Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Turley P, Meyer MN, Wang N, et al. Problems with the use of polygenic scores in embryo selection. N Engl J Med 2021;385:78-86. DOI: 10.1056/NEJMsr2105065

Supplementary Appendix to
"Challenges with Embryo Selection Using Polygenic Scores"

Table of Contents

List of Investigators

Patrick Turley, Ph.D.,[1,†] Michelle N. Meyer, Ph.D., J.D.,[2,†] Nancy Wang, S.B.,[3] David Cesarini, Ph.D.,[3,4] Evelynn Hammonds, Ph.D.,[5] Alicia R. Martin, Ph.D.,[6,7,8] Ben M. Neale, Ph.D.,[6,7,8] Heidi L. Rehm, Ph.D.,[6,7,8] Louise Wilkins-Haug, M.D., Ph.D.,[8,9] Daniel J. Benjamin, Ph.D.,[3,10,*,‡] Steven Hyman, M.D.,[5,6,‡] David Laibson, Ph.D.,[3,5,‡] Peter M. Visscher, Ph.D.[11,‡]

Affiliations:

[1]University of Southern California.

[2]Geisinger Health System.

[3]National Bureau of Economic Research.

[4]New York University.

[5]Harvard University.

[6]Broad Institute of Harvard and MIT.

[7]Massachusetts General Hospital.

[8]Harvard Medical School.

[9]Brigham and Women's Hospital.

[10]University of California Los Angeles.

[11]University of Queensland.

*Correspondence to: daniel.benjamin@gmail.com.

[†]These authors contributed equally to this work.

[‡]These authors, in alphabetical order, jointly supervised this work.

Supplementary Text

1. Theoretical Framework for Quantitative Traits

We begin by describing the framework for quantitative traits These are traits that can take on a continuum of values (e.g., height). This is in contrast to clinical outcomes, which are binary (e.g., type 1 diabetes). Our simulations are based on an additive model for a polygenic trait within families:

$$y_i = a_i + \varepsilon_i,$$

where y_i is the potential value of the trait for some individual (or embryo) i, a_i is the additive genetic factor for the trait, and ε_i is the residual. By construction of the additive genetic factor, ε_i is uncorrelated with a_i. Without loss of generality, we standardize y_i to have variance one. We assume that there is no non-additive genetic factor for the trait (i.e., no dominance or epistasis). (For educational attainment, the additive model in fact provides a reasonably good fit to observed correlations between relatives; see Rietveld et al.,[1] SOM section 2a). Therefore, the within-family broad-sense heritability of the trait coincides with the within-family narrow-sense heritability and equals $h^2 \equiv \mathrm{Var}(a_i)$.

Note that the value of y_i would not be observed for most embryos. It corresponds to the phenotypic value of the embryo if it were the embryo chosen to transfer and if the embryo survived until the time that the phenotype could be measured. Though this can't be observed in real data, we are able to simulate this value to assess the theoretical consequences of choosing one embryo over another.

We define c_i to be the common component of the non-genetic factor ε_i for all embryos who share the same parents. We can then decompose the non-genetic factor as $\varepsilon_i = c_i + u_i$, where u_i is the non-genetic component that is unique to the embryo.

In our framework, c_i and u_i are analogous to the shared and non-shared environment components in a standard twin model, but they are unlikely to be equal to those components. The reason is that twins co-exist as siblings, while the embryos replace each other as potential children. As a result, we expect c_i to be larger and u_i to be smaller than their analogs in a twin model. For example, twins may be assigned to different classrooms or be assigned to sit near different peers. In our setting, however, the embryo that is selected could be assigned to exactly the same environments as the other would be if it had been selected.

We define g_i as the best linear predictor of a_i using the measured SNPs. Thus, $a_i = g_i + \xi_i$, where ξ_i is the residual and is uncorrelated with g_i (and ε_i) by construction. We can therefore write $h^2 = \mathrm{Var}(g_i) + \mathrm{Var}(\xi_i)$. Moreover, the trait's within-family SNP heritability is $h^2_{SNP} \equiv \mathrm{Var}(g_i)$.

We assume that y_i is polygenic in both the part of the genetic component captured by measured SNPs, g_i, and the part not captured by measured SNPs, ξ_i. Consequently, both g_i and ξ_i can be well approximated as normally distributed in the population.

Thus, the maternal components of the additive genetic factor, $g_{m,i}$ and $\xi_{m,i}$, and the paternal components of the additive genetic factor, $g_{p,i}$ and $\xi_{p,i}$, are distributed as:

$$g_{m,i} \sim N(0, h^2_{SNP}) \quad (1)$$

$$g_{p,i} \sim N(0, h^2_{SNP}) \quad (2)$$
$$\xi_{m,i} \sim N(0, h^2 - h^2_{SNP}) \quad (3)$$
$$\xi_{p,i} \sim N(0, h^2 - h^2_{SNP}). \quad (4)$$

We assume random mating so that the maternal and paternal genetic components are independent. This implies that, conditional on the parental genetic components, the genetic components of each embryo are independently and identically distributed as

$$g_i | g_{m,i}, g_{p,i} \sim N\left(\frac{g_{m,i}+g_{p,i}}{2}, \frac{h^2_{SNP}}{2}\right) \quad (5)$$
$$\xi_i | \xi_{m,i}, \xi_{p,i} \sim N\left(\frac{\xi_{m,i}+\xi_{p,i}}{2}, \frac{h^2-h^2_{SNP}}{2}\right). \quad (6)$$

Assortative mating would reduce the within-family variance of g_i and ξ_i (that is, with assortative mating, $\mathrm{Var}\left(g_i \middle| g_{m,i}, g_{p,i}\right) < h^2_{SNP}/2$ and $\mathrm{Var}\left(g_i \middle| g_{m,i}, g_{p,i}\right) < (h^2_{SNP} - h^2)/2$). This would, in turn, reduce the expected gain from embryo selection. Therefore, as pointed out by Karavani et al.[2], these calculations represent upper bounds on the expected gain from embryo selection.

All of the calculations and simulations in this paper are independent of the distribution of c_i, so we make no assumptions about it. We do, however, make the strong assumption that variation in u_i is negligible between embryos conceived by the same parents. As discussed below (in the section "Comparison to Framework in Karavani et al.[2]"), assumptions about u_i are irrelevant for the expected gain from embryo selection. Our assumption implies that our calculation of the amount of unpredictable variation is a lower bound for the true amount of unpredictable variation.

We model an embryo's polygenic score, $\hat{g}_i$, as:

$$\hat{g}_i \equiv g_i + e_i, \quad (7)$$

where e_i has mean zero and is independent of g_i. (The error e_i comes from the estimation error in the SNP weights used for constructing the polygenic score. The variance of this error shrinks with the sample size of the GWAS from which the weights are estimated.) We denote the variance of e_i by $\sigma^2_e \equiv \mathrm{Var}(e_i)$.

Daetwyler et al.[3] showed that the predictive power of the polygenic score, as measured by R^2, can be approximated as

$$R^2 = \frac{(h^2_{SNP})^2}{h^2_{SNP} + \sigma^2_e}.$$

Rearranging, this implies that for a polygenic score with predictive power R^2,

$$\sigma^2_e = \frac{h^2_{SNP}(h^2_{SNP} - R^2)}{R^2}.$$

Since we have estimates of h^2_{SNP} and R^2, we will use this formula to obtain our estimate of σ^2_e, as we discuss below in the section "Parameter Values."

2. Theoretical Framework for Binary Traits

Our framework for binary traits is an extension of the continuous framework described above using a liability-scale model. Let $y_i^\star \in \{0,1\}$ denote an indicator variable for whether embryo i would have the corresponding condition during their life if it were selected. We assume that there is an unobservable factor y_i corresponding to $y_i^\star$ such that

$$y_i^\star = \begin{cases} 0 & \text{if } y_i \leq c_0 \\ 1 & \text{otherwise.} \end{cases} \quad (8)$$

where c_0 is a constant that is fixed by the lifetime risk of $y_i^\star$ in the population. In this way, y_i is a quantitative trait and it can follow the same framework described in Section 1 of this Appendix.

3. Comparison to Framework in Karavani et al.[2]

The model for quantitative traits above differs from that of Karavani et al.[2] in one way. Recall that the polygenic score only captures some of the variation in the phenotype, the amount denoted R^2. Karavani et al. assume that the residual variation beyond what is captured by the polygenic score, the amount of which is $1 - R^2$, is uncorrelated between embryos. This implicitly assumes two things. First, it implies that the polygenic score captures all of the additive genetic variation; if it didn't, the residual genetic variation would each have a correlation of ½ within a family, not 1. Second, Karavani et al.'s assumption also requires that c_i, the common-environment component of ε_i, be zero, whereas as we argued above, this component is likely to be substantial in an embryo selection setting (at least, larger than the "shared environment" component estimated in a twin study).

That said, these differences do not affect the calculations for the expected gain from embryo selection. As shown in Karavani et al., the only parameters that matter for this calculation are the predictive power of the polygenic score and the number of embryos tested. However, the differences matter for the calculation of the amount of unpredictable variation. Allowing for correlation between pairs of embryos (as we do) shrinks the prediction interval for embryo selection. For this reason, in our simulation we assume that $u_i = 0$ for all embryos because that minimizes the unexplained variation in the phenotype within a family. Thus, our prediction intervals represent a lower bound on the unpredictable variation in the gain from embryo selection.

Our framework also contains a number of features that are extensions of Karavani et al. For example, we directly model the imperfect genetic correlation between the environment faced by embryos and the environment of GWAS participants. We also derive a simple extension of their model for binary traits. Finally, in the following section, we derive a framework to model the effect of ESPS on pleiotropic phenotypes.

4. Pleiotropy

We also would like to model the effect of embryo selection for trait y_i on some pleiotropic trait z_i. Analogously to above, we assume that, within a family,

$$z_i = g_{z,i} + \xi_{z,i} + c_{z,i} + u_{z,i}$$

where $g_{z,i}$ is the portion of the additive genetic factor captured by SNPs, $\xi_{z,i}$ is the residual additive genetic factor, $c_{z,i}$ is the common environmental component, and $u_{z,i}$ is the unique environmental component. We make the corresponding assumptions for $g_{z,i}$, $\xi_{z,i}$, $c_{z,i}$, and $u_{z,i}$ that we make for g_i, ξ_i, c_i, and u_i. Specifically, $c_{z,i} = c_{z,j}$ when embryos i and j have the same parents (and hence $u_{z,i} = 0$), and our assumptions imply that $g_{z,i}$, $\xi_{z,i}$, $c_{z,i}$, and $u_{z,i}$ are pairwise uncorrelated. We also assume that z_i has been standardized to be mean zero and variance one. This means that

$$g_{z,i} \sim N(0, h^2_{z,SNP})$$

and

$$\xi_{z,i} \sim N\left(0, h^2_z - h^2_{z,SNP}\right),$$

where h^2_z is the within-family broad-sense (and narrow-sense) heritability and $h^2_{z,SNP}$ is the within-family SNP heritability of z_i.

To model the pleiotropy between y_i and z_i, we assume that $\mathrm{Corr}\left(g_i, g_{z,i}\right) = r_g$ and that $\mathrm{Corr}\left(\xi_i, \xi_{z,i}\right) = r_\xi$. We assume that $r_g = r_\xi$. Thus, by the properties of bivariate normal distributions,

$$\left(g_{z,i} + \xi_{z,i}\right) | (g_i + \xi_i) \sim N\left[r_g \sqrt{\frac{h^2_z}{h^2}}\, g_i, h^2_z\left(1 - r_g^2\right)\right]. \quad (8)$$

5. Simulation Framework

For quantitative traits, we begin by simulating independent genetic components for N parent pairs according to equations (1)-(4). Next, for each parent pair, we simulate genetic components for M embryos according to equations (5)-(6) and the embryos' polygenic scores using equation (7). For the pleiotropy analysis, we calculate the additive genetic factor for the secondary trait using equation (8).

To calculate the likelihood that a single set of parents would have at least one embryo with a polygenic score in the top decile of the polygenic score distribution and at least one embryo with a polygenic score in the bottom decile, we simply evaluate the fraction of such parents in the simulation. To obtain a conservative estimate of this fraction, we also restrict the simulation to parents who have a mean polygenic score of zero. Doing so maximizes the above probability.

To calculate the expected gain from embryo selection, we identify the embryo with the highest polygenic score as the one that would be chosen under embryo selection. (An alternative approach that has been shown to have smaller gains is to omit a certain number of embryos with the smallest polygenic scores and to choose randomly among the remaining embryos.[4]) We also choose one embryo uniformly at random to be the embryo that would be

chosen if there were no embryo selection. This embryo may be the same as the embryo with the maximum polygenic score. For all variables corresponding to these embryos, we use the subscripts "max" and "rand," respectively.

The gain from embryo selection for a specific parent pair is

$$\begin{aligned} gain &= g_{\max} + \xi_{\max} + \varepsilon_{\max} - g_{\text{rand}} - \xi_{\text{rand}} - \varepsilon_{\text{rand}} \\ &= g_{\max} + \xi_{\max} - g_{\text{rand}} - \xi_{\text{rand}} \end{aligned}$$

since $u_{\max} = u_{\text{rand}} = 0$ and therefore $\varepsilon_{\max} = \varepsilon_{\text{rand}}$. To calculate the expected gain, we take the average gain across all M parent pairs. To assess the unpredictable variation in the gain from embryo selection, we also calculate and report the 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentile of the gain across all simulated parent pairs.

To estimate the expected effect of pleiotropy on embryo selection, we evaluate the change in the second phenotype for each family as

$$\begin{aligned} change &= g_{\text{z,max}} + \xi_{\text{z,max}} + \varepsilon_{\text{z,max}} - g_{\text{z,rand}} - \xi_{\text{z,rand}} - \varepsilon_{\text{z,rand}} \\ &= g_{\text{z,max}} + \xi_{\text{z,max}} - g_{\text{z,rand}} - \xi_{\text{z,rand}}, \end{aligned}$$

where the "max" subscript does not correspond to the embryo with the maximum polygenic score for z_i but rather the embryo that has the maximum polygenic score for y_i. We then take the mean change across families to calculate the expected change in the secondary trait for parents using embryo selection on the primary trait.

For traits where z_i is on the liability scale, we must convert z_i into the probability that the individual will have the binary trait. To do this, we convert our simulated value of z_i for each embryo into the potentially observed phenotype $z_i^\star$ using

$$z_i^\star = \begin{cases} 0 & \text{if } z_i \leq \Phi^{-1}(1-p) \\ 1 & \text{otherwise,} \end{cases}$$

where p is the prevalence of trait $z_i^\star$ in a population without embryo selection. We then can measure the prevalence of $z_i^\star$ in the set of embryos that have the maximum polygenic score of y_i.

For binary traits, we follow the same simulation procedure described above for quantitative traits to simulate the latent variable, $y_i^\star$. We then use (8) to convert $y_i^\star$ into the value of the trait for the embryo if it were selected. The value of c_0 is set such that the expected value of $y_i^\star$ is fixed at some level. For the results reported in Table 1, we fix c_0 such that the expected lifetime risk for a randomly selected embryo is equal to the US lifetime reported risk for the condition. For the results reported in Figure 2 and S1-S8, we vary c_0 to obtain estimates of the expected effect of ESPS for parent pairs with different lifetime risk of the clinical trait (e.g., due to environmental differences).

To calculate the absolute risk reduction due to ESPS, we calculate select one embryo at random from each parent pair and calculate the mean value of $y_i^\star$ across the parent pairs. We use $y_{\text{rand}}^\star$ to denote this value. We then select the embryo with the lowest polygenic score for the phenotype from each parent pair and calculate the mean value of $y_i^\star$ across the parent

pairs. We use $y^{\star}_{\min}$ to denote this value. The absolute risk reduction with ESPS is therefore $y^{\star}_{\text{rand}} - y^{\star}_{\min}$. The relative risk reduction is $(y^{\star}_{\text{rand}} - y^{\star}_{\min})/y^{\star}_{\text{rand}}$.

6. Between-Family Simulation

To create a between-family benchmark for the within-family "expected gain," we imagine taking M unrelated European-ancestry individuals from the population and measuring the expected difference in the phenotype between the individuals with the largest polygenic score and a randomly chosen individual from that group. More precisely, we first simulate an additive genetic factor explained by SNPs for each set of M individuals and the residual:

$$g_{BF,i} \sim N\left(0, h^2_{BF,SNP}\right)$$
$$\varepsilon_{BF,i} \sim N\left(0, 1 - h^2_{BF,SNP}\right).$$

These will differ from the additive genetic factor in the within-family framework in two ways. First, the between-family SNP heritability will be larger than the within-family SNP heritability since, between families, the polygenic score captures indirect effects from parents and siblings. Second, individuals within a group have uncorrelated non-genetic factors because they are unrelated. Note that we have not separately simulated the component of the additive genetic factor that is unexplained by SNPs. Because the members of each group are unrelated, this term is drawn independently across individuals and can therefore be ignored.

Phenotypes $y_{BF,i}$ and polygenic scores $\hat{g}_{BF,i}$ for each individual are constructed in a parallel way as in the within-family framework:

$$y_{BF,i} = g_{BF_i} + \varepsilon_{BF,i}$$
$$\hat{g}_{BF,i} = g_{BF,i} + e_{BF,i},$$

where

$$e_{BF,i} \sim N\left(0, \sigma^2_{BF,e}\right).$$

Analogously to the within-family framework, we set $\sigma^2_{BF,e} = \frac{h^2_{BF,SNP}\left(h^2_{BF,SNP} - R^2_{BF}\right)}{R^2_{BF}}$ assuming some value of R^2_{BF} (as described in the next section, "Parameter Values").

7. Parameter Values

We simulate $N = 1{,}000{,}000$ parent pairs. For Type 1 Diabetes we simulated 10,000,000 parent pairs because the lifetime risk is so low. Following Karavani et al.[2], we assume each parent pair is selecting among $M = 10$ viable embryos.

We assume that the between-family broad-sense (and narrow-sense) heritability of educational attainment is $h^2 = 0.4$, which is consistent with evidence reviewed in Branigan et al.[5]. For the five largest cohorts ($N \geq 3000$) with the most precise measures of educational attainment (each of which has 20 categories of educational attainment), Lee et al.[6] estimate the mean SNP heritability to be 0.2. Thus, we set $h^2_{BF,SNP} = 0.2$. We use this estimate of SNP

heritability in the between-family simulation, but this estimate is biased upward in a within-family framework due to indirect effects from parents. For the within-family SNP heritability parameter in parents with European ancestries, we assume $h^2_{SNP} = 0.10$. This number is based on Kong et al.[7], who find that roughly half of the SNP heritability for educational attainment is due to direct effects.

Next, we calibrate σ^2_e such that predictive power is half of the SNP heritability, giving us a between-family R^2 of 0.1, which is used in the between-family simulation, and a within-family R^2 of 0.05. This is consistent with the relative size of SNP heritability and predictive power in Lee et al.[6]

For the calculations of all the solid bars in Figure 1, we further attenuate the predictive power of the polygenic score by 0.87^2 to account for potential imperfect genetic correlation between the GWAS sample population and the population from which parents are drawn, as per the formula derived by de Vlaming et al.[8] The parameter 0.87 is the estimated genetic correlation between the two largest cohorts in Lee et al.[6], the UK Biobank and 23andMe. This leaves us with a within-family R^2 for parents with European ancestries of 0.038.

For the clinical traits considered other than idiopathic short stature and intellectual disability, we begin with the within-family area-under-the-curve (AUC) estimates for polygenic scores from Lello et al.[9] We can convert these AUC estimates into R^2 estimates on the liability scale if we have an estimate of the lifetime risk of the clinical outcome in the same population.[10] Because these AUC estimates correspond to a sample of individuals with European ancestries, we use the lifetime risk for each clinical outcome for the U.S. White, non-Hispanic population.[11] While the population of individuals with European ancestries and the White, non-Hispanic population are not the same group, they were as close an approximation we could find with reliable current information on the lifetime risk of the clinical outcomes we considered. To obtain an estimate of the lifetime risk of type 2 diabetes, we subtracted the lifetime risk of type 1 diabetes[12] from the lifetime risk for all types of diabetes.[13] For coronary artery disease, hypercholesterolemia, and hypertension, we could not identify estimates of the lifetime risk, so population prevalence was used instead.[14] In these simulations, we defined idiopathic short stature and intellectual disability as having a height or cognitive performance, respectively, two standard deviations below the mean.[15,16]

In the cases of idiopathic short stature and intellectual disability, we use estimates of the within-family R^2 of the polygenic score for their underlying continuous outcomes (height and cognitive performance, respectively). For height, Lloyd-Jones et al.[17] report a between-family R^2 of the most recent polygenic score of 34.2% in the Health and Retirement Study and of 35.2% in the Estonian Biobank. We use 35.2% to obtains an upper bound risk reduction of idiopathic short stature. Selzam et al.[18] find that the coefficient associated with the polygenic score in a within-family analysis is 94.9% as large as the coefficient from an analysis of unrelated individuals. This implies that the within-family R^2 is 31.7%$= 35.2\% \times 0.949^2$. For cognitive performance, Lee et al.[6] report a between family R^2 for cognitive performance of 6.9% in the Add Health Cohort[19] and 9.7% in the Wisconsin Longitudinal Study.[20] As with idiopathic short stature, we use 9.7% to obtain an upper bound risk reduction of intellectual disability. Selzam et al.[18] find that the coefficient associated with the polygenic score in a within-family analysis is 73.3% as large as the coefficient from an analysis of unrelated individuals. This implies that the within-family R^2 is $5.2\% = 9.7\% \times 0.733^2$.

Finally, because all of these polygenic scores were trained on individuals with European ancestries, the predictive power of the polygenic score will be smaller for those with non-European ancestries. Therefore, for parents with AMR, EAS, or AFR ancestries, we reduce the predictive power of the polygenic score by a factor of 1.6, 2, or 4.5, respectively, which are the average degrees of attenuation as estimated across multiple traits by Martin et al.[21]

To obtain the upper and lower bounds of the confidence intervals reported in Table 1, we follow the same simulation procedure, but this time using the upper and lower bound of the within-family predictive power estimates. For all the traits except idiopathic short stature and intellectual disability, we take the upper and lower bounds of the AUC from Lello et al.[9] and convert them to into within-family R^2 using the same procedure as above. For intellectual disability, we use the confidence bounds for the between-family R^2 reported in Lee et al.[6] and shrink those estimates by a factor of 0.733^2 (as above) to obtain approximate bounds of the within-family predictive power. For idiopathic short stature, Lloyd-Jones et al.[17] do not report confidence intervals on their between-family R^2 estimates. However, their R^2 estimate is based on a sample of 32,594 individuals. The sampling variance of a squared-correlation estimate is $\mathrm{Var}(\hat{R}^2) = 4R^2(1-R^2)^2/N$.[22,23] Using the sample estimates of R^2 and the known sample size, we can therefore obtain standard errors of the between-family R^2. We shrink these standard errors by a factor of 0.949^2 (as above) to obtain standard errors for our within-family predictive power estimates. As our 95% confidence interval, we therefore use $\hat{R}^2_{WF} \pm 1.96\ \mathrm{SE}(\hat{R}^2_{WF})$, where $\hat{R}^2_{WF}$ is the within-family predictive power.

For the pleiotropy analysis for bipolar disorder, we use a between-family, liability-scale, broad-sense heritability of 0.75, based on the twin estimates found in Sullivan et al.[24] We use a prevalence of 1%, reported in the same paper. For the genetic correlation between educational attainment and liability to bipolar disorder, we use 0.25, reported in Okbay et al.[25]

8. Code

On publication, Python code implementing the above simulations can be found https://github.com/JonJala/ESPS_sim.

9. Ancestry Group Labels

In this manuscript, we refer to four continental ancestry groups and the labels for each of them used by the 1000 Genomes Project: European (EUR), Admixed American (AMR), East Asian (EAS), and African (AFR). In the 1000 Genome Project data, the EUR sample consists of five groups: Utah residents with Northern and Western European ancestry, British in England and Scotland, Iberian populations in Spain, Finnish in Finland, and Toscani in Italy. The AMR population consists of four groups: Puerto Rican in Puerto Rico; Colombian in Medellin, Columbia; Peruvian in Lima, Peru; and Mexican ancestry in Los Angeles, California. The EAS population consists of five groups: Han Chinese South; Kinh in Ho Chi Minh City, Vietnam; Han Chinese in Beijing, China; Japanese in Tokyo, Japan; and Chinese Dai in Xishuangbanna, China. The AFR population consists of seven groups: African Caribbean in Barbados; African Ancestry in Southwest US; Gambian in Western Division, The Gambia (Wolof, Mandinka, Fula, and Jola); Mende in Sierra Leone; Esan in Nigeria; Yoruba in Ibadan, Nigeria; and Luhya in Webuye, Kenya.

Supplementary Figures

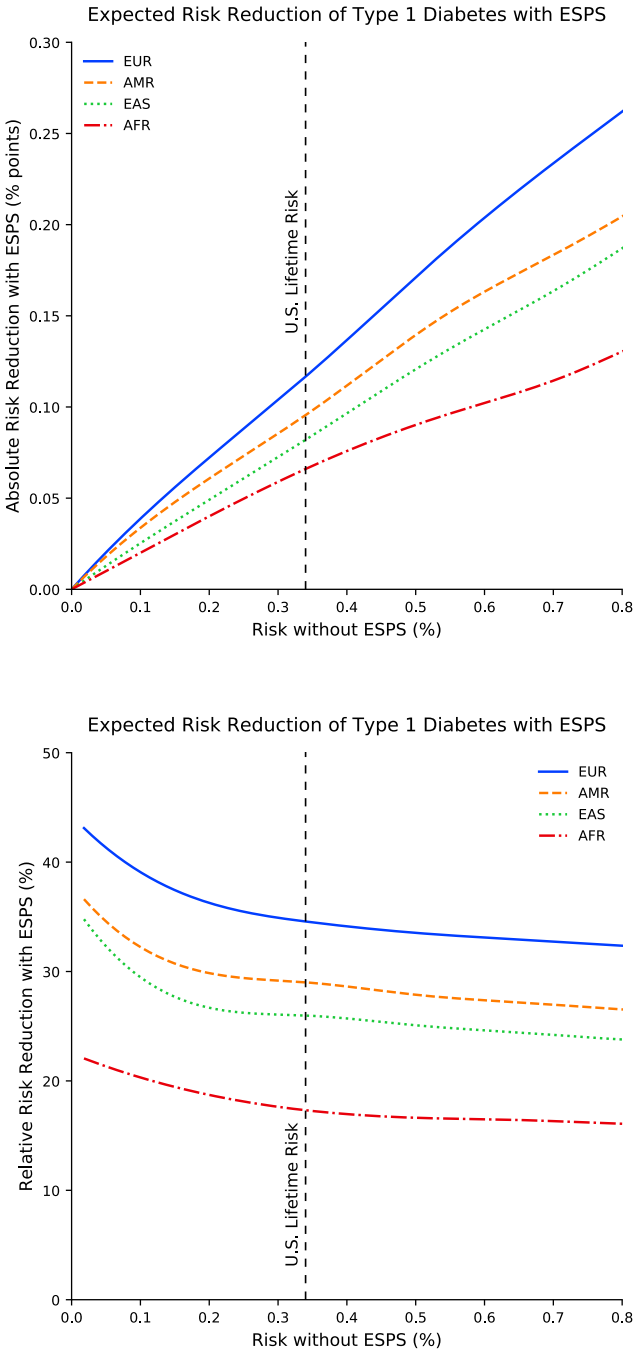

Figure S1. Absolute and relative risk reductions from ESPS for type 1 diabetes among different ancestry groups and for groups of embryos with levels of baseline risk due to family history or environmental conditions. The lifetime risk in the US for type 1 diabetes is marked with a vertical dashed line.[12] The calculations underlying these figures are found in the Supplementary Note of this appendix.

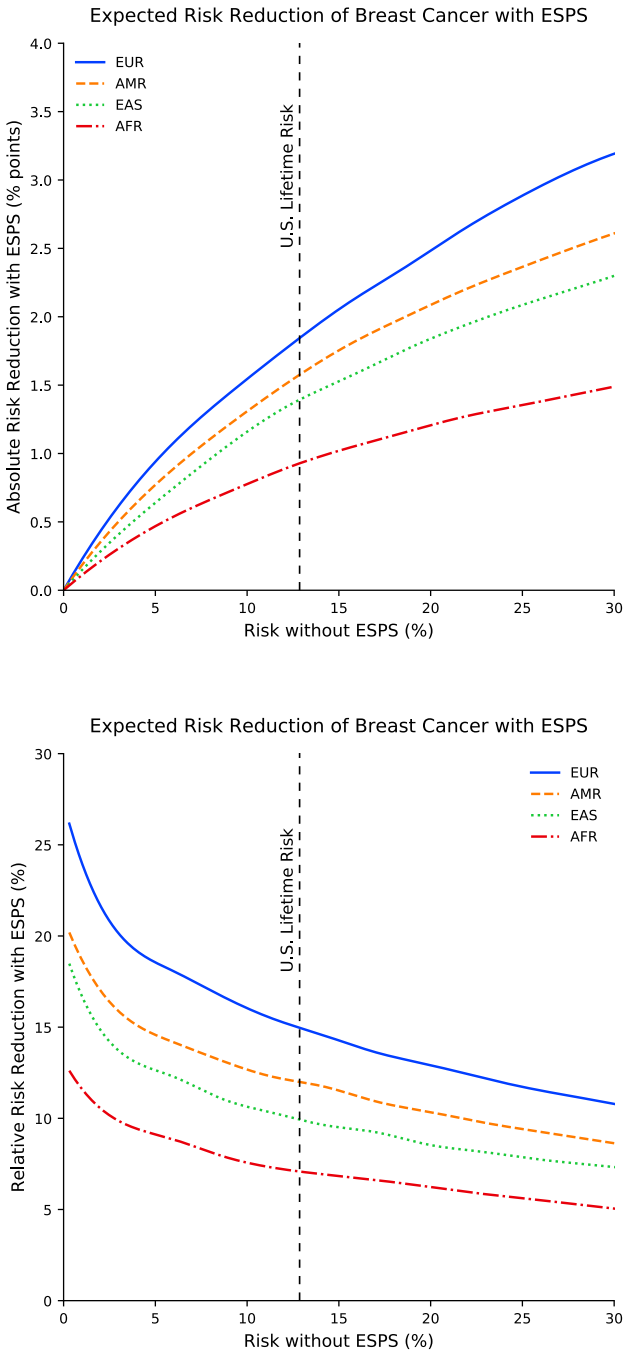

Figure S2. Absolute and relative risk reductions from ESPS for breast cancer in women among different ancestry groups and for groups of embryos with levels of baseline risk due to family history or environmental conditions. The lifetime risk for women in the US for breast cancer is marked with a vertical dashed line.[11] The calculations underlying these figures are found in the Supplementary Note of this appendix.

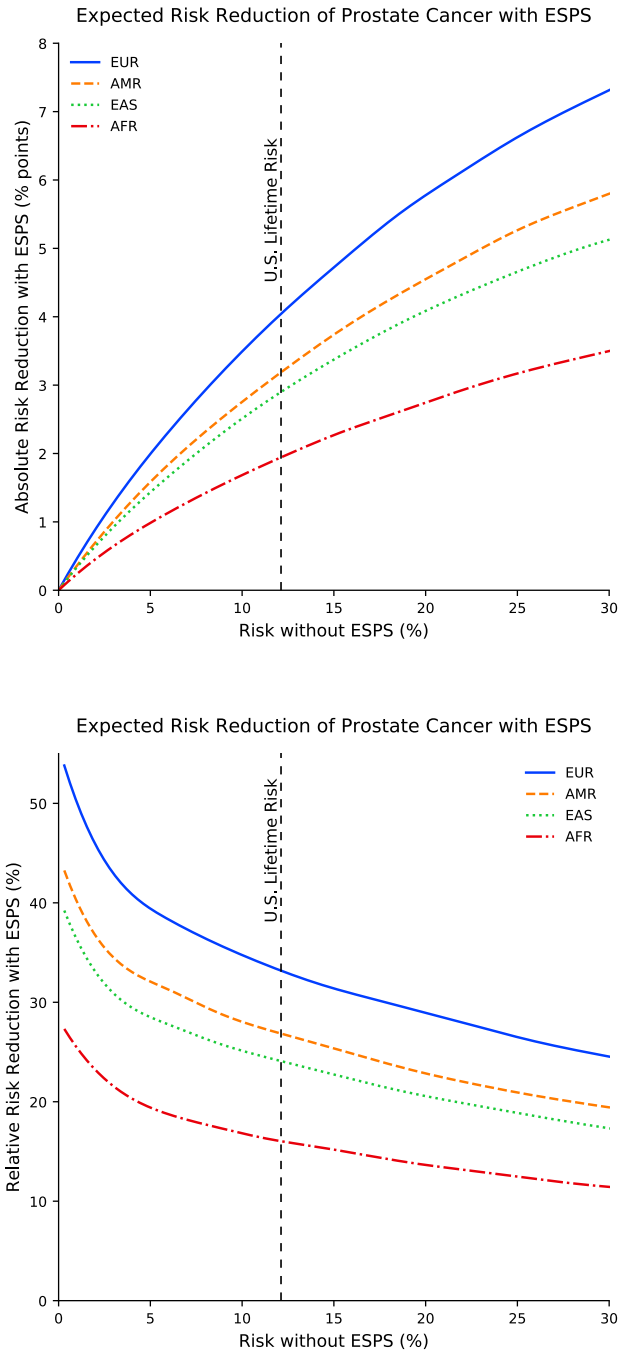

Figure S3. Absolute and relative risk reductions from ESPS for prostate cancer in men among different ancestry groups and for groups of embryos with levels of baseline risk due to family history or environmental conditions. The lifetime risk for men in the US for prostate cancer is marked with a vertical dashed line.[11] The calculations underlying these figures are found in the Supplementary Note of this appendix.

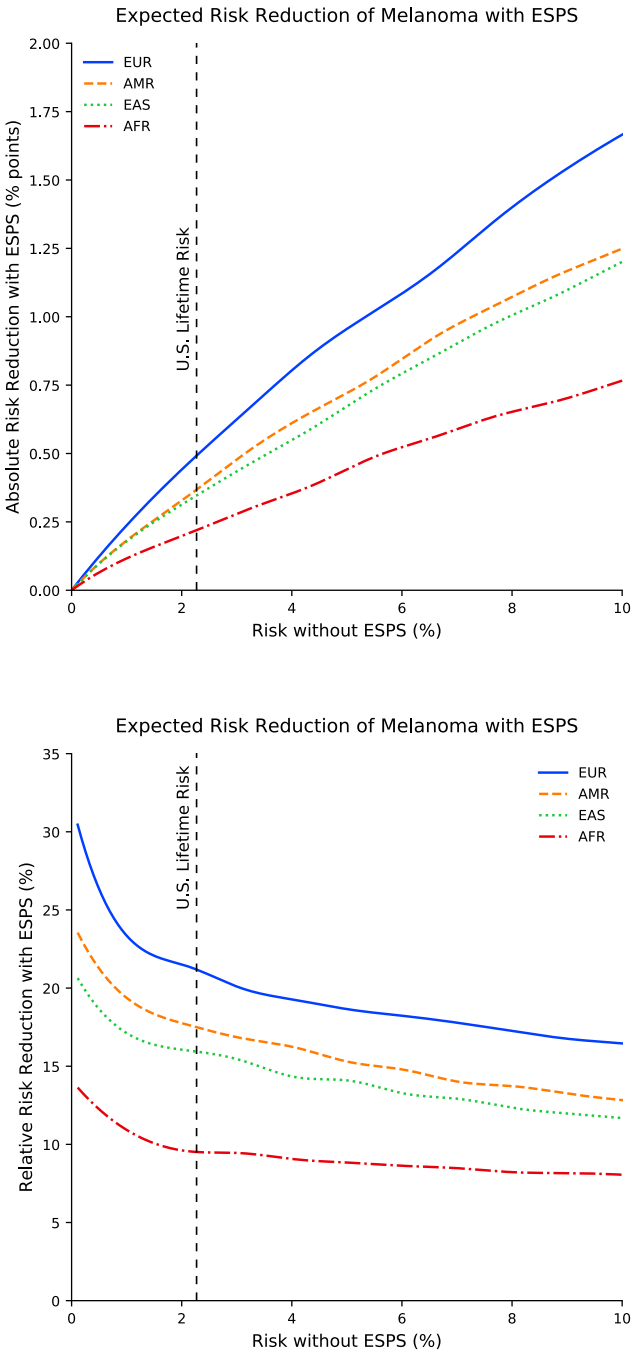

Figure S4. Absolute and relative risk reductions from ESPS for malignant melanoma among different ancestry groups and for groups of embryos with levels of baseline risk due to family history or environmental conditions. The lifetime risk in the US for malignant melanoma is marked with a vertical dashed line.[11] The calculations underlying these figures are found in the Supplementary Note of this appendix.

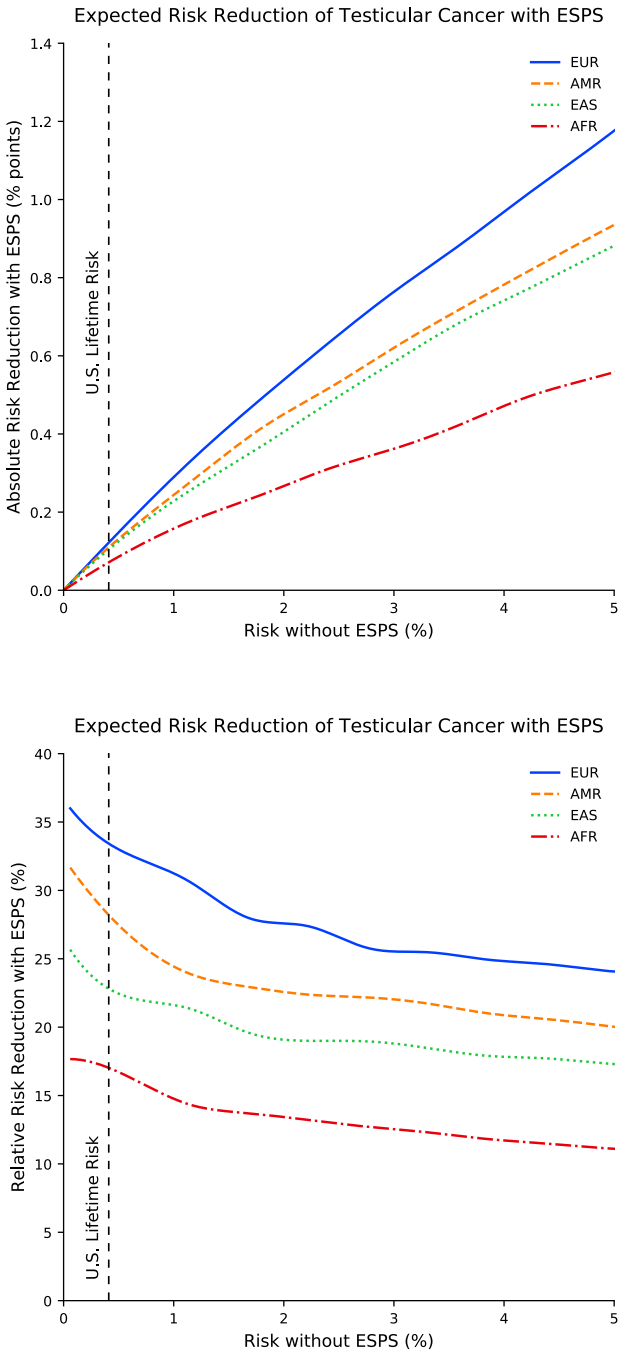

Figure S5. Absolute and relative risk reductions from ESPS for testicular cancer in men among different ancestry groups and for groups of embryos with levels of baseline risk due to family history or environmental conditions. The lifetime risk for men in the US for testicular cancer is marked with a vertical dashed line.[11] The calculations underlying these figures are found in the Supplementary Note of this appendix.

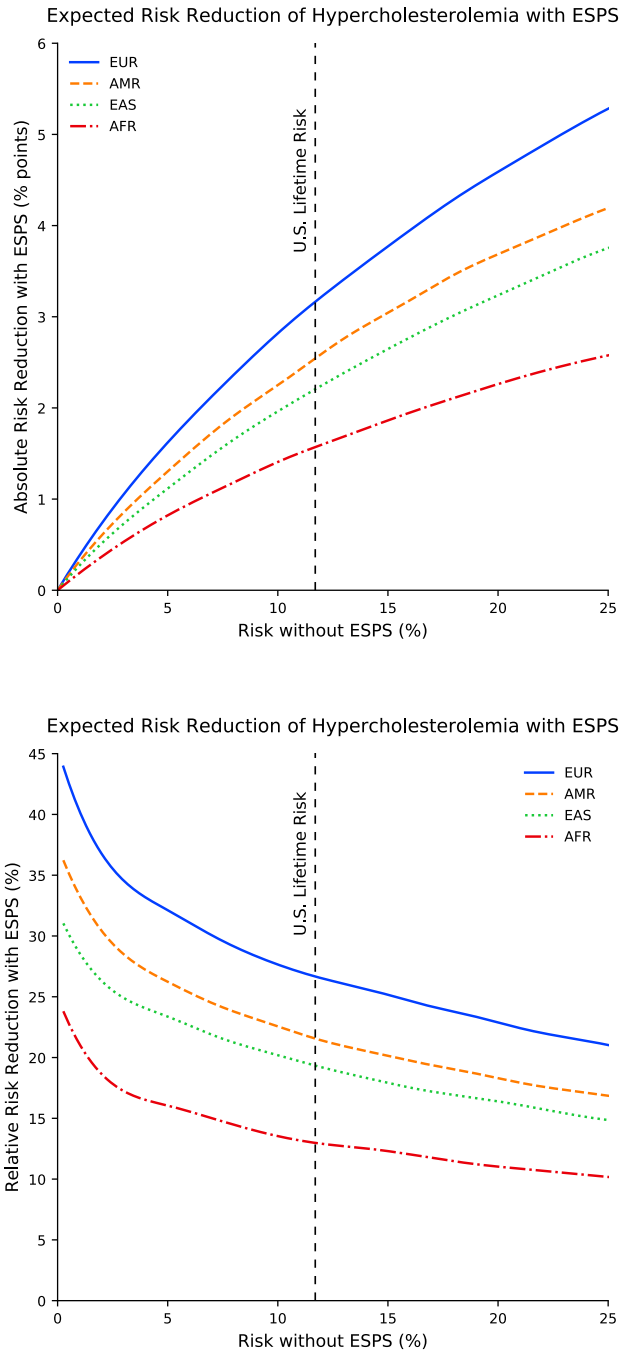

Figure S6. Absolute and relative risk reductions from ESPS for hypercholesterolemia among different ancestry groups and for groups of embryos with levels of baseline risk due to family history or environmental conditions. The lifetime risk in the US for hypercholesterolemia is marked with a vertical dashed line.[14] The calculations underlying these figures are found in the Supplementary Note of this appendix.

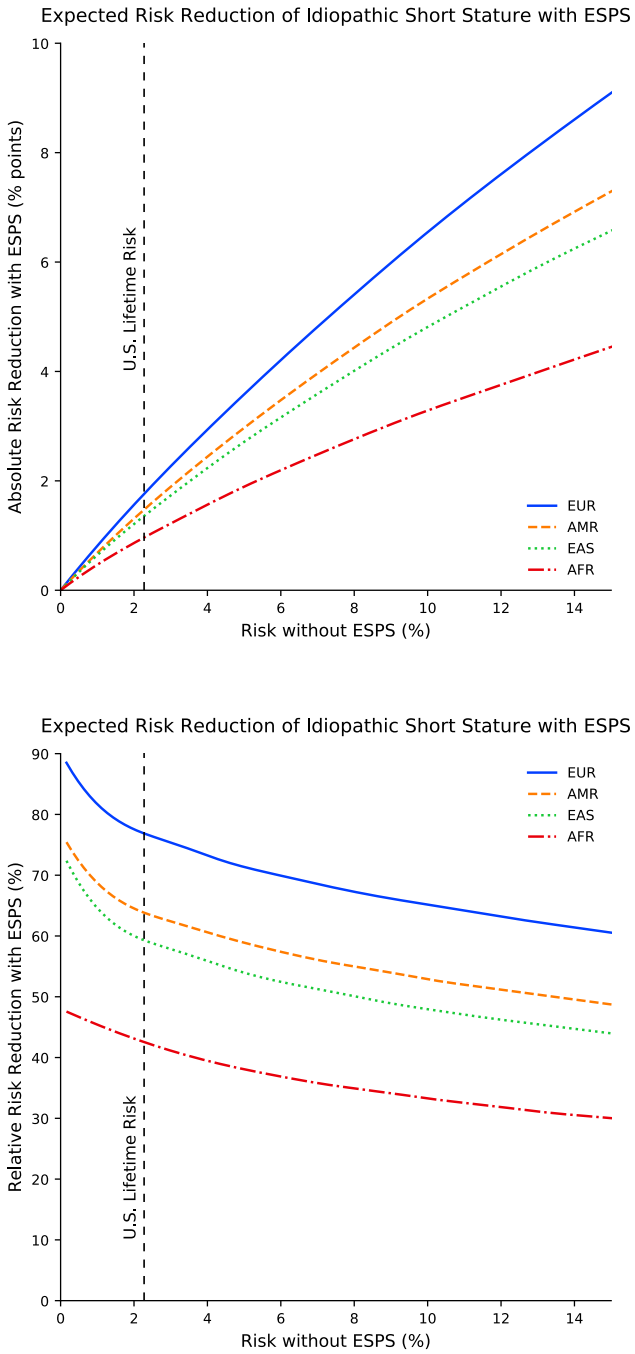

Figure S7. Absolute and relative risk reductions from ESPS for idiopathic short stature among different ancestry groups and for groups of embryos with levels of baseline risk due to family history or environmental conditions. The lifetime risk in the US for idiopathic short stature is marked with a vertical dashed line.[15] The calculations underlying these figures are found in the Supplementary Note of this appendix.

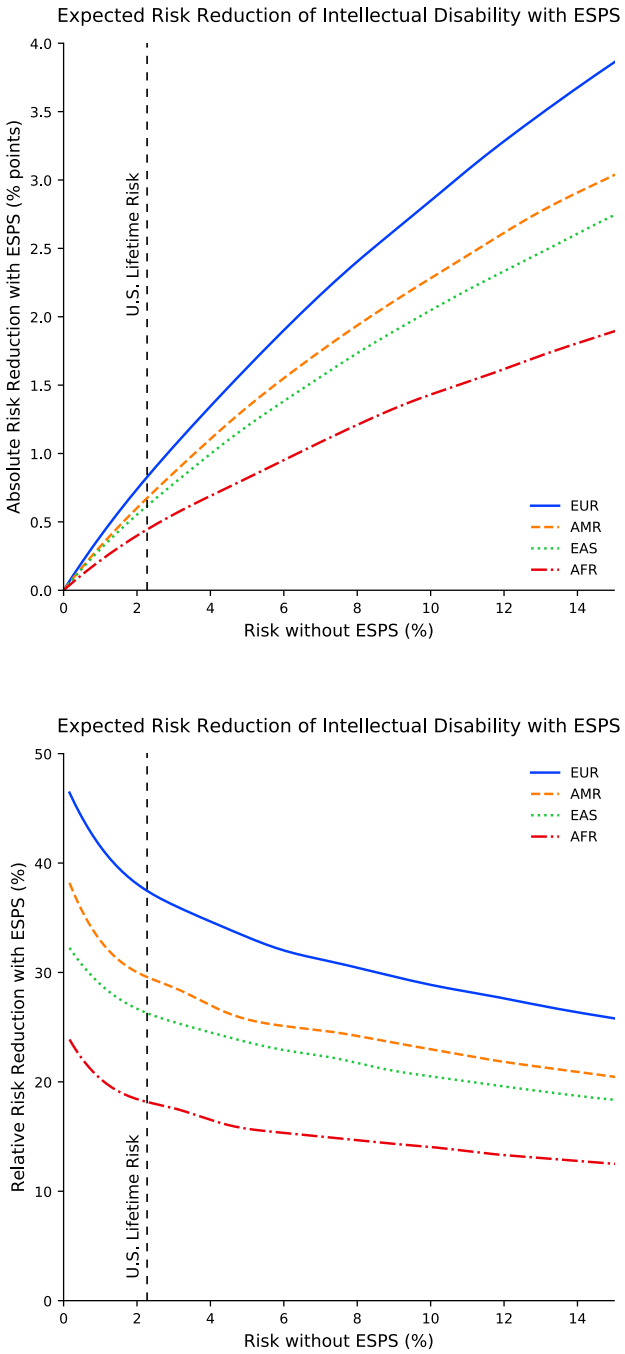

Figure S8. Absolute and relative risk reductions from ESPS for intellectual disability among different ancestry groups and for groups of embryos with levels of baseline risk due to family history or environmental conditions. The lifetime risk in the US for intellectual disability is marked with a vertical dashed line.[16] The calculations underlying these figures are found in the Supplementary Note of this appendix.

Supplementary Tables

Table S1. Within-family AUC and R^2 of Polygenic Scores for Various Clinical Outcomes

				Within-family R^2			
	AUC[8]	LTR/Prev	Source	EUR	AMR	EAS	AFR
Type 1 diabetes	0.643	0.37%	Menke et al., 2013[11]	0.029	0.018	0.015	0.007
Type 2 diabetes	0.599	28.58%	Narayan et al., 2003[12]	0.038	0.024	0.019	0.008
Breast Cancer (women)	0.567	13.13%	Howlader et al., 2020[10]	0.015	0.009	0.007	0.003
Prostate Cancer (men)	0.654	11.28%	Howlader et al., 2020[10]	0.077	0.048	0.038	0.017
Malignant Melanoma	0.585	2.66%	Howlader et al., 2020[10]	0.016	0.010	0.008	0.004
Testicular Cancer (men)	0.631	0.48%	Howlader et al., 2020[10]	0.026	0.016	0.013	0.006
Coronary artery disease	0.570	6.9%	Benjamin et al, 2019[13]	0.014	0.009	0.007	0.003
Hypercholesterolemia	0.622	12.1%	Benjamin et al, 2019[13]	0.048	0.030	0.024	0.011
Hypertension	0.635	44.8%	Benjamin et al, 2019[13]	0.073	0.046	0.037	0.016
Idiopathic short stature	--	--	--	0.317	0.198	0.159	0.070
Intellectual disability	--	--	--	0.052	0.033	0.026	0.012

Note: This table reports the within-family AUC, lifetime risk/prevalence (LTR/Prev), and implied liability-scale, within-family R^2 values for several clinical outcomes and four ancestry groups. AUC and LTR/Prev values were drawn from the literature, and the R^2 values were inferred from the AUC and LTR/Prev. Because idiopathic short stature and intellectual disability correspond to clinical cut-offs of continuous outcomes (height and cognitive performance, respectively), we directly use the R^2 estimates for the corresponding continuous outcome in those cases.[6,18,26] We shrink the R^2 values for the AMR, EAS, and AFR populations relative to the estimates for the EUR population by a constant factor based on Martin et al.[21] For more details on these calculations, see the Supplementary Note.

References

1. Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science (80-.).* **340**, 1467–1471 (2013).
2. Karavani, E. *et al.* Screening human embryos for polygenic traits has limited utility. *Cell* **179**, 1424–1435 (2019).
3. Daetwyler, H. D., Villanueva, B. & Woolliams, J. A. Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach. *PLoS One* **3**, e3395 (2008).
4. Lencz, T. *et al.* Utility of polygenic embryo screening for disease depends on the selection strategy. *bioRxiv* 2020.11.05.370478 (2020). doi:10.1101/2020.11.05.370478
5. Branigan, A. R. *et al.* Variation in the Heritability of Educational Attainment: An International Meta-Analysis. *Soc. Forces* **92**, 109–140 (2013).
6. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
7. Kong, A. *et al.* The nature of nurture: Effects of parental genotypes. *Science (80-.).* **359**, 424–428 (2018).
8. de Vlaming, R. *et al.* Meta-GWAS Accuracy and Power (MetaGAP) Calculator Shows that Hiding Heritability Is Partially Due to Imperfect Genetic Correlations across Studies. *PLOS Genet.* **13**, e1006495 (2017).
9. Lello, L., Raben, T. & Hsu, S. D. H. Within-Family Validation of Polygenic Risk Scores and Complex Trait Prediction. *bioRxiv* (2020).
10. Wray, N. R., Yang, J., Goddard, M. E. & Visscher, P. M. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet* **6**, e1000864 (2010).
11. Howlader, N. *et al. SEER Cancer Statistics Review, 1975-2017*. (2020).
12. Menke, A. *et al.* The prevalence of type 1 diabetes in the United States. *Epidemiology* **24**, 773 (2013).
13. Narayan, K. M. V., Boyle, J. P., Thompson, T. J., Sorensen, S. W. & Williamson, D. F. Lifetime risk for diabetes mellitus in the United States. *Jama* **290**, 1884–1890 (2003).
14. Benjamin, E. J. *et al.* Heart disease and stroke Statistics-2019 update a report from the American Heart Association. *Circulation* (2019).
15. Wit, J. M. *et al.* Idiopathic short stature: definition, epidemiology, and diagnostic evaluation. *Growth Horm. IGF Res.* **18**, 89–110 (2008).
16. Luckasson, R. Intellectual disability. in *Encyclopedia of Mental Health* (ed. Friedman, H. S.) 395–399 (Elsevier, 2016).
17. Lloyd-Jones, L. R. *et al.* Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* **10**, 1–11 (2019).
18. Selzam, S. *et al.* Comparing within-and between-family polygenic score prediction. *Am. J. Hum. Genet.* **105**, 351–363 (2019).
19. Harris, K. M. The Add Health Study : Design and Accomplishments. *Chapel Hill Carolina Popul. Center, Univ. North Carolina Chapel Hill* 1–22 (2013).
20. Herd, P., Carr, D. & Roan, C. Cohort Profile: Wisconsin longitudinal study (WLS). *Int. J. Epidemiol.* **43**, 34–41 (2014).
21. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health

disparities. *Nat. Genet.* **51**, 584–591 (2019).
22. Wishart, J., Kondo, T. & Elderton, E. M. The mean and second moment coefficient of the multiple correlation coefficient, in samples from a normal population. *Biometrika* 353–376 (1931).
23. Hall, P. Multiple and partial correlation coefficients in the case of an n-fold variate system. *Biometrika* 100–109 (1927).
24. Sullivan, P. F., Daly, M. J. & O'donovan, M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat. Rev. Genet.* **13**, 537–551 (2012).
25. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
26. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).