

REVIEW

Open Access

Polygenic risk scores: from research tools to clinical instruments



Cathryn M. Lewis^{1,2*}  and Evangelos Vassos¹

Abstract

Genome-wide association studies have shown unequivocally that common complex disorders have a polygenic genetic architecture and have enabled researchers to identify genetic variants associated with diseases. These variants can be combined into a polygenic risk score that captures part of an individual's susceptibility to diseases. Polygenic risk scores have been widely applied in research studies, confirming the association between the scores and disease status, but their clinical utility has yet to be established. Polygenic risk scores may be used to estimate an individual's lifetime genetic risk of disease, but the current discriminative ability is low in the general population. Clinical implementation of polygenic risk score (PRS) may be useful in cohorts where there is a higher prior probability of disease, for example, in early stages of diseases to assist in diagnosis or to inform treatment choices. Important considerations are the weaker evidence base in application to non-European ancestry and the challenges in translating an individual's PRS from a percentile of a normal distribution to a lifetime disease risk. In this review, we consider how PRS may be informative at different points in the disease trajectory giving examples of progress in the field and discussing obstacles that need to be addressed before clinical implementation.

Keywords: Genetics, Common disorders, Polygenic risk scores, Prediction, Pharmacogenetics, Risk

Background

Over the last decade, genome-wide association studies (GWAS) have uncovered the contribution of inherited variants to common complex disorders. Our current understanding is that most non-communicable disorders with a major public health impact have a genetic underpinning that is highly polygenic, comprising hundreds or thousands of genetic variants (or polymorphisms), each having a small effect on disease risk. Each genetic variant associated with a disease is valuable in indicating a gene or pathway of biological relevance to the disorder, but there are also expectations that the genetic data could be used to predict disease risk, with potential clinical utility.

In a polygenic disorder, a single variant is not informative for assessing disease risk. Instead, a genetic loading conferred by the combined set of risk variants is necessary to obtain a measure that has sufficient information to identify those at high risk. There are many possible approaches to combine information across loci; the genetic risk is most often assessed through the polygenic risk score (PRS), a weighted sum of the number of risk alleles an individual carries. Despite methodological concerns about construct, content, and criterion validity of PRS [1], many studies have shown that PRSs can predict disease status in research-based case-control studies [2–4]. More convincingly, the prediction is also valid in population-based cohort studies and in electronic health record-based studies [5–7].

In this review, we consider how polygenic risk scores may be informative at different points in the disease trajectory, from unaffected individuals being tested for future disease risk to diagnosed patients, assessing how genetic information might inform their treatment or

* Correspondence: cathryn.lewis@kcl.ac.uk

¹Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, de Crespigny Park, London SE5 8AF, UK

²Department of Medical and Molecular Genetics, Faculty of Life Sciences and Medicine, King's College London, London, UK



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

provide prognostic information on disease course. For each stage, we illustrate how PRS might be used and give examples of the current progress in the field.

Properties of polygenic risk scores

The PRS is formed from a set of independent risk variants associated with a disorder, based on the current evidence from the largest or most informative genome-wide association studies. For each individual, the number of risk alleles carried at each variant (0, 1, or 2) is summed, weighted by its effect size (i.e. log (OR) for binary traits or beta coefficient for continuous traits). The outcome is a single score of each individual's genetic loading for a disease or for a continuous trait (Fig. 1).

Summing across variants assumes an additive genetic architecture, with independence of risk variants. Although simplistic, this reflects our best estimate of genetic architecture of common complex disorders, where little evidence of interaction between genetic variants is detected. These additive polygenic risk scores do not model any gene-gene or gene-environment interactions [9]; however, the largest meta-analysis of heritability from twin studies supports a simple additive model in most of the traits examined [10].

Several methods can be used to calculate polygenic risk scores. These include 'clumping/pruning and thresholding' methods, where a reduced set of genetic variants is identified through pruning on linkage disequilibrium, and accounting for evidence of association with the trait being studied (clumping). Polygenic risk scores are then

calculated summing over all SNPs meeting a p value threshold, or set of thresholds, as implemented in PRSice [11] and PLINK [12]. In contrast, other methods assess the best prediction genome-wide by explicitly modelling the correlation structure between variants without attempting to identify a minimal subset of SNPs for prediction; the most widely used implementation is the Bayesian LDpred approach [13]. Many novel risk score methods are under development and may have increased power in comparison with our current methods (for example, SBayesR [14]). We will use the term 'polygenic risk scores' to cover all methods that sum genetic data to provide individual risk measures and will assume that these are transformed to have a standard normal distribution. The measures used to assess the predictive ability of a PRS are summarised in Table 1.

The PRS measure is beguiling in its simplicity, but it is limited in its ability to capture the full genetic loading for a disorder. We currently have an incomplete list of genetic variants associated with a disorder, and the effect sizes used to construct the score are imprecise. The use of tagging SNPs in place of the (unknown) causal variant or variants also limits precision, but a novel methodology has been developed, for example, extensions of the LDpred, in order to address this issue [18, 19]. Further, it is becoming clear that genetic risk scores also capture information from the environment. Evidence for this is seen in family-based studies which show that contributions from the PRS computed from non-transmitted alleles of parents also affect offspring phenotypes. This

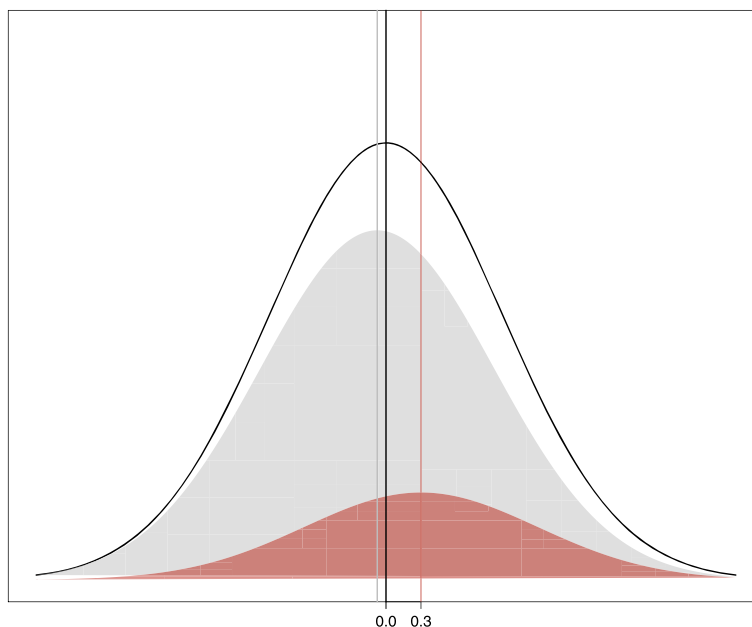


Fig. 1 Normal distribution of polygenic risk scores, for a disorder of prevalence 20% (prev), with cases having a mean PRS of $t = 0.3$. Black line: population $N(0,1)$ distribution. Grey shaded area: controls, unaffected with disorder, with mean PRS $= -\text{prev} \times t / (1 - \text{prev}) = -0.075$. Red shaded area: cases, mean PRS $t = 0.3$. AUC = 0.605, calculated from $\Phi(\text{Cohen's } d / \sqrt{2})$, where Φ is the normal distribution cumulative distribution function, and Cohen's d is the difference between mean PRSs for cases and controls [8]

Table 1 Assessing the clinical utility of polygenic risk scores

A: Population level

The predictive ability of polygenic risk scores can be measured in research studies, where differences between cases and controls (Fig. 1) or of a continuous trait in a population are assessed. Here, the disease status or trait is pre-established, and the studies measure the extent to which this is determined by the PRS. Outcome measures from such studies include:

- (1) R^2 from linear regression, which quantifies the proportion of variance in a continuous trait captured by the PRS, or equivalently Nagelkerke's R^2 for logistic regression for case-control disease status.
- (2) R^2 on a liability scale, which transforms Nagelkerke's R^2 to reflect disease prevalence, instead of the case-control ratio of the research study [15].
- (3) The area under the receiver operating characteristic curve (AUC) [16], which takes a value from 0.5 to 1. This gives an overall summary of the predictive ability of the model. It is most easily interpreted as the probability that a randomly selected case will have a higher polygenic risk score than a randomly selected control. Such models can also include risk factors such as age and sex, which will increase the AUC values above that based on PRS alone.
- (4) The proportion of the population that has a k -fold increased odds ($k = 2, 3, \dots$), compared to the population disease risk.
- (5) Odds ratio of disease risk conferred by a 1-standard deviation increase in PRS.
- (6) Odds ratio of disease for an individual in the top PRS decile (or other quantiles) compared to individuals in a different part of the PRS distribution. The high-risk group may be compared to the lowest decile, a mid-quintile (e.g. 40–60%), or those outside the high-risk group (0–90%). Comparing the upper and lower tails maximises the odds ratio for impact but raises concerns about the arbitrariness of the quantile used.

B: Individual level

In a clinical setting, the focus is on a single person: what information does their PRS give about their risk of disease? Possible outcome measures that are relevant at an individual level include:

- (a) At what percentile in the distribution of PRS does this individual lie? This is between 0 and 100%, with scores having a normal distribution.
- (b) What is this person's relative risk of disease compared to the average risk in the population?
- (c) What is this person's absolute risk of disease, and by what age [17]?

'genetic nurture' phenomenon indicates, for example, that educational attainment is influenced by both offspring genetics and the non-transmitted parental genetics, which may determine the family environment [20].

There are four important considerations of the information content of a polygenic score, and how it can be interpreted:

- 1) The known information, which shows where an individual lies compared to others on the risk scale
- 2) The unknown information from incomplete genetics or unmodelled environment
- 3) The potential for incorrect information, for example, where the individual differs from

characteristics of the research study used to estimate the effect size of each genetic variant by genetic ancestry, age, environmental load, or disease definition, or where there is a technical bias in data collection

- 4) The intended use of the PRS, for example, more complete information would be required for justifying a pharmacological intervention than for using the PRS to motivate behaviour change

The first two properties of known and unknown information are summarised by the proportion of disease liability that the polygenic risk score captures, whilst our understanding of the incorrect information is still evolving.

Applicability of PRS across ethnic groups

One of the most challenging aspects of moving PRS to the clinical arena is ensuring that they are equally applicable to all health care users across ethnic groups to limit exacerbating health disparities [21]. This is an important issue both for minority ethnic groups within high-income countries, who may be under-represented in research studies, and for low- and middle-income countries, where genetic studies of the relevant ancestry may not exist because of limited research infrastructure. Current PRS methods rely on an individual's genetic ancestry being similar to the large GWAS study from which reference effect sizes are taken for PRS calculation and may require access to an ancestry-matched genotype-level reference panel. Such studies are currently only widely available in European ancestries [22, 23], so polygenic risk scores are applicable to only a small proportion of the world's population; in this paper, unless otherwise stated, study participants are of European ancestries. Transferability of PRS across populations is limited, with PRS generated from GWAS in one population usually providing attenuated predictive accuracy in other populations [21, 24]. Reasons for this include the use of tagging SNPs, differences in the patterns of linkage disequilibrium between populations, and SNP arrays biased to variants of European descent [25]. More importantly, differential genetic drift can cause unpredictable biases when scores inferred from one population are applied in another [26]. At an individual level, it is crucially important that an individual's PRS is compared to a population-specific distribution so that the interpretation is valid. Progress in performing GWAS on non-European ancestries has been slow, with, for example, < 3% of study participants in the GWAS Catalog were of African ancestry [22]. Large-scale GWAS of diabetes and schizophrenia have been performed in African and East Asian populations [27–29], and novel initiatives of the collection in worldwide populations like the Human Heredity and Health in Africa (H3Africa) Initiative (<https://h3africa>).

org/) and the African Mental Health Research Initiative (<https://amari-africa.org>) are underway. Key methodological considerations for GWAS in ancestrally diverse populations have been recently discussed, including the choice between performing a meta-analysis stratified by ethnic groups and performing a joint mixed-model across all participants [23]. Novel methods for ‘polyethnic’ scores, like XP-BLUP and Multi-ethnic PRS, which improve predictive accuracy by combining transethnic with ethnic-specific information, are being developed [30–32].

Substantial investment will be needed to achieve the equivalence of genetic information required for equity of access when polygenic risk scores are applied in the clinic [21, 33].

Clinical utility of PRS

The potential value of polygenic scores is supported by the increasing number of research studies that show a highly significant association between PRS and disease status, but their clinical utility has yet to be established. Can PRS be used by clinicians for disease prediction or stratification, either now or in the future? For this to be achieved, the focus must shift from association with case-control status to the information in the PRS for a single individual. Furthermore, to translate PRS to clinical tools, relative risks that compare individuals across the PRS continuum with a baseline group will eventually need to be transformed to absolute risks for the disease [34, 35]. Unlike monogenic disorders caused by high-penetrance mutations, in complex disorders, the discriminative ability of PRS is compromised by the multifactorial contributors to the disease, the imperfect measurement of the full genetic signal, and the potentially incorrect measurement.

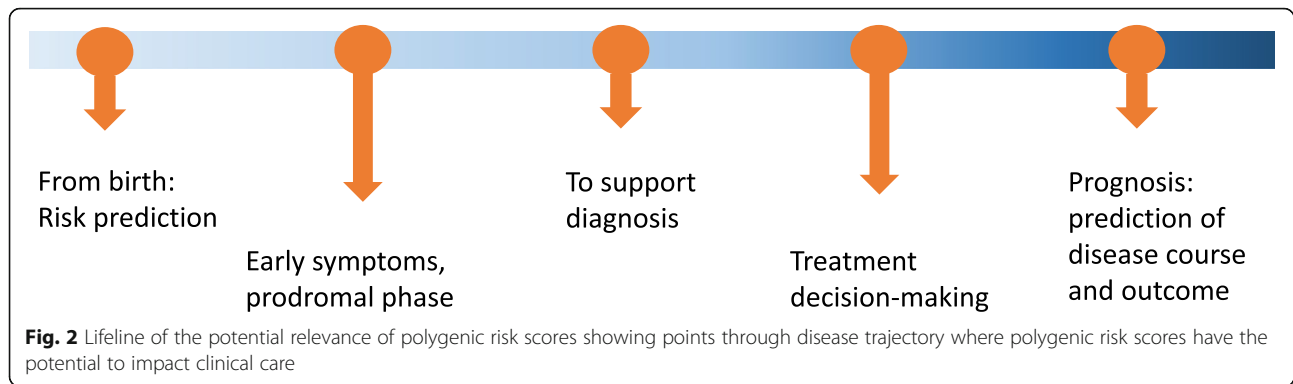
Risk prediction models, including a combination of clinical, biochemistry, lifestyle, and historical risk factors, are currently used to predict 10-year risk of cardiovascular disease and diabetes [36–39]. These models combining risk factors achieve a good prediction (AUCs of 80–85%) and are included in clinical guidelines for prevention and public health [40]. Polygenic risk scores have much lower AUCs, as expected from a single risk factor, and should not be considered as an alternative to these clinical risk models but as a possible addition. With the established polygenic architecture of complex disorders, the improvement of genetic and statistical methodology, and the increase of global genotyped samples, it is reasonable to anticipate that genetic prediction will improve. In the meantime, it may be timely to consider the use of PRS in specific cohorts where there is a higher prior probability of disease.

The current focus is on identifying individuals at high genetic risk of disease for risk stratification. This information could be useful in decisions about participation in screening programmes, lifestyle modifications, or preventive

treatment, when available and appropriate. PRS may also be relevant at different points along disease diagnosis and course (Fig. 2). An important consideration is the need to avoid presenting a false impression of genetic determinism (the notion that genes alone define biology). This could otherwise detrimentally impact personal choices, harming physical and mental well-being (e.g. diet, exercise, lifestyle), and possibly even education, employment, or family planning. Research on measuring the beliefs of the public in genetic determinism [41] should expand from single-gene disorders to include polygenic prediction of complex diseases. The widespread interest in PRS is illustrated by their use by direct-to-consumer genetic testing companies; for example, 23andMe now offers polygenic risk scores for T2D. The Polygenic Score Catalog (<http://pgscatalog.org/>) curates data and tables extracted from polygenic risk scores for common disorders, capturing performance metrics of the PRS developed.

Disease risk prediction

Although one’s genetic liability is fixed from conception, the risk arising from one’s genes is dynamic, depending on changing factors such as age, environmental exposures, and previous illnesses. For example, if someone is at high genetic risk for alcohol/drug dependence but is never exposed to alcohol or drugs, the genetic risk is irrelevant. Even if sequencing at birth were to become standard clinical practice [42], communicating risk scores at birth is neither appropriate nor useful, and it is likely that genetic data would be stored and interrogated throughout life for both single-gene Mendelian disorders and common polygenic disorders. The decision to assess PRS might be triggered by age, onset of symptoms, family history, or presence of relevant environmental factors. The role that PRS will play in clinical care is currently unclear, and any use of PRS must be predicated on clear clinical utility, with a specific outcome activated by the score. For example, a PRS for coronary artery disease assessed in early adulthood may be useful to encourage healthy behaviour throughout life, although we still lack experience of how to use genetic data to motivate behaviour change [43–45]. Not all preventive strategies are so benign; pharmacological interventions or surgical procedures are more controversial. For example, it would be very difficult to consider prophylactic mastectomy for breast cancer prevention. Even simple decisions like increased screening may result in false positives with significant economic cost to society and unnecessary stress of the individual. At its simplest, PRS may be used to estimate an individual’s lifetime risk of disease. This application follows the design of most genome-wide association studies, differentiating between cases and controls. From the studies performed in different clinical areas assessing the predictive ability of PRS, we discuss below



the strongest evidence for potential clinical utility, with special reference to our research area, psychiatric disorders [46].

Medical conditions

Much of the research to motivate moving polygenic risk scores from research studies to clinical implementation comes from cardiovascular disease, type 2 diabetes, breast and prostate cancers, and Alzheimer's disease [47]. Khera et al. [2] recently demonstrated in the UK Biobank that PRS can identify which percentage of the sample have at least 3-fold increased risk for coronary artery disease, atrial fibrillation, type 2 diabetes, inflammatory bowel disease, and breast cancer, with the proportion of individuals identified varying between 1.5 and 8% depending on the disorder. Although these effects appear modest, PRS can identify substantial larger fractions of the population at high disease risk than monogenic mutations, making PRS potentially more clinically relevant. Apart from the generic prediction of case-control status, specific applications of PRS have been proposed. For example, the PRSs of left ventricular cardiovascular magnetic resonance phenotypes are predictive of heart failure events independently of clinical risk factors in the UK Biobank [48], whilst elevated genetic score for albuminuria is strongly associated with increased risk of hypertension [49].

In breast cancer, pioneering work in risk prediction, modelling genetic susceptibility based on two genes, *BRCA1* and *BRCA2*, [50] has been expanded to the use of polygenic scores, which have improved predictive ability. In a recent study, a PRS based on 303 genetic variants had an AUC of 0.63, with an odds ratio of 1.61 (95% confidence interval, 1.57–1.65) per unit increase in PRS [3]. Whilst these figures are modest, they translate into a substantial spread of risk in the population: women in the top 1% of PRS have a 4-fold increased risk of developing ER-positive breast cancer and a corresponding 6-fold decreased risk for those in the lowest 1% PRS (both compared to women in the mid-quintile of the PRS distribution). Despite the modest discriminative ability,

PRS could be utilised in improving screening programmes, including defining the age at which breast cancer screening should start and the screening interval. In the UK, mammogram screening is offered to women over age 47, when the average 10-year risk of breast cancer is ~2.6%. However, Mavaddat et al. [3] show that PRS is informative in stratifying risk, with the 20% of women with the highest polygenic risk scores reaching this level of risk before age 40 and the 20% of women with the lowest scores never reaching this level of risk. This study shows that breast cancer polygenic risk scores already capture sufficient information to identify a high-risk subgroup of women who could be offered mammogram screening at an earlier age. Commercial breast cancer risk tests based on polygenic risk scores are already offered by Myriad Genetics (riskScore™) and Ambry Genetics (AmbryScore).

These PRS studies were all performed in European-ancestry populations, and expansion to worldwide populations is essential. For example, a study of lung cancer-associated variants in Chinese populations [51] demonstrated that PRS was an independent risk stratification indicator for lung cancer beyond age and smoking pack-years. Polygenic risk scores have been less widely applied to auto-immune disorders. In type 1 diabetes, strong genetic risk scores have been created, with the large effect of the HLA haplotypes increasing AUC values to >0.8 in European ancestry and Hispanic populations [52]. An ancestry-specific genetic score using 7 SNPs outperformed a European-based genetic score in African ancestry participants (AUC 0.87 and 0.80, respectively) [53].

PRS also can be used to predict continuous traits in the population like BMI, which is important as a risk factor for cardiometabolic traits. A recent large study showed that participants in the highest BMI PRS decile have a BMI that is 2.6 kg/m² higher than those in the mid-quintile PRS—to put in context, this is half the width of the 'overweight category' of BMI (from 25 to 29.9 mg/m²)—and corresponds to 31% of individuals with a BMI > 40 (obesity class 3) [54]. Genetically dissected trajectories of BMI across the lifespan help us identify where genetic prediction starts to be relevant.

BMI PRS is only minimally associated with birthweight; it has increasing prediction through childhood, and by age 18, the differences in BMI by PRS quartile are similar to that seen in adulthood [54]. Despite the significant polygenic prediction of BMI in the general population, it remains to be determined whether BMI PRS has clinical utility in high-risk populations. For example, a common adverse effect of antipsychotics or antidepressants is weight gain; however, we do not know whether this is determined or modified by the BMI PRS.

Psychiatric disorders

Polygenic risk scores for major depression can be calculated using results from recent genome-wide association studies [4, 55]. Wray et al. identified 44 SNPs at genome-wide significance and showed that a PRS built from SNPs with a p value < 0.05 had the highest predictive ability. Individuals in the upper PRS decile had approximately 2.5-fold increased risk of disease compared to those in the lowest decile, which translates into a substantial change in absolute risk, given an approximate 15% lifetime risk of major depression. However, this score has an AUC of 0.57 and captures only 2% of variance in disease risk (R^2 on the liability scale), whilst the remaining 98% is uncaptured by the PRS. An individual's risk of depression therefore comprises 2% of measurable genetic risk score and 98% unaccounted variation from unmodelled genetic and environmental factors. Even for an individual at very high genetic risk, the PRS signal would be overpowered by the unmodelled component. Depression PRS is therefore not yet useful, and any future utility would be based on substantially increasing the variance explained by the PRS or by joint modelling of genetic and environmental risk factors. Increasing the sample size with the inclusion of broader self-reported definitions of depression [55] resulted in a modest increase of the variance explained by PRS, albeit at the cost of specificity for major depression. Phenotypic refinement has been proposed as an alternative to produce more clinically relevant findings [56].

For schizophrenia, the predictive ability is higher, with the current score accounting for 7% of trait variance and an AUC of 0.61, but these values are still far below that needed for an individual's score to have sufficient signal for interpretation or for clinical utility [57]. There is greater potential for using risk prediction from genetics in schizophrenia, since the heritability of 65–80% [58, 59] is much higher than 37% for major depression [60], but the substantially different disease lifetime risks ($< 1\%$ for schizophrenia vs. 15% for major depressive disorder) is also relevant. Even though polygenic scores are not meaningful for general prediction [57], there are points in the clinical care pathway where PRS could be useful in achieving an earlier or a more precise diagnosis. For

example, in first-episode psychosis, we have shown that schizophrenia PRS can differentiate schizophrenia from other psychosis diagnoses (Nagelkerke's R^2 of 9% and those in the top quintile of PRS having an approximately 2-fold increased risk of being subsequently diagnosed with schizophrenia) [24]. This is a low predictive ability, but the setting within first-episode psychosis cases makes it more appealing because (1) it does not require genotyping of the general population, only people with psychosis, and (2) it is not relevant to major decisions (like treat/not treat), but could provide additional information potentially useful for the care plan. In addition to assisting diagnosis, genotype data could be used to calculate other PRS in secondary screening, for example, cardiovascular disease, since psychosis cases are already at high cardiometabolic risk.

In Alzheimer's disease (AD), the three *APOE* variants ($\epsilon 2$, $\epsilon 3$, and $\epsilon 4$) have been consistently associated with disease risk, making it the strongest single-gene predictor at a population level in neuropsychiatry. However, additional risk variants summarised in a PRS improve the prediction model further. For example, Desikan et al. [61] showed that amongst *APOE* $\epsilon 3/\epsilon 3$ carriers, PRS modified the expected age of AD onset by more than 10 years between the lowest and highest deciles (hazard ratio 3.34, $p = 10^{-22}$).

PRS by environment interaction

All the common disorders considered here have both genetic and environmental risk factors. Another area of exploration in PRS is possible gene-environment interaction, which implies that the effect of the disease PRS would depend on the level of the environmental risk factor. This interaction model contrasts with an additive model, where PRS and environment contribute independently to disease risk. With a positive interaction, the effect of a high PRS would be amplified in the presence of an environmental risk factor (E), putting this subgroup of the population (G+, E+) at particularly high risk of disease. These individuals could form a specific target group for interventions, identified by either PRS or environment, or both.

Despite the attraction of identifying PRS-E interactions, currently, there is no strong evidence supporting these interaction models.

In depression, childhood maltreatment is an important risk factor for later diagnosis with depression. A meta-analysis of 6000 individuals confirmed strong effects for both PRS and childhood maltreatment contributing to the risk of depression but showed no evidence of an interaction [62]. In cardiovascular disease, an extensive study of lifestyle effects (diet, exercise, BMI, smoking) with polygenic risk scores showed strong effects from both sources, but no evidence of an interaction [63].

Similarly, a recent cohort study of UK Biobank participants aged > 60 without cognitive impairment, followed up for 8 years, both genetic risk and lifestyle factors predicted incidence of dementia, but no interaction was found [64]. Even though this paucity of confirmed interactions in large samples is not helpful in identifying individuals at very high risk, testing and disproving interactions are essential for correct joint modelling of genes and environment for risk prediction [65].

PRS in treatment choice

Pharmacogenetic studies test how genetic variants affect response to treatment, with the aim of assisting treatment choices to maximise efficacy and minimise side effects. Most progress has been made in identifying rare high-risk variants that increase risk of adverse drug events (for example, abacavir and HLA-B*57:01, carbamazepine and HLA-B*15:02), whilst prediction of treatment efficacy has largely evaded genetic dissection.

The potential impact of PRS in treatment response is unknown, but an easy first target is to test whether genetic disease susceptibility also plays a role in treatment outcome. Currently, the strongest evidence for a role of PRS in treatment response is in statin use to reduce the risk of first coronary event, where studies have shown that the relative risk reduction is higher in those at high genetic risk for cardiovascular disease [66, 67]. These results are in line with the previous reporting of better efficacy of statins in high-risk samples, for example, due to diabetes, hypertension, or high CRP concentrations [68]. A recent study demonstrated a potential role of PRS for electrocardiogram parameters in predicting the cardiac electrical response to sodium channel blockade [69].

In psychiatric disorders, only weak evidence exists to suggest that the PRS for disorder susceptibility might be predictive of treatment response in depression [70, 71] or psychosis [72]. Further studies to identify specific treatment response polygenic risk scores are in progress in these disorders, but it is challenging to achieve sufficiently large sample sizes, with accurately captured response measures. Meta-analysis studies are underway, pooling clinical trials and observational studies of response to anti-depressants and to anti-psychotics. These would identify polygenic predictors for treatment response that might be useful in, for example, deciding between pharmacological and psychological treatment for depression [73]. Only one third of patients respond to the first anti-depressant prescribed [74], so a polygenic predictor might be useful to guide treatment; even a modest increase in the proportion of patients responding could have a substantial impact on the effectiveness and time to recovery. An important perspective in genetic testing for treatment response is in identifying patients who are unlikely to respond to a specific drug, as Gibson

highlights [75, 76]. This could reduce the time taken by clinicians to find efficacious treatment, improve treatment response, and prevent treatment-related adverse effects, which is cost-effective for both the patient and the healthcare system. When the choice of treatment is not dictated by different effectiveness, but from personal experience, preference, or intuition on a trial-and-error basis, PRS can potentially give some quantifiable information to be considered along other lines of evidence.

PRS to refine penetrance of high-risk variants

Evidence is accruing that polygenic risk scores have a role in both the general population and carriers of rare, high-risk genetic variants. In disorders as diverse as breast cancer, developmental disorders, and schizophrenia, polygenic risk scores affect penetrance, acting as moderators for high-risk variants or structural variation. This highlights a possible role for PRS within the well-established framework of high-risk genetic testing. For example, the Deciphering Developmental Disorders study showed that in 7000 children with severe neurodevelopmental disorders expected to be monogenic in aetiology, common variation affects the overall risk of severe neurodevelopmental disorders. It explains over 7% of the variance and affects the individual presentation of symptoms [77]. The role of common variation in moderating expressivity was confirmed in a large electronic health record study, where, in addition to the large effect from the rare pathogenic variants, PRSs for height and BMI were associated with clinical outcome [78]. In breast cancer, the absolute risk increase in carriers of *BRCA1* and *BRCA2* pathogenic variants depends on breast cancer polygenic risk scores, which might influence clinical decision-making [79]. The joint modelling of common and rare variants for breast cancer risk prediction can now be performed in the risk calculation tool, BOADICEA [80]. Similarly, in schizophrenia, both structural variation and PRS contribute to risk: schizophrenia cases that carry confirmed copy number variants (CNVs) have higher PRS than cases which do not; within carriers of CNVs, schizophrenia cases have higher PRS than controls [81, 82]. Hence, even in the presence of CNVs with high penetrance [83], polygenic scores affect the overall risk of disease and may be relevant to the clinical expression in CNVs associated with multiple phenotypes like the 22q11.2 deletion [84].

Role of direct-to-consumer testing

Direct-to-consumer (DTC) genetic testing companies give consumers easy access to their genetic data, specifically genotyping on genome-wide chips of up to 1 million variants. Estimates suggest that 26 million people had used online DTC companies such as [Ancestry.com](https://www.ancestry.com) and 23andMe up to the end of 2018 (<https://www.ancestry.com> and <https://www.23andme.com>).

[technologyreview.com/s/612880/more-than-26-million-people-have-taken-an-at-home-ancestry-test/](https://www.technologyreview.com/s/612880/more-than-26-million-people-have-taken-an-at-home-ancestry-test/)). Whilst many purchasers are initially interested in ancestry testing, customers may then move on to analyse their genetic data for health [85], downloading their raw genotype data to explore in third-party interpretation programmes. These programmes are unregulated and differ in the genetic risks provided, the explanatory information provided, and the cautions given over interpretation. Some sites allow users to calculate polygenic risk scores; for example, Impute.me (<https://www.impute.me/>) shows users where their polygenic risk score lies against a population-specific distribution of scores. Allelica provides an online service calculating polygenic risk scores [86]. In direct-to-consumer genetic testing, MyHeritage (<https://www.myheritage.com/health/genetic-risk-reports>) provides polygenic risk scores on four traits, ‘for people who are of mainly European ancestry’. The most detailed assessment of PRS in a DTC setting is from 23andMe, whose white paper presents their epidemiological modelling and the challenges of deriving individual-level absolute disease risks from PRS [67]. 23andMe provides polygenic risk scores for type 2 diabetes; based on external validation, their models have AUC values of between 59 and 65%, similar to those obtained from research studies [87]. Their customer reports give an estimate of the remaining lifetime risk of T2D based on genetics, age, and ancestry, with additional information on how BMI, diet, and exercise habits affect T2D prevalence.

The accuracy and generalisability of any PRS model need to be validated with external data, but even when the scientific basis is robust, the correct, unbiased interpretation of risk profiles by the consumers will need to be evaluated. The extent to which DTC genetic testing will move polygenic risk scores into the clinical arena is unknown. For example, patients bringing their DTC results may motivate conversations between primary care physicians and patients on health education.

Conclusions and future directions

Polygenic risk scores have moved from research discovery studies to clinical research studies (for example, a trial aiming to assess the impact of PRS reporting on breast cancer risk management recommendations NCT03688204) (<https://clinicaltrials.gov/ct2/show/NCT03688204>) and have started on the slow path to clinical implementation. This review discusses some of the disorders where this is likely to occur and highlights the obstacles that remain in harnessing the information contained in PRS. The strongest evidence for PRS currently comes from cardiovascular diseases and breast cancer, where risk stratification of those at high polygenic risk has clinical utility [2, 3, 47]. Other disorders are likely to follow; however, there is still a long route to be

covered before PRSs become useful tools for clinicians (Table 2).

One challenge of exploring the value of PRS within the clinical setting to predict the outcome, or determine the treatment, is that the sample sizes from case-only clinical studies with relevant phenotypic data related to the course of illness, treatment response, or adverse effects are substantially lower than those from case-control disease susceptibility studies. The latter requires minimal phenotypic information—a clinical diagnosis, or self-report—whilst determining prognosis or treatment outcome requires longitudinal follow-up across sustained periods of time. This is expensive and challenging to collect, and such studies often have much smaller sample sizes. Electronic health records (EHR) may provide longitudinal data, but ‘treatment response’ is often poorly recorded and needs to be captured laterally through prescription records. We highlight how applying PRS in treatment response may better facilitate clinical utility, as the genetic data will complement the clinician’s choice of treatment. We envisage that the role of PRS in informing treatment choices, for example, prioritising pharmaceutical or psychosocial interventions or providing quantitative information on the benefit to harm ratio for each treatment, rather than treat/not treat decisions, may be the low-hanging fruit where the clinical utility of PRS will become apparent.

An ultimate goal might be to have genotype data—and later whole-genome sequence data—integrated into our clinical record; this could then be interrogated at each clinical encounter for relevant information on risk prediction, treatment response, and disorder prognosis. For polygenic risk scores, this is not yet scientifically justified and is technically challenging, particularly since an individual score must be built and interpreted against the

Table 2 A brief overview of the steps required to make PRS relevant in a clinical setting

1. Realistic estimation of predictive ability in clinical populations, which may differ from research samples in disease severity, ancestral diversity, and exposure to environmental risk
2. Identification of the intended purpose of the PRS, which may affect its design and validation, and relevant clinical questions that can be answered, for example, prediction of severity, course of illness, or response to treatment
3. Recognition that even though not useful for the majority of the population with PRS in the middle of the distribution, the outcome may be relevant for those with high or low PRS, in the tails of the distribution
4. Clarification if PRS has an additive or interaction effect with established epidemiological or biological risk factors before combining in joint prediction models [88]
5. Engagement of clinicians and service users, to ensure that any application of polygenic risk scores avoids deterministic interpretations and is based on the understanding that PRS is an indicator, not a precise measure

appropriate genomic reference population, which may not be available. Both these restrictions are likely to change with continuing scientific progress in uncovering the genetic contribution to common diseases and with expanding capabilities of electronic health records [7]. Projects such as the eMERGE network (<https://emerge-network.org/>) are leading the way in these initiatives, although also highlighting clinicians' concerns about the role of unsolicited genetic results in their practice [89].

We focus on two limitations to the implementation of polygenic risk scores in clinical practice: firstly, the weaker evidence base in application to non-European ancestry; this needs substantial research investment in study collection worldwide and in methodological research to improve genetic prediction in admixed individuals. Secondly, major challenges exist with the interpretation of polygenic risk scores. At its simplest, an individual can be placed on the distribution, 'your PRS lies at the 22.8th percentile', which gives limited information on their lifetime risk. But a more nuanced interpretation is needed, for example, a lifetime risk of disease that combines genetic information with their current age, sex, and environmental and clinical risk factors.

In summary, we have made astounding biological advances in uncovering the genetic component to common complex disorders since the advent of genome-wide association studies in 2007. This is slowly moving from research discovery to clinical implementation, but much work remains in acquiring the necessary research base for polygenic risk scores and in establishing how the information can be best be used and communicated.

Abbreviations

AUC: Area under the receiver operating characteristic curve; BMI: Body mass index; CNV: Copy number variant; DTC: Direct-to-consumer; EHR: Electronic health records; OR: Odds ratio; PRS: Polygenic risk score

Acknowledgements

Not applicable

Authors' contributions

Cathryn Lewis and Evangelos Vassos co-wrote the paper. The authors read and approved the final manuscript.

Funding

This paper was funded by the MRC (MR/N015746/1) and the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

Availability of data and materials

Not applicable

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

Cathryn Lewis is a member of the Research and Development SAB at Myriad Neuroscience. The remaining author declares that there are no competing interests.

Received: 23 October 2019 Accepted: 1 May 2020

Published online: 18 May 2020

References

- Janssens AC. Validity of polygenic risk scores: are we measuring what we think we are? *Hum Mol Genet.* 2019;28(R2):R143–50.
- 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.
- Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, Tyrer JP, Chen TH, Wang Q, Bolla MK, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet.* 2019;104:21–34.
- Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, Adams MJ, Agerbo E, Air TM, Andlauer TMF, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet.* 2018;50:668–81.
- Musliner KL, Mortensen PB, McGrath JJ, Suppli NP, Hougaard DM, Bybjerg-Grauholm J, Baekvad-Hansen M, Andreassen O, Pedersen CB, Pedersen MG, et al. Association of polygenic liabilities for major depression, bipolar disorder, and schizophrenia with risk for depression in the Danish population. *JAMA Psychiatry.* 2019;76:516–25.
- Lewis CM, Hagenaars SP. Progressing polygenic medicine in psychiatry through electronic health records. *JAMA Psychiatry.* 2019;76:470–2.
- Zheutlin AB, Dennis J, Karlsson Linner R, Moscatti A, Restrepo N, Straub P, Ruderfer D, Castro VM, Chen CY, Ge T, et al. Penetrance and pleiotropy of polygenic risk scores for schizophrenia in 106,160 patients across four health care systems. *Am J Psychiatry.* 2019;176(10):846–55.
- Swets JA. Indices of discrimination or diagnostic accuracy: their ROCs and implied models. *Psychol Bull.* 1986;99:100–17.
- Aschard H. A perspective on interaction effects in genetic association studies. *Genet Epidemiol.* 2016;40:678–88.
- Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, Posthuma D. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet.* 2015;47:702–9.
- Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience.* 2019;8(7). <https://doi.org/10.1093/gigascience/gjz082>.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
- Vilhjalmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, Genovese G, Loh PR, Bhatia G, Do R, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet.* 2015;97:576–92.
- Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, Wang H, Zheng Z, Magi R, Esko T, et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Commun.* 2019;10:5086.
- Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol.* 2012;36:214–24.
- Wray NR, Yang J, Goddard ME, Visscher PM. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet.* 2010;6:e1000864.
- Mars NJ, Koskela JT, Ripatti P, Kiiskinen TTJ, Havulinna AS, Lindbohm JV, Ahola-Olli A, Kurki M, Karjalainen J, Palta P, et al. Polygenic and clinical risk scores and their impact on age at onset of cardiometabolic diseases and common cancers. *Nat Med* 2020;26(4):549–557. doi: <https://doi.org/10.1038/s41591-020-0800-0>. Epub 2020 Apr 7.
- Hu Y, Lu Q, Liu W, Zhang Y, Li M, Zhao H. Joint modeling of genetically correlated diseases and functional annotations increases accuracy of polygenic risk prediction. *PLoS Genet.* 2017;13:e1006836.
- Hu Y, Lu Q, Powles R, Yao X, Yang C, Fang F, Xu X, Zhao H. Leveraging functional annotations in genetic risk prediction for human complex diseases. *PLoS Comput Biol.* 2017;13:e1005589.

20. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsdottir BJ, Young AL, Thorgeirsson TE, Benonisdottir S, Oddsson A, Halldorsson BV, Masson G, et al. The nature of nurture: effects of parental genotypes. *Science*. 2018;359:424–8.
21. 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.
22. Morales J, Welter D, Bowler EH, Cerezo M, Harris LW, McMahon AC, Hall P, Junkins HA, Milano A, Hastings E, et al. A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biol*. 2018;19:21.
23. Peterson RE, Kuchenbaecker K, Walters RK, Chen C-Y, Popejoy AB, Periyasamy S, Lam M, Iyegbe C, Strawbridge RJ, Brick L, et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell*. 2019;179(3):589–603.
24. Vassos E, Di Forti M, Coleman J, Iyegbe C, Prata D, Euesden J, O'Reilly P, Curtis C, Koliakou A, Patel H, et al. An examination of polygenic score risk prediction in individuals with first-episode psychosis. *Biol Psychiatry*. 2017;81:470–7.
25. Kim MS, Patel KP, Teng AK, Berens AJ, Lachance J. Genetic disease risks can be misestimated across global populations. *Genome Biol*. 2018;19:179.
26. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet*. 2017;100:635–49.
27. Adeyemo AA, Zaghoul NA, Chen G, Doumatey AP, Leitch CC, Hostelley TL, Nesmith JE, Zhou J, Bentley AR, Shriner D, et al. ZRANB3 is an African-specific type 2 diabetes locus associated with beta-cell mass and insulin response. *Nat Commun*. 2019;10:3195.
28. Suzuki K, Akiyama M, Ishigaki K, Kanai M, Hosoe J, Shojima N, Hozawa A, Kadota A, Kuriki K, Naito M, et al. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nat Genet*. 2019;51:379–86.
29. Lam M, Chen C-Y, Li Z, Martin AR, Bryois J, Ma X, Gaspar H, Ikeda M, Benyamin B, Brown BC, et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet*. 2019;51:1670–8.
30. Marquez-Luna C, Loh PR, South Asian Type 2 Diabetes C, Consortium STD, Price AL. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol*. 2017;41:811–23.
31. Coram MA, Fang H, Candille SI, Assimes TL, Tang H. Leveraging multi-ethnic evidence for risk assessment of quantitative traits in minority populations. *Am J Hum Genet*. 2017;101:218–26.
32. Grinde KE, Qi Q, Thornton TA, Liu S, Shadyab AH, Chan KHK, Reiner AP, Sofer T. Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. *Genet Epidemiol*. 2019;43:50–62.
33. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, Highland HM, Patel YM, Sorokin EP, Avery CL, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature*. 2019;570:514–8.
34. Sugrue LP, Desikan RS. What are polygenic scores and why are they important? *JAMA*. 2019;321:1820–1.
35. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet*. 2018;19:581–90.
36. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837–47.
37. Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study. *Circulation*. 2002;105:310–5.
38. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, Brindle P. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ*. 2008;336:1475–82.
39. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med*. 2007;167:1068–74.
40. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, Lassale CM, Siontis GC, Chiochia V, Roberts C, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ*. 2016;353:i2416.
41. Carver RB, Castera J, Gericke N, Evangelista NA, El-Hani CN. Young adults' belief in genetic determinism, and knowledge and attitudes towards modern genetics and genomics: the PUGGS Questionnaire. *PLoS One*. 2017;12:e0169808.
42. Berg JS, Agrawal PB, Bailey DB Jr, Beggs AH, Brenner SE, Brower AM, Kacici JA, Ceyhan-Birsoy O, Chan K, Chen F, et al. Newborn sequencing in genomic medicine and public health. *Pediatrics*. 2017;139(2). <https://doi.org/10.1542/peds.2016-2252>. Epub 2017 Jan 17.
43. Hollands GJ, French DP, Griffin SJ, Prevost AT, Sutton S, King S, Marteau TM. The impact of communicating genetic risks of disease on risk-reducing health behaviour: systematic review with meta-analysis. *BMJ*. 2016;352:i1102.
44. Silarova B, Sharp S, Usher-Smith JA, Lucas J, Payne RA, Shefer G, Moore C, Girling C, Lawrence K, Tolkien Z, et al. Effect of communicating phenotypic and genetic risk of coronary heart disease alongside web-based lifestyle advice: the INFORM Randomised Controlled Trial. *Heart*. 2019;105:982–9.
45. Austin J. The effect of genetic test-based risk information on behavioral outcomes: a critical examination of failed trials and a call to action. *Am J Med Genet A*. 2015;167A:2913–5.
46. Martin AR, Daly MJ, Robinson EB, Hyman SE, Neale BM. Predicting polygenic risk of psychiatric disorders. *Biol Psychiatry*. 2019;86:97–109.
47. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. *Hum Mol Genet*. 2019;28(R2):R133–42.
48. Aung N, Vargas JD, Yang C, Cabrera CP, Warren HR, Fung K, Tzannis E, Barnes MR, Rotter JI, Taylor KD, et al. Genome-wide analysis of left ventricular image-derived phenotypes identifies fourteen loci associated with cardiac morphogenesis and heart failure development. *Circulation*. 2019;140(16):1318–30.
49. Haas ME, Aragam KG, Emdin CA, Bick AG, International Consortium for Blood P, Hemani G, Davey Smith G, Kathiresan S. Genetic association of albuminuria with cardiometabolic disease and blood pressure. *Am J Hum Genet*. 2018;103:461–73.
50. Antoniou AC, Cunningham AP, Peto J, Evans DG, Lalloo F, Narod SA, Risch HA, Eyfjord JE, Hopper JL, Southey MC, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer*. 2008;98:2015.
51. Dai J, Lv J, Zhu M, Wang Y, Qin N, Ma H, He YQ, Zhang R, Tan W, Fan J, et al. Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese populations. *Lancet Respir Med*. 2019;7:881–91.
52. Oram RA, Redondo MJ. New insights on the genetics of type 1 diabetes. *Curr Opin Endocrinol Diabetes Obes*. 2019;26:181–7.
53. Onengut-Gumuscu S, Chen WM, Robertson CC, Bonnie JK, Farber E, Zhu Z, Oksenberg JR, Brant SR, Bridges SL Jr, Edberg JC, et al. Type 1 diabetes risk in African-ancestry participants and utility of an ancestry-specific genetic risk score. *Diabetes Care*. 2019;42:406–15.
54. Khera AV, Chaffin M, Wade KH, Zahid S, Brancale J, Xia R, Distefano M, Senol-Cosar O, Haas ME, Bick A, et al. Polygenic prediction of weight and obesity trajectories from birth to adulthood. *Cell*. 2019;177:587–96 e589.
55. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shiralil M, Coleman JRI, Hagenars SP, Ward J, Wigmore EM, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22:343–52.
56. Cai N, Revez JA, Adams MJ, Andlauer TFM, Breen G, Byrne EM, Clarke TK, Forstner AJ, Grabe HJ, Hamilton SP, et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nat Genet*. 2020;52:437–47.
57. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–7.
58. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 2003;60:1187–92.
59. Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet*. 2009;373:234–9.
60. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry*. 2000;157:1552–62.
61. Desikan RS, Fan CC, Wang Y, Schork AJ, Cabral HJ, Cupples LA, Thompson WK, Besser L, Kukull WA, Holland D, et al. Genetic assessment of age-associated Alzheimer disease risk: development and validation of a polygenic hazard score. *PLoS Med*. 2017;14:e1002258.
62. Peyrot WJ, Van der Auwera S, Milaneschi Y, Dolan CV, Madden PAF, Sullivan PF, Strohmaier J, Ripke S, Rietschel M, Nivard MG, et al. Does childhood trauma moderate polygenic risk for depression? A meta-analysis of 5765 subjects from the psychiatric genomics consortium. *Biol Psychiatry*. 2018;84:138–47.
63. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, Chasman DJ, Baber U, Mehran R, Rader DJ, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N Engl J Med*. 2016;375:2349–58.

64. Lourida I, Hannon E, Littlejohns TJ, Langa KM, Hypponen E, Kuzma E, Llewellyn DJ. Association of lifestyle and genetic risk with incidence of dementia. *JAMA*. 2019;322(5):430–37.
65. Gillett AC, Vassos E, Lewis CM. Transforming summary statistics from logistic regression to the liability scale: application to genetic and environmental risk scores. *Hum Hered*. 2018;83:210–24.
66. Natarajan P, Young R, Stitzel NO, Padmanabhan S, Baber U, Mehran R, Sartori S, Fuster V, Reilly DF, Butterworth A, et al. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. *Circulation*. 2017;135:2091–101.
67. Mega JL, Stitzel NO, Smith JG, Chasman DI, Caulfield M, Devlin JJ, Nordio F, Hyde C, Cannon CP, Sacks F, et al. 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.
68. Schunkert H, Samani NJ. Statin treatment: can genetics sharpen the focus? *Lancet*. 2015;385:2227–9.
69. Tadors R, Tan HL, Investigators E-N, El Mathari S, Kors JA, Postema PG, Lahrouchi N, Beekman L, Radivojkov-Blagojevic M, Amin AS, et al. Predicting cardiac electrical response to sodium-channel blockade and Brugada syndrome using polygenic risk scores. *Eur Heart J*. 2019;40:3097–107.
70. Garcia-Gonzalez J, Tansey KE, Hauser J, Henigsberg N, Maier W, Mors O, Placentino A, Rietschel M, Souery D, Zagar T, et al. Pharmacogenetics of antidepressant response: a polygenic approach. *Prog Neuro-Psychopharmacol Biol Psychiatry*. 2017;75:128–34.
71. Ward J, Graham N, Strawbridge RJ, Ferguson A, Jenkins G, Chen W, Hodgson K, Frye M, Weinshilboum R, Uher R, et al. Polygenic risk scores for major depressive disorder and neuroticism as predictors of antidepressant response: meta-analysis of three treatment cohorts. *PLoS One*. 2018;13:e0203896.
72. Zhang JP, Robinson D, Yu J, Gallego J, Fleischhacker WW, Kahn RS, Crespo-Facorro B, Vazquez-Bourgon J, Kane JM, Malhotra AK, Lencz T. Schizophrenia polygenic risk score as a predictor of antipsychotic efficacy in first-episode psychosis. *Am J Psychiatry*. 2019;176:21–8.
73. Ramanuj P, Ferenchick EK, Pincus HA. Depression in primary care: part 2—management. *BMJ*. 2019;365:1835.
74. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry*. 2006;163:1905–17.
75. Gibson G. On the utilization of polygenic risk scores for therapeutic targeting. *PLoS Genet*. 2019;15:e1008060.
76. Gibson G. Going to the negative: genomics for optimized medical prescription. *Nat Rev Genet*. 2019;20:1–2.
77. Niemi MEK, Martin HC, Rice DL, Gallone G, Gordon S, Kelemen M, McAloney K, McRae J, Radford EJ, Yu S, et al. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature*. 2018;562:268–71.
78. Oetjens MT, Kelly MA, Sturm AC, Martin CL, Ledbetter DH. Quantifying the polygenic contribution to variable expressivity in eleven rare genetic disorders. *Nat Commun*. 2019;10:4897.
79. Kuchenbaecker KB, McGuffog L, Barrowdale D, Lee A, Soucy P, Dennis J, Domchek SM, Robson M, Spurdle AB, Ramus SJ, et al. Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. *J Natl Cancer Inst*. 2017;109(7). <https://doi.org/10.1093/jnci/djw302>.
80. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, Babb de Villiers C, Izquierdo A, Simard J, Schmidt MK, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med*. 2019;21:1708–18.
81. Bergen SE, Ploner A, Howrigan D, Group CNVA, the Schizophrenia Working Group of the Psychiatric Genomics C, O'Donovan MC, Smoller JW, Sullivan PF, Sebat J, Neale B, Kendler KS. Joint contributions of rare copy number variants and common SNPs to risk for schizophrenia. *Am J Psychiatry*. 2019;176:29–35.
82. Tansey KE, Rees E, Linden DE, Ripke S, Chambert KD, Moran JL, McCarroll SA, Holmans P, Kirov G, Walters J, et al. Common alleles contribute to schizophrenia in CNV carriers. *Mol Psychiatry*. 2016;21:1085–9.
83. Vassos E, Collier DA, Holden S, Patch C, Rujescu D, St Clair D, Lewis CM. Penetrance for copy number variants associated with schizophrenia. *Hum Mol Genet*. 2010;19:3477–81.
84. Cleyne I, Engchuan W, Hestand MS, Heung T, Holleman AM, Johnston HR, Monfeuga T, DM MD-MG, Gur RE, Morrow BE, et al. Genetic contributors to risk of schizophrenia in the presence of a 22q11.2 deletion. *Mol Psychiatry*. 2020. <https://doi.org/10.1038/s41380-020-0654-3>. [Epub ahead of print].
85. Nelson SC, Bowen DJ, Fullerton SM. Third-party genetic interpretation tools: a mixed-methods study of consumer motivation and behavior. *Am J Hum Genet*. 2019;105:122–31.
86. Bolli A, Di Domenico P, Bottà G. Software as a service for the genomic prediction of complex diseases. *bioRxiv*. 2019:763722. <https://doi.org/10.1101/763722>.
87. Multhaup ML, Kita R, Krock B, Eriksson N, Fontanillas P, Aslibekyan S, Del Gobbo L, Shelton JF, Tennen RI, Lehman A, et al. The science behind 23andMe's Type 2 Diabetes report. In: White paper 23–19: 23andMe; 2019.
88. Dudbridge F, Pashayan N, Yang J. Predictive accuracy of combined genetic and environmental risk scores. *Genet Epidemiol*. 2018;42:4–19.
89. Pet DB, Holm IA, Williams JL, Myers MF, Novak LL, Brothers KB, Wiesner GL, Clayton EW. Physicians' perspectives on receiving unsolicited genomic results. *Genet Med*. 2019;21:311–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.