Quantitative Biology 2021, 9(2): 133–140
https://doi.org/10.15302/J-QB-021-0238

REVIEW

Polygenic risk scores: effect estimation and model optimization

Zijie Zhao[1], Jie Song[2], Tuo Wang[1], Qiongshi Lu[1,2,3,*]

[1] Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison WI 53726, USA
[2] Department of Statistics, University of Wisconsin-Madison, Madison WI 53726, USA
[3] Center for Demography of Health and Aging, University of Wisconsin-Madison, Madison WI 53726, USA
* Correspondence: qlu@biostat.wisc.edu

Received October 13, 2020; Revised January 12, 2021; Accepted January 13, 2021

***Background*: Polygenic risk score (PRS) derived from summary statistics of genome-wide association studies (GWAS) is a useful tool to infer an individual's genetic risk for health outcomes and has gained increasing popularity in human genetics research. PRS in its simplest form enjoys both computational efficiency and easy accessibility, yet the predictive performance of PRS remains moderate for diseases and traits.**
***Results*: We provide an overview of recent advances in statistical methods to improve PRS's performance by incorporating information from linkage disequilibrium, functional annotation, and pleiotropy. We also introduce model validation methods that fine-tune PRS using GWAS summary statistics.**
***Conclusion*: In this review, we showcase methodological advances and current limitations of PRS, and discuss several emerging issues in risk prediction research.**

Keywords: GWAS; polygenic risk score; summary statistics; model selection

Author summary: The prosperity of powerful genome-wide association studies (GWASs) has facilitated rapid development of polygenic risk score (PRS). Many post-GWAS PRS methods have been introduced to directly address the mediocre prediction accuracy of traditional PRS built upon marginal estimates from GWAS. This review first summarizes PRS methods inspired by different biological concepts including LD, functional annotation, and pleiotropy to better quantify SNP effects. Then we introduce recent PRS frameworks that enable model optimization using summary statistics. Finally, we point out current pitfalls of risk prediction research. We expect emerging methods that address current challenges in the near future.

INTRODUCTION

Genetic prediction of complex disease is a major goal in human genetics research. Accurate stratification of genetic risk requires a quantitative understanding of the genetic architecture underlying the trait of interest. The success of genome-wide association studies (GWAS) in the past decade has shed important light on the etiology of numerous complex diseases. Common single-nucleotide polymorphisms (SNPs) typically have weak to moderate effects on the phenotypic outcome individually. However, when effects of hundreds of thousands of SNPs are aggregated, they explain a substantial proportion of the phenotypic variability [1–3]. By aggregating the risk burden of numerous SNPs, including those that fail to surpass the genome-wide significance threshold, polygenic risk scores (PRSs) have demonstrated to yield greater predictive performance of various diseases and traits. Compared with genetic risk models built upon statistical learning methods that require individual-level genotype and phenotype data [4–6], PRS methods directly model publicly available GWAS summary statistics and enjoy superior computational efficiency and broader applications. Consequently, construction and evaluation of PRS have become a routine follow-up

analysis for many recent GWASs [7–9].

Despite the success, PRS derived from summary level data still only has moderate predictive capability. In particular, simple PRSs based on independent genetic markers and marginal effect estimates may not be sufficient to fully appreciate the genetic architecture of complex traits and provide accurate prediction results [10]. To improve the predictive power of PRSs derived from GWAS summary statistics, it is crucial to select the best set of genetic markers with accurate effect sizes estimation [11]. Despite recent advances in biobankscale GWASs, it remains challenging to precisely identify causal variants and quantify variants' true effects, in part due to the presence of linkage disequilibrium (LD).

In this review, we introduce recent advancements in PRS methodology. We summarize commonly used PRS approaches (Table 1), including simple yet popular methods and recent, more sophisticated models, with a focus on post-GWAS estimation of SNP weights and selection of optimal tuning parameter. We also discuss methods that achieve both objectives simultaneously. Specifically, we first focus on PRS frameworks that integrates GWAS data with LD, functional genome annotation, and pleiotropy information to estimate variant's effect. Then, we introduce recent methods for PRS model fine-tuning. Finally, we discuss the limitation of PRS applications and layout some future directions for PRS research.

THE BASICS OF PRS MODELS

A PRS is a weighted sum of effective allele counts at a set of pre-selected genetic markers.

$$S = \sum_{i \in R} X_i w_i$$

Here, S denotes the PRS; X_i indicates the i^{th} SNP in the dataset; w_i is the weight value assigned to the i^{th} SNP; R is the set of SNPs included in the model. The standard PRS approach uses marginal association coefficients obtained from GWAS as weights and applies arbitrary thresholds on the association strength of genetic variants (*e.g.*, $p < 5e{-}8$ or no threshold at all). It is not always ideal to include all SNPs in the prediction model, especially when GWAS is underpowered and coefficient estimates are noisy [24]. PRSs with stringent p-value cutoffs may outperform other models if few causal variants exist for the phenotype of interest or if the GWAS does not have sufficient statistical power. However, PRSs based on a genome-wide significance threshold have been demonstrated to underperform on polygenic traits [11]. In general, SNPs that are strongly associated with the phenotype should be considered with priority but it remains an empirical challenge to properly select the thresholds in real data applications.

Due to pervasive LD in the genome, SNPs at the same genomic locus may be strongly correlated. Therefore, it is a common approach to prune SNPs in the data so that only independent predictors are included in the model. LD-pruning is an iterative algorithm that removes SNPs having strong correlations with an 'index SNP' in each designated LD block. In this case, the 'index SNP' is randomly chosen. Additionally, a threshold (*e.g.*, for LD strength r^2) needs to be selected to decide whether a pair of SNPs are in LD. A similar but more popular method is LD-clumping (also known as informed LD-pruning). This approach incorporates information from GWAS

Table 1 A list of PRS methods discussed in this review

Name	URL	Year	Ref.
PRSice-2	https://www.prsice.info	2019	Choi and O'Reilly [12]
LDpred	https://github.com/bvilhjal/ldpred	2015	Vilhjálmsson *et al.* [13]
PRS-CS	https://github.com/getian107/PRScs	2019	Ge *et al.* [14]
Annopred	https://github.com/yiminghu/AnnoPred	2017	Hu *et al.* [15]
lassosum	https://github.com/tshmak/lassosum	2017	Mak *et al.* [16]
PANPRS	https://github.com/lsncibb/PANPRS	2020	Chen *et al.* [17]
PleioPred	https://github.com/yiminghu/PleioPred	2017	Hu *et al.* [18]
wMT-SBLUP	https://github.com/uqrmaie1/smtpred	2018	Maier *et al.* [19]
CTPR	https://github.com/wonilchung/CTPR	2019	Chung *et al.* [20]
MTAG	https://github.com/JonJala/mtag	2018	Turley *et al.* [21]
GenomicSEM	https://github.com/MichelNivard/GenomicSEM	2019	Grotzinger *et al.* [22]
SummaryAUC	https://github.com/lsncibb/SummaryAUC	2019	Song *et al.* [23]
PUMAS	https://github.com/qlu-lab/PUMAS	2019	Zhao *et al.* [24]
SBayesR	https://cnsgenomics.com/software/gctb/	2019	Lloyd-Jones *et al.* [25]
sBLUP	https://cnsgenomics.com/software/gcta/#SBLUP	2017	Robinson *et al.* [26]
DBSLMM	https://biostat0903.github.io/DBSLMM/index.html	2020	Yang and Zhou [27]

associations and always selects the most significant marker within an LD range to be the 'index SNP'. Combined with *p*-value thresholding, these approaches are referred to as "pruning + thresholding" and "clumping + thresholding" (P + T and C + T) and can be implemented using software tools such as PLINK [28] and PRSice-2 [12].

A common application of PRS is to predict individual trait values or disease outcomes on an external dataset independent from the training GWAS samples. It is a common practice to regress sample phenotypes on PRS values in the testing data and report R^2 and area under the ROC curve (AUC) to quantify the predictive performance on continuous and dichotomized outcomes. Note that there does not exist a universal "best" subset of SNPs that always achieves the highest prediction accuracy. Depending on the true underlying genetic architecture of the traits being studied and the quality of GWAS data, the LD and *p*-value thresholds need be optimized by comparing PRSs' predictive power on independent samples.

PENALIZED REGRESSION MODELS BASED ON GWAS SUMMARY STATISTICS

PRS models that clump SNPs and use marginal GWAS weights may not achieve optimal prediction accuracy when there are multiple causal SNPs at a single locus or when the causal SNPs are not included in the data. Several methods that incorporate LD into the PRS model have been shown to improve predictive performance.

LDpred is one of the first summary-statistics-based PRS frameworks that explicitly incorporates LD [13]. Following recent advances in polygenic modeling and heritability estimation, LDpred was built upon a random effects model. It uses an empirical Bayesian approach to estimate SNP effects. More specifically, the SNP weights can be denoted as:

$$w = (w_1,...,w_m)^T = \mathbb{E}(\beta|\hat{\beta},V),$$

where β denotes the vector of SNP causal effects on the phenotype; $\hat{\beta}$ is the marginal effect estimates obtained from GWAS summary statistics; V is the LD matrix which is typically estimated from an external reference panel in practice; m is the number of SNPs in GWAS. Combining marginal GWAS coefficients and LD, LDpred re-estimates the effects of all SNPs using the posterior expectation without pruning the SNPs. Of note, this framework has been adopted by multiple PRS methods that were developed later [14,15]. The main difference of these approaches is the choice of prior for β.

LDpred models SNP effects with two types of priors: an infinitesimal prior that assumes equal contribution from all SNPs in the data, *i.e.*,

$$\beta_i \sim N\left(0, \frac{h^2}{m}\right),$$

where h^2 denotes the heritability of the phenotype; or a non-infinitesimal (point-normal) prior that assumes non-zero effects for a proportion of causal SNPs, *i.e.*,

$$\beta_i \sim pN\left(0, \frac{h^2}{mp}\right) + (1-p)\delta_0$$

where δ_0 is a point mass at 0 and p is the proportion of causal SNPs. Here, p is treated as a tuning parameter and needs to be selected using an independent validation dataset. LDpred enjoys the flexibility of being able to handle both sparse and polygenic genetic architecture. Under a non-infinitesimal model, LDpred uses a Gibbs sampler approach to estimate the posterior expectation $\mathbb{E}(\beta|\hat{\beta},V)$ in each LD block. Under the infinitesimal model where all SNPs are considered as causal, the posterior expectation has a closed form solution.

Similar in terms of the choice of post-GWAS estimator for SNP effects, PRS-CS is a recently developed Bayesian high-dimensional PRS framework that employs a continuous shrinkage prior on β [14]. Unlike the normal or point-normal priors implemented in LDpred, the prior in PRS-CS is a continuous distribution that is jointly determined by a global scaling parameter and a marker-specific local shrinkage parameter.

$$\beta_i \sim N\left(0, \frac{\sigma^2}{N}\phi\psi_i\right), \ \ \psi_i \sim g,$$

where ϕ is the global shrinkage parameter; ψ_i is the SNP-specific local shrinkage factor; σ^2 is the residual variance; N is the sample size. While the global scaling factor represents the uniform shrinkage applied to all markers, the local shrinkage parameter is unique for each variant and can be drawn from an appropriate absolute continuous distribution (*e.g.*, a gamma-gamma distribution):

$$\psi_i \sim G(a,\delta_i), \ \ \delta_i \sim G(b,1),$$

where $G(a,\beta)$ is the gamma distribution and δ_i is the unknown scale parameter of the distribution for the local shrinkage factor ψ_i at each SNP. Such a prior design facilitates the algorithm's capability of shrinking noisy estimates towards zero while maintaining the large effects of SNPs demonstrating stronger signals, therefore making PRS-CS adaptive for diverse genetic architecture. The optimal global shrinkage factor can be obtained either through external validation on an independent dataset, or estimated through a full Bayesian approach that assigns a half-Cauchy prior on the global scaling parameter.

We note that frequentist methods have also been developed to fit penalized regression models on GWAS

summary statistics. These methods (*e.g.*, lassosum [16] and PANPRS [17]) provide a different perspective on re-estimating SNP effects using marginal association results as inputs. A traditional penalized regression model (*e.g.*, lasso) requires both the phenotype and genotype data measured on each individual to optimize the objective function and estimate regression coefficients. However, it can be shown that the penalized loss function can be iteratively optimized using the inner product of genotypes and the inner product of genotypes and phenotypes:

$$f(\beta) = y^T y + \beta^T X^T X \beta - 2\beta^T X^T y + 2\lambda||\beta||_1^1,$$

where X and y are the standardized genotype and phenotype and λ is the tuning parameter. Notably, the inner product of genotypes can be obtained from the LD correlation matrix and the inner product of genotypes and phenotypes can be obtained from marginal association results. Regarding the selection of the tuning parameter λ that controls the penalty strength, lassosum implemented a "pseudo-validation" approach that replaced validation genotypes and phenotypes in PRS-phenotype correlation with terms that can be obtained from GWAS summary statistics:

$$Corr(PGS(\lambda), y_t) = \frac{\beta_\lambda^T X_t^T P y_t}{\sqrt{\beta_\lambda^T X_t^T P X_t \beta_\lambda y_t^T P y_t}},$$

where X_t and y_t are the genotype and phenotype of the testing samples, and $P = I - \frac{11^T}{n}$ is the mean-centering matrix. Built upon a similar idea, PANPRS is a new method that also trains penalized regression using GWAS summary statistics. Compared with lassosum, PANPRS can model not only quantitative but also binary traits and is capable of integrating information from external annotation information.

INCORPORATING FUNCTIONAL ANNOTATIONS IN PRS MODELS

Another major limitation of genetic prediction methods is the lack of biological interpretation. A plethora of transcriptomic and epigenomic annotation data have been generated and made available by large consortia including GTEx [29], ENCODE [30], and Roadmap Epigenomics Project [31]. Integrating these data with GWAS has provided insights into functional variant fine-mapping, heritability enrichment, and risk gene prioritization [32]. Methods have also been developed to incorporate functional annotation data in PRS models.

Most Bayesian shrinkage PRS methods use noninformative priors for SNP effect sizes. Leveraging advances in statistical methods to partition heritability by functional annotation [33,34], AnnoPred [15] uses an empirically estimated informative prior to prioritize SNP predictors in PRS:

$$\beta_i \sim N(0, \sigma_i) \text{ or } \beta_i \sim pN\left(0, \frac{\sigma_i}{p}\right) + (1-p)\delta_0.$$

Here, σ_i is the per-SNP heritability estimate in an annotation-stratified heritability model. That is,

$$\sigma_i = \sum_{j: i \in A_j} \tau_j,$$

where A_j denotes the j^{th} annotation category in the analysis and τ_j is the corresponding regression coefficient from stratified LD score regression which quantifies per-SNP heritability for variants in the j^{th} annotation [33]. Similar to LDpred, AnnoPred also uses the posterior expectation of SNP effects, *i.e.*, $\mathbb{E}(\beta|\hat{\beta}, V)$, as weights in PRS. This approach appreciates the trait-specific genetic architecture by empirically and adaptively prioritizing functional SNPs with greater impacts. We also note that functional annotations have been similarly incorporated into the penalty terms of frequentist approaches such as PANPRS [17]. Using continuous traits as an example, when modeling functional annotation, PANPRS calculates the penalized regression coefficients by

$$\beta = \text{argmin}_\beta \frac{1}{2n} \sum_{j=1}^{n} (y_j - \sum_{i=1}^{M} X_{ji}\beta_i)^2$$

$$+ \sum_{i=1}^{M} (\lambda_0 + \sum_{s=1}^{r} \lambda_s R_{is}) |\beta_i|,$$

where λ_0 is the baseline penalty, λ_s is the penalty for the s^{th} annotation category and R_{is} is a binary variable that equals 1 when the i^{th} SNP is not annotated in the s^{th} annotation category and 0 otherwise. In this way, SNPs that are enriched in more categories of functional annotation receive less penalty and are therefore weighted more in the final PRS.

INTEGRATING MULTIPLE PHENOTYPES

Genetic correlation analysis and pleiotropic association mapping have revealed concordant genetic associations across many traits [35–37]. The same genetic variant may yield correlated effects to several diseases or traits. Aggregating association results from multiple genetically correlated GWASs may increase the effective sample size, improve SNP effects estimation, and enhance the prediction accuracy of PRS. We introduce several methods that jointly model summary statistics from multiple GWASs in genetic risk prediction.

PleioPred [18] is a multi-trait extension of both LDpred [13] and AnnoPred [15]. PleioPred models each SNP's effects on two different traits with a bivariate normal

distribution, *i.e.,*

$$\begin{pmatrix}\beta_i\\ \gamma_i\end{pmatrix}\sim N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\sigma_{1i}^2 & \rho_g\sigma_{1i}\sigma_{2i}\\ \rho_g\sigma_{1i}\sigma_{2i} & \sigma_{2i}^2\end{pmatrix}\right),$$

where β_i and γ_i denote the i^{th} SNP's effects on two traits; σ_{1i}^2 and σ_{2i}^2 are the per-SNP heritability for two traits; ρ_g is the genetic correlation. Conditioning on marginal association statistics and an external LD reference panel, the method provides an empirical Bayesian estimator of SNP effects for both traits.

$$w=\begin{pmatrix}w^{(1)}\\ w^{(2)}\end{pmatrix}=\mathbb{E}\left(\begin{pmatrix}\beta\\ \gamma\end{pmatrix}|\hat{\beta},\hat{\gamma},V\right).$$

The PRSs are constructed as $Xw^{(1)}$ and $Xw^{(2)}$, respectively. Like Annopred, PleioPred can be generalized to non-infinitesimal models and it upweights the effects of SNPs in annotation regions with strong heritability enrichment. Under the infinitesimal assumption, genetic correlation can be estimated using existing methods [36,37] and SNP effect has a closed form solution. A Gibbs sampler approach is used to estimate effects in a non-infinitesimal model.

Since PleioPred, multiple methods based on similar ideas have been developed (*e.g.*, wMT-SBLUP [19] and CTPR [20]), showing various degrees of improvement. MTAG is another approach designed for multi-trait GWAS meta-analysis [21]. For a given SNP, MTAG approaches its true effects on a number of traits by the random effects model:

$$E(\boldsymbol{\beta}_i)=\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix},\mathrm{Var}(\boldsymbol{\beta}_i)=\begin{pmatrix}\omega_{11} & \cdots & \omega_{1T}\\ \vdots & \ddots & \vdots\\ \omega_{T1} & \cdots & \omega_{TT}\end{pmatrix},$$

where $\boldsymbol{\beta}_i$ are the random effects of i^{th} SNP on T traits. Then, MTAG can solve for the generalized method of moment estimator by:

$$E\left(\hat{\boldsymbol{\beta}}_i-\frac{\omega_t}{\omega_{tt}}\beta_{i,t}\right)=\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix}$$

where ω_t and ω_{tt} are the t^{th} column and diagonal element of the covariance matrix of $\boldsymbol{\beta}_i$. Although genetic prediction is not its primary purpose, effect sizes estimated by MTAG can be used to construct PRS. Due to improved precision in effect estimates, these PRSs generally outperform single-trait PRS approaches. Finally, a recent approach, GenomicSEM [22], leverages the pair-wise genetic correlations to quantify the genetic basis of the underlying psychopathological factor shared by multiple psychiatric traits. Borrowing information from multiple GWAS, the PRS for the latent factor outperforms predictive models constructed using each GWAS.

MODEL TUNING

Most existing PRS methods include tuning parameters. Examples include the *p*-value and LD thresholds in a C + T PRS, the proportion of causal variants in LDpred, and the penalty strength in penalized regressions. These parameters add flexibility to the models. When properly selected, they effectively improve the predictive performance of PRS. When individual-level data are available, it is straightforward to select the optimal tuning parameters through cross-validation [38]. For a k-fold cross-validation, we equally partition the entire dataset into k folds. Each time, we fix the tuning parameter values and hold out one fold of data as the validation set to evaluate PRS performance. We use the remaining $(k-1)$ folds as the training set to conduct GWAS and estimate SNP weights. After repeating this procedure k times, we obtain the optimal PRS model by comparing the average predictive performance for different values of tuning parameters.

Despite its broad applications, cross-validation is almost never used to fine-tune PRS models in practice since individual-level data are rarely available for the full GWAS sample. Instead, a standard approach is to use summary statistics from a large GWAS to train PRS models based on different tuning parameter values, and use an external dataset independent from the GWAS to select tuning parameters that yield the best performance. However, in practice, most datasets with sufficient and easily accessible samples will most likely have been included in the large GWAS. Even if such a dataset is available, researchers may prefer to use it as the testing dataset to report the optimized PRS's predictive performance, rather than holding it out as a validation dataset for model selection. To address this challenge, SummaryAUC is a new approach designed for assessing PRS's prediction accuracy using summary-level data as the validation set [23]. For a case-control GWAS, SummaryAUC estimates the area under the ROC curve (AUC) by calculating the probability that a randomly selected case has a higher PRS than a randomly selected control. When individual-level PRS can be calculated, this is simply a two-sample Z-test. When individual-level validation samples are not available, SummaryAUC can approximate the standardized Z score using summary statistics and minor allele frequencies from a GWAS. This method brings an advancement to the field such that we no longer require individual-level validation data to evaluate the performance of PRSs on binary traits.

However, SummaryAUC does not fully resolve challenges in model tuning. In many cases, even an

independent summary statistics dataset may not be available. PUMAS is a new method that can fine-tune PRS models using only GWAS summary statistics [24]. This method uses a resampling approach to create down-sampled training and validation GWAS summary statistics from the input GWAS, and then applies a procedure similar to cross-validation to fine-tune PRS models. PUMAS provides an alternative to approximate the predictive R2 of regressing phenotypes of the validation dataset on PRS without accessing individual-level genetic and phenotypic data. PUMAS can be applied to not only traditional PRSs where the tuning parameter is simply association p-value cutoffs, but also more sophisticated PRS frameworks including LDpred that uses GWAS summary statistics as input. With this approach, it has become possible to systematically benchmark and optimize PRS models for various traits using publicly available GWAS summary data.

PRS COMPUTATION IN PRACTICE

There are other summary-data-based PRS methods that we did not cover in this review (*e.g.*, SBayesR [25], sBLUP [26], and DBSLMM [27]). These PRS frameworks generalize regression models that typically require access to individual-level genotype and phenotype data to be able to use GWAS summary statistics as inputs and construct PRS henceforth. The abundance of PRS methods requires researchers to not only optimize PRS within a single model framework but also benchmark the most predictive PRS model across different phenotypes. A recent study has conducted a systematic comparison between the most predictive PRSs of 15 PRS methods on 25 disease phenotypes from the UK Biobank cohort and found that complicated PRS methods do not necessarily outperform simple PRS models in terms of AUC [39]. Besides prediction accuracy, computational complexity is another important aspect of PRS application in practice. Compared with more sophisticated methods such as LDpred or lassosum, C + T is the most scalable PRS method for biobank-scale datasets and requires less computational time and memory space [27]. In practice, researchers may need to consider the tradeoff between predictive performance and computational burden when selecting PRS algorithms [4].

DISCUSSION

PRS methods and applications have reached prosperity in recent years, largely owing to the fast-growing sample sizes in biobank cohorts and widely available GWAS summary statistics. These methods provide a practical alternative to build genetic prediction models when individual-level data cannot be directly accessed. In general, PRS methods based on GWAS summary statistics often have substantially lower computational burden compared with models that need to be trained on individual-level data; meanwhile, as the GWAS sample size continues to grow, PRS methods have also achieved comparable prediction accuracy for some traits. In this review, we have discussed recent advances in PRS approaches with a focus on methods that estimate SNP effects and optimize tuning parameters. These methods have laid the ground for developing accurate and robust genetic prediction models for a variety of diseases and traits and may have broad applications in disease diagnosis and precision medicine.

However, existing PRS methods still have limitations. First, the sensitivity and specificity of PRS still remain too low to become immediately useful in clinical intervention for most diseases. Despite the methodological advances, improvement in prediction accuracy is usually incremental for most PRS approaches compared with a simple PRS based on GWAS effects. However, a recent study has convincingly demonstrated that PRS can be used to identify individuals with substantially elevated risk for coronary artery diseases and a few other phenotypes despite the low AUC [40], which hints at a need to develop better metrics to quantify the performance of PRS. Second, even if a PRS is predictive, its effect may be mediated through environment and needs to be interpreted with caution. Since a person's genotypes are correlated with other family members' genotypes (*e.g.*, parents), which are subsequently correlated with the family environment, effect estimates obtained from a GWAS are mixtures of both direct and indirect genetic effects [41,42]. A recent study has pointed out that the predictive performance of PRS substantially reduced for many traits when GWAS was conducted on sibling pairs which controls for the shared environment [43]. Blindly trusting the PRS without understanding the underlying etiology may lead to bias and misinterpretation of results in PRS applications. Another major limitation of current PRS methods is the lack of portability. It has been extensively discussed in multiple studies that existing PRSs cannot accurately predict the disease risk of individuals from a population that is different from the GWAS cohorts, possibly due to differences in LD patterns, causal effect sizes, allele frequencies, and environmental mediators across populations [44–47]. This has become a major hurdle in PRS application especially because major GWAS cohorts lack diversity [48] – over 70% of GWAS samples came from three countries: the United States, the United Kingdom, and Iceland. While the number of participants with European ancestry increases exponentially over the years, the proportion of other ethnicities in GWAS samples has declined since 2014 [44]. Methods have been developed

to incorporate GWAS data from multiple populations to improve transethnic portability of predictive performance [49], but few approaches could achieve improvements using summary statistics alone [50]. Overcoming the poor portability of PRS will greatly benefit risk prediction research. We anticipate emerging PRS methodologies in the near future that offer novel insights and solutions to these challenges.

ACKNOWLEDGEMENTS

We acknowledge research support from the University of Wisconsin-Madison Office of the Chancellor and the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Zijie Zhao, Jie Song, Tuo Wang, and Qiongshi Lu declare that they have no conflict of interests or financial conflicts to disclose.

This article is a review article and does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., Sklar, P., and the International Schizophrenia Consortium. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature, 460, 748–752
2. Wray, N. R., Goddard, M. E. and Visscher, P. M. (2007) Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res., 17, 1520–1528
3. Bush, W. S., Sawcer, S. J., de Jager, P. L., Oksenberg, J. R., McCauley, J. L., Pericak-Vance, M. A., Haines, J. L., and the International Multiple Sclerosis Genetics Consortium (IMSGC). (2010) Evidence for polygenic susceptibility to multiple sclerosis−the shape of things to come. Am. J. Hum. Genet., 86, 621–625
4. Zhou, X., Carbonetto, P. and Stephens, M. (2013) Polygenic modeling with Bayesian sparse linear mixed models. PLoS Genet., 9, e1003264
5. Maier, R., Moser, G., Chen, G. B., Ripke, S., Coryell, W., Potash, J. B., Scheftner, W. A., Shi, J., Weissman, M. M., Hultman, C. M., *et al.* (2015) Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. Am. J. Hum. Genet., 96, 283–294
6. Speed, D. and Balding, D. J. (2014) MultiBLUP: improved SNP-based prediction for complex traits. Genome Res., 24, 1550–1557
7. Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., *et al.* (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat. Genet., 50, 1112–1121
8. Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., Frayling, T. M., Hirschhorn, J., Yang, J. and Visscher, P. M., *et al.* (2018) Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum. Mol. Genet., 27, 3641–3649
9. Warrington, N. M., Beaumont, R. N., Horikoshi, M., Day, F. R., Helgeland, Ø., Laurin, C., Bacelis, J., Peng, S., Hao, K., Feenstra, B., *et al.* (2019) Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. Nat. Genet., 51, 804–814
10. Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E. and Visscher, P. M. (2013) Pitfalls of predicting complex traits from SNPs. Nat. Rev. Genet., 14, 507–515
11. Chatterjee, N., Shi, J. and García-Closas, M. (2016) Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat. Rev. Genet., 17, 392–406
12. Choi, S. W. and O'Reilly, P. F. (2019) PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience, 8, giz082
13. Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., Genovese, G., Loh, P. R., Bhatia, G., Do, R., *et al.* (2015) Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am. J. Hum. Genet., 97, 576–592
14. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. and Smoller, J. W. (2019) Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat. Commun., 10, 1776
15. Hu, Y., Lu, Q., Powles, R., Yao, X., Yang, C., Fang, F., Xu, X. and Zhao, H. (2017) Leveraging functional annotations in genetic risk prediction for human complex diseases. PLOS Comput. Biol., 13, e1005589
16. Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X. and Sham, P. C. (2017) Polygenic scores via penalized regression on summary statistics. Genet. Epidemiol., 41, 469–480
17. Chen, T.-H., Chatterjee, N., Landi, M. T. and Shi, J. (2020) A penalized regression framework for building polygenic risk models based on summary statistics from genome-wide association studies and incorporating external information. J. Am. Stat. Assoc., 116, 133–143
18. Hu, Y., Lu, Q., Liu, W., Zhang, Y., Li, M. and Zhao, H. (2017) Joint modeling of genetically correlated diseases and functional annotations increases accuracy of polygenic risk prediction. PLoS Genet., 13, e1006836
19. Maier, R. M., Zhu, Z., Lee, S. H., Trzaskowski, M., Ruderfer, D. M., Stahl, E. A., Ripke, S., Wray, N. R., Yang, J., Visscher, P. M., *et al.* (2018) Improving genetic prediction by leveraging genetic correlations among human diseases and traits. Nat. Commun. 9, 989
20. Chung, W., Chen, J., Turman, C., Lindstrom, S., Zhu, Z., Loh, P.-R., Kraft, P. and Liang, L. (2019) Efficient cross-trait penalized regression increases prediction accuracy in large cohorts using secondary phenotypes. Nat. Commun. 10, 569
21. Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., Nguyen-Viet, T. A., Wedow, R., Zacher, M., Furlotte, N. A., *et al.* (2018) Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat. Genet., 50, 229–237
22. Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S.J.,

Mallard, T.T., Hill, W.D., Ip, H. F., Marioni, R. E., McIntosh, A. M., Deary, I. J., *et al.* (2019) Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. Nat. Hum. Behav. 3, 513–525

23. Song, L., Liu, A., Shi, J., Gejman, P. V., Sanders, A. R., Duan, J., Cloninger, C. R., Svrakic, D. M., Buccola, N. G., Levinson, D. F., *et al.* (2019) SummaryAUC: a tool for evaluating the performance of polygenic risk prediction models in validation datasets with only summary level statistics. Bioinformatics, 35, 4038–4044
24. Zhao, Z., Yi, Y., Wu, Y., Zhong, X., Lin, Y., Hohman, T. J., Fletcher, J. (2019) Fine-tuning polygenic risk scores with GWAS summary statistics. bioRxiv, doi: https://doi.org/10.1101/810713
25. Lloyd-Jones, L.R., Zeng, J., Sidorenko, J., Yengo, L., Moser, G., Kemper, K.E.Wang, H., Zheng, Z., Magi, R., Esko, T., *et al.* (2019) Improved polygenic prediction by Bayesian multiple regression on summary statistics. Nat. Commun., 10, 5086
26. Robinson, M.R., Kleinman, A., Graff, M., Vinkhuyzen, A.A.E., Couper, D., Miller, M.B., Peyrot, W. J., Abdellaoui, A., Zietsch, B. P., Nolte, I. M., *et al.* (2017) Genetic evidence of assortative mating in humans. Nat. Hum. Behav., 1, 0016
27. Yang, S. and Zhou, X. (2020) Accurate and scalable construction of polygenic scores in large biobank data sets. Am. J. Hum. Genet., 106, 679–693
28. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet., 81, 559–575
29. GTEx Consortium (2017) Genetic effects on gene expression across human tissues. Nature, 550, 204–213
30. The ENCODE Project Consortium (2020) Perspectives on ENCODE. Nature, 583, 693–698
31. Roadmap Epigenomics Consortium (2015)Integrative analysis of 111 reference human epigenomes. Nature, 518, 317–330
32. Pasaniuc, B. and Price, A. L. (2017) Dissecting the genetics of complex traits using summary association statistics. Nat. Rev. Genet., 18, 117–127
33. Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., *et al.* (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet., 47, 1228–1235
34. Lu, Q., Powles, R. L., Wang, Q., He, B. J. and Zhao, H. (2016) Integrative tissue-specific functional annotations in the human genome provide novel insights on many complex traits and improve signal prioritization in genome wide association studies. PLoS Genet., 12, e1005947
35. Yang, C., Li, C., Wang, Q., Chung, D. and Zhao, H. (2015) Implications of pleiotropy: challenges and opportunities for mining Big Data in biomedicine. Front. Genet., 6, 229
36. Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., Duncan, L., Perry, J. R., Patterson, N., Robinson, E. B., *et al.* (2015) An atlas of genetic correlations across human diseases and traits. Nat. Genet., 47, 1236–1241
37. Lu, Q., Li, B., Ou, D., Erlendsdottir, M., Powles, R. L., Jiang, T., Hu, Y., Chang, D., Jin, C., Dai, W., *et al.* (2017) A powerful approach to estimating annotation-stratified genetic covariance via GWAS summary statistics. Am. J. Hum. Genet., 101, 939–964
38. Zhang, P. (1993) Model selection via multifold cross validation. Ann. Stat., 21, 299–313
39. Kulm, S., Marderstein, A., Mezey, J. and Elemento, O. (2020) Benchmarking the accuracy of polygenic risk scores and their generative methods. medRxiv, 2020.04.06.20055574
40. Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T., *et al.* (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat. Genet., 50, 1219–1224
41. Wu, Y., Zhong, X., Lin, Y., Zhao, Z., Chen, J., Zheng, B., Li, J. J., Fletcher, J. M. and Lu, Q. (2020) Estimating genetic nurture with summary statistics of multi-generational genome-wide association studies. bioRxiv, 2020.10.06.328724
42. Young, A. I., Benonisdottir, S., Przeworski, M. and Kong, A. (2019) Deconstructing the sources of genotype-phenotype associations in humans. Science, 365, 1396–1400
43. Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K. and Przeworski, M. (2020) Variable prediction accuracy of polygenic scores within an ancestry group. eLife, 9, e48376
44. Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M. and Daly, M. J. (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. Nat. Genet., 51, 584–591
45. Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., Daly, M. J., Bustamante, C. D. and Kenny, E. E. (2017) Human demographic history impacts genetic risk prediction across diverse populations. Am. J. Hum. Genet., 100, 635–649
46. Rosenberg, N. A., Edge, M. D., Pritchard, J. K. and Feldman, M. W. (2019) Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. Evol. Med. Public Health, 2019, 26–34
47. Adhikari, K., Mendoza-Revilla, J., Sohail, A., Fuentes-Guajardo, M., Lampert, J., Chacón-Duque, J. C., Hurtado, M., Villegas, V., Granja, V., Acuña-Alonzo, V., *et al.* (2019) A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. Nat. Commun., 10, 358
48. Mills, M.C. and Rahal, C. (2019) A scientometric review of genome-wide association studies. Commun. Biol., 2, 9
49. Coram, M. A., Fang, H., Candille, S. I., Assimes, T. L. and Tang, H. (2017) Leveraging multi-ethnic evidence for risk assessment of quantitative traits in minority populations. Am. J. Hum. Genet., 101, 218–226
50. Amariuta, T., Ishigaki, K., Sugishita, H., Ohta, T., Matsuda, K., Murakami, Y., Price, A. L., Kawakami, E.,Terao, C. and Raychaudhuri, S. (2020) In silico integration of thousands of epigenetic datasets into 707 cell type regulatory annotations improves the trans-ethnic portability of polygenic risk scores. bioRxiv, 2020.02.21.959510