the PGS after covariate adjustment), net classification index and the relative risk (highest percentile vs lowest percentile). A clear understanding on how to interpret the performance within various settings is important in determining which model is most suitable.[44]

As per normal practice when developing any prediction model, polygenic models with the optimal performance in a testing/training data set should be further validated in external data sets. External data sets are critical in validation of models and assessment of generalisability, hence must also conform to the desired situations in which a model is to be used. The goal is to find a model with suitable parameters of predictive performance in data sets outside of those in which it was developed. Ideally, external validation requires replication in independent data sets. Few existing polygenic models have been validated to this extent, the focus being rather on the development of new models rather than evaluation of existing ones. One example where replication has been carried out is in the field of CAD, where the GPS$_{CAD}$[45] and metaGRS$_{CAD}$[10] polygenic models (both developed using UK Biobank data) were evaluated in a Finnish population cohort.[46] Predictive ability was found to be lower in the Finnish population. This is likely to be due to the differences in genetic structure of this population and the population of the data set used for polygenic model development. Research is ongoing to evaluate polygenic models in other populations and strategies are being developed to ensure the same performance when used more widely, possibly through reweighting and adjustment of the scores.[47]

MOVING TOWARDS CLINICAL APPLICATIONS

PGS are thought to be useful information that could improve risk estimation and provide an avenue for disease prevention and deciding treatment strategies. There are indications from a number of fields that genetic information in the form of PGS can act as independent biomarkers and aid stratification.[11 16 48]

However, the clinical benefits of stratification using a PGS and the implications for clinical practice are only just beginning to be examined. The use of PGS as part of existing risk prediction tools or as a stand-alone predictor has been suggested. This latter option may be true for diseases where knowledge or predictive ability with other risk factors is limited, such as in prostate cancer.[49] In either case, polygenic models need to be individually examined to determine suitability and applicability for the specific clinical question.[50] Despite some commercial companies developing PGS,[51 52] currently PGS are not an established part of clinical practice.

Integration into clinical practice requires evaluation of a PGS-based test. An important concept to consider in this regard is the distinction between an assay and a test. This has been previously discussed with respect to genetic test evaluation.[53 54] It is worth examining this concept as applied to PGS, as their evaluation is reliant on a clear understanding of the test to be offered. As outlined by Zimmern and Kroese,[54] the method used to analyse a substance in a sample is considered the assay, whereas a test is the use of an assay within a specific context. With respect to PGS, the process of developing a model to derive a score can be considered the assay, while the use of this model for a particular disease, population and purpose can be considered the test. This distinction is important when assessing if studies are reporting on assay performance as opposed to test performance. It is our view that, with respect to polygenic models, progress has been made with respect to assay development, but PGS-based tests are yet to be developed and evaluated. This can enable a clearer understanding of their potential clinical utility and issues that may arise for clinical implementation.[11 18 55] It is clear that this is still an evolving field, and going forward different models may be required for different traits due to their underlying genetic architecture,[26] different clinical contexts and needs.

Clinical contexts where risk stratification is already established practice are most likely where implementation of PGS will occur first. Risk prediction models based on non-genetic factors have been developed for many conditions and are used in clinical care, for example, in cardiovascular disease over 100 such models exist.[56] In such contexts, how a PGS and its ability to predict risk compared with, or improves on, these existing models is being investigated.[3 44 57–61] The extent to which PGS improves prediction, as well as the cost implications of including this, is likely to impact on implementation.

Integration of PGS into clinical practice, for any application, requires robust and validated mechanisms to generate these scores. Therefore, given the numerous models available, an assessment of their suitability as part of a test is required. Parameters or guidelines with respect to aspects of model performance and metrics that could assist in selecting the model to take forward as a PGS-based test are limited and need to be addressed. Currently, there are different mechanisms to generate PGS and have arisen in response to the challenges in aggregating large-scale genomic data for prediction. For example, a review reported 29 PGS models for breast cancer from 22 publications.[62] Due to there being a number of different methodologies to generate a score, numerous models may exist for the same condition and each of the resulting models could perform differently. Models may perform differently because the population, measured outcome or context of the development data sets used to generate the models is diverse, for example, a score for risk of breast cancer versus a breast cancer subtype.[44 63] This diversity, alongside the lack of established best practice and standardised reporting in publications, makes comparison and evaluation of polygenic models for use in clinical settings challenging. It

is clear that moving the field forward is reliant on transparent reporting and evaluation. Recommendations for best practices on the reporting of polygenic models in literature have been proposed[14 64] as well as a database,[65 66] which could allow for such comparisons. Statements and guidelines for risk prediction model development, such as the Genetic Risk Prediction Studies and Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD), already exist, but are not consistently used. TRIPOD explicitly covers the development and validation of prediction models for both diagnosis and prognosis, for all medical domains.

One clear issue is generalisability and drop in performance of polygenic models once they are applied in a population group different from the one in which they were developed.[22 46 67–70] This is an ongoing challenge in genomics as most GWAS, from which most PGS are being developed, have been conducted in European-Caucasian populations.[71] Efforts to improve representation are underway[72] and there are attempts to reweight/adjust scores when applied to different population groups which are showing some potential but need further research.[47] Others have demonstrated that models developed in more diverse population groups have improved performance when applied to external data sets in different populations.[24 73] It is important to consider this issue when moving towards clinical applications as it may pose an ethical challenge if the PGS is not generalisable.

A greater understanding of different complex traits and the impact of pleiotropy is only beginning to be investigated.[74] There is growing appreciation of the role of pleiotropy as multiple variants have been identified to be associated with multiple traits and exert diverse effects, providing insight into overlapping mechanisms.[75 76] This, together with the impact of population stratification, genetic relatedness, ascertainment and other sources of heterogeneity leading to spurious signals and reduced power in genetic association studies, all impacting on the predictive ability of PGS in different populations and for different diseases.

While many publications report on model development and evaluation, often there is a lack of clarity on intended purpose,[50 77] leading to uncertainties as to the clinical pathways in which implementation is envisaged. A clear description of intended use within clinical pathways is a central component in evaluating the use of an application with any form of PGS and in considering practical implications, such as mechanisms of obtaining the score, incorporation into health records, interpretation of scores, relevant cut-offs for intervention initiation, mechanisms for feedback of results and costs, among other issues. These parameters will also be impacted by the polygenic model that is taken forward for implementation. Meaning that there are still some important questions that need to be addressed to determine how and where PGS could work within current healthcare systems, particularly at a population level.[78]

It is widely accepted that genotyping using arrays is a lower cost endeavour in comparison to genome sequencing, making the incorporation of PGS into routine healthcare an attractive proposition. However, we were unable to find any studies reporting on the use or associated costs of such technology for population screening. Studies are beginning to examine use case scenarios and model cost-effectiveness, but this has only been in very few, specific investigations.[79 80] Costs will also be influenced by the testing technology and by the downstream consequences of testing, which is likely to differ depending on specific applications that are developed and the pathways in which such tests are incorporated. This is particularly the case in screening or primary care settings, where such testing is currently not an established part of care pathways and may require additional

resources, not least as a result of the volume of testing that could be expected. Moving forward, the clinical role of PGS needs to be developed further, including defining the clinical applications as well as supporting evidence, for example, on the effect of clinical outcomes, the feasibility for use in routine practice and cost-effectiveness.

CONCLUSION

There is a large amount of diversity in the PGS field with respect to model development approaches, and this continues to evolve. There is rapid progress which is being driven by the availability of larger data sets, primarily from GWAS and concomitant developments in statistical methodologies. As understanding and knowledge develops, the usefulness and appropriateness of polygenic models for different diseases and contexts are being explored. Nevertheless, this is still an emerging field, with a variable evidence base demonstrating some potential. The validity of PGS needs to be clearly demonstrated, and their applications evaluated prior to clinical implementation.

ORCID iD
Chantal Babb de Villiers http://orcid.org/0000-0003-1334-1819

REFERENCES

1 Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant…or not? *Hum Mol Genet* 2002;11:2417–23.
2 Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 2007;17:1520–8.
3 Seibert TM, Fan CC, Wang Y, Zuber V, Karunamuni R, Parsons JK, Eeles RA, Easton DF, Kote-Jarai ZSofia, Al Olama AA, Garcia SB, Muir K, Grönberg H, Wiklund F, Aly M, Schleutker J, Sipeky C, Tammela TL, Nordestgaard BG, Nielsen SF, Weischer M, Bisbjerg R, Røder MA, Iversen P, Key TJ, Travis RC, Neal DE, Donovan JL, Hamdy FC, Pharoah P, Pashayan N, Khaw K-T, Maier C, Vogel W, Luedeke M, Herkommer K, Kibel AS, Cybulski C, Wokolorczyk D, Kluzniak W, Cannon-Albright L, Brenner H, Cuk K, Saum K-U, Park JY, Sellers TA, Slavov C, Kaneva R, Mitev V, Batra J, Clements JA, Spurdle A, Teixeira MR, Paulo P, Maia S, Pandha H, Michael A, Kierzek A, Karow DS, Mills IG, Andreassen OA, Dale AM, PRACTICAL Consortium*. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. *BMJ* 2018;360:j5757.
4 Fritsche LG, Beesley LJ, VandeHaar P, Peng RB, Salvatore M, Zawistowski M, Gagliano Taliun SA, Das S, LeFaive J, Kaleba EO, Klumpner TT, Moser SE, Blanc VM, Brummett CM, Kheterpal S, Abecasis GR, Gruber SB, Mukherjee B. Exploring various polygenic risk scores for skin cancer in the phenomes of the Michigan genomics initiative and the UK Biobank with a visual catalog: PRSWeb. *PLoS Genet* 2019;15:e1008202.
5 Vassos E, Di Forti M, Coleman J, Iyegbe C, Prata D, Euesden J, O'Reilly P, Curtis C, Kolliakou A, Patel H, Newhouse S, Traylor M, Ajnakina O, Mondelli V, Marques TR, Gardner-Sood P, Aitchison KJ, Powell J, Atakan Z, Greenwood KE, Smith S, Ismail K, Pariante C, Gaughran F, Dazzan P, Markus HS, David AS, Lewis CM, Murray RM, Breen G. An examination of polygenic score risk prediction in individuals with First-Episode psychosis. *Biol Psychiatry* 2017;81:470–7.
6 Chasioti D, Yan J, Nho K, Saykin AJ. Progress in polygenic composite scores in Alzheimer's and other complex diseases. *Trends Genet* 2019;35:371–82.
7 Mullins N, Bigdeli TB, Børglum AD, Coleman JRI, Demontis D, Mehta D, Power RA, Ripke S, Stahl EA, Starnawska A, Anjorin A, Corvin A, Sanders AR, Forstner AJ, Reif A,

Koller AC, Świątkowska B, Baune BT, Müller-Myhsok B, Penninx BWJH, Pato C, Zai C, Rujescu D, Hougaard DM, Quested D, Levinson DF, Binder EB, Byrne EM, Agerbo E, Streit F, Mayoral F, Bellivier F, Degenhardt F, Breen G, Morken G, Turecki G, Rouleau GA, Grabe HJ, Völzke H, Jones I, Giegling I, Agartz I, Melle I, Lawrence J, Walters JTR, Strohmaier J, Shi J, Hauser J, Biernacka JM, Vincent JB, Kelsoe J, Strauss JS, Lissowska J, Pimm J, Smoller JW, Guzman-Parra J, Berger K, Scott LJ, Jones LA, Azevedo MH, Trzaskowski M, Kogevinas M, Rietschel M, Boks M, Ising M, Grigoroiu-Serbanescu M, Hamshere ML, Leboyer M, Frye M, Nöthen MM, Alda M, Preisig M, Nordentoft M, Boehnke M, O'Donovan MC, Owen MJ, Pato MT, Renteria ME, Budde M, Weissman MM, Wray NR, Bass N, Craddock N, Smeland OB, Andreassen OA, Mors O, Gejman PV, Sklar P, McGrath P, Hoffmann P, McGuffin P, Lee PH, Mortensen PB, Kahn RS, Ophoff RA, Adolfsson R, Van der Auwera S, Djurovic S, Kloiber S, Heilmann-Heimbach S, Jamain S, Hamilton SP, McElroy SL, Lucae S, Cichon S, Schulze TG, Hansen T, Werge T, Air TM, Nimgaonkar V, Appadurai V, Cahn W, Milaneschi Y, Fanous AH, Kendler KS, McQuillin A, Lewis CM. Gwas of suicide attempt in psychiatric disorders and association with major depression polygenic risk scores. *Am J Psychiatry* 2019;176:651–60.
8 McCarthy MI, Mahajan A. The value of genetic risk scores in precision medicine for diabetes. *Expert Rev Precis Med Drug Dev* 2018;3:279–81.
9 Belsky DW, Moffitt TE, Sugden K, Williams B, Houts R, McCarthy J, Caspi A. Development and evaluation of a genetic risk score for obesity. *Biodemography Soc Biol* 2013;59:85–100.
10 Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, Lai FY, Kaptoge S, Brozynska M, Wang T, Ye S, Webb TR, Rutter MK, Tzoulaki I, Patel RS, Loos RJF, Keavney B, Hemingway H, Thompson J, Watkins H, Deloukas P, Di Angelantonio E, Butterworth AS, Danesh J, Samani NJ, UK Biobank CardioMetabolic Consortium CHD Working Group. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. *J Am Coll Cardiol* 2018;72:1883–93.
11 Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* 2018;19:581–90.
12 Choi SW, TSH M, O'Reilly P. A guide to performing polygenic risk score analyses. *Biorxiv* 2018.
13 Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM. Research review: polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry* 2014;55:1068–87.
14 Mistry S, Harrison JR, Smith DJ, Escott-Price V, Zammit S. The use of polygenic risk scores to identify phenotypes associated with genetic risk of schizophrenia: systematic review. *Schizophr Res* 2018;197:2-8.
15 De La Vega FM, Bustamante CD. Polygenic risk scores: a biased prediction? *Genome Med* 2018;10:100.
16 Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet* 2016;17:392–406.
17 Dudbridge F, Epidemiology P. Polygenic epidemiology. *Genet Epidemiol* 2016;40:268–72.
18 Moorthie S, Babb de Villiers C, Brigden T, Gaynor L, Hall A, Johnson E, Sanderson S, Kroese M. *Polygenic scores risk and cardiovascular disease: PHG Foundation*, 2019.
19 Juran BD, Lazaridis KN. Genomics in the post-GWAS era. *Semin Liver Dis* 2011;31:215–22.
20 Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, Guiducci C, Perola M, Jula A, Sinisalo J, Lokki M-L, Nieminen MS, Melander O, Salomaa V, Peltonen L, Kathiresan S. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* 2010;376:1393–400.
21 Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt NP, Roos C, Hirschhorn JN, Berglund G, Hedblad B, Groop L, Altshuler DM, Newton-Cheh C, Orho-Melander M. Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med* 2008;358:1240–9.
22 Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, Genovese G, Loh P-R, Bhatia G, Do R, Hayeck T, Won H-H, Kathiresan S, Pato M, Pato C, Tamimi R, Stahl E, Zaitlen N, Pasaniuc B, Belbin G, Kenny EE, Schierup MH, De Jager P, Patsopoulos NA, McCarroll S, Daly M, Purcell S, Chasman D, Neale B, Goddard M, Visscher PM, Kraft P, Patterson N, Price AL, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE) study. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet* 2015;97:576–92.
23 Shi J, Park J-H, Duan J, Berndt ST, Moy W, Yu K, Song L, Wheeler W, Hua X, Silverman D, Garcia-Closas M, Hsiung CA, Figueroa JD, Cortessis VK, Malats N, Karagas MR, Vineis P, Chang I-S, Lin D, Zhou B, Seow A, Matsuo K, Hong Y-C, Caporaso NE, Wolpin B, Jacobs E, Petersen GM, Klein AP, Li D, Risch H, Sanders AR, Hsu L, Schoen RE, Brenner H, Stolzenberg-Solomon R, Gejman P, Lan Q, Rothman N, Amundadottir LT, Landi MT, Levinson DF, Chanock SJ, Chatterjee N, MGS (Molecular Genetics of Schizophrenia) GWAS Consortium, GECCO (The Genetics and Epidemiology of Colorectal Cancer Consortium), GAME-ON/TRICL (Transdisciplinary Research in Cancer of the Lung) GWAS Consortium, PRACTICAL (PRostate cancer AssoCiation group To Investigate Cancer Associated aLterations) Consortium, PanScan Consortium, GAME-ON/ELLIPSE Consortium. Winner's curse correction and variable Thresholding improve performance of polygenic risk modeling based on genome-wide association study summary-level data. *PLoS Genet* 2016;12:e1006493.

24 Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, Peterson R, Domingue B. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun* 2019;10:3328.

25 Privé F, Aschard H, Blum MGB. Efficient implementation of penalized regression for genetic risk prediction. *Genetics* 2019;212:65–74.

26 So H-C, Sham PC. Improving polygenic risk prediction from summary statistics by an empirical Bayes approach. *Sci Rep* 2017;7:41262.

27 Akter S, Kashino I, Mizoue T, Matsuo K, Ito H, Wakai K, Nagata C, Nakayama T, Sadakane A, Tanaka K, Tamakoshi A, Sugawara Y, Sawada N, Inoue M, Tsugane S, Sasazuki S. Coffee drinking and colorectal cancer risk: an evaluation based on a systematic review and meta-analysis among the Japanese population. *Jpn J Clin Oncol* 2016;46:781–7.

28 Cherlin S, Howey RAJ, Cordell HJ. Using penalized regression to predict phenotype from SNP data. *BMC Proc* 2018;12:38.

29 Wang L, Jia P, Wolfinger RD, Chen X, Grayson BL, Aune TM, Zhao Z. An efficient hierarchical generalized linear mixed model for pathway analysis of genome-wide association studies. *Bioinformatics* 2011;27:686–92.

30 Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, Derks EM. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int J Methods Psychiatr Res* 2018;27:e1608.

31 Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc* 2010;5:1564–73.

32 Khera AV, Chaffin M, Zekavat SM, Collins RL, Roselli C, Natarajan P, Lichtman JH, D'Onofrio G, Mattera J, Dreyer R, Spertus JA, Taylor KD, Psaty BM, Rich SS, Post W, Gupta N, Gabriel S, Lander E, Ida Chen Y-D, Talkowski ME, Rotter JI, Krumholz HM, Kathiresan S. Whole-Genome sequencing to characterize monogenic and polygenic contributions in patients hospitalized with early-onset myocardial infarction. *Circulation* 2019;139:1593–602.

33 Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association statistics. *Nat Rev Genet* 2017;18:117–27.

34 Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3*

35 Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinformatics* 2015;31:782–4.

36 Rüeger S, McDaid A, Kutalik Z. Evaluation and application of summary statistic imputation to discover new height-associated loci. *PLoS Genet* 2018;14:e1007371.

37 Mega JL, Stitziel NO, Smith JG, Chasman DI, Caulfield M, Devlin JJ, Nordio F, Hyde C, Cannon CP, Sacks F, Poulter N, Sever P, Ridker PM, Braunwald E, Melander O, Kathiresan S, Sabatine MS. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *Lancet* 2015;385:2264–71.

38 Yang X, Leslie G, Gentry-Maharaj A, Ryan A, Intermaggio M, Lee A, Kalsi JK, Tyrer J, Gaba F, Manchanda R, Pharoah PDP, Gayther SA, Ramus SJ, Jacobs I, Menon U, Antoniou AC. Evaluation of polygenic risk scores for ovarian cancer risk prediction in a prospective cohort study. *J Med Genet* 2018;55:546–54.

39 Mostafavi H, Harpak A, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. *BioRxiv* 2019;25.

40 Duncan L, Shen H, Gelaye B, Ressler K, Feldman M, Peterson R, Domingue B. Analysis of polygenic score usage and performance across diverse human populations. *bioRxiv* 2018.

41 Kuchenbaecker K, Telkar N, Reiker T, Walters RG, Lin K, Eriksson A, Gurdasani D, Gilly A, Southam L, Tsafantakis E, Karaleftheri M, Seeley J, Kamali A, Asiki G, Millwood IY, Holmes M, Du H, Guo Y, Kumari M, Dedoussis G, Li L, Chen Z, Sandhu MS, Zeggini E, Understanding Society Scientific Group. The transferability of lipid loci across African, Asian and European cohorts. *Nat Commun* 2019;10:4330.

42 Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 2013;9:e1003348.

43 Janssens ACJW, Martens FK. Reflection on modern methods: revisiting the area under the ROC curve. *Int J Epidemiol* 2020. doi:10.1093/ije/dyz274. [Epub ahead of print: 22 Jan 2020].

44 Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, Tyrer JP, Chen T-H, Wang Q, Bolla MK, Yang X, Adank MA, Ahearn T, Aittomäki K, Allen J, Andrulis IL, Anton-Culver H, Antonenkova NN, Arndt V, Aronson KJ, Auer PL, Auvinen P, Barrdahl M, Beane Freeman LE, Beckmann MW, Behrens S, Benitez J, Bermisheva M, Bernstein L, Blomqvist C, Bogdanova NV, Bojesen SE, Bonanni B, Børresen-Dale A-L, Brauch H, Bremer M, Brenner H, Brentnall A, Brock IW, Brooks-Wilson A, Brucker SY, Brüning T, Burwinkel B, Campa D, Carter BD, Castelao JE, Chanock SJ, Chlebowski R, Christiansen H, Clarke CL, Collée JM, Cordina-Duverger E, Cornelissen S, Couch FJ, Cox A, Cross SS, Czene K, Daly MB, Devilee P, Dörk T, Dos-Santos-Silva I, Dumont M, Durcan L, Dwek M, Eccles DM, Ekici AB, Eliassen AH, Ellberg C, Engel C, Eriksson M, Evans DG, Fasching PA, Figueroa J, Fletcher O, Flyger H, Försti A, Fritschi L, Gabrielson M, Gago-Dominguez M, Gapstur SM, García-Sáenz JA, Gaudet MM, Georgoulias V, Giles GG, Gilyazova IR, Glendon G, Goldberg MS, Goldgar DE, González-Neira A, Grenaker Alnæs GI, Grip M, Gronwald J, Grundy A, Guénel P, Haeberle L, Hahnen E, Haiman CA, Håkansson N, Hamann U, Hankinson SE, Harkness EF, Hart SN, He W, Hein A, Heyworth J, Hillemanns P, Hollestelle A, Hooning MJ, Hoover RN, Hopper JL, Howell A, Huang G, Humphreys K, Hunter DJ, Jakimovska M, Jakubowska A, Janni W,

John EM, Johnson N, Jones ME, Jukkola-Vuorinen A, Jung A, Kaaks R, Kaczmarek K, Kataja V, Keeman R, Kerin MJ, Khusnutdinova E, Kiiski JI, Knight JA, Ko Y-D, Kosma V-M, Koutros S, Kristensen VN, Krüger U, Kühl T, Lambrechts D, Le Marchand L, Lee E, Lejbkowicz F, Lilyquist J, Lindblom A, Lindström S, Lissowska J, Lo W-Y, Loibl S, Long J, Lubiński J, Lux MP, MacInnis RJ, Maishman T, Makalic E, Maleva Kostovska I, Mannermaa A, Manoukian S, Margolin S, Martens JWM, Martinez ME, Mavroudis D, McLean C, Meindl A, Menon U, Middha P, Miller N, Moreno F, Mulligan AM, Mulot C, Muñoz-Garzon VM, Neuhausen SL, Nevanlinna H, Neven P, Newman WG, Nielsen SF, Nordestgaard BG, Norman A, Offit K, Olson JE, Olsson H, Orr N, Pankratz VS, Park-Simon T-W, Perez JIA, Pérez-Barrios C, Peterlongo P, Peto J, Pinchev M, Plaseska-Karanfilska D, Polley EC, Prentice R, Presneau N, Prokofyeva D, Purrington K, Pylkäs K, Rack B, Radice P, Rau-Murthy R, Rennert G, Rennert HS, Rhenius V, Robson M, Romero A, Ruddy KJ, Ruebner M, Saloustros E, Sandler DP, Sawyer EJ, Schmidt DF, Schmutzler RK, Schneeweiss A, Schoemaker MJ, Schumacher F, Schürmann P, Schwentner L, Scott C, Scott RJ, Seynaeve C, Shah M, Sherman ME, Shrubsole MJ, Shu X-O, Slager S, Smeets A, Sohn C, Soucy P, Southey MC, Spinelli JJ, Stegmaier C, Stone J, Swerdlow AJ, Tamimi RM, Tapper WJ, Taylor JA, Terry MB, Thöne K, Tollenaar RAEM, Tomlinson I, Truong T, Tzardi M, Ulmer H-U, Untch M, Vachon CM, van Veen EM, Vijai J, Weinberg CR, Wendt C, Whittemore AS, Wildiers H, Willett W, Winqvist R, Wolk A, Yang XR, Yannoukakos D, Zhang Y, Zheng W, Ziogas A, Dunning AM, Thompson DJ, Chenevix-Trench G, Chang-Claude J, Schmidt MK, Hall P, Milne RL, Pharoah PDP, Antoniou AC, Chatterjee N, Kraft P, García-Closas M, Simard J, Easton DF, ABCTB Investigators, kConFab/AOCS Investigators, NBCS Collaborators. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet* 2019;104:21–34.

45 Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, Kathiresan S. Genome-Wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018;50:1219–24.

46 Kerminen S, Martin AR, Koskela J, Ruotsalainen SE, Havulinna AS, Surakka I, Palotie A, Perola M, Salomaa V, Daly MJ, Ripatti S, Pirinen M. Geographic variation and bias in the polygenic scores of complex diseases and traits in Finland. *Am J Hum Genet* 2019;104:1169–81.

47 Kim MS, Patel KP, Teng AK, Berens AJ, Lachance J. Genetic disease risks can be misestimated across global populations. *Genome Biol* 2018;19:179.

48 Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. *Hum Mol Genet* 2019;28:R133–42.

49 Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, Dadaev T, Leongamornlert D, Anokian E, Cieza-Borrella C, Goh C, Brook MN, Sheng X, Fachal L, Dennis J, Tyrer J, Muir K, Lophatananon A, Stevens VL, Gapstur SM, Carter BD, Tangen CM, Goodman PJ, Thompson IM, Batra J, Chambers S, Moya L, Clements J, Horvath L, Tilley W, Risbridger GP, Gronberg H, Aly M, Nordström T, Pharoah P, Pashayan N, Schleutker J, Tammela TLJ, Sipeky C, Auvinen A, Albanes D, Weinstein S, Wolk A, Håkansson N, West CML, Dunning AM, Burnet N, Mucci LA, Giovannucci E, Andriole GL, Cussenot O, Cancel-Tassin G, Koutros S, Beane Freeman LE, Sorensen KD, Orntoft TF, Borre M, Maehle L, Grindedal EM, Neal DE, Donovan JL, Hamdy FC, Martin RM, Travis RC, Key TJ, Hamilton RJ, Fleshner NE, Finelli A, Ingles SA, Stern MC, Rosenstein BS, Kerns SL, Ostrer H, Lu Y-J, Zhang H-W, Feng N, Mao X, Guo X, Wang G, Sun Z, Giles GG, Southey MC, MacInnis RJ, FitzGerald LM, Kibel AS, Drake BF, Vega A, Gómez-Caamaño A, Szulkin R, Eklund M, Kogevinas M, Llorca J, Castaño-Vinyals G, Penney KL, Stampfer M, Park JY, Sellers TA, Lin H-Y, Stanford JL, Cybulski C, Wokolorczyk D, Lubinski J, Ostrander EA, Geybels MS, Nordestgaard BG, Nielsen SF, Weischer M, Bisbjerg R, Røder MA, Iversen P, Brenner H, Cuk K, Holleczek B, Maier C, Luedeke M, Schnoeller T, Kim J, Logothetis CJ, John EM, Teixeira MR, Paulo P, Cardoso M, Neuhausen SL, Steele L, Ding YC, De Ruyck K, De Meerleer G, Ost P, Razack A, Lim J, Teo S-H, Lin DW, Newcomb LF, Lessel D, Gamulin M, Kulis T, Kaneva R, Usmani N, Singhal S, Slavov C, Mitev V, Parliament M, Claessens F, Joniau S, Van den Broeck T, Larkin S, Townsend PA, Aukim-Hastie C, Gago-Dominguez M, Castelao JE, Martinez ME, Roobol MJ, Jenster G, van Schaik RHN, Menegaux F, Truong T, Koudou YA, Xu J, Khaw K-T, Cannon-Albright L, Pandha H, Michael A, Thibodeau SN, McDonnell SK, Schaid DJ, Lindstrom S, Turman C, Ma J, Hunter DJ, Riboli E, Siddiq A, Canzian F, Kolonel LN, Le Marchand L, Hoover RN, Machiela MJ, Cui Z, Kraft P, Amos CI, Conti DV, Easton DF, Wiklund F, Chanock SJ, Henderson BE, Kote-Jarai Z, Haiman CA, Eeles RA. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* 2018;50:928–36.

50 Martens FK, Janssens ACJW. How the intended use of polygenic risk scores guides the design and evaluation of prediction studies. *Current Epidemiology Reports* 2019;6:184–90.

51 Black MH, Li S, LaDuca H, Chen J, Hoiness R, Gutierrez S, Lu H-M, Dolinsky JS, Xu J, Vachon C, Couch F, Tippin Davis B, Davis BT. Polygenic risk score for breast cancer in high-risk women. *JCO* 2018;36:1508–08.

52 Hughes E, Judkins T, Wagner S, Wenstrup RJ, Lanchbury JS, Gutin A. Development and validation of a residual risk score to predict breast cancer risk in unaffected women negative for mutations on a multi-gene hereditary cancer panel. *JCO* 2017;35:1579–79.

53 Wright CF, Kroese M. Evaluation of genetic tests for susceptibility to common complex diseases: why, when and how? *Hum Genet* 2010;127:125–34.

54 Zimmern RL, Kroese M. The evaluation of genetic tests. *J Public Health* 2007;29:246–50.

55 Khoury MJ, Feero WG, Chambers DA, Brody LC, Aziz N, Green RC, Janssens ACJW, Murray MF, Rodriguez LL, Rutter JL, Schully SD, Winn DM, Mensah GA. A collaborative translational research framework for evaluating and implementing the appropriate use of human genome sequencing to improve health. *PLoS Med* 2018;15:e1002631.

56 Allan GM, Garrison S, McCormack J. Comparison of cardiovascular disease risk calculators. *Curr Opin Lipidol* 2014;25:254–65.

57 Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, Babb de Villiers C, Izquierdo A, Simard J, Schmidt MK, Walter FM, Chatterjee N, Garcia-Closas M, Tischkowitz M, Pharoah P, Easton DF, Antoniou AC. Boadicea: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med* 2019.

58 Timmerman N, Haitjema S, de Kleijn DPV, de Borst GJ, Ruijter HMden, Asselbergs FW, Pasterkamp G, van der Laan SW. Family history and polygenic risk of cardiovascular disease are associated with a worse secondary cardiovascular outcome in patients undergoing carotid endarterectomy. *Atherosclerosis* 2019;287:e87.

59 Lakeman IMM, Hilbers FS, Rodríguez-Girondo M, Lee A, Vreeswijk MPG, Hollestelle A, Seynaeve C, Meijers-Heijboer H, Oosterwijk JC, Hoogerbrugge N, Olah E, Vasen HFA, van Asperen CJ, Devilee P. 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* 2019;56:581–9.

60 Dudbridge F, Pashayan N, Yang J. Predictive accuracy of combined genetic and environmental risk scores. *Genet Epidemiol* 2018;42:4–19.

61 Abraham G, Havulinna AS, Bhalala OG, Byars SG, De Livera AM, Yetukuri L, Tikkanen E, Perola M, Schunkert H, Sijbrands EJ, Palotie A, Samani NJ, Salomaa V, Ripatti S, Inouye M. Genomic prediction of coronary heart disease. *Eur Heart J* 2016;37:3267–78.

62 Willoughby A, Andreassen PR, Toland AE. Genetic testing to guide risk-stratified screens for breast cancer. *J Pers Med* 2019;9:22.

63 Mavaddat N, Pharoah PDP, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M, Luben R, Brown J, Bojesen SE, Nordestgaard BG, Nielsen SF, Flyger H, Czene K, Darabi H, Eriksson M, Peto J, Dos-Santos-Silva I, Dudbridge F, Johnson N, Schmidt MK, Broeks A, Verhoef S, Rutgers EJ, Swerdlow A, Ashworth A, Orr N, Schoemaker MJ, Figueroa J, Chanock SJ, Brinton L, Lissowska J, Couch FJ, Olson JE, Vachon C, Pankratz VS, Lambrechts D, Wildiers H, Van Ongeval C, van Limbergen E, Kristensen V, Grenaker Alnæs G, Nord S, Borresen-Dale A-L, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Chang-Claude J, Rudolph A, Seibold P, Flesch-Janys D, Fasching PA, Haeberle L, Ekici AB, Beckmann MW, Burwinkel B, Marme F, Schneeweiss A, Sohn C, Trentham-Dietz A, Newcomb P, Titus L, Egan KM, Hunter DJ, Lindstrom S, Tamimi RM, Kraft P, Rahman N, Turnbull C, Renwick A, Seal S, Li J, Liu J, Humphreys K, Benitez J, Pilar Zamora M, Arias Perez JI, Menéndez P, Jakubowska A, Lubinski J, Jaworska-Bieniek K, Durda K, Bogdanova NV, Antonenkova NN, Dörk T, Anton-Culver H, Neuhausen SL, Ziogas A, Bernstein L, Devilee P, Tollenaar RAEM, Seynaeve C, van Asperen CJ, Cox A, Cross SS, Reed MWR, Khusnutdinova E, Bermisheva M, Prokofyeva D, Takhirova Z, Meindl A, Schmutzler RK, Sutter C, Yang R, Schürmann P, Bremer M, Christiansen H, Park-Simon T-W, Hillemanns P, Guénel P, Truong T, Menegaux F, Sanchez M, Radice P, Peterlongo P, Manoukian S, Pensotti V, Hopper JL, Tsimiklis H, Apicella C, Southey MC, Brauch H, Brüning T, Ko Y-D, Sigurdson AJ, Doody MM, Hamann U, Torres D, Ulmer H-U, Försti A, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Andrulis IL, Knight JA, Glendon G, Marie Mulligan A, Chenevix-Trench G, Balleine R, Giles GG, Milne RL, McLean C, Lindblom A, Margolin S, Haiman CA, Henderson BE, Schumacher F, Le Marchand L, Eilber U, Wang-Gohrke S, Hooning MJ, Hollestelle A, van den Ouweland AMW, Koppert LB, Carpenter J, Clarke C, Scott R, Mannermaa A, Kataja V, Kosma V-M, Hartikainen JM, Brenner H, Arndt V, Stegmaier C, Karina Dieffenbach A, Winqvist R, Pylkäs K, Jukkola-Vuorinen A, Grip M, Offit K, Vijai J, Robson M, Rau-Murthy R, Dwek M, Swann R, Annie Perkins K, Goldberg MS, Labrèche F, Dumont M, Eccles DM, Tapper WJ, Rafiq S, John EM, Whittemore AS, Slager S, Yannoukakos D, Toland AE, Yao S, Zheng W, Halverson SL, González-Neira A, Pita G, Rosario Alonso M, Álvarez N, Herrero D, Tessier DC, Vincent D, Bacot F, Luccarini C, Baynes C, Ahmed S, Maranian M, Healey CS, Simard J, Hall P, Easton DF, Garcia-Closas M. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst* 2015;107. doi:10.1093/jnci/djv036. [Epub ahead of print: 08 Apr 2015].

64 Ware EB, Schmitz LL, Faul JD, Gard A, Mitchell C, Smith JA, Zhao W, Weir D, Kardia SLR. Heterogeneity in polygenic scores for common human traits. *bioRxiv* 2017;13.

65 Lambert SA, Jupp S, Abraham G, Parkinson H, Danesh J, MacArthur JAL, Inouye M. The polygenic score (PGs) catalog: a database of published PGs to enable reproducibility and uniform evaluation, 2019. Available: wwwpgscatalogorg

66 Fritsche LG, Patil S, Beesley LJ, VandeHaar P, Salvatore M, Peng RB, Taliun D, Zhou X, Mukherjee B. Cancer PRSweb – an online Repository with polygenic risk scores (PRS) for major cancer traits and their Phenome-wide exploration in two independent biobanks. *bioRxiv* 2020;49.

67 Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 2019;51:584–91.

68 Wojcik G, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, Highland HM, Patel YM, Sorokin EP, Avery CL. The page study: how genetic diversity improves our understanding of the architecture of complex traits. *bioRxiv* 2018;188094.

69 Rosenberg NA, Edge MD, Pritchard JK, Feldman MW. Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. *Evol Med Public Health* 2019;2019:26–34.

70 Mostafavi H, Harpak A, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. *BioRxiv* 2019;629949.

71 Mills MC, Rahal C. A scientometric review of genome-wide association studies. *Commun Biol* 2019;2.

72 Scherr CL, Aufox S, Ross AA, Ramesh S, Wicklund CA, Smith M. What people want to know about their genes: a critical review of the literature on large-scale genome sequencing studies. *Healthcare* 2018;6. doi:10.3390/healthcare6030096. [Epub ahead of print: 08 Aug 2018].

73 Márquez-Luna C, Loh P-R, Price AL, South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol* 2017;41:811–23.

74 Watanabe K, Stringer S, Frei O, Umićević Mirkov M, de Leeuw C, Polderman TJC, van der Sluis S, Andreassen OA, Neale BM, Posthuma D. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* 2019;51:1339–48.

75 Gamazon ER, Segrè AV, van de Bunt M, Wen X, Xi HS, Hormozdiari F, Ongen H, Konkashbaev A, Derks EM, Aguet F, Quan J, Nicolae DL, Eskin E, Kellis M, Getz G, McCarthy MI, Dermitzakis ET, Cox NJ, Ardlie KG, GTEx Consortium. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet* 2018;50:956–67.

76 Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Blondal T, Stacey SN, Helgason A, Gunnarsdottir S, Olafsdottir A, Kristinsson KT, Birgisdottir B, Ghosh S, Thorlacius S, Magnusdottir D, Stefansdottir G, Kristjansson K, Bagger Y, Wilensky RL, Reilly MP, Morris AD, Kimber CH, Adeyemo A, Chen Y, Zhou J, So W-Y, Tong PCY, Ng MCY, Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Tres A, Fuertes F, Ruiz-Echarri M, Asin L, Saez B, van Boven E, Klaver S, Swinkels DW, Aben KK, Graif T, Cashy J, Suarez BK, van Vierssen Trip O, Frigge ML, Ober C, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Palmer CNA, Rotimi C, Chan JCN, Pedersen O, Sigurdsson G, Benediktsson R, Jonsson E, Einarsson GV, Mayordomo JI, Catalona WJ, Kiemeney LA, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. Two variants on chromosome 17 confer prostate cancer risk, and the one in Tcf2 protects against type 2 diabetes. *Nat Genet* 2007;39:977–83.

77 Burke W, Zimmern RL, Kroese M. Defining purpose: a key step in genetic test evaluation. *Genet Med* 2007;9:675–81.

78 Briggs S, Slade I. Evaluating the integration of genomics into cancer screening programmes: challenges and opportunities. *Curr Genet Med Rep* 2019;7:63–74.

79 Pashayan N, Morris S, Gilbert FJ, Pharoah PDP. Cost-Effectiveness and Benefit-to-Harm ratio of risk-stratified screening for breast cancer: a life-table model. *JAMA Oncol* 2018;4:1504–10.

80 Naber SK, Kundu S, Kuntz KM, Dotson WD, Williams MS, Zauber AG, Calonge N, Zallen DT, Ganiats TG, Webber EM, Goddard KAB, Henrikson NB, van Ballegooijen M, Janssens ACJW, Lansdorp-Vogelaar I. Cost-Effectiveness of risk-stratified colorectal cancer screening based on polygenic risk: current status and future potential. *JNCI Cancer Spectr* 2020;4:pkz086.