

Predictive accuracy of combined genetic and environmental risk scores

Frank Dudbridge^{1,2,3,4}  | Nora Pashayan⁵ | Jian Yang^{6,7}

¹Department of Health Sciences, University of Leicester, Leicester, United Kingdom

²Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom

³Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom

⁴MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom

⁵Department of Applied Health Research, University College London, London, United Kingdom

⁶Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia

⁷Queensland Brain Institute University of Queensland, Brisbane, Queensland, Australia

Correspondence

Frank Dudbridge, Department of Health Sciences, University of Leicester, Leicester LE1 7RH, UK.

Email: frank.dudbridge@leicester.ac.uk

Funding information

Grant sponsor: MRC; Grant numbers: MR/K006215/1 and MR/M026175/1.

ABSTRACT

The substantial heritability of most complex diseases suggests that genetic data could provide useful risk prediction. To date the performance of genetic risk scores has fallen short of the potential implied by heritability, but this can be explained by insufficient sample sizes for estimating highly polygenic models. When risk predictors already exist based on environment or lifestyle, two key questions are to what extent can they be improved by adding genetic information, and what is the ultimate potential of combined genetic and environmental risk scores? Here, we extend previous work on the predictive accuracy of polygenic scores to allow for an environmental score that may be correlated with the polygenic score, for example when the environmental factors mediate the genetic risk. We derive common measures of predictive accuracy and improvement as functions of the training sample size, chip heritabilities of disease and environmental score, and genetic correlation between disease and environmental risk factors. We consider simple addition of the two scores and a weighted sum that accounts for their correlation. Using examples from studies of cardiovascular disease and breast cancer, we show that improvements in discrimination are generally small but reasonable degrees of reclassification could be obtained with current sample sizes. Correlation between genetic and environmental scores has only minor effects on numerical results in realistic scenarios. In the longer term, as the accuracy of polygenic scores improves they will come to dominate the predictive accuracy compared to environmental scores.

KEYWORDS

polygenic score, reclassification, risk prediction, risk score, ROC curve

1 | INTRODUCTION

Predicting the individual risk of disease is one of the major goals of epidemiology and, because most common diseases have a heritable component, there has long been interest in using genotype data to inform prediction of disease onset, prognosis, or treatment response. Indeed for Mendelian disorders genetic prediction has established clinical applications in counselling, prophylactic intervention, and embryonic screen-

ing. For the common, complex disorders however, progress has to date been slow (Abraham & Inouye, 2015; Chatterjee, Shi, & Garcia-Closas, 2016; Jostins & Barrett, 2011). Despite the success of genome-wide association studies (GWAS) in identifying numerous risk variants for many disorders, these variants typically explain a small proportion of the variation in risk; a number of studies have examined the predictive accuracy of GWAS “hits” and generally found limited utility for risk prediction (Eriksson et al., 2015; Talmud et al., 2015;

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2017 The Authors. *Genetic Epidemiology* Published by Wiley Periodicals, Inc.

Weissfeld et al., 2015), although there have been some successes (Maas et al., 2016; Patel et al., 2016). The reasons are that the associated markers individually have small effects on risk, and the markers discovered to date are a small fraction of the total complement of risk variants. The realization that most, perhaps all, complex traits are polygenic, that is, determined by thousands of genetic variants with small effects, has motivated a shift toward thinking about the genetic basis of disease as a single entity (Dudbridge, 2016). Under this paradigm, the whole genome should be regarded as a risk predictor, but attempts in this direction have so far also yielded modest results (Evans, Visscher, & Wray, 2009; Locke et al., 2015; Maier et al., 2015).

However, given the substantial heritability of many common disorders, it remains true that we should improve risk prediction if we could measure the heritable component accurately (Pharoah et al., 2002; Wray, Goddard, & Visscher, 2007; Wray, Yang, Goddard, & Visscher, 2010). Indeed, a crude measure of polygenic risk, namely family history, is already widely used as a risk predictor (Valdez, Yoon, Qureshi, Green, & Khoury, 2010). An important insight is that the currently disappointing performance of polygenic prediction can be explained by insufficient sample sizes available for estimating genetic effects (Chatterjee et al., 2013; Dudbridge, 2013). Essentially, sampling variation accumulates across thousands of variants so that the polygenic predictor has high measurement error even if standard errors are low on each individual variant. However, international consortium efforts and national biobank projects are now approaching the sizes at which accurate genetic predictors could be derived, so the performance of genetic prediction may well improve toward useful levels in the next few years.

Nevertheless, the heritability imposes a limit on the accuracy of genetic prediction, even if we knew all the variants affecting risk and their precise effect sizes (Clayton, 2009). For many disorders, this limit falls short of levels usually regarded as clinically useful, and the only way we could improve predictive accuracy would be by adding further non-genetic factors into the risk score. Indeed for some conditions, for example, cardiovascular disease (CVD) and breast cancer, epidemiological risk scores are already available and in clinical use. The question then is whether inclusion of genetic information could improve the predictive accuracy to a useful degree. As with studies of genetic prediction alone, efforts to improve existing predictors by adding in genetic data have met with only moderate success (Morris et al., 2016; Talmud et al., 2015; Wacholder et al., 2010).

Theoretical treatments of genetic prediction have mainly considered prediction from genotypes alone, but not the common scenario in which both genetic and nongenetic predictors are combined. An exception is work of Garcia-Closas, Gunsoy, and Chatterjee (2014), who argue that addition of genetic data to questionnaire-based risk factors could pro-

vide useful levels of risk stratification for breast cancer. Here, we extend previous work on the predictive accuracy of polygenic risk scores (Dudbridge, 2013) to include a nonpolygenic component. For convenience, we term all nonpolygenic risk factors as “environmental,” although such factors could include major genes, behavioral and other factors such as age that arguably do not fit that description. A key concern is that many environmental risk factors are themselves heritable, and may therefore mediate some or all of the genetic risk. Indeed, comparable resources have been directed toward GWAS of heritable risk factors such as lipid fractions (Willer et al., 2013) and body mass index (BMI) (Locke et al., 2015) as have been toward clinical outcomes. On the other hand, for environments that are difficult to measure well (e.g., alcohol consumption), it is conceivable that their heritable component may act as a more accurate predictor of the outcome than the measured environment itself (Verhulst, Neale, & Kendler, 2015).

Here, we show how under a quantitative genetics model, commonly used measures of predictive accuracy and improvement can be expressed in terms of the genetic variance of an outcome of interest (its “chip heritability”), genetic variance of an environmental risk score, variance in the outcome explained by the environmental score, and genetic correlation between the outcome and the environmental score. These are quantities that are readily estimated by existing methods (Bulik-Sullivan et al., 2015; Lee, Yang, Goddard, Visscher, & Wray, 2012), allowing examples to be drawn from recent literature to illustrate the prospects for prediction from combined genetic and environmental factors. As in previous work (Dudbridge, 2013), we allow for the estimation of a polygenic risk score from finite training data, but in contrast to other work (Garcia-Closas et al., 2014; So & Sham, 2010), we explicitly consider the genetic correlation between environment and outcome and its implications for the computation of combined risk scores.

2 | MATERIALS AND METHODS

2.1 | Quantitative Model

Consider two linear models for an outcome Y , the first in terms of a scalar environmental factor X :

$$E(Y) = X$$

the second in terms of m genetic markers:

$$E(Y) = \beta'G = \sum_{i=1}^m \beta_i G_i,$$

where β is an m -vector of coefficients β_i . Assume that the effect of the environmental factor X on Y is known precisely,

so that without loss of generality X may be assumed to have a unit effect on Y . Denote by R_{XY}^2 the proportion of variance in Y explained by X . Assuming also that Y is standardized, then $\text{var}(X) = R_{XY}^2$.

Assume that genotypes G are independent (i.e., in linkage equilibrium) and standardized. Thus, for single nucleotide polymorphisms (SNPs) in Hardy–Weinberg equilibrium, the usual additive coding gives $G_i = (A_i - 2f_i)/(2f_i(1 - f_i))^{1/2}$, where A_i is the number of minor alleles and f_i is the minor allele frequency at SNP i . The genetic effects β_i are regarded as fixed across samples, but random over $i = 1, \dots, m$ with $E(\beta_i) = 0$, $\text{var}(\beta_i) = m^{-1}\sigma_G^2$. Then the genetic variance of Y is σ_G^2 , and because Y is standardized the chip heritability of markers G , denoted by R_{GY}^2 , is equal to σ_G^2 .

The separate marginal genetic and environmental models for Y reflect common practice in fitting such models. Many environmental risk factors have been extensively studied prior to the genomics era, and accurate estimates of their effects have been obtained from large epidemiological studies without adjustment for genetic factors. Here it is assumed that the environmental effect size is known with negligible standard error so that we may ignore sampling error in the effect of X .

Genetic effect sizes are often estimated by large consortia in which it may be impractical to adjust for any but the simplest risk factors (e.g., age and sex) across contributing studies. However, estimation of the marginal genetic effects is a key determinant of predictive accuracy. Let the genetic effects be estimated from a training sample of size n with the polygenic risk score then defined by:

$$\hat{S} = \hat{\beta}'G.$$

For ordinary least squares estimates of β , Dudbridge (Dudbridge, 2013) showed that:

$$\text{cov}(\hat{S}, Y) = 2R_{GY}^2 [\Phi(r_0) - \Phi(r_1) + r_1\varphi(r_1) - r_0\varphi(r_0)], \quad (1)$$

$$\begin{aligned} \text{var}(\hat{S}) &= 2m\pi_0 n^{-1} [\Phi(q_0) - \Phi(q_1) + q_1\varphi(q_1) - q_0\varphi(q_0)] \\ &\quad + 2m(1 - \pi_0)((1 - \pi_0)^{-1}m^{-1}R_{GY}^2 + n^{-1}) \\ &\quad [\Phi(r_0) - \Phi(r_1) + r_1\varphi(r_1) - r_0\varphi(r_0)], \end{aligned} \quad (2)$$

where π_0 is a proportion of markers assumed to have no effect ($\beta_i = 0$), with the rest assumed to have effects following a normal distribution; $q_0 = \Phi^{-1}(1 - \frac{1}{2}p_0)$, $q_1 = \Phi^{-1}(1 - \frac{1}{2}p_1)$, where p_0, p_1 are lower and upper two-tailed P -values to select markers into the polygenic score; and $r_0 = q_0(n(1 - \pi_0)^{-1}m^{-1}R_{GY}^2 + 1)^{-\frac{1}{2}}$, $r_1 = q_1(n(1 - \pi_0)^{-1}m^{-1}R_{GY}^2 + 1)^{-\frac{1}{2}}$.

Now consider a polygenic component to the environmental factor X . Denote the chip heritability of X by R_{GX}^2 and the

genetic correlation between X and Y by ρ . That is, ρ is the correlation between the β_i and the corresponding genetic effects on X : thus it is a property of the markers G and should properly be called a chip correlation. Then the chip covariance between X and Y is:

$$\rho\sqrt{m^{-2}R_{GY}^2R_{XY}^2R_{GX}^2}$$

and so

$$\text{cov}(S, X) = m\rho\sqrt{m^{-2}R_{GY}^2R_{XY}^2R_{GX}^2} = \rho\sqrt{R_{GY}^2R_{XY}^2R_{GX}^2}.$$

The covariance of the estimated polygenic score \hat{S} with the environmental score X also follows from the results of Dudbridge:

$$\begin{aligned} \text{cov}(\hat{S}, X) &= 2\text{cov}(S, X) [\Phi(r_0) - \Phi(r_1) + r_1\varphi(r_1) - r_0\varphi(r_0)] \\ &= 2\rho\sqrt{R_{GY}^2R_{XY}^2R_{GX}^2} [\Phi(r_0) - \Phi(r_1) + r_1\varphi(r_1) \\ &\quad - r_0\varphi(r_0)]. \end{aligned} \quad (3)$$

Now consider a combined score formed as the weighted sum of polygenic and environmental scores:

$$\hat{S}_{comb} = w_1\hat{S} + w_2X. \quad (4)$$

The coefficient of determination for the combined score is:

$$R_{\hat{S}_{comb}Y}^2 = \frac{\text{cov}(\hat{S}_{comb}, Y)^2}{\text{var}(\hat{S}_{comb})}, \quad (5)$$

where

$$\text{cov}(\hat{S}_{comb}, Y) = w_1\text{cov}(\hat{S}, Y) + w_2R_{XY}^2, \quad (6)$$

$$\text{var}(\hat{S}_{comb}) = w_1^2\text{var}(\hat{S}) + w_2^2R_{XY}^2 + 2w_1w_2\text{cov}(\hat{S}, X). \quad (7)$$

A simple choice of weights just adds the two scores together, $w_1 = w_2 = 1$. Although this seems naive when there is genetic correlation between X and Y , it is a very commonly used approach. A better choice is to take the ordinary least squares solution to the linear model:

$$E(Y) = \hat{S}_{comb},$$

which is

$$\begin{aligned} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} &= \begin{bmatrix} \text{var}(\hat{S}) & \text{cov}(\hat{S}, X) \\ \text{cov}(\hat{S}, X) & R_{XY}^2 \end{bmatrix}^{-1} \begin{bmatrix} \text{cov}(\hat{S}, Y) \\ \text{cov}(X, Y) \end{bmatrix} \\ &= \begin{bmatrix} \text{var}(\hat{S}) & \text{cov}(\hat{S}, X) \\ \text{cov}(\hat{S}, X) & R_{XY}^2 \end{bmatrix}^{-1} \begin{bmatrix} \text{cov}(\hat{S}, Y) \\ R_{XY}^2 \end{bmatrix}. \end{aligned}$$

To summarize thus far, the coefficient of determination for the combined score is expressed in terms of the genetic

TABLE 1 Parameters and Notation of Polygenic Model

Design Parameters	Interpretation
n	Training sample size
m	Total number of independent markers in genotyping panel
p_0, p_1	Lower and upper P values to select markers into polygenic score
Genetic model parameters	
R_{GY}^2	Variance in Y explained by genotypes \mathbf{G} ; chip heritability of Y
R_{XY}^2	Variance in Y explained by environment X
R_{GX}^2	Variance in X explained by genotypes \mathbf{G} ; chip heritability of X
ρ	Genetic (chip) correlation between X and Y
π_0	Proportion of markers with no effect on Y

model parameters listed in Table 1 and the derived quantities $\text{cov}(\hat{S}, Y)$, $\text{var}(\hat{S})$, and $\text{cov}(\hat{S}, X)$ (Equations (1), (2), and (3)). For continuous Y , this coefficient of determination is the R^2 between predicted and observed traits, a common measure of predictive accuracy.

2.2 | Binary Outcomes

For binary outcomes, the liability threshold (or probit) model closely approximates the linear model for the small effects expected on polygenic traits, and allows several measures of predictive accuracy to be expressed analytically (So & Sham, 2010). Let Y now be an unobserved liability, distributed as standard normal, with the binary outcome $D = 1$ when Y exceeds a fixed threshold τ , 0 otherwise. Denote by K the prevalence of the trait and by P the case/control sampling ratio in the training data, defined to equal K in a prospective study. The variance of the observed D in the training data is $P(1 - P)$ and effects on Y are transformed to effects on D by the factor $c = \varphi(\tau) \frac{P(1-P)}{K(1-K)}$ (Lee, Wray, Goddard, & Visscher, 2011). Let \hat{S} now be the polygenic score with effects estimated on the observed D and transformed back to the liability scale by c^{-1} . Then analogous to Equations (1), (2), and (3),

$$\text{cov}(\hat{S}, Y) = 2R_{GY}^2 [\Phi(r_0) - \Phi(r_1) + r_1\varphi(r_1) - r_0\varphi(r_0)],$$

$$\text{var}(\hat{S}) = c^{-2}2m\pi_0P(1 - P)n^{-1}$$

$$\begin{aligned} & [\Phi(q_0) - \Phi(q_1) + q_1\varphi(q_1) - q_0\varphi(q_0)] \\ & + 2mc^{-2}(1 - \pi_0)((1 - \pi_0)^{-1}m^{-1}c^2R_{GY}^2 \\ & + P(1 - P)n^{-1}) [\Phi(r_0) - \Phi(r_1) + r_1\varphi(r_1) \\ & - r_0\varphi(r_0)], \end{aligned}$$

where now

$$r_0 = q_0(nP^{-1}(1 - P)^{-1}(1 - \pi_0)^{-1}m^{-1}c^2R_{GY}^2 + 1)^{-\frac{1}{2}},$$

$$r_1 = q_1(nP^{-1}(1 - P)^{-1}(1 - \pi_0)^{-1}m^{-1}c^2R_{GY}^2 + 1)^{-\frac{1}{2}}.$$

The combined score \hat{S}_{comb} is calculated as before from Equation (4) and its coefficient of determination $R_{\hat{S}_{comb}Y}^2$ from Equations (5), (6), and (7).

2.3 | Area under ROC Curve

The accuracy of predicting a binary outcome is often assessed by the area under the receiver operator characteristic curve (AUC), which measures the discrimination concordance between risk scores and outcomes. That is, AUC is the probability that a subject with the trait has a higher risk score than a subject without the trait. The central limit theorem implies that the polygenic score \hat{S} is normally distributed. Assuming that the environmental score X also is normally distributed:

$$AUC = \Phi \left(\frac{E(\hat{S}_{comb}|D=1) - E(\hat{S}_{comb}|D=0)}{\sqrt{\text{var}(\hat{S}_{comb}|D=1) + \text{var}(\hat{S}_{comb}|D=0)}} \right), \quad (8)$$

where

$$E(\hat{S}_{comb}|D=1) = \frac{\varphi(\tau)}{K} R_{\hat{S}_{comb}Y}^2,$$

$$\begin{aligned} & \text{var}(\hat{S}_{comb}|D=1) \\ & = R_{\hat{S}_{comb}Y}^2 \left[1 - \frac{\varphi(\tau)}{K} R_{\hat{S}_{comb}Y}^2 \left(\frac{\varphi(\tau)}{K} - \tau \right) \right], \end{aligned}$$

$$E(\hat{S}_{comb}|D=0) = \frac{-\varphi(\tau)}{1 - K} R_{\hat{S}_{comb}Y}^2,$$

$$\begin{aligned} & \text{var}(\hat{S}_{comb}|D=0) \\ & = R_{\hat{S}_{comb}Y}^2 \left[1 - \frac{\varphi(\tau)}{1 - K} R_{\hat{S}_{comb}Y}^2 \left(\frac{\varphi(\tau)}{K} + \tau \right) \right]. \end{aligned}$$

Full derivations of these equalities have been provided previously (Dudbridge, 2013; So & Sham, 2010; Wray et al., 2010).

2.4 | Net Reclassification Index

The AUC has some limitations that have been well documented (Cook, 2007), notably that it reflects only the ranking of subjects, but not their absolute risk levels. Furthermore, substantial increases in AUC may be hard to achieve

even with the addition of informative markers. This has led to the proposal of new measures of incremental predictive accuracy, of which the net reclassification improvement (NRI) has proved popular (Pencina, D'Agostino, D'Agostino, & Vasan, 2008). Given discrete risk categories defined by absolute risk thresholds, the NRI compares the classification of subjects under an initial risk predictor to their reclassification under an augmented predictor including an additional marker. Specifically, the NRI is calculated in cases as the proportion correctly reclassified to a higher risk category minus the proportion incorrectly reclassified to a lower category. Similarly, the NRI is calculated in controls as the proportion correctly reclassified to a lower category minus the proportion incorrectly reclassified to a higher category. The case and control NRI could be combined into an overall NRI (Kerr et al., 2014; Pencina, D'Agostino, & Steyerberg, 2011).

Currently, there is substantial interest in whether a polygenic risk score adds worthwhile information to an existing environmental score. Under the liability threshold model, the NRI can be calculated from tail probabilities of the joint distribution of liability, environmental, and combined scores (So & Sham, 2010). Here the calculation is adjusted to allow for chip correlation between environment X and liability Y .

Consider two categories of risk defined by a threshold a . (This calculation can be extended if desired to a greater number of categories.) Under the liability threshold model, the risk for environmental score X is:

$$1 - \Phi\left(\frac{\tau - X}{\sqrt{1 - R_{XY}^2}}\right) \quad (9)$$

and so the risk equals the threshold a when

$$X = a_X = \tau - \sqrt{1 - R_{XY}^2} \Phi^{-1}(1 - a).$$

The combined score, though constructed as an estimate of liability, may have variance differing from 1 depending on the choice of the weights w_1, w_2 . To obtain absolute risks from the combined score, it is therefore calibrated to the liability by rescaling by its regression coefficient on Y :

$$\gamma = \frac{\text{cov}(\hat{S}_{comb}, Y)}{\text{var}(\hat{S}_{comb})}$$

calculated from Equations (6) and (7), so that

$$\text{var}(\gamma \hat{S}_{comb}) = \frac{\text{cov}(\hat{S}_{comb}, Y)^2}{\text{var}(\hat{S}_{comb})} = R_{\hat{S}_{comb}Y}^2.$$

(The least squares solutions of w_1, w_2 do give $\gamma = 1$.) The risk equals the threshold a when

$$\gamma \hat{S}_{comb} = a_S = \tau - \sqrt{1 - R_{\hat{S}_{comb}Y}^2} \Phi^{-1}(1 - a).$$

Again assuming that both polygenic and environmental scores are normally distributed, the joint distribution of liability, X , and the calibrated combined score is trivariate normal with mean zero and variance-covariance matrix:

$$\begin{bmatrix} 1 & R_{XY}^2 & R_{\hat{S}_{comb}Y}^2 \\ R_{XY}^2 & R_{XY}^2 & \text{cov}(\gamma \hat{S}_{comb}, X) \\ R_{\hat{S}_{comb}Y}^2 & \text{cov}(\gamma \hat{S}_{comb}, X) & R_{\hat{S}_{comb}Y}^2 \end{bmatrix},$$

where

$$\text{cov}(\gamma \hat{S}_{comb}, X) = \gamma(\text{cov}(\hat{S}, X) + R_{XY}^2)$$

is calculated using Equation (3).

The case NRI is:

$$\begin{aligned} & \Pr(\gamma \hat{S}_{comb} > a_S, X \leq a_X | D = 1) \\ & - \Pr(\gamma \hat{S}_{comb} \leq a_S, X > a_X | D = 1), \end{aligned}$$

where

$$\begin{aligned} & \Pr(\gamma \hat{S}_{comb} > a_S, X \leq a_X | D = 1) \\ & = \int_{\tau}^{\infty} \int_{-\infty}^{a_X} \int_{a_S}^{\infty} \varphi_3(l, x, s) ds dx dl / K, \end{aligned}$$

$$\begin{aligned} & \Pr(\gamma \hat{S}_{comb} \leq a_S, X > a_X | D = 1) \\ & = \int_{\tau}^{\infty} \int_{a_X}^{\infty} \int_{-\infty}^{a_S} \varphi_3(l, x, s) ds dx dl / K, \end{aligned}$$

φ_3 being the trivariate normal density function just defined.

Similarly the control NRI is:

$$\begin{aligned} & \Pr(\gamma \hat{S}_{comb} \leq a_S, X > a_X | D = 0) \\ & - \Pr(\gamma \hat{S}_{comb} > a_S, X \leq a_X | D = 0), \end{aligned}$$

where

$$\begin{aligned} & \Pr(\gamma \hat{S}_{comb} \leq a_S, X > a_X | D = 0) \\ & = \int_{-\infty}^{\tau} \int_{a_X}^{\infty} \int_{-\infty}^{a_S} \varphi_3(l, x, s) ds dx dl / (1 - K), \end{aligned}$$

$$\begin{aligned} & \Pr(\gamma \hat{S}_{comb} > a_S, X \leq a_X | D = 0) \\ &= \int_{-\infty}^{\tau} \int_{-\infty}^{a_X} \int_{a_S}^{\infty} \varphi_3(l, x, s) ds dx dl / (1 - K). \end{aligned}$$

2.5 | Continuous NRI and IDI

The discrete categories used in the NRI may be a limitation when such thresholds have not been fixed in practice or when comparing predictors between studies with different characteristics. Two alternatives that do not rely on discrete categories are the continuous NRI (Pencina et al., 2011) and the integrated discrimination improvement (IDI) (Pencina et al., 2008).

In cases, the continuous NRI is the proportion whose risk score increases under the augmented predictor minus the proportion whose score decreases. Similarly in the controls it is the proportion whose risk score decreases minus that whose score increases. Therefore, in cases the continuous NRI is:

$$\begin{aligned} & \Pr(\gamma \hat{S}_{comb} > X | D = 1) - \Pr(\gamma \hat{S}_{comb} \leq X | D = 1) \\ &= 2 \Pr(\gamma \hat{S}_{comb} > X | D = 1) - 1, \end{aligned}$$

where

$$\Pr(\gamma \hat{S}_{comb} > X | D = 1) = \int_{\tau}^{\infty} \int_0^{\infty} \varphi_2(l, x) dx dl / K$$

with φ_2 the bivariate normal density of liability and $\gamma \hat{S}_{comb} - X$. This distribution has mean 0 and variance-covariance matrix:

$$\begin{bmatrix} 1 & R_{\hat{S}_{comb}Y}^2 - R_{XY}^2 \\ R_{\hat{S}_{comb}Y}^2 - R_{XY}^2 & R_{\hat{S}_{comb}Y}^2 + R_{XY}^2 - 2\text{cov}(\gamma \hat{S}_{comb}, X) \end{bmatrix}.$$

Similarly in controls, the continuous NRI is:

$$2 \int_{-\infty}^{\tau} \int_{-\infty}^0 \varphi_2(l, x) dx dl / (1 - K) - 1.$$

The IDI is the mean change in risk among cases minus the mean change in controls:

$$\begin{aligned} & E_{\hat{S}_{comb}} \left[1 - \Phi \left(\frac{\tau - \gamma \hat{S}_{comb}}{\sqrt{1 - R_{\hat{S}_{comb}Y}^2}} \right) | D = 1 \right] - E_X \left[1 - \Phi \left(\frac{\tau - X}{\sqrt{1 - R_{XY}^2}} \right) | D = 1 \right] \\ & - E_{\hat{S}_{comb}} \left[1 - \Phi \left(\frac{\tau - \gamma \hat{S}_{comb}}{\sqrt{1 - R_{\hat{S}_{comb}Y}^2}} \right) | D = 0 \right] + E_X \left[1 - \Phi \left(\frac{\tau - X}{\sqrt{1 - R_{XY}^2}} \right) | D = 0 \right], \\ &= \int_{-\infty}^{\infty} \Phi \left(\frac{\tau - l}{\sqrt{1 - R_{\hat{S}_{comb}Y}^2}} \right) \left[\varphi \left(\frac{l - E(\gamma \hat{S}_{comb} | D=0)}{\sqrt{\text{var}(\gamma \hat{S}_{comb} | D=0)}} \right) - \varphi \left(\frac{l - E(\gamma \hat{S}_{comb} | D=1)}{\sqrt{\text{var}(\gamma \hat{S}_{comb} | D=1)}} \right) \right] dl \\ & - \int_{-\infty}^{\infty} \Phi \left(\frac{\tau - l}{\sqrt{1 - R_{XY}^2}} \right) \left[\varphi \left(\frac{l - E(X | D=0)}{\sqrt{\text{var}(X | D=0)}} \right) - \varphi \left(\frac{l - E(X | D=1)}{\sqrt{\text{var}(X | D=1)}} \right) \right] dl \end{aligned}$$

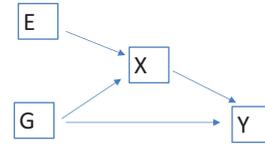


FIGURE 1 Directed acyclic graph showing correlation between polygenic and environmental scores arising from mediation

Note: G: polygenic score; E: nongenetic risk factors; X: environmental risk score; Y: outcome.

where the conditional means and variances have been obtained previously for calculation of AUC.

3 | RESULTS

3.1 | Cardiovascular Disease

The analytic results are now compared to some published studies of the predictive improvement from genotype data. In many countries, lipid lowering medication is advised for subjects whose medium term risk of a coronary event exceeds a threshold (Goff et al., 2014; JBS3 Board, 2014). Environmental risk scores have been developed using data from longitudinal studies, including the American College of Cardiology/American Heart Association Score (Goff et al., 2014), Framingham Risk Score (FRS) (D'Agostino et al., 2008), and QRISK-2 equation (Hippisley-Cox, Coupland, Robson, & Brindle, 2010). Risk factors commonly included are age, gender, components of total cholesterol, smoking, BMI, and blood pressure. A 10-year risk of 20% has been traditionally used as a threshold although recent guidelines have suggested lower thresholds, thus widening the prescription of such medications, primarily statins (Hippisley-Cox et al., 2010). Figure 1 is a directed acyclic graph illustrating how the environmental risk score mediates some of the genetic risk, while also including nongenetic factors.

Several studies have assessed the incremental value of a limited number of SNPs compared to traditional risk factors (Hughes et al., 2012; Morrison et al., 2007; Thanassoulis et al., 2012; Tikkanen, Havulinna, Palotie, Salomaa, & Ripatti, 2013). In general, very small improvements in AUC have been observed, but some promise has been identified for reclassification, particularly for individuals at intermediate risk. It is indeed those individuals who may be reluctant to adopt a medication and thus for whom accurate risk prediction is desirable. Two recent studies have used genome-wide significant SNPs from the CARDIoGRAMplusC4D consortium (Deloukas et al., 2013), the largest study yet conducted in CVD, representing the state of the art in terms of the number of significantly associated SNPs and the precision of their estimated effects.

de Vries et al. (2015) considered prediction of coronary heart disease (CHD) in the Rotterdam Study using 49 SNPs significant at $P < 5 \times 10^{-8}$ and with 152 SNPs significant at approximately $P < 10^{-3}$. Improvements over traditional risk factors were small: for example, in comparison to an environmental score including established risk factors, the AUC improved from 0.716 to 0.718 and the total NRI was 0.014 across four risk categories using the 49 SNP score. For the 152 SNP score the AUC improved to 0.719 and the NRI was 0.022.

Morris et al. (2016) considered prediction of CVD (defined as CHD or stroke) in the UCLEB consortium using the same 49 SNPs with $P < 5 \times 10^{-8}$ and three additional SNPs likewise associated with stroke. In comparison with the QRISK-2 score, the AUC actually decreased from 0.635 to 0.623, whereas the NRI was 0.0025 for a 10-year risk of 10%.

In a further recent study, Abraham et al. (2016) used all the SNPs genotyped in CARDIoGRAMplusC4D with some pruning by linkage disequilibrium. Over two target samples, AUC improved by approximately 0.016 and the continuous NRI over the FRS was between 0.147 and 0.195 in cases, and between 0.102 and 0.175 in controls.

The present model is difficult to fit to these studies because the full CARDIoGRAMplusC4D sample was genotyped on the MetaboChip, a targeted array of approximately 200,000 SNPs chosen on the basis of prior association with several cardio-metabolic disorders. The genomic coverage and chip heritability of this array falls short of a full GWAS array; the proportion of null SNPs may be lower than in a full array, but the distribution of effect sizes will exhibit selection bias. The published list of SNP effects used by these recent studies consists of 79,138 SNPs present on both GWAS and MetaboChip products. To simplify the presentation, it is assumed that a full GWAS array was used, which could be approximated by 100,000 independent SNPs. A genetic model was identified that closely matches the results of Morris et al. (2016), which were more readily compared to the present theory than were those of the other studies. Although this model may depart from the truth, it will serve the main aim of illustrating the effects of marker selection and sample size on the incremental accuracy.

The training data from CARDIoGRAMplusC4D include 63,746 cases and 130,681 controls. Taking the chip heritability of CHD as $R_{GY}^2 = 0.3$, proportion of null SNPs $\pi_0 = 0.8$ and disease prevalence $K = 0.15$ gives an expected AUC of 0.536 when SNPs are selected by $P < 5 \times 10^{-8}$, similar to the result of Morris et al. For the environmental score, the chip heritability of FRS, which includes similar risk factors to QRISK-2, is approximately 0.3 (Simonson, Wills, Keller, & McQueen, 2011). To obtain an AUC of 0.635, as per QRISK-2 in Morris et al., Equation (8) implies $\sigma_X^2 = 0.052$.

With these parameters, and selecting SNPs into the score by $P < 5 \times 10^{-8}$, Equation (8) gives AUC of 0.622 for the unweighted combined score when $\rho = 0.1$, or AUC of 0.621

when $\rho = 0.4$, both lower than the AUC of 0.635 for the environmental score alone and consistent with the results of Morris et al. The least squares weighted score gives AUC of 0.623 and 0.622, respectively. Such a decrease in AUC is a known effect when adding in a weakly predictive marker (Pepe, Fan, Feng, Gerds, & Hilden, 2015). However, the present theory suggests that the predictive accuracy of the unweighted score would be maximized when selecting SNPs by $P < 0.033$, giving AUC of 0.693; for the weighted score, the AUC would be 0.700 when $P < 0.053$. Clearly, in contrast to common practice, predictive accuracy is improved by more liberal selection of SNPs into the polygenic score. These results are summarized in Table 2, which also shows that under these parameters, AUC of nearly 0.8 could be achieved by the polygenic score with a very large training sample and improved only slightly by the environmental score. Supplementary Table S1 shows results with the prevalence reduced to 0.06, and Table 3 with the proportion of null SNPs increased to 0.95 (Palla & Dudbridge, 2015), both cases showing modest increase in AUC with similar qualitative conclusions.

Figures 2 and 3 show the expected AUC as a function of training sample size for the unweighted and least squares weighted scores. Notably, the results are almost identical for unweighted and weighted scores except at low training sample sizes. Although at a given sample size the degree of genetic correlation has small effects on AUC, it can have a strong bearing on the sample size required to reach a critical level of AUC. For example, to reach AUC of 0.75, the required sample size is approximately 159,000 when $\rho = 0.1$ but 210,000 when $\rho = 0.4$, and 284,000 for the polygenic score alone. The curves level out at around 200,000 cases, beyond which further gains are small. The CARDIoGRAMplusC4D consortium has recently exceeded this sample size and its prediction studies, when completed, will allow further refinement of these projections.

Table 3 shows expected NRI for $P < 5 \times 10^{-8}$ and for optimal thresholds at a single risk threshold of 10%. Morris et al. reported a case NRI of -0.0207 and control NRI of 0.0233 for an unweighted combined score, each slightly larger in magnitude than predicted here but compatible in direction and with sampling error. Substantially larger NRI is possible with more liberal selection of SNPs, and also with a larger training sample. The improvements are largely within the controls, unaffected individuals correctly reclassified to lower risk by addition of the polygenic score. Supplementary Tables S3 and S4 show results for alternative values of the prevalence and proportion of null SNPs.

Table 4 shows results for the weighted score and $\rho = 0.4$ for the higher risk threshold of 20% and for the continuous NRI and IDI. Results (not shown) are similar for the unweighted score or $\rho = 0.1$. Again, substantial improvements are possible, though the gains now appear concentrated among cases. Supplementary Tables S5 and S6 show results

TABLE 2 AUC for environmental score, polygenic score, and combined scores based on a genetic model matching results for CVD reported by Morris et al

N cases	Environment	Polygenic	Unweighted Sum		Weighted Sum	
			$\rho = 0.1$	$\rho = 0.4$	$\rho = 0.1$	$\rho = 0.4$
63,746	0.635	0.536 (5×10^{-8})	0.622 (5×10^{-8})	0.621 (5×10^{-8})	0.623 (5×10^{-8})	0.622 (5×10^{-8})
63,746	0.635	0.666 (0.053)	0.693 (0.033)	0.688 (0.036)	0.701 (0.053)	0.693 (0.053)
∞	0.635	0.782	0.800	0.784	0.800	0.788

In parentheses, P -value thresholds to select SNPs into polygenic score; N cases, number of cases in training sample with 2.05 controls per case as in CARDIoGRAMplusC4D.

TABLE 3 NRI for a single risk threshold of 10% for combined scores based on a genetic model matching results for CVD reported by Morris et al

N cases	Unweighted Sum				Weighted Sum			
	$\rho = 0.1$		$\rho = 0.4$		$\rho = 0.1$		$\rho = 0.4$	
	Case	Control	Case	Control	Case	Control	Case	Control
63,746	-0.0049	0.012	-0.0038	0.0095	-0.0060	0.015	0.0053	0.013
63,784	-0.046 (0.79)	0.186 (0.031)	-0.045 (0.76)	0.176 (0.031)	-0.052 (0.85)	0.2 (0.056)	-0.049 (0.86)	0.186 (0.059)
∞	-0.046	0.367	-0.050	0.344	-0.046	0.368	-0.049	0.349

In parentheses, P -value thresholds to select SNPs into polygenic score; N cases, number of cases in training sample with 2.05 controls per case as in CARDIoGRAMplusC4D.

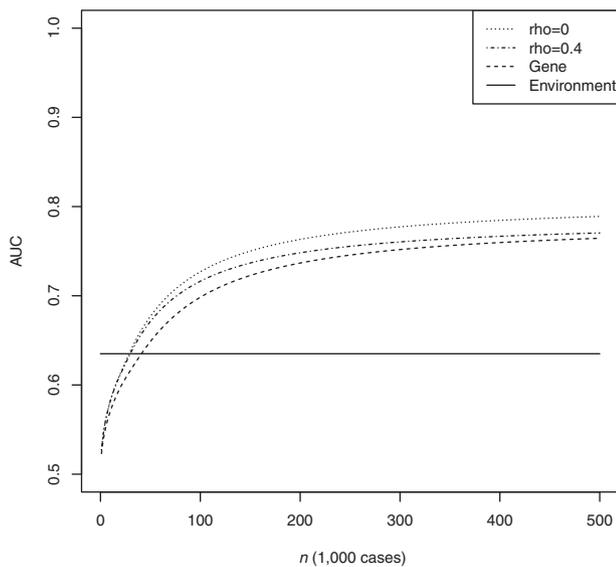


FIGURE 2 AUC of unweighted combined score as a function of training sample size

Note: Genetic model chosen to match results for CVD reported by Morris et al., with 2.05 controls per case as in the CARDIoGRAMplusC4D consortium. Rho: chip correlation between environment and outcome. Gene: polygenic score alone. Environment: environmental score alone.

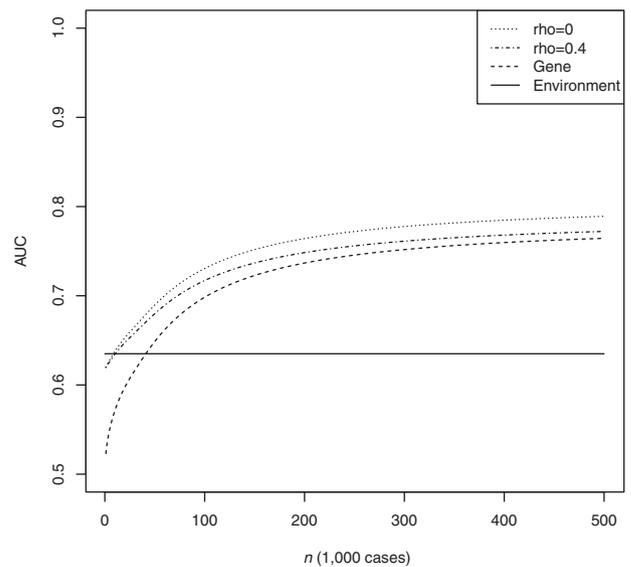


FIGURE 3 AUC of weighted combined score as a function of training sample size

Note: Genetic model chosen to match results for CVD reported by Morris et al., with 2.05 controls per case as in the CARDIoGRAMplusC4D consortium. Rho: chip correlation between environment and outcome. Gene: polygenic score alone. Environment: environmental score alone.

for alternative values of the prevalence and proportion of null SNPs.

Note that at finite sample size, the optimal P -value threshold varies according to the risk threshold and whether the case or control NRI is maximized, and those thresholds are not the ones that maximize AUC. Although the values of NRI are not greatly changed by using the threshold that maximizes AUC, this shows that the optimal genetic predictor can depend on the

chosen measure of accuracy. However, the optimal threshold is the same for continuous NRI and IDI as it is for AUC.

For this model, the optimal threshold to select SNPs at the CARDIoGRAMplusC4D sample size is at approximately nominal significance. Although this is well short of genome-wide significance, it results in fewer than 5,000 expected type-1 errors from 100,000 tests and an expected false discovery rate of about 0.4. It is this false discovery rate, rather than

TABLE 4 NRI for a single-risk threshold of 20%, continuous NRI and IDI for combined scores based on a genetic model matching results for CVD reported by Morris et al

<i>N</i> cases	20% Risk		Continuous NRI		IDI
	Case	Control	Case	Control	
63,784	0.014	− 0.006	0.080	0.014	0.0018
63,784	0.187 (0.048)	− 0.054 (0.76)	0.369 (0.053)	0.065 (0.053)	0.042 (0.053)
∞	0.347	− 0.058	0.631	0.111	0.135

In parentheses, *P*-value thresholds to select SNPs into polygenic score; *N* cases, number of cases in training sample with 2.05 controls per case as in CARDIoGRAM-plusC4D.

TABLE 5 AUC for environmental score, polygenic score and combined scores based on a genetic model matching results for breast cancer reported by Mavaddat et al

<i>N</i> cases	Environment	Polygenic	$R^2_{GX} = 0.1$		$R^2_{GX} = 0.8$	
			$\rho = 0.1$	$\rho = 0.4$	$\rho = 0.1$	$\rho = 0.4$
33,673 (5×10^{-8})	0.618	0.621	0.682	0.679	0.680	0.673
33,673 (0.0035)	0.618	0.728	0.762	0.755	0.757	0.742
∞	0.618	0.820	0.840	0.832	0.835	0.821

In parentheses, *P*-value thresholds to select SNPs into polygenic score; *N* cases, number of cases in training sample with 0.99 controls per case as in the Breast Cancer Association Consortium.

the family-wise type-1 error, that influences the explanatory power of the polygenic score.

3.2 | Breast cancer

Breast cancer is another disorder for which there is substantial interest in improving environmental risk scores with SNPs. Here, the general intention is to more effectively target enrolment on screening programs for early detection, rather than to directly treat risk factors. Chemoprevention may be recommended for high risk groups. A number of prediction models have been proposed, including BCRAT (commonly known at the Gail model) (Gail et al., 1989), IBIS (Tyrer-Cuzick) (Tyrer, Duffy, & Cuzick, 2004), Barlow (Barlow et al., 2006), Rosner-Colditz (Rosner & Colditz, 1996), and BOADICEA (Antoniou, Pharoah, Smith, & Easton, 2004). Common risk factors include age, age at menarche, mutations in *BRCA1* or *BRCA2*, age at first live birth, first degree family history, and history of breast biopsy. Generally speaking these models can perform well in women at high risk but are less accurate in the population at large.

In an early effort, Wacholder et al. (2010) estimated the improvement in the AUC of the Gail model with the addition of 10 SNPs associated from GWAS. The Gail model achieved AUC of 0.580, improving to 0.618 with the addition of SNPs. Subsequently, Darabi et al. (2012) used 18 SNPs to improve AUC from 0.548 to 0.615 in a Swedish sample. Most recently, Mavaddat et al. (2015) constructed a polygenic score from 77 SNPs identified in the largest consortium yet assembled (Michailidou et al., 2013), obtaining AUC of 0.622 with a training sample of 33,673 cases and 33,381 controls. In a simulation study, Garcia-Closas et al. (2014) suggested this could be improved to 0.670 in combination with an environmental

score with similar components to the Gail model. The environmental score alone had AUC of 0.618.

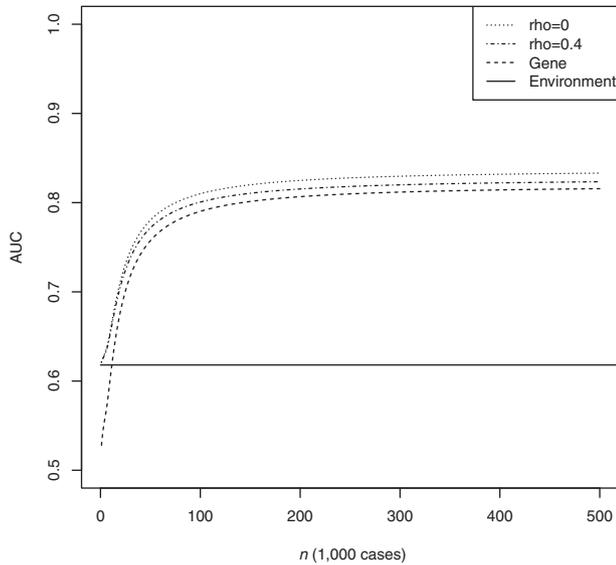
These previous studies did not consider the heritability of environmental factors such as age at menarche (Elks et al., 2010) and family history, and thus not the correlation between genetic and environmental scores. With this sample size, and again for simplicity assuming a genome-wide panel of 100,000 independent SNPs, the 77 SNP AUC of 0.622 would be achieved when the chip heritability of breast cancer $R^2_{GY} = 0.3$, proportion of null SNPs $\pi_0 = 0.95$ and disease prevalence $K = 0.05$ with SNPs selected by $P < 5 \times 10^{-8}$. With this prevalence, an AUC of 0.618 would be achieved by an environmental score with $R^2_{XY} = 0.0375$.

Under these parameters, Table 5 shows expected AUC for low and high chip heritabilities of the environmental score, and low and moderate genetic correlation between disease and environment. The results for selection by $P < 5 \times 10^{-8}$ agree well with Garcia-Closas et al. Again it is clear that substantial improvements are possible by more liberal selection of SNPs, and that further gains will be possible with larger training samples. Interestingly the results are quite robust to the environmental chip heritability R^2_{GX} and the genetic correlation ρ . Supplementary Table S7 shows results for a higher prevalence of 0.1, and Supplementary Table S8 for a lower proportion of null SNPs of 0.8. Figure 4 shows expected AUC as a function of training sample size. Again, the gains are small even if the two scores are assumed independent, but the sample size required to reach a given AUC can depend more strongly on the genetic correlation. For example, to reach AUC of 0.8, the required sample size is about 81,000 when $R^2_{GX} = 0.1$ and $\rho = 0.1$, but 140,000 when $R^2_{GX} = 0.8$ and $\rho = 0.4$. The Breast Cancer Association Consortium has recently grown to about 120,000 cases, which according to Figure 4 will yield an AUC

TABLE 6 NRI for a single-risk threshold of 8%, continuous NRI and IDI for combined scores based on a genetic model matching results for breast cancer reported by Mavaddat et al

N cases	8% Risk		Continuous NRI		IDI
	Case	Control	Case	Control	
33,673 (5×10^{-8})	0.119	-0.046	0.286	0.015	0.008
33,673	0.303 (0.0034)	-0.059 (0.89)	0.544 (0.0035)	0.029 (0.0035)	0.034 (0.0035)
∞	0.441	-0.094	0.771	0.041	0.089

In parentheses, P -value thresholds to select SNPs into polygenic score; N cases, number of cases in training sample with 0.99 controls per case as in the Breast Cancer Association Consortium. $\rho = 0.4$, $R_{GX}^2 = 0.8$.

**FIGURE 4** AUC of weighted combined score as a function of training sample size

Note: Genetic model chosen to match results for breast cancer reported by Mavaddat et al., with 0.99 controls per case as in the Breast Cancer Association Consortium and $R_{GX}^2 = 0.1$. Rho: chip correlation between environment and outcome. Gene: polygenic score alone. Environment: environmental score alone.

not far from the large sample limit. Again, in the near future these projections will be refined by further prediction studies.

Table 6 shows NRI at a single risk threshold of 8%, which is the 10-year absolute risk between ages 40–50 above which chemoprevention is advised in the United Kingdom (National Institute for Health and Care Excellence, 2013). Again the optimal SNP selection depends on the criterion optimized. Addition of SNPs to the Gail model would result in case NRI of 0.119 and control NRI of -0.046 when selecting SNPs by $P < 5 \times 10^{-8}$, improving to case NRI of 0.303 and control NRI of -0.059 with more liberal selection. The negative control NRI implies that more women would be unnecessarily recommended to receive chemoprevention, and because most women do not develop breast cancer, this translates to a large absolute number of women. Indeed the specificity for the development of cancer is 92% for the environmental score alone, but is 86% for the combined score, whereas the sensitivities are 16% and 32% respectively. Thus, while AUC and

NRI appear encouraging, a large number of women would in fact be misclassified under either score. The continuous NRI shows that over half of cases are expected to increase their risk score. Supplementary Tables S9 and S10 show results for alternative values of the prevalence and proportion of null SNPs.

Reflecting applications in screening, the breast cancer literature emphasises the proportion of cases present within some highest-risk proportion of the population (Pharoah, Antoniou, Easton, & Ponder, 2008). This is a point on a Lorenz curve, which resembles the receiver-operator characteristic curve with specificity replaced by a population proportion. For a proportion of the population q at highest risk according to score X , the corresponding threshold of liability is $\sqrt{\sigma_X^2} \Phi^{-1}(1 - q)$ and so the proportion of cases selected by that threshold is:

$$1 - \Phi \left(\frac{\sqrt{\sigma_X^2} \Phi^{-1}(1 - q) - E(X|Y = 1)}{\sqrt{\text{var}(X|Y = 1)}} \right).$$

Table 7 shows the proportion of cases within the top 10%, 20%, and 50% of the population at highest risk according to the environmental score X and the combined score \hat{S}_{comb} . These results suggest, for example, that at current sample sizes nearly half of cases could be detected by screening the 20% of the population with highest combined scores. With larger training samples, over 90% of cases might be detected by screening the half of the population with highest scores. These results are compatible with those of previous studies (Garcia-Closas et al., 2014; Pharoah et al., 2008). Supplementary Tables S11 and S12 show results for alternative values of the prevalence and proportion of null SNPs. Similar to other sensitivity analyses in the supplementary tables, these yield modest quantitative changes with similar qualitative conclusions.

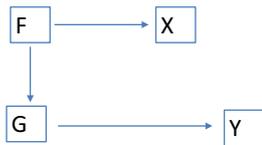
3.3 | Height

Here the example of height is used to illustrate a relationship between a polygenic score and family history. Of course, family history is an environmental risk factor for any heri-

TABLE 7 Proportion of cases present among highest risk quantiles in the population under a genetic model matching results reported in Mavaddat et al

<i>N</i> cases	Top 10%			Top 20%			Top 50%		
	Polygenic	Env	Comb	Polygenic	Env	Comb	Polygenic	Env	Comb
33,673 (5×10^{-8})	18.8	18.5	22.2	33.0	32.7	37.7	66.2	65.7	71.1
33,673 (0.0035)	29.8	18.5	30.7	47.6	32.7	48.7	79.9	65.7	80.9
∞	43.0	18.5	43.0	63.1	32.7	63.1	90.5	65.7	90.5

Polygenic: polygenic score alone; Env: environmental score alone; Comb: least squares weighted sum; in parentheses: *P*-value thresholds to select SNPs into polygenic score; *N* cases: number of cases in training sample with 0.99 controls per case as in the Breast Cancer Association Consortium.

**FIGURE 5** Directed acyclic graph showing correlation between polygenic and environmental scores arising from family history

A simplified scenario is shown in which there are no shared environmental effects. G: polygenic score; F: polygenic scores of other family members; X: family history; Y: outcome.

table condition, but it will be correlated with the measured genetic risk. An initial challenge for any genetic predictor is to exceed the predictive accuracy of family history (Do, Hinds, Francke, & Eriksson, 2012). For example, Aulchenko et al. (2009) showed that 54 associated SNPs could not predict an individual's height as accurately as could the mean height of its parents. Through simulation they showed that a gene score explaining all the heritability could predict better than the family history. A natural question is: what sample size would allow development of a gene score with predictive accuracy better than the family history?

Figure 5 is a directed acyclic graph showing a simplified situation in which the family history is entirely explained by genetics. Although in many cases shared environment also contributes to family history, this graph may be fairly appropriate for height because it is highly heritable and parents have often reached their full stature before producing offspring. Under this structure, the parameters of the quantitative model may be derived exactly in terms of the heritability.

Assume the commonly used value of 80% for the heritability of height. Under an additive model, the height of a child is the sum of genetic contributions from mother and father and of unique environmental contributions. It is easy to show that the genetic covariance between the child and the mean-parents is half the heritability, and then that the genetic correlation, $\rho = \frac{1}{\sqrt{2}}$. Furthermore the variance in child explained by the mean-parents is twice the square of the covariance, so $R_{XY}^2 = 0.32$, which is taken as the predictive accuracy of the family history. Finally, the chip heritability of the mean-parents is half the chip heritability, $R_{GX}^2 = \frac{1}{2}R_{GY}^2$.

The GIANT consortium has conducted discovery meta-analyses for height, which despite very large samples have yet to exceed the predictive accuracy of family history. A discovery sample of approximately 134,000 achieved R^2 of 13.3% by selecting SNPs with $P < 5 \times 10^{-4}$ (Lango Allen et al., 2010); a larger study of approximately 250,000 improved only to R^2 of 17% with $P < 5 \times 10^{-5}$ (Wood et al., 2014). These studies used a densely imputed panel of 2.5 M SNPs with chip heritability estimated within contributing studies at 60% of the total heritability, so $R_{GY}^2 = 0.6 \times 0.8 = 0.48$. Stepwise regressions were used to estimate SNP weights, so the estimated effects may be considered approximately independent across all 2.5 M SNPs. Across six polygenic scores with different selection thresholds, the best fitting model (Palla & Dudbridge, 2015) has chip heritability $R_{GY}^2 = 0.34$ and proportion of null SNPs $\pi_0 = 0.996$. This consortium level estimate of R_{GY}^2 is lower than the study-level estimates of 0.48, as has previously been observed (Yang et al., 2015). This is likely due to heterogeneity in data management across studies, if not actual genetic heterogeneity, and can be accounted for by estimating the genetic covariance between training and target data as a free parameter separately from the chip heritability in the training data (Palla & Dudbridge, 2015). This yields a genetic covariance of 0.46 with $\pi_0 = 0.995$; the estimate of R_{GY}^2 is 1, but this is a known artifact of this method, which has little bearing on estimates of predictive accuracy. Under this model, the prediction R^2 is 0.21 with an infinite training sample, implying that the consortium-based estimate can never predict as well as the family history.

This alarming result is mitigated by considering a homogeneous training sample in which we assume $R_{GY}^2 = 0.48$, with $\pi_0 = 0.995$ as estimated above and perfect genetic correlation with the target sample. Then $R^2 > 0.32$ when the training sample exceeds 301,000 subjects. Emerging datasets such as UK Biobank will soon allow such predictions to be tested empirically.

Under this model, Figure 6 shows the accuracy of a combined score as a function of training sample size, showing that the family history continues to provide useful information even as the polygenic risk becomes more informative. This is not surprising, as the chip heritability falls short of the total

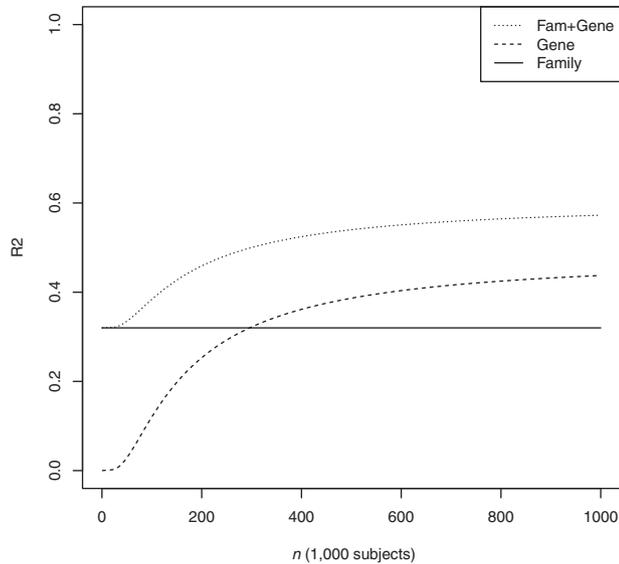


FIGURE 6 Prediction R^2 as a function of training sample size
Note: Genetic model chosen to match results for height reported by the GIANT consortium. Family: mean-parental height. Gene: polygenic score alone. Fam + Gene: least squares weighted combination of polygenic score and mean-parental height

heritability, which can be accessed, albeit imperfectly, by the family history. Nevertheless, this vignette demonstrates that an independent effect of family history is consistent with it being entirely genetic in origin, as long as the genetic predictor is not complete. Furthermore, it is apparent that for strongly heritable traits, the high threshold of predictive accuracy set by family history can only be overcome through very large and homogeneous genetic discovery studies.

4 | DISCUSSION

Accurate genetic prediction of disease risk remains an elusive goal, but its importance is evident from an increasing number of studies evaluating its potential for many traits. In conjunction with existing environmental predictors, the incremental benefit of polygenic scores has appeared modest to date. The results presented here allow an informed interpretation of observed results and anticipation of future results as larger training datasets are assembled. In particular, it is clear that at current sample sizes predictive accuracy, and incremental benefit, could be substantially improved by including SNPs at liberal significance thresholds. Furthermore, for chip correlation up to 0.4, the degree of correlation and the heritability of the environmental score have little bearing on predictive accuracy. These results are encouraging for practice because they reduce concern about accounting for the heritable components of environmental scores. In particular, it seems acceptable to combine consortium estimates of marginal SNP effects with environmental risk scores, rather than jointly estimating

conditional SNP and environmental effects in training samples that may be much smaller than those of consortia.

The majority of studies to date have restricted their polygenic scores to genome-wide significant SNPs, but this is sub-optimal. One justification has been that genotyping a limited number of SNPs is more cost-effective than a whole-genome panel; but this should be viewed in the context of potentially predicting multiple conditions from a single DNA sample, for which genome-wide typing will be more efficient. Another view is that it is difficult to argue for inclusion of individual risk factors (here SNPs) that have but weak evidence of association. However, by viewing the polygenic risk as a single entity, this argument becomes irrelevant; the role of individual SNPs then applies only to how that risk is calculated.

The present work unifies previous work of So and Sham (2010), who showed how many common measures of predictive accuracy can be expressed in terms of liability R^2 , with that of Dudbridge (2013), who allowed for finite training samples and selection of SNPs into the polygenic score. Only the most commonly reported measures are considered here, namely AUC, NRI, and IDI, but other measures, such as those related to predictiveness curves (Pepe et al., 2008) and net benefit (Baker, 2009), could be treated in the same manner. In contrast to some previous approaches, the genetic covariance between environmental score and predicted trait is considered here explicitly, along with unweighted and weighted combinations of genetic and environmental scores. Although the genetic covariance is the key property, the results are presented in terms of genetic correlation and variances, as these quantities can be estimated readily by existing methods (Bulik-Sullivan et al., 2015; Lee et al., 2012).

The modest effect of genetic correlation on AUC is not too surprising, because AUC depends on ranking the scores rather than their absolute values, but more surprisingly the NRI, which does depend on absolute risk, is also quite robust to correlation. Higher levels of correlation will of course tend toward no incremental benefit, but such scenarios are arguably unlikely because most environmental scores include nonheritable factors such as age, as well as the environmental component of the heritable factors. Nevertheless, the predictive accuracy can be improved by accounting for covariance between genetic and environmental predictors. Furthermore, the level of correlation can have more substantial effects on the sample size required to reach a prespecified level of accuracy.

The NRI has been criticized on theoretical and practical grounds (Kerr et al., 2014; Pepe et al., 2015), yet it has an intuitive appeal that should not be overlooked. An important recommendation is that it should be reported separately for cases and controls, especially if they are associated with different costs (Pencina et al., 2011). Here, it has been shown that for given training data, the optimal polygenic score differs for case NRI and control NRI, and also varies with risk thresholds. Because in a prospective setting one cannot know

who is a case, the polygenic score must be defined according to a criterion based on relative costs of case and control NRI. Such considerations will further complicate reporting of NRI in evaluation studies, although this problem does not apply to the continuous NRI or IDI.

We have focused on two commonly reported measures, AUC and NRI, yet both measures have been severely criticized and there is a recognized need for more appropriate measures of incremental benefit. For example, stratification of the population by genetic risk may allow more efficient application of environmental risk scores. In CVD, a group at high genetic risk attained 20% estimated risk at an age up to 18 years younger than the low genetic risk group (Abraham et al., 2016). In breast cancer, the absolute risk associated with modifiable risk factors was significantly higher among women in the top decile of nonmodifiable risk (including 92 SNPs) compared to those in the lowest decile (Maas et al., 2016). These results suggest that even limited genetic data can provide useful stratification for identifying subjects who would benefit most from intervention. Such perspectives offer a more optimistic view of genetic prediction that will only become stronger with larger training samples and more liberal selection of genetic markers.

Consortium studies are now approaching the sizes at which useful levels of predictive accuracy are predicted by theory. However, such results have not yet been achieved in practice. One clear reason is that many studies have restricted their polygenic scores to genome-wide significant SNPs. Given the large size of the CARIODoGRAMplusC4D consortium, higher levels of accuracy should be possible by including more SNPs in the risk score. This has been demonstrated by Abraham et al. (2016), although their approach of pruning the full set of SNPs may be suboptimal. Furthermore, accurate odds ratios are currently available only for a targeted array product, the MetaboChip, further limiting the accuracy that can be achieved. Therefore their results, which are intermediate between those using genome-wide significant SNPs and those predicted here, seem consistent with the present theory. The present work suggests that as genome-wide typing is completed on a larger number of studies, more accurate prediction will be achieved by highly polygenic scores.

The Breast Cancer Association Consortium too is approaching a sufficiently large number of cases. To date, consortium-wide genotypes have only been generated on the iCOGS platform, another targeted array focused on candidate genes for certain cancers. Again therefore, prediction studies have focused on genome-wide significant SNPs, but the next generation of targeted arrays, which include a genome-wide backbone, offer greater promise for developing more accurate polygenic scores.

Given the large size of the GIANT consortium, the continuing modest prediction of height is surprising but could be explained by heterogeneity between consortium-level train-

ing data and individual target studies. The source of this heterogeneity remains unclear (Yang et al., 2015), but with the emergence of large, homogeneous datasets such as UK Biobank, more accurate genetic predictors could be developed that exceed the predictive accuracy of the family history.

An alternative explanation is that the present model for genetic effects, consisting of a mixture of a normal distribution and a mass at zero, is incorrect for this phenotype. If in fact the true distribution had a sharp peak at zero, then the present model could fit the data well, but the individual SNP effects would be more dispersed and thus harder to estimate *en masse*. Such a model could apply, for example, to schizophrenia, for which fine-scale heritability analyses suggest an extremely polygenic architecture (Loh et al., 2015), while the best fitting normal-null mixture estimates a null proportion of around 90% (Palla & Dudbridge, 2015). Although theory based on a mixture of exponential distributions (Chatterjee et al., 2013) has roughly agreed with the normal-null mixture, there is a notable exception in the case of height, for which a mixture of three exponentials fits GIANT data well with chip heritability at 45%. This fitted model projects that prediction R^2 will not exceed 0.32 until the training sample size exceeds one million, in contrast to our projection of 301,000 for a homogeneous sample. Finally, departures from the additive model may lead to a poor fit of the polygenic score.

The substantial heritability of most common disorders implies that clinically useful prediction can be achieved when the heritable risk is accurately measured. The present results suggest that such results may not be too far off in terms of discovery sample size. As that level is approached, the polygenic score dominates the prediction compared to existing environmental scores, although the dynamic and potentially modifiable nature of environmental factors ensures that they will continue to play a crucial role in assessing absolute risk. Such predictors may include epigenetic or other “omic” factors that themselves are selected from high dimensional panels. The present work assumes a known, low dimensional risk factor whose effect on outcome is known precisely. Extension to high dimensional, dynamic nongenetic predictors, allowing for simple model selection as in the polygenic score, will be pursued in subsequent work.

ACKNOWLEDGEMENTS

This work is supported by the MRC (MR/K006215/1, MR/M026175/1). We thank Peter Kraft, John Perry, Peter Visscher, and Naomi Wray for discussions.

ORCID

Frank Dudbridge  <http://orcid.org/0000-0002-8817-8908>

REFERENCES

- Abraham, G., Havulinna, A. S., Bhalala, O. G., Byars, S. G., De Livera, A. M., Yetukuri, L., ... Inouye, M. (2016). Genomic prediction of coronary heart disease. *European Heart Journal*, *37*, 3267–3278.
- Abraham, G., & Inouye, M. (2015). Genomic risk prediction of complex human disease and its clinical application. *Current Opinion in Genetics and Development*, *33*, 10–16.
- Antoniou, A. C., Pharoah, P. P., Smith, P., & Easton, D. F. (2004). The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *British Journal of Cancer*, *91*, 1580–1590.
- Aulchenko, Y. S., Struchalin, M. V., Belonogova, N. M., Axenovich, T. I., Weedon, M. N., Hofman, A., ... Borodin, P. M. (2009). Predicting human height by Victorian and genomic methods. *European Journal of Human Genetics*, *17*, 1070–1075.
- Baker, S. G. (2009). Putting risk prediction in perspective: Relative utility curves. *Journal of the National Cancer Institute*, *101*, 1538–1542.
- Barlow, W. E., White, E., Ballard-Barbash, R., Vacek, P. M., Titus-Ernstoff, L., Carney, P. A., ... Kerlikowske, K. (2006). Prospective breast cancer risk prediction model for women undergoing screening mammography. *Journal of the National Cancer Institute*, *98*, 1204–1214.
- Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., ... Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*, 1236–1241.
- Chatterjee, N., Shi, J., & Garcia-Closas, M. (2016). Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature Reviews Genetics*, *17*, 392–406.
- Chatterjee, N., Wheeler, B., Sampson, J., Hartge, P., Chanock, S. J., & Park, J. H. (2013). Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nature Genetics*, *45*, 400–405, 405e401–403.
- Clayton, D. G. (2009). Prediction and interaction in complex disease genetics: Experience in type 1 diabetes. *PLoS Genetics*, *5*, e1000540.
- Cook, N. R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, *115*, 928–935.
- D'Agostino, R. B., Sr., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., & Kannel, W. B. (2008). General cardiovascular risk profile for use in primary care: The Framingham Heart Study. *Circulation*, *117*, 743–753.
- Darabi, H., Czene, K., Zhao, W., Liu, J., Hall, P., & Humphreys, K. (2012). Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. *Breast Cancer Research*, *14*, R25.
- de Vries, P. S., Kavousi, M., Ligthart, S., Uitterlinden, A. G., Hofman, A., Franco, O. H., & Dehghan, A. (2015). Incremental predictive value of 152 single nucleotide polymorphisms in the 10-year risk prediction of incident coronary heart disease: The Rotterdam Study. *International Journal of Epidemiology*, *44*, 682–688.
- Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T. L., Thompson, J. R., ... Samani, N. J. (2013). Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genetics*, *45*, 25–33.
- Do, C. B., Hinds, D. A., Francke, U., & Eriksson, N. (2012). Comparison of family history and SNPs for predicting risk of complex disease. *PLoS Genetics*, *8*, e1002973.
- Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genetics*, *9*, e1003348.
- Dudbridge, F. (2016). Polygenic epidemiology. *Genetic Epidemiology*, *40*, 268–272.
- Elks, C. E., Perry, J. R., Sulem, P., Chasman, D. I., Franceschini, N., He, C., ... Murray, A. (2010). Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nature Genetics*, *42*, 1077–1085.
- Eriksson, J., Evans, D. S., Nielson, C. M., Shen, J., Srikanth, P., Hochberg, M., ... Ohlsson, C. (2015). Limited clinical utility of a genetic risk score for the prediction of fracture risk in elderly subjects. *Journal of Bone and Mineral Research*, *30*, 184–194.
- Evans, D. M., Visscher, P. M., & Wray, N. R. (2009). Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Human Molecular Genetics*, *18*, 3525–3531.
- Gail, M. H., Brinton, L. A., Byar, D. P., Corle, D. K., Green, S. B., Schairer, C., & Mulvihill, J. J. (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute*, *81*, 1879–1886.
- Garcia-Closas, M., Gunsoy, N. B., & Chatterjee, N. (2014). Combined associations of genetic and environmental risk factors: Implications for prevention of breast cancer. *Journal of the National Cancer Institute*, *106*:dju305.
- Goff, D. C., Jr., Lloyd-Jones, D. M., Bennett, G., Coady, S., D'Agostino, R. B., Gibbons, R., ... Tomaselli, G. F. (2014). 2013 ACC/AHA guideline on the assessment of cardiovascular risk: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*, *129*, S49–S73.
- Hippisley-Cox, J., Coupland, C., Robson, J., & Brindle, P. (2010). Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: Cohort study using QResearch database. *British Medical Journal*, *341*, c6624.
- Hughes, M. F., Saarela, O., Stritzke, J., Kee, F., Silander, K., Klopp, N., ... Schunkert, H. (2012). Genetic markers enhance coronary risk prediction in men: The MORGAM prospective cohorts. *PLoS One*, *7*, e40922.
- JBS3 Board. (2014). Joint British Societies' consensus recommendations for the prevention of cardiovascular disease (JBS3). *Heart*, *100*(2), ii1–ii67.
- Jostins, L., & Barrett, J. C. (2011). Genetic risk prediction in complex disease. *Human Molecular Genetics*, *20*, R182–R188.
- Kerr, K. F., Wang, Z., Janes, H., McClelland, R. L., Psaty, B. M., & Pepe, M. S. (2014). Net reclassification indices for evaluating risk prediction instruments: A critical review. *Epidemiology*, *25*, 114–121.
- Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., ... Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, *467*, 832–838.
- Lee, S. H., Wray, N. R., Goddard, M. E., & Visscher, P. M. (2011). Estimating missing heritability for disease from genome-wide association studies. *American Journal of Human Genetics*, *88*, 294–305.

- Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, *28*, 2540–2542.
- Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., ... Speliotes, E. K. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, *518*, 197–206.
- Loh, P. R., Bhatia, G., Gusev, A., Finucane, H. K., Bulik-Sullivan, B. K., Pollack, S. J., ... Price, A. L. (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature Genetics*, *47*, 1385–1392.
- Maas, P., Barrdahl, M., Joshi, A. D., Auer, P. L., Gaudet, M. M., Milne, R. L., ... Chatterjee, N. (2016). Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA Oncology*, *2*, 1295–1302.
- Maier, R., Moser, G., Chen, G. B., Ripke, S., Coryell, W., Potash, J. B., ... Lee, S. H. (2015). Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *American Journal of Human Genetics*, *96*, 283–294.
- Mavaddat, N., Pharoah, P. D., Michailidou, K., Tyrer, J., Brook, M. N., Bolla, M. K., ... Garcia-Closas, M. (2015). Prediction of breast cancer risk based on profiling with common genetic variants. *Journal of the National Cancer Institute*, *107*:dju036.
- Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R. L., ... Easton, D. F. (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature Genetics*, *45*, 353–361.
- Morris, R. W., Cooper, J. A., Shah, T., Wong, A., Drenos, F., Engmann, J., ... Humphries, S. E. (2016). Marginal role for 53 common genetic variants in cardiovascular disease prediction. *Heart*, *102*, 1640–1647.
- Morrison, A. C., Bare, L. A., Chambless, L. E., Ellis, S. G., Malloy, M., Kane, J. P., ... Boerwinkle, E. (2007). Prediction of coronary heart disease risk using a genetic risk score: The atherosclerosis risk in communities study. *American Journal of Epidemiology*, *166*, 28–35.
- National Institute for Health and Care Excellence. (2013). *Familial Breast Cancer: Classification and Care of People at Risk of Familial Breast Cancer and Management of Breast Cancer and Related Risks in People with a Family History of Breast Cancer*. Cardiff (UK): National Collaborating Centre for Cancer (UK).
- Palla, L., & Dudbridge, F. (2015). A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *American Journal of Human Genetics*, *97*, 250–259.
- Patel, K. A., Oram, R. A., Flanagan, S. E., De Franco, E., Colclough, K., Shepherd, M., ... Hattersley, A. T. (2016). Type 1 diabetes genetic risk score: A novel tool to discriminate monogenic and type 1 diabetes. *Diabetes*, *65*, 2094–2099.
- Pencina, M. J., D'Agostino, R. B. Sr., D'Agostino, R. B. Jr., & Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, *27*, 157–172; discussion 207–112.
- Pencina, M. J., D'Agostino, R. B., Sr., & Steyerberg, E. W. (2011). Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine*, *30*, 11–21.
- Pepe, M. S., Fan, J., Feng, Z., Gerds, T., & Hilden, J. (2015). The net reclassification index (NRI): A misleading measure of prediction improvement even with independent test data sets. *Statistics in Biostatistics*, *7*, 282–295.
- Pepe, M. S., Feng, Z., Huang, Y., Longton, G., Prentice, R., Thompson, I. M., & Zheng, Y. (2008). Integrating the predictiveness of a marker with its performance as a classifier. *American Journal of Epidemiology*, *167*, 362–368.
- Pharoah, P. D., Antoniou, A., Bobrow, M., Zimmern, R. L., Easton, D. F., & Ponder, B. A. (2002). Polygenic susceptibility to breast cancer and implications for prevention. *Nature Genetics*, *31*, 33–36.
- Pharoah, P. D., Antoniou, A. C., Easton, D. F., & Ponder, B. A. (2008). Polygenes, risk prediction, and targeted prevention of breast cancer. *New England Journal of Medicine*, *358*, 2796–2803.
- Rosner, B., & Colditz, G. A. (1996). Nurses' health study: Log-incidence mathematical model of breast cancer incidence. *Journal of the National Cancer Institute*, *88*, 359–364.
- Simonson, M. A., Wills, A. G., Keller, M. C., & McQueen, M. B. (2011). Recent methods for polygenic analysis of genome-wide data implicate an important effect of common variants on cardiovascular disease risk. *BMC Medical Genetics*, *12*, 146.
- So, H. C., & Sham, P. C. (2010). A unifying framework for evaluating the predictive power of genetic variants based on the level of heritability explained. *PLoS Genetics*, *6*, e1001230.
- Talmud, P. J., Cooper, J. A., Morris, R. W., Dudbridge, F., Shah, T., Engmann, J., ... Humphries, S. E. (2015). Sixty-five common genetic variants and prediction of type 2 diabetes. *Diabetes*, *64*, 1830–1840.
- Thanassoulis, G., Peloso, G. M., Pencina, M. J., Hoffmann, U., Fox, C. S., Cupples, L. A., ... O'Donnell, C. J. (2012). A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: The Framingham Heart Study. *Circulation: Cardiovascular Genetics*, *5*, 113–121.
- Tikkanen, E., Havulinna, A. S., Palotie, A., Salomaa, V., & Ripatti, S. (2013). Genetic risk prediction and a 2-stage risk screening strategy for coronary heart disease. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *33*, 2261–2266.
- Tyrer, J., Duffy, S. W., & Cuzick, J. (2004). A breast cancer prediction model incorporating familial and personal risk factors. *Statistics in Medicine*, *23*, 1111–1130.
- Valdez, R., Yoon, P. W., Qureshi, N., Green, R. F., & Khoury, M. J. (2010). Family history in public health practice: A genomic tool for disease prevention and health promotion. *Annual Review of Public Health*, *31*, 69–87; 61 p following 87.
- Verhulst, B., Neale, M. C., & Kendler, K. S. (2015). The heritability of alcohol use disorders: A meta-analysis of twin and adoption studies. *Psychological Medicine*, *45*, 1061–1072.
- Wacholder, S., Hartge, P., Prentice, R., Garcia-Closas, M., Feigelson, H. S., Diver, W. R., ... Hunter, D. J. (2010). Performance of common genetic variants in breast-cancer risk models. *New England Journal of Medicine*, *362*, 986–993.
- Weissfeld, J. L., Lin, Y., Lin, H. M., Kurland, B. F., Wilson, D. O., Fuhrman, C. R., ... Diergaarde, B. (2015). Lung cancer risk prediction using common SNPs located in GWAS-identified susceptibility regions. *Journal of Thoracic Oncology*, *10*, 1538–1545.

- Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., ... Abecasis, G. R. (2013). Discovery and refinement of loci associated with lipid levels. *Nature Genetics*, *45*, 1274–1283.
- Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., ... Frayling, T. M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, *46*, 1173–1186.
- Wray, N. R., Goddard, M. E., & Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Research*, *17*, 1520–1528.
- Wray, N. R., Yang, J., Goddard, M. E., & Visscher, P. M. (2010). The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genetics*, *6*, e1000864.
- Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A., Nolte, I. M., ... Visscher, P. M. (2015). Genome-wide genetic homogene-

ity between sexes and populations for human height and body mass index. *Human Molecular Genetics*, *24*, 7445–7449.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Dudbridge F, Pashayan N, Yang J. Predictive accuracy of combined genetic and environmental risk scores. *Genet Epidemiol.* 2018; 42:4–19. <https://doi.org/10.1002/gepi.22092>