SMART Topic

Polygenic Risk Scores (PRS): Development, Interpretation and Utility

Phil J Greer, MS[1] and Talia DeFrancesco Oranburg, PhD[1]

1 *Ariel Precision Medicine, Pittsburgh, Pennsylvania, USA.*

1. Overview

Polygenic Risk Scores (PRS) use dozens to hundreds of genetic variants from disease populations to add insights to an individual's risk of that disease. Many PRSs are undergoing clinical trials to assess their ability to enhance clinical risk assessments, making the assessments more personal and accurate while enabling earlier implementation of disease mitigation strategies.[3-7] This paper focuses on the development, utility, interpretation, and limitations of PRS in clinical practice. A PRS is the weighted sum of disease risk contributed by multiple, common, single nucleotide variants (SNVs) across the genome. PRSs are developed using output from one or more genome-wide association studies (GWAS) and are effective for both binary and continuous traits. The accuracy and utility of a PRS depend on the SNVs used in the GWAS, the selection of SNVs for the PRS, the population tested, the characteristics of the disorder, and the accuracy of phenotyping. Interpretation must be done in the context of overall patient risk, such as age, sex, ancestry, past medical history, family history, lifestyle, and biomarkers. Better tools are needed for translation of PRS into the clinical practice of healthcare professionals.

2. Key points about PRS

- PRS are most useful for disorders that are complex (e.g., involving multiple cell types, organs, or systems with variable etiologies, trajectories, complications, or responses to different treatments).
- PRS are key to precision medicine, where the combination of specific genetic risks and biomarkers helps make a more exact, mechanistic diagnosis of syndromes characterized by insensitive and non-specific early signs and symptoms.
- Over 5,500 PRS for more than 800 human phenotypes and disorders have been published (PGS Catalog).[2]
- Most PRS are developed with subjects of European ancestry, but new, larger, multi-ancestry GWAS studies are facilitating broader applications.
- The optimal number of SNVs in a PRS varies widely depending on the specificity of the phenotype, the complexity of the disease, the burden of common variants and the criteria used for selecting specific SNVs in the score.
- Large PRS for disorders such as diabetes (over 1000 SNVs) can be subdivided into mechanisms to help understand the specific cells and mechanisms leading to the phenotype in an individual patient.[8]
- Very large PRS panels (e.g., >1,000,000 variants) may be less useful due to inclusion of rare SNVs, overfitting, and limited applicability.
- At a population level, PRS are useful for public health planning, identifying high-risk groups, and prioritization for surveillance programs.[9]
- For individualized care, PRS can enhance disease risk predictions, diagnostic refinement from the differential diagnosis of early signs and symptoms, prediction of progression and recurrence of disease, and for precision therapeutics.[9]

3. Introduction

Most diseases are influenced by the underlying genetics of the individual. These diseases are classified as monogenic (due to variations in one gene, aka Mendelian genetics) and polygenic (due to the contribution of variants in multiple genes).

Monogenic diseases and their associated genes like Cystic Fibrosis [*CFTR*], Huntington's disease [*HTT*], and familial hypercholesterolemia [*LDLR*] are rare, inherited disorders with prevalence of less than 1%. Polygenic disorders, such as type 2 diabetes mellitus (T2DM) and cardiovascular disease (CVD), are complex with multiple etiologies, sporadic-appearing penetrance, and they are far more common - with prevalence >5%. Monogenic diseases occur when a mutation within a gene has an effect size sufficient to cause the disease. Differences in the effect size of causal variants contribute to variation in penetrance across monogenic diseases.".

Abbreviations used in this paper: PRS, Polygenic Risk Scores; SNVs, single nucleotide variants; GWAS, genome-wide association studies; CFTR; Cystic Fibrosis; HTT, Huntington's disease; LDLR, familial hypercholesterolemia; T2DM, Type 2 Diabetes Mellitus; CVD, Cardiovascular disease; PV, pathogenic Variant; RV, risk variants; LD, linkage disequilibrium; AUROC, Area Under the Receiver Operating Characteristic curve; PDAC, pancreatic ductal adenocarcinoma; CD, Celiac disease; PGx, Pharmacogenetics.

ISSN 2997-2876 (online)
ISSN 2997-2868 (print)
DOI: https://doi.org/10.69734/sqe1mp91
Website: www.SMART-MD.org

Disease penetrance is the proportion of individuals with a particular *genotype* who express the associated *phenotype*. High penetrance diseases are those with up to 100% of carriers developing the phenotype, such as Huntingdon disease. Incomplete penetrance means that there are asymptomatic carriers of a pathogenic genetic variant (PV). The reasons for incomplete penetrance include the fact that genes do not act in isolation and may require other genetic or environmental conditions that may develop over time (e.g. breast cancer with *BRCA2* PVs) or because of protective genes.

Polygenic diseases develop through the effects of multiple genetic variants across many genes or their regulatory elements, with each variant having a very small, independent effect, that contribute additive effects on the overall risk of disease when occurring together in an individual. Variants that contribute to specific diseases but are not necessary nor sufficient to cause the disease alone are called risk variants (RV). Importantly, researchers have shown that there are also polygenic modifiers to monogenic disease affecting both penetrance[15] and age of onset[16-18] Detailed methods beyond assessing presence of disease-causing gene mutations are needed to identify specific risk of disease in individuals carrying such mutations.

4. Generation of Polygenic Risk Scores

How is a PRS generated?

Polygenic Risk Scores are calculated as the weighted sum of the disease risk contributed by multiple, common, disease-associated SNVs across the genome. The associated SNVs and their weights are collected from prior GWAS. Prior to the widespread use of GWAS, genetic risk scores (GRS) were based on a selective small set of known SNVs and consideration of family history. The terms GRS and PRS have since merged with PRS becoming the dominant term. The SNVs that make up a PRS can be manually curated from prior GWAS (i.e. the GWAS Catalog) or they can be created directly from GWAS results (i.e. a PRS that includes all of the significantly associated SNVs from a GWAS on T2DM). The optimal number of SNVs included in a PRS will be disease dependent with trade-offs between using relatively few, well-characterized SNVs with precise effect size estimates and large collections of SNVs with weaker or noisier effect size estimates. Regardless of the provenance of SNVs included in the PRS, care must be taken to ensure that the included SNVs are not in linkage disequilibrium (LD) with one another.

For SNVs to be in LD means that they are on the same chromosome and close to each other on the same DNA strand or allele. In this case they are in *cis*, and the presence of one variant predicts the other variant since they are physically "linked". A combination of two or more linked SNVs is a haplotype.

Including multiple SNVs that are in LD will lead to an overrepresentation of risk at that locus, an inflated PRS, and an overestimation of the genetic risk for disease. For developing new PRSs, see the PGS Catalog Data submission website (PGS Submission) The interpretation and utility of various PRSs is discussed below.

What does a PRS measure?

A PRS is a measurement of genetic similarity of an individual based on a set of SNVs associated with a phenotype compared to controls. This can be either a discrete phenotype (diabetes, yes/no) or a continuous phenotype (HDL cholesterol level). The PRS represents a non-modifiable risk factor based on current available knowledge and allows for classification of individuals into similar risk ranges/rankings. Some of the variants included in a PRS are not directly causal to the phenotype, but instead are indirectly associated due to a gene x gene or gene x environment interaction that cannot be easily measured. An example of this would be variants associated with T2DM through an increase in obesity due to a specific diet/lifestyle. An indirectly associated variant would have a weaker effect on the phenotype (e.g. linked to a type of obesity), but would be more likely to be found significant in large scale GWAS results where the individuals share the risk variant and similar dietary habits. Most PRS explain between 5% and 30% of the variance in a phenotype and are often independent of reported family history.[19] This independence may be due to the sporadic nature of complex diseases, the frequency of high-risk variants in the population, shared environmental exposures, or incomplete capture of familial history.

The Area Under the Receiver Operating Characteristic curve (AUROC, or AUC) is a statistical measure generated to quantify the performance of a PRS as a classifier of individuals with and without a disorder. The score ranges from 0 to 1 where 0.5 indicates no discriminative power and 1.0 indicates a perfect classifier. As seen in Table 1, AUROC values typically range between 0.6 and 0.8, indicating that the PRS analysis helps in determining the likelihood that an individual has or will develop a specific disease, given the contribution of genetics to the disease and population that was analyzed in developing the PRS. Note that in the pancreatic ductal adenocarcinoma (PDAC) PRS, the AUROC is higher (more accurate) in subsets of patients who are otherwise at some risk for PDAC.

5. Utility of PRS

While a PRS is a continuous variable, most comparisons are performed on population-based quintiles (5 categories) or deciles (10 categories) ranked from lowest to highest score with either the lowest or middle category as the baseline. Splitting the PRS into quintiles allows easier calculations of odds ratios, risk ratios and hazard ratios. Individuals with PRS in the top quintile are at highest risk for disease while those in the bottom quintile are at lowest risk for disease. When used in this fashion, a PRS can be used to stratify a population into high, average, and low risk categories for disease screening purposes.

Example PRSs

PRS ID	Trait	Ontology	# variants	AUROC	Comments	ref
PGS001256	Gallstones	gallstones	876	EUR: 0.67227	Most of the risk is driven by the *ABCG8* SNV rs4245791. Lim used a 6-SNV PRS with sex and BMI for clinical risk. (1)	(2)
PGS001297	T1DM	type 1 diabetes mellitus	69	EUR: 0.797		(2)
PGS002243	T2DM	type 2 diabetes mellitus	6,431,973	EUR: 0.794 AFR: 0.849 AMR: 0,851 ASN: 0.810	Used to develop a trans-ancestry PRS. Top 2% used for global PRS with 2.5-4.5-fold risk of T2DM	(8)
PGS000063	HTG	triglyceride measurement	32			(9)
PGS000001	Breast cancer	breast carcinoma	77		BOADICEA/CanRisk tool (10, 11)	(12)
PGS002264	PDAC	pancreatic ductal adenocarcinoma	49	EUR: 0.605 EUR: 0.830	AUROC is improved in patients at risk of PDAC.	(13)
PGS000316	CD	Celiac Disease	14 HLA-DQ 38	EUR:0.81-0.90		(14)

Table 1. AFR: African ancestry. AMR: Admixed American ancestry (reflecting a mix of Indigenous American, European, and African ancestries, common in populations from the Americas); ASN, East Asian ancestry; AUROC (Area Under the Receiver Operating Characteristic curve); EUR, European ancestry; HTG, hypertriglyceridemia; PDAC, pancreatic ductal adenocarcinoma; T1D, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

T2DM example.

For screening purposes, the utility of the PRS is directly related to the prevalence and incidence of the disease. For example, in a common disease like T2DM with a prevalence of >13% worldwide and a 1% incidence per year, having a T2DM PRS score in the top (5th) quintile (top 20%) makes an individual 2x more likely to have a T2DM diagnosis than if their PRS score was in the bottom quintile. Put another way, 28.3% of all individuals with T2DM are in the top quintile compared to 13.4% in the lowest. This effectively increases the prevalence of T2DM from 13% in the entire population to 19% for individuals with a T2DM PRS in the top quintile and reduces the prevalence to 9% for individuals in the lowest quintile. T2DM has at least 170 PRS that are designed for global, ancestral, mechanistic or other phenotypic descriptors that range in the number of variants from 3 to ~7,000,000. PRS for T2DM can be used to distinguish clinical subtypes of DM (T1DM, T2DM, T3cDM etc.) mechanistical subtypes (e.g. 8 subtypes of T2DM[20]), and risk of complications such as diabetic retinopathy[21], cardiovascular disease[22], renal disease, etc.[23], as well as targeted treatment selection.[24] In addition, the T2DM PRS can be used to further estimate the likelihood of T2DM in older patients with risk of occult pancreatic ductal adenocarcinoma (PDAC – see PDAC PRS below) causing unexplained, new-onset diabetes.

PDAC example.

While this is helpful for common diseases, it is less so for rare diseases such as pancreatic cancer (PDAC). PDAC has a 1% lifetime risk of disease and a peak incidence of 0.11% in males, >75 years old, and of European ancestry. Let's assume a population of 500,000 individuals and 1% will eventually develop PDAC for 5,000 cases. Further assume our PDAC PRS has similar performance to the T2DM with 30% in the top quintile and 13% in the bottom quintile such that 1500 (5000*0.3) cases are in the top quintile and 650 (5000*0.13) are in the bottom quintile. By definition each quintile will have close to 100,000 individuals resulting in 1.5% of those in the top quintile and 0.65% in the bottom quintile eventually developing PDAC. While this is a significant enrichment from the larger population, the PRS is not sufficient for identifying cases on its own. This level of enrichment may be sufficient for a research study or a clinical trial where the surveillance test or intervention is much more costly than the generation of the PRS, but may not be as useful for normal clinical screening. In other words, a PRS is not useful for risk stratification in low

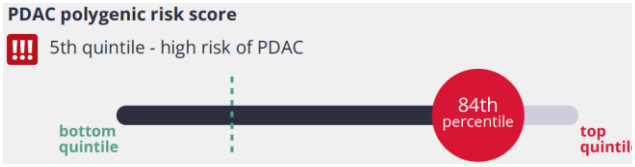

Figure 1. PRS for PDAC translated into percentiles and quintiles. The interpretation must be done within the context of the individuals other risk factors such as age, sex, ancestry, family history, smoking history, body mass index, medical history such presence and timing of onset of diabetes, pancreatitis, etc., and biomarkers (e.g. elevated CA-19-9), unexplained weight loss and/or abdominal imaging findings. (Courtesy of Ariel Precision Medicine, Pittsburgh, PA)

prevalent disorders such as PDAC when the individual is otherwise at low risk because of younger age (e.g. <50 years old), no first- or second-degree relatives with high-risk cancers, minimal environmental risk (e.g. smoking, pancreatitis) and absence of emerging signs or symptoms such as unexplained weight loss or new-onset diabetes mellitus.

Celiac example

Celiac disease (CD) is an autoimmune disease with a worldwide prevalence of ~1.4% and has a long association with HLA-DQ locus. Over 95% of all CD subjects have the HLA-DQ2.5, HLA-DQ2.2, HLA-DQ8, or HLA-DQ7.5 heterodimers but these same heterodimers appear in 25-30% of the general population.[25] GWAS studies have found additional SNVs associated with CD leading to multiple proposed PRS. [14, 26] The AUC for these CD PRSs range from 0.81 to 0.90 for the PRS alone depending on the study population and diagnostic criteria. Due to the heavy HLA-DQ influence these PRS distributions are bimodal with a distinct inflection point between those at high risk and those at minimal risk. With this cutoff point near the 75% percentile, you have ~25% of the entire population but >85% of all CD patients. Using the same formula from the PDAC PRS example, [prevalence*(%cases/%controls) or 1.4*(.85/.25)) brings the prevalence in this high-risk quartile to >4.75% and an odds ratio for CD of 16.1 compared to the lower 75%.

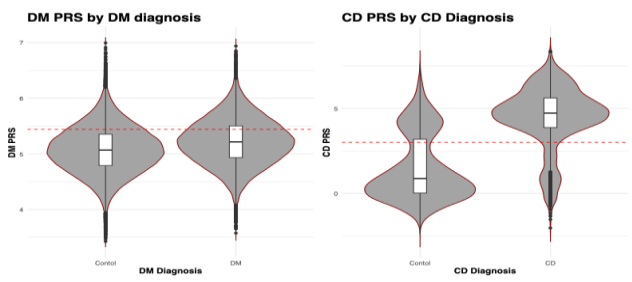

Figure 2. Plots comparing PRS distributions for Type 2 DM (left) and Celiac Disease (right)compared to age, sex, and ancestry matched controls from the UK Biobank (CD 1596 cases vs 15960 controls, DM 50553 cases vs 204094 controls) The area above the red dashed line represents the top quintile in T2DM and the top 26% for CD. T2DM PRS is more typical, normal distribution in both cases and controls while CD PRS shows bi-modal distribution due to the strong influence of HLA heterodimers on the PRS.

6. PRS Interpretation

A PRS is most useful when combined with clinical risk models of biomarkers and modifiable behaviors, such as smoking, alcohol use, diet, and BMI, or that have an effective treatment in some individuals such as high cholesterol. This step accounts for the environmental and lifestyle impacts on the genetic predisposition to disease estimated by the PRS. Often, modifying these non-genetic factors can reduce the risk of disease. This is especially true for individuals that are at greater risk (i.e. those with a PRS in the 5th quintile). For example, it was recently shown that "healthy lifestyles" (non-smoking, healthy weight, regular exercise, and healthy diet) showed a greater reduction in risk of early mortality in those with a high PRS for Prostate Cancer than in those with a low PRS.[27] This may be due to the unmeasured gene x environment interaction mentioned earlier but more research is needed to determine the reason for this effect. Although individuals with the highest PRS seem to benefit more from these behavioral modifications, anyone, including those with the lowest PRS, should consider similar disease-reducing lifestyle changes.

Because having a high disease PRS has been shown to be associated with earlier age of onset, individuals with a high PRS may need to initiate screening procedures and preventative interventions/therapies earlier and possibly increase surveillance routines from what the typical guidelines suggest. Despite having a below average risk for disease, those with a low PRS should not forego screening procedures recommended at a population level because it is possible that certain environmental exposures will overwhelm any protective genetic traits. A low PRS does not mean *no* chance of disease, rather a lower disease prevalence than average. Therefore, it is important to understand that a high PRS does not confirm nor does a low PRS exclude a diagnosis, it only contributes to the baseline, risk of a phenotype.

7. Limitations

As with most genetic studies, the majority of PRS were created using individuals of European ancestry. Because disease prevalence and allele frequencies differ across ancestries, any given PRS will be less effective in an off-target ancestry, and the performance decreases as the differences in allele frequencies increases.

PRS are built from GWAS summary statistics and have many of the same limitations as the GWAS from which they are constructed. Each SNV is independently tested and no advanced gene x gene or gene x environment interactions are included. Many of these GWAS will have very different genotyping technologies, analysis pipelines, and covariate structures. Because of this, there is the possibility of hidden confounders that were not measured in one or more of the original GWAS analysis. While SNVs make up the majority of variation in the genome, other types of variants (e.g. structural, copy number, multi-nucleotide, tandem repeats, etc) that may have a greater effect on disease are not currently included in GWAS and PRS.

PRS risk should not be equated with lifetime risk of developing a phenotype at this time. Greater amounts of surveillance data, (genetic, disease, exposures, etc.) across all ancestries and complete lifetimes are needed to compute lifetime risk with more accuracy.

Germline genetic testing cannot be performed on samples from immunocompromised individuals with recent blood transfusions, bone marrow transplants or in those undergoing chemotherapy treatments.[28-30]

8. Future Directions

Additions of PRS are actively undergoing clinical trials in CVD to assess their utility.[3-7] Early results look promising, but specific details on intervention and counseling remain.

PRS are becoming more mechanistically nuanced by combining SNVs from different pathophysiologic pathways of disease. Suzuki et al. recently showed a multi-ancestry T2DM PRS built around 8 mechanistic clusters.[20] Understanding the risk for each cluster could guide therapy to address the underlying mechanism of disease and inform more specifically appropriate disease screening recommendations.[31] Furthermore, modeling interactions between PRS mechanisms may be able to differentiate between diseases with similar, non-specific presentations.

The acceptance that drug response is primarily polygenic is leading to the active development of Pharmacogenetics (PGx) PRS but this requires individual PRS for each drug.[32] Until these expanded PGx PRS are created and validated, current PGx methods should still be applied.

9. Conclusion

A PRS measures the genetic similarity of a cohort of patients with a phenotype based on a subset of SNVs from GWAS results. Due to study designs, ancestral makeup, and cultural variation of cohorts, PRS also account for some amount of gene x environment interaction. It is this gene x environment interaction that provides possible interventions to reduce the genetic risk for disease, whether through pharmacologic or lifestyle interventions. In addition to providing possible interventions, PRS can also alert the clinician to individuals at risk for early onset and possibly more severe disease trajectories. These patients should be screened earlier and perhaps more often for signs and symptoms of the disease.

Everyone interpreting a PRS needs to recognize that a PRS is a probability score and not deterministic. There are ways to reduce your personal risk no matter your PRS score but that may require difficult lifestyle changes. The only certainty is that making no changes and continuing unhealthy lifestyle habits will not reduce your risk of that disease. Referral to genetic counselors well versed in PRS may be necessary.

Finally, PRS may not be useful to everyone, but as genetic data from more ancestral populations become available, the scores will become more generalizable.

Further Reading

1. AHA PRS statement (Ref. 33)
2. ACMG PRS Statement (Ref. 34)
3. ACMG Points to consider statement (Ref. 35)

References

1. Lim J, Wirth J, Wu K, Giovannucci E, Kraft P, Turman C, et al. Obesity, Adiposity, and Risk of Symptomatic Gallstone Disease According to Genetic Susceptibility. *Clin Gastroenterol Hepatol*. 2022;20(5):e1083-e120 D.O.I: 10.1016/j.cgh.2021.06.044. PMID: 34217876

2. Tanigawa Y, Qian J, Venkataraman G, Justesen JM, Li R, Tibshirani R, et al. Significant sparse polygenic risk scores across 813 traits in UK Biobank. *PLoS Genet*. 2022;18(3):e1010105 D.O.I: 10.1371/journal.pgen.1010105. PMID: 35324888

3. Pastorino R, Pezzullo AM, Agodi A, de Waure C, Mazzucco W, Russo L, et al. Efficacy of polygenic risk scores and digital technologies for INNOvative personalized cardiovascular disease PREVention in high-risk adults: protocol of a randomized controlled trial. *Front Public Health*. 2024;12:1335894 D.O.I: 10.3389/fpubh.2024.1335894. PMID: 38947346

4. Vassy JL, Brunette CA, Lebo MS, MacIsaac K, Yi T, Danowski ME, et al. The GenoVA study: Equitable implementation of a pragmatic randomized trial of polygenic-risk scoring in primary care. *Am J Hum Genet*. 2023;110(11):1841-52 D.O.I: 10.1016/j.ajhg.2023.10.001. PMID: 37922883

5. Naderian M, Hamed ME, Vaseem AA, Norland K, Dikilitas O, Teymourzadeh A, et al. Effect of Disclosing a Polygenic Risk Score for Coronary Heart Disease on Adverse Cardiovascular Events. *Circ Genom Precis Med*. 2025;18(2):e004968 D.O.I: 10.1161/CIRCGEN.124.004968. PMID: 40151934

6. Kullo IJ, Jouni H, Austin EE, Brown SA, Kruisselbrink TM, Isseh IN, et al. Incorporating a Genetic Risk Score Into Coronary Heart Disease Risk Estimates: Effect on Low-Density Lipoprotein Cholesterol Levels (the MI-GENES Clinical Trial). *Circulation*. 2016;133(12):1181-8 D.O.I: 10.1161/CIRCULATIONAHA.115.020109. PMID: 26915630

7. Christensen KD, Schonman EF, Robinson JO, Roberts JS, Diamond PM, Lee KB, et al. Behavioral and psychological impact of genome sequencing: a pilot randomized trial of primary care and cardiology patients. *NPJ Genom Med*. 2021;6(1):72 D.O.I: 10.1038/s41525-021-00236-2. PMID: 34429410

8. Ge T, Irvin MR, Patki A, Srinivasasainagendra V, Lin YF, Tiwari HK, et al. Development and validation of a trans-ancestry polygenic risk score for type 2 diabetes in diverse populations. *Genome medicine*. 2022;14(1):70 D.O.I: 10.1186/s13073-022-01074-2. PMID: 35765100

9. Johnson L, Zhu J, Scott ER, Wineinger NE. An Examination of the Relationship between Lipid Levels and Associated Genetic Markers across Racial/Ethnic Populations in the Multi-Ethnic Study of Atherosclerosis. *PLoS One*. 2015;10(5):e0126361 D.O.I: 10.1371/journal.pone.0126361. PMID: 25951326

10. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med*. 2019;21(8):1708-18 D.O.I: 10.1038/s41436-018-0406-9. PMID: 30643217

11. Ficorella L, Yang X, Mavaddat N, Carver T, Hassan H, Dennis J, et al. Adapting the BOADICEA breast and ovarian cancer risk models for the ethnically diverse UK population. *Br J Cancer*. 2025 D.O.I: 10.1038/s41416-025-03117-y. PMID: 40676226

12. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst*. 2015;107(5) D.O.I: 10.1093/jnci/djv036. PMID: 25855707

13. Sharma S, Tapper WJ, Collins A, Hamady ZZR. Predicting Pancreatic Cancer in the UK Biobank Cohort Using Polygenic Risk Scores and Diabetes Mellitus. *Gastroenterology*. 2022;162(6):1665-74 e2 D.O.I: 10.1053/j.gastro.2022.01.016. PMID: 35065983

14. Sharp SA, Jones SE, Kimmitt RA, Weedon MN, Halpin AM, Wood AR, et al. A single nucleotide polymorphism genetic risk score to aid diagnosis of coeliac disease: a pilot study in clinical care. *Aliment Pharmacol Ther*. 2020;52(7):1165-73 D.O.I: 10.1111/apt.15826. PMID: 32790217

15. Khan A, Shang N, Nestor JG, Weng C, Hripcsak G, Harris PC, et al. Polygenic risk alters the penetrance of monogenic kidney disease. *Nature communications*. 2023;14(1):8318 D.O.I: 10.1038/s41467-023-43878-9. PMID: 38097619

16. Mars N, Koskela JT, Ripatti P, Kiiskinen TTJ, Havulinna AS, Lindbohm JV, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med*. 2020;26(4):549-57 D.O.I: 10.1038/s41591-020-0800-0. PMID: 32273609

17. Fahed AC, Wang M, Homburger JR, Patel AP, Bick AG, Neben CL, et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nature communications*. 2020;11(1):3635 D.O.I: 10.1038/s41467-020-17374-3. PMID: 32820175

18. Lazarte J, Dron JS, McIntyre AD, Skanes AC, Gula LJ, Tang AS, et al. Role of Common Genetic Variation in Lone Atrial Fibrillation. *Circ Genom Precis Med*. 2021;14(1):e003179 D.O.I: 10.1161/CIRCGEN.120.003179. PMID: 33517663

19. Mars N, Lindbohm JV, Della Briotta Parolo P, Widen E, Kaprio J, Palotie A, et al. Systematic comparison of family history and polygenic risk across 24 common diseases. *Am J Hum Genet*. 2022;109(12):2152-62 D.O.I: 10.1016/j.ajhg.2022.10.009. PMID: 36347255

20. Suzuki K, Hatzikotoulas K, Southam L, Taylor HJ, Yin X, Lorenz KM, et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature*. 2024;627(8003):347-57 D.O.I: 10.1038/s41586-024-07019-6. PMID: 38374256

21. Forrest IS, Chaudhary K, Paranjpe I, Vy HMT, Marquez-Luna C, Rocheleau G, et al. Genome-wide polygenic risk score for retinopathy of type 2 diabetes. *Hum Mol Genet*. 2021;30(10):952-60 D.O.I: 10.1093/hmg/ddab067. PMID: 33704450

22. Yao Z, Zhang X, Deng L, Zhang J, Wen Y, Zheng D, et al. Exploring the Genetic Relationship Between Type 2 Diabetes and Cardiovascular Disease: A Large-Scale Genetic Association and Polygenic Risk Score Study. *Biomolecules*. 2024;14(11) D.O.I: 10.3390/biom14111467. PMID: 39595643

23. Yu G, Tam CHT, Lim CKP, Shi M, Lau ESH, Ozaki R, et al. Type 2 diabetes pathway-specific polygenic risk scores elucidate heterogeneity in clinical presentation, disease progression and diabetic complications in 18,217 Chinese individuals with type 2 diabetes. *Diabetologia*. 2025;68(3):602-14 D.O.I: 10.1007/s00125-024-06309-y. PMID: 39531041

24. Singh O, Verma M, Dahiya N, Senapati S, Kakkar R, Kalra S. Integrating Polygenic Risk Scores (PRS) for Personalized Diabetes Care: Advancing Clinical Practice with Tailored Pharmacological Approaches. *Diabetes Ther*. 2025;16(2):149-68 D.O.I: 10.1007/s13300-024-01676-6. PMID: 39688777

25. van Heel DA, West J. Recent advances in coeliac disease. *Gut*. 2006;55(7):1037-46 D.O.I: 10.1136/gut.2005.075119. PMID: 16766754

26. Abraham G, Rohmer A, Tye-Din JA, Inouye M. Genomic prediction of celiac disease targeting HLA-positive individuals. *Genome medicine*. 2015;7(1):72 D.O.I: 10.1186/s13073-015-0196-5. PMID: 26244058

27. Plym A, Zhang Y, Stopsack KH, Ugalde-Morales E, Seibert TM, Conti DV, et al. Early Prostate Cancer Deaths Among Men With Higher vs Lower Genetic Risk. *JAMA Netw Open*. 2024;7(7):e2420034 D.O.I: 10.1001/jamanetworkopen.2024.20034. PMID: 38958976

28. Tran SD, Pillemer SR, Dutra A, Barrett AJ, Brownstein MJ, Key S, et al. Differentiation of human bone marrow-derived cells into buccal epithelial cells in vivo: a molecular analytical study. *Lancet*. 2003;361(9363):1084-8 D.O.I: 10.1016/S0140-6736(03)12894-2. PMID: 12672312

29. DiGuardo MA, Kester SJ, Mahaffey VJ, Hammel SA, Heaser KK, Hofich CD, et al. Does Transfusion of Red Blood Cells Impact Germline Genetic Test Results? *J Pers Med*. 2020;10(4) D.O.I: 10.3390/jpm10040268. PMID: 33316904

30. van den Boogaard WMC, Komninos DSJ, Vermeij WP. Chemotherapy Side-Effects: Not All DNA Damage Is Equal. *Cancers (Basel)*. 2022;14(3) D.O.I: 10.3390/cancers14030627. PMID: 35158895

31. Polygenic Risk Score Task Force of the International Common Disease A. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nat Med*. 2021;27(11):1876-84 D.O.I: 10.1038/s41591-021-01549-6. PMID: 34782789

32. Zhai S, Zhang H, Mehrotra DV, Shen J. Pharmacogenomics polygenic risk score for drug response prediction using PRS-PGx methods. *Nature communications*. 2022;13(1):5278 D.O.I: 10.1038/s41467-022-32407-9. PMID: 36075892

33. O'Sullivan JW, Raghavan S, Marquez-Luna C, Luzum JA, Damrauer SM, Ashley EA, et al. Polygenic Risk Scores for Cardiovascular Disease: A Scientific Statement From the American Heart Association. *Circulation*. 2022;146(8):e93-e118 D.O.I: 10.1161/CIR.0000000000001077. PMID: 35862132

34. Abu-El-Haija A, Reddi HV, Wand H, Rose NC, Mori M, Qian E, et al. The clinical application of polygenic risk scores: A points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2023;25(5):100803 D.O.I: 10.1016/j.gim.2023.100803. PMID: 36920474

35. Reddi HV, Wand H, Funke B, Zimmermann MT, Lebo MS, Qian E, et al. Laboratory perspectives in the development of polygenic risk scores for disease: A points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2023;25(5):100804 D.O.I: 10.1016/j.gim.2023.100804. PMID: 36971772

Corresponding Author:

Phil J Greer, MS
Ariel Precision Medicine,
Pittsburgh, Pennsylvania, USA
pgreer@arielmedicine.com

Contributions:

PJG developed the background and wrote the first draft. PJG and TDO codeveloped the ideas and further drafts. All authors reviewed and approved the final version.

Conflicts of interest:

The author(s) declare that they have no competing interests.

Funding:

None.