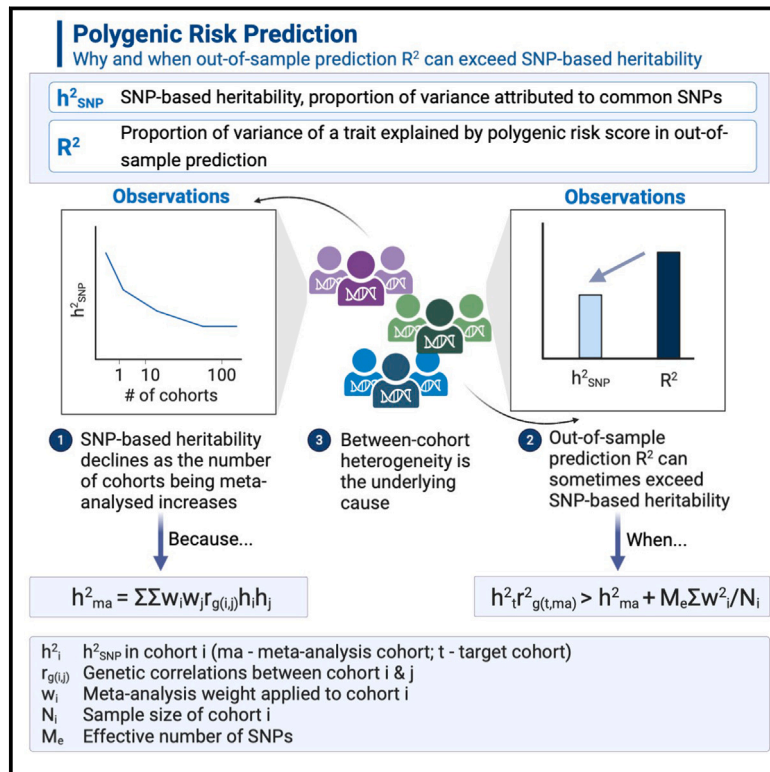


Polygenic risk prediction: why and when out-of-sample prediction R^2 can exceed SNP-based heritability

Graphical abstract



Authors

Xiaotong Wang, Alicia Walker, Joana A. Revez, ..., Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Peter M. Visscher, Naomi R. Wray

Correspondence

peter.visscher@uq.edu.au (P.M.V.),
naomi.wray@uq.edu.au (N.R.W.)

SNP-based heritability estimates tend to decline and then plateau as the number of cohorts being meta-analyzed in a GWAS increase, and the out-of-sample prediction R^2 in "target" cohorts can sometimes exceed its theoretical upper limit. Here, we provide theory to explain these observations that reflect heterogeneity between cohorts in meta-analyses.



Polygenic risk prediction: why and when out-of-sample prediction R^2 can exceed SNP-based heritability

Xiaotong Wang,¹ Alicia Walker,¹ Joana A. Revez,¹ Guiyan Ni,¹ Mark J. Adams,² Andrew M. McIntosh,^{2,3} Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Peter M. Visscher,^{1,*} and Naomi R. Wray^{1,4,*}

Summary

In polygenic score (PGS) analysis, the coefficient of determination (R^2) is a key statistic to evaluate efficacy. R^2 is the proportion of phenotypic variance explained by the PGS, calculated in a cohort that is independent of the genome-wide association study (GWAS) that provided estimates of allelic effect sizes. The SNP-based heritability (h_{SNP}^2 , the proportion of total phenotypic variances attributable to all common SNPs) is the theoretical upper limit of the out-of-sample prediction R^2 . However, in real data analyses R^2 has been reported to exceed h_{SNP}^2 , which occurs in parallel with the observation that h_{SNP}^2 estimates tend to decline as the number of cohorts being meta-analyzed increases. Here, we quantify why and when these observations are expected. Using theory and simulation, we show that if heterogeneities in cohort-specific h_{SNP}^2 exist, or if genetic correlations between cohorts are less than one, h_{SNP}^2 estimates can decrease as the number of cohorts being meta-analyzed increases. We derive conditions when the out-of-sample prediction R^2 will be greater than h_{SNP}^2 and show the validity of our derivations with real data from a binary trait (major depression) and a continuous trait (educational attainment). Our research calls for a better approach to integrating information from multiple cohorts to address issues of between-cohort heterogeneity.

Complex human traits (such as educational attainment) or complex diseases (such as major depression) are polygenic.¹ Trait-associated alleles can be identified in genome-wide association studies (GWASs). Polygenic scores (PGSs), estimates of the genetic contribution to a trait or disease liability for individuals, are calculated as an aggregate score of associated variants (with weights derived from GWAS results). The coefficient of determination (R^2) is a key statistic to evaluate the efficacy of PGS. R^2 is the proportion of phenotypic variance explained by the PGS in a “target” cohort independent of the GWAS used to identify risk alleles and estimate their effect sizes. By definition, the SNP-based heritability (h_{SNP}^2 , the proportion of total phenotypic variance attributable to all common SNPs²) is the upper limit of the out-of-sample prediction R^2 . The difference between h_{SNP}^2 and R^2 is attributed to measurement errors of SNP effects that decrease as sample sizes increase.³ h_{SNP}^2 can be estimated from individual-level genotype data using methods such as GREML⁴ implemented in software such as GCTA.⁵ Increased power for GWASs is achieved through meta-analysis of GWAS summary statistics from multiple cohorts, and methods to estimate h_{SNP}^2 from summary statistics are available. The LD Score Regression (LDSC)⁶ is an example of such methods that is commonly used in practice owing to its computational efficiency.⁷

In practice, a decrease in estimates of h_{SNP}^2 is often noted as the number of cohorts included in the GWAS meta-anal-

ysis increases, until the estimate reaches a plateau. For example, in the GWAS of major depression, the h_{SNP}^2 of a cohort with ~18,000 samples (9,041 affected individuals and 9,381 control subjects) was 0.21 (standard error [SE] 0.021),⁸ while in a subsequent GWAS meta-analysis of more than half million samples (135,458 affected individuals and 344,901 control subjects), the h_{SNP}^2 estimate declined to 0.087 (SE 0.004).⁹ Similar trends were also observed in educational attainment¹⁰ and Alzheimer disease.¹¹ At the same time, as opposed to the standard narrative, out-of-sample prediction R^2 can sometimes approximate or even exceed the h_{SNP}^2 estimated from the GWAS meta-analysis. For example, in studies of educational attainment, h_{SNP}^2 of years of education was 0.122 (SE 0.003), but the out-of-sample prediction R^2 in the National Longitudinal Study of Adolescent to Adult Health (Add Health) cohort was 0.158, with the lower limit of 95% confidence interval (C.I.) of 0.143.^{10,12} Here, we provide the theory to explain these observations.

Previously, de Vlaming et al.¹³ demonstrated that heterogeneity in genetic effects across cohorts attenuates the statistical power of GWASs, i.e., the empirical power from a GWAS meta-analysis is less than the power from a single-cohort GWAS of the same sample size. Their conclusions thus focused on the reduced performance of PGSs from meta-analysis compared to expectation from a sample of equal size constructed under idealized conditions of equal h_{SNP}^2 and genetic correlations (r_g) between cohorts of 1 (both conditions are expected if all cohorts

¹Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia; ²Division of Psychiatry, University of Edinburgh, Edinburgh, UK; ³Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK; ⁴Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia

*Correspondence: peter.visscher@uq.edu.au (P.M.V.), naomi.wray@uq.edu.au (N.R.W.)

<https://doi.org/10.1016/j.ajhg.2023.06.006>

© 2023 American Society of Human Genetics.



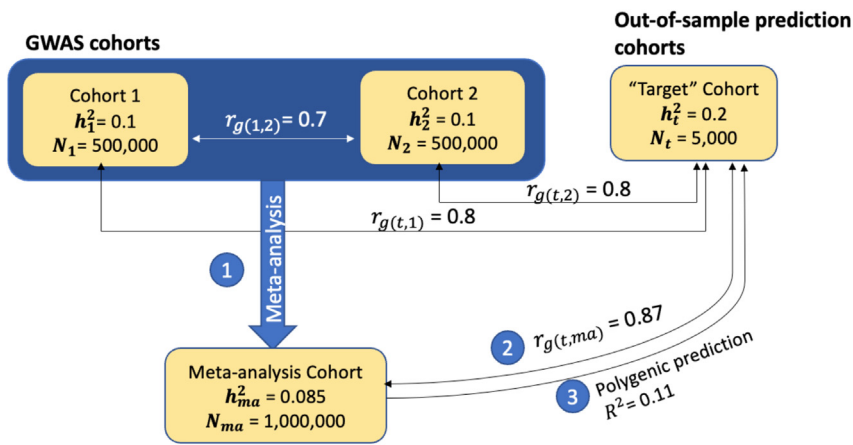


Figure 1. A schematic illustration of the report in a simplified scenario

In this made-up scenario, there are two large GWAS cohorts, each with a sample size of 500,000 and SNP-based heritability of 0.1. After meta-analysis, the results are used to generate genetic predictors in an independent cohort, the “target” cohort. (1) The h_{ma}^2 can be expressed as a function of per-cohort h_{SNP}^2 , the between-cohorts genetic correlations, and meta-analysis weights (Equation 2) or directly estimated from the meta-analysis summary statistics (these two estimates should be consistent). (2) Similarly, the genetic correlation between the meta-analyzed cohort and the target sample ($r_{g(ma,t)}$) can be calculated from Equation 4 (which should be consistent with that estimated from summary statistics of the

“target” and the meta-analysis cohort using LD Score Regression). (3) From Equation 6, $h_t^2 r_{g(ma,t)}^2 > h_{ma}^2 + \frac{M_t}{N_t}$, the out-of-sample prediction R^2 (0.11) is greater than the SNP-based heritability in the meta-analysis cohort (0.085) that is used to generate genetic predictors.

are random samples of the same population and with phenotype measured in the same way). Using similar principles, we derive an equation for the expected value of the SNP-based heritability of the meta-analysis GWAS (h_{ma}^2) as a function of cohort-specific SNP-based heritabilities and between-cohort genetic correlations and show this explains the observed decrease in h_{ma}^2 estimates as the number of cohorts in a GWAS meta-analysis increases. We show the validity of our derivations using simulation and empirical data. Building on the work from de Vlaming et al.,¹³ Dudbridge,¹⁴ and Daetwyler et al.,¹⁵ we derive theoretical conditions when out-of-sample prediction R^2 can exceed h_{ma}^2 (it will occur only when the SNP-based heritability of the target sample is greater than h_{ma}^2) and test our theory with major depression and educational attainment datasets (Figure 1).

Our derivations require recognition of analytical approaches taken in practice and the assumptions of the underlying true model on which they depend, which contrast to an alternative model that likely operates when bringing together real datasets. An underlying assumption of standard GWAS meta-analyses is that each contributing GWAS cohort is a random sample from a homogeneous idealized population, such that the “true” SNP effects and the “true” h_{SNP}^2 of each cohort are the same and that the “true” genetic correlations between cohorts are 1. Hence, the assumption is that differences in estimated h_{SNP}^2 between cohorts and that genetic correlations less than 1 simply reflect statistical sampling. However, with real data these assumptions may be violated. For example, genetic ancestry between cohorts can be different even when labeled the same continent-based ancestry, and/or there may be differences in experimental settings and/or measured phenotypes.¹⁶ Notably, GWAS results derived from population-based databases may use “proxy” phenotypes in place of formal clinical phenotypes. For example, genetic understanding of major depressive disorder has

been facilitated by use of datasets that record major depression proxy phenotypes (e.g., from different UK Biobank data fields, multiple major depression phenotypes have been derived and significant variabilities in h_{SNP}^2 have been reported in these phenotypes, and genetic correlations are significantly less than 1¹⁷).

The expected value of the parameter h_{ma}^2 can be expressed as a function of cohort-specific SNP-based heritability (i.e., true h_{SNP}^2 of the “population” from which the cohort is sampled) and between-cohorts genetic correlations (i.e., the true genetic correlations between the cohort populations) (supplemental note). Here “population” reflects genetic ancestry, phenotype definition and sampling frame of the phenotype:

$$h_{ma}^2 = \sum_{i=1}^C \sum_{j=1}^C w_i w_j r_{g(i,j)} h_i h_j \quad (\text{Equation 1})$$

Here,

- w_i is the meta-analysis weight applied to the i^{th} cohort
- h_i^2 is the true h_{SNP}^2 of the i^{th} population from which the i^{th} cohort is sampled, $h_i = \sqrt{h_i^2}$. In practice, h_i is commonly replaced by \hat{h}_i , the h_{SNP}^2 estimated in the i^{th} cohort
- $r_{g(i,j)}$ is the true genetic correlation between the i^{th} and j^{th} populations from which the i^{th} and j^{th} cohorts are sampled. Similarly, $r_{g(i,j)}$ is commonly replaced by $\hat{r}_{g(i,j)}$, genetic correlations estimated between i^{th} and j^{th} cohorts
- C is the number of cohorts included in the meta-analysis

Notably, for the purpose of our study we have defined h_{ma}^2 as parameter whose definition depends on the specific cohorts and their sample sizes that contributed to the GWAS meta-analysis.

In practice, the per cohort weights (w_i) in Equation 1 are derived from the fixed-effect inverse-variance meta-analysis (IVM) method that is commonly used in meta-analysis of GWASs from multiple cohorts. Under this model, the h_{ma}^2 can be written as:

$$h_{ma}^2 = \frac{1}{N_T^2} \sum_{i=1}^C \sum_{j=1}^C N_i N_j r_{g(i,j)} h_i h_j \quad (\text{Equation 2})$$

where,

- N_i is the effective sample size for the i^{th} cohort in the meta-analysis
- N_T is the total effective sample size

The estimate of genetic correlation between the meta-analysis cohort (subscript ma) and the target cohort used in out-of-sample prediction (subscript t) is:

$$r_{g(ma,t)} = \frac{\sum_{i=1}^C w_i r_{g(i,t)} h_i}{\sqrt{\sum_{i=1}^C \sum_{j=1}^C w_i w_j r_{g(i,j)} h_i h_j}} \quad (\text{Equation 3})$$

which under the fixed-effect IVM model is,

$$r_{g(ma,t)} = \frac{\sum_{i=1}^C N_i r_{g(i,t)} h_i}{\sqrt{\sum_{i=1}^C \sum_{j=1}^C N_i N_j r_{g(i,j)} h_i h_j}} \quad (\text{Equation 4})$$

Building on these results, we extended work from de Vlaming et al.,¹³ Dudbridge,¹⁴ and Daetwyler et al.¹⁵ and derived theoretical conditions for when out-of-sample prediction R^2 can exceed the h_{SNP}^2 estimated from the meta-analyzed cohorts that provide the PGS weights (supplemental note) and found that this occurs when the product of the SNP-based heritability of the left-out sample (h_t^2) with the squared genetic correlation of the left-out sample and the meta-analyzed sample ($r_{g(ma,t)}^2$) exceeds the sum of the estimated SNP-based heritability of the meta-analyzed sample used to generate the PGS (Equation 2) and a term associated with the error of the estimates in the meta-analyzed sample ($\frac{M_e}{N_T}$), i.e.,

$$\begin{aligned} h_t^2 r_{g(ma,t)}^2 &> \sum_{i=1}^C \sum_{j=1}^C w_i w_j r_{g(i,j)} h_i h_j + M_e \sum_{i=1}^C \frac{w_i^2}{N_i} \\ &= h_{ma}^2 + M_e \sum_{i=1}^C \frac{w_i^2}{N_i} \end{aligned} \quad (\text{Equation 5})$$

Note, this inequality also explains why in idealized settings (where $h_t^2 = h_{ma}^2$ and $r_{g(ma,t)} = 1$), h_{ma}^2 should be the upper limit of out-of-sample prediction R^2 . In idealized settings, the inequality above will never hold (i.e., h_{ma}^2 is an upper limit of the out-of-sample coefficient of variation) because the term $\sum_{i=1}^C w_i^2 \frac{M_e}{N_i}$ is always greater than 0, but will approximate 0 with decreasing standard errors (increasing large sample sizes).

Under the fixed-effect inverse variance assumptions, the inequality can be expressed as:

$$h_t^2 r_{g(ma,t)}^2 > \frac{1}{N_T^2} \sum_{i,j} N_i N_j r_{g(i,j)} h_i h_j + \frac{M_e}{N_T} \quad (\text{Equation 6})$$

where i and j are the i^{th} and j^{th} cohort in the meta-analysis, respectively. M_e is the effective number of SNPs, which is defined as¹⁸:

$$M_e = \frac{M_T^2}{\sum_{k=1}^{M_T} \sum_{j=1}^{M_T} r_{jk}^2}$$

M_T is the total number of SNPs included in the GWAS and r_{jk}^2 is a standard measurement of the LD between the SNP j and SNP k in the study.¹⁹ M_e in European populations for common SNPs on a standard GWAS chip array is approximately 60,000.²⁰

To illustrate how heterogeneities in r_g and h_i^2 will affect h_{ma}^2 and to explain empirical observations, we use simulations (supplemental methods S1) to investigate the impact of varying r_g and h_{SNP}^2 on estimates of h_{ma}^2 . To reflect common practice, the meta-analysis weights are determined under the fixed-effect IVM model. In brief, we simulated h_{SNP}^2 of 100 cohorts and pairwise r_g between these 100 cohorts. The underlying true h_{SNP}^2 is set to be either the same across cohorts (where differences in h_i^2 estimates are purely attributed to sampling variation) or set to be different across cohorts (where differences in both the underlying true h_i^2 and sampling variation contribute to variation in estimates h_i^2). We simulated between-cohort r_g under similar assumptions. We arbitrarily chose 0.2, 0.5, and 0.8 as true underlying h_{SNP}^2 and r_g . Cohort sample sizes were simulated under four different settings. For each combination of h_{SNP}^2 , r_g , and sample sizes settings, the simulation was replicated 100 times. As shown in Figure 2 (sample sizes between 5,000 and 10,000) and Figures S1–S3 (other sample sizes), when r_g and h_{SNP}^2 vary, the h_{ma}^2 drops initially with an increasing number of cohorts being meta-analyzed, but eventually reaches a plateau. Note that in a single simulation, h_{ma}^2 can both increase and decrease with an increasing number of cohorts, but the average over simulations shows the clear trend to a decreased plateau value (Figure S4). The overall trends are consistent across different h_{SNP}^2 , r_g , and sample size assumptions, and the main difference is the increased standard error with increased heterogeneity or decreased sample sizes.

To show empirical validity of our derivations, we obtained GWAS summary statistics of 21 cohorts for major depression²¹ (Table S1; supplemental method S2.1). The h_{SNP}^2 estimates of these cohorts, along with genetic correlations between them, were estimated using LDSC,⁶ following standard practice of the Psychiatric Genomics Consortium (PGC). We randomly ordered the 21 selected cohorts, and meta-analyzed (using IVM and the software METAL²²) adding one cohort at a time. With the 21 resulting meta-analysis summary statistics we estimated their

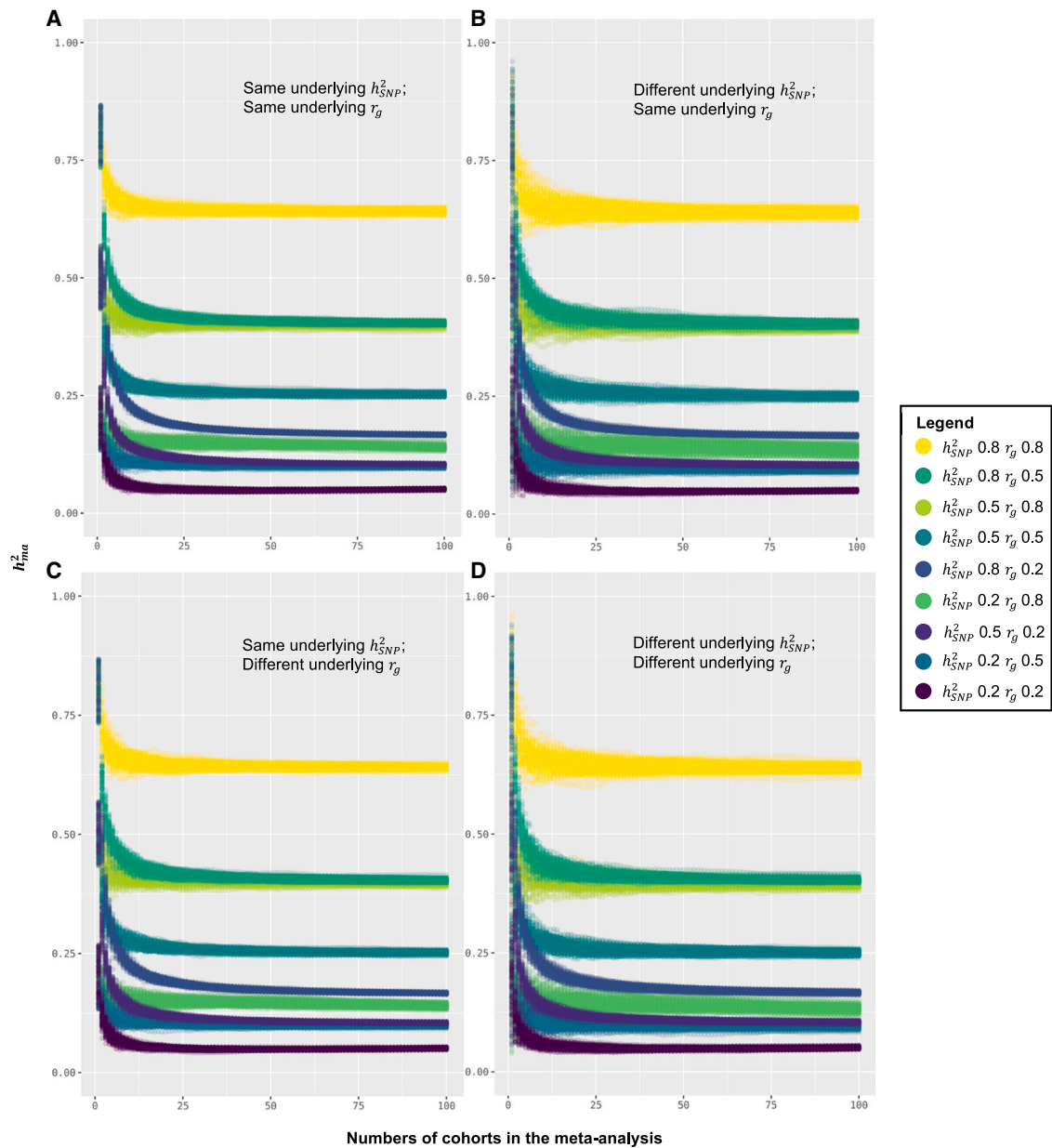


Figure 2. Estimates of h_{ma}^2 as the number of cohorts being meta-analyzed increases under different r_g and h_{SNP}^2 simulation settings. X axes are the number of cohorts being meta-analyzed, and y axes are the estimates of h_{ma}^2 . Each cohort being meta-analyzed is simulated to have a sample size between 5,000 and 10,000, and the cohort with the largest effective sample size was meta-analyzed first. (A) True r_g and h_{SNP}^2 are the same in all cohorts so variation between simulation replicates represents sampling variation. (B) True r_g are the same but h_{SNP}^2 are different between cohorts. (C) True h_{SNP}^2 are the same but r_g are different between cohorts. (D) True r_g and h_{SNP}^2 are both different in each cohort. Each scenario (depicted by color) has been repeated 100 times and so the graph shows the variation across replicates. [Figures S1–S3](#) show similar simulation results but for different sample sizes.

h_{ma}^2 using LDSC⁶ (supplemental method S2.2) and show good agreement with estimates of h_{ma}^2 from Equation 2 (Figure 3; supplemental method S2.3). Notably, in some meta-analyses addition of a cohort generates an increase in the magnitude of h_{ma}^2 , both estimated from the data and predicted from Equation 2. However, the clear trend is a decrease in the estimated h_{ma}^2 as more cohorts are added, a reflection of the estimated genetic correlations be-

ing less than 1. To demonstrate consistent estimates of $r_{g(ma,t)}$ estimated directly and from Equation 4, we held out the last cohort being meta-analyzed and estimated genetic correlations between the meta-analyzed sample and target samples with the LDSC⁶ (20 resulting $r_{g(ma,t)}$ estimates) (Figure 3). We repeated these analyses with different orders of the cohorts and show our derivations are valid regardless of this (Figure 3).

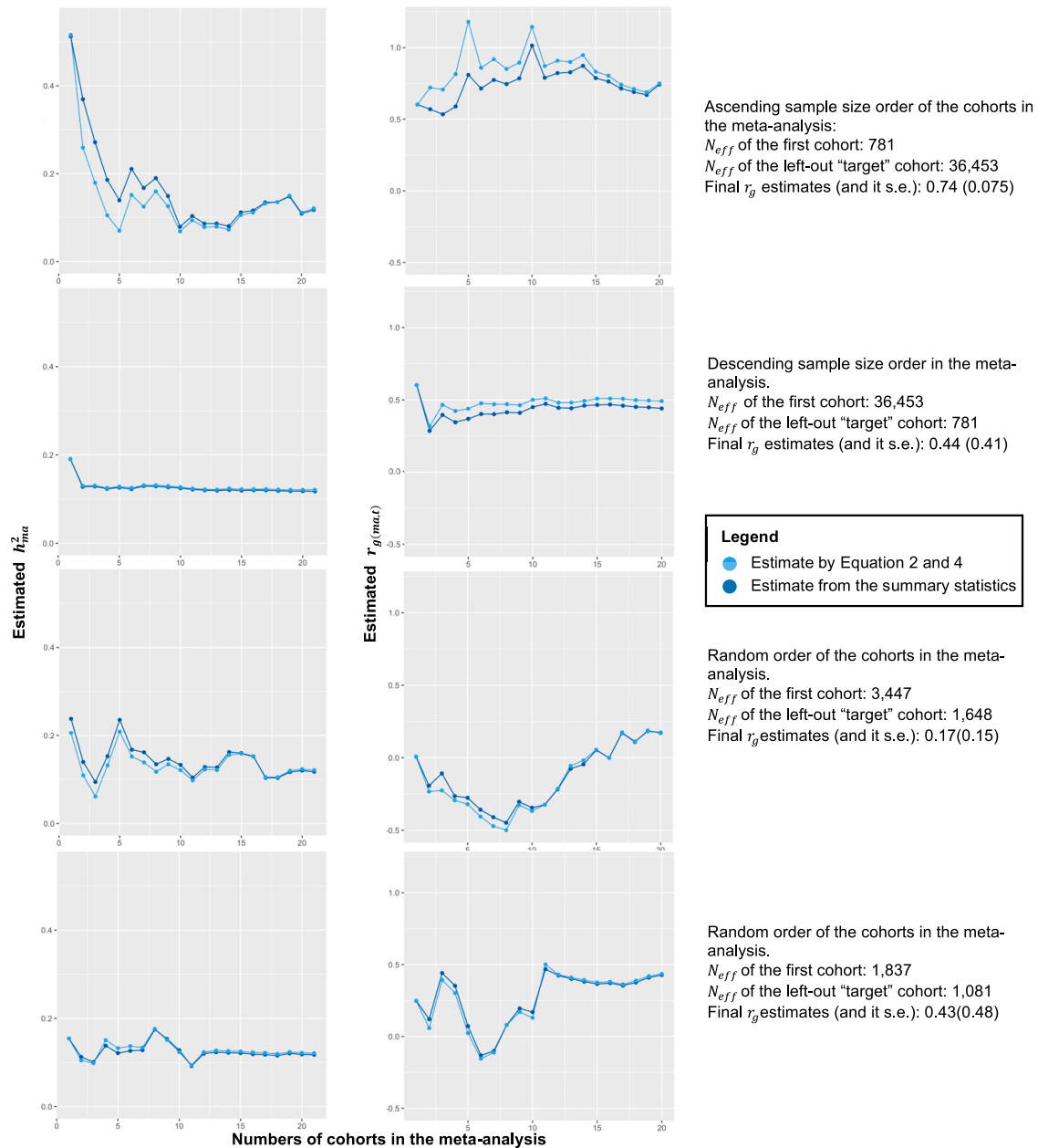


Figure 3. Comparison of h_{ma}^2 and $r_{g(ma,t)}^2$ estimated by the derived equations with those directly estimated from the meta-analysis summary statistics

Left: 21 major depression cohorts were meta-analyzed, adding one cohort at a time, in different orders. The first meta-analysis results are simply the GWAS summary statistics of the first cohort. Right: We held out the last cohort being meta-analyzed and calculated genetic correlations between the left-out cohort and each of 20 remaining meta-analyses. X axes are the number of cohorts included in the meta-analysis. In general, h_{ma}^2 (first column) and the $r_{g(ma,t)}^2$ (second column) estimated with the formulae are consistent with those directly estimated from the meta-analysis summary statistics. Notably, when the sample size of the left-out cohort is small, $r_{g(ma,t)}^2$ estimated with two methods could have some insignificant differences because of large standard errors.

To show that out-of-sample prediction R^2 can be higher than the estimated h_{ma}^2 under the conditions outlined by the inequality in Equation 6 above, we chose a binary trait (major depression, MD²³) and a continuous trait (educational attainment, EA¹⁰) where cohort-specific SNP-based heritabilities and between-cohort genetic correlations were available, as proof-of-principle examples.

For MD (supplemental method S3.1), we obtained access to results of leave-one-cohort out (LOO) analysis.^{21,24} In

brief, for the 26 cohorts with individual-level genotype data, one cohort at a time was left out and the remaining cohorts meta-analyzed, together with 9 additional cohorts whose GWAS summary statistics (but not individual-level data) were available. PGS were calculated for all individuals in the left-out sample with SNP weights derived from the LOO meta-analysis summary statistics using the SBayesR method with default settings.²⁵ The h_t^2 of the left-out sample, h_{ma}^2 of the LOO meta-analysis, and the $r_{g(ma,t)}^2$ between

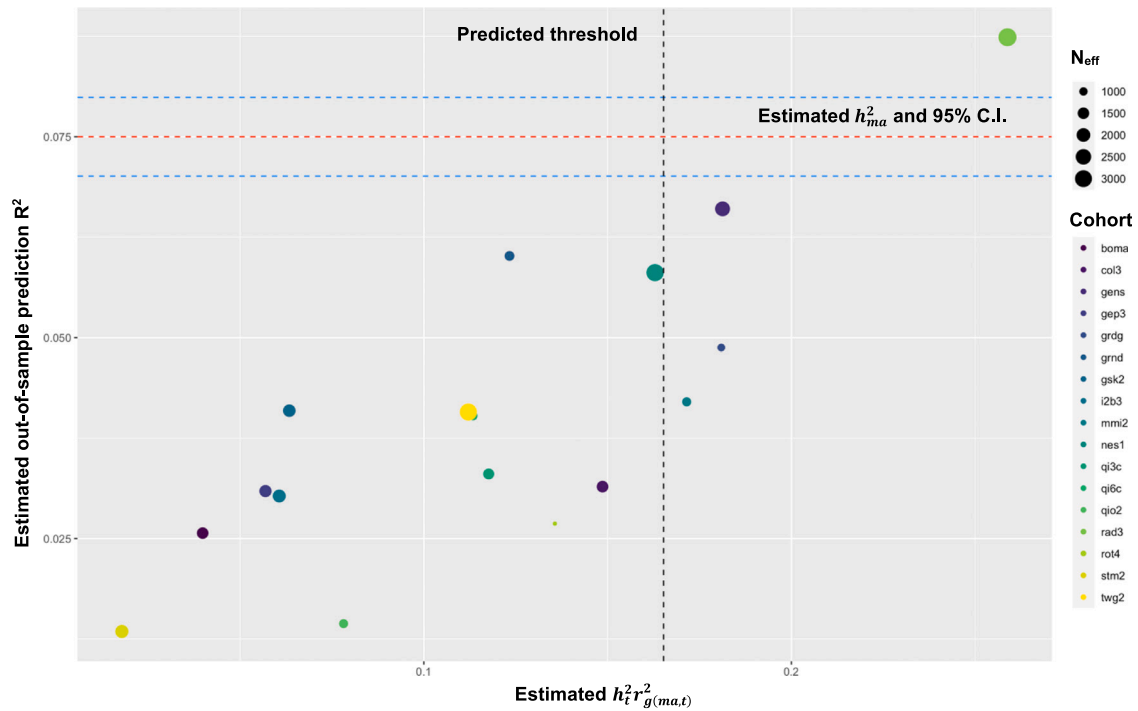


Figure 4. Empirical investigation of Equation 6 for major depression datasets

In this figure, the x axis is the product of h_t^2 (estimates of SNP-based heritability on the liability scale of target cohorts being predicted into) and $r_g^2(ma,t)$ (squared genetic correlations between the leave-one-cohort out meta-analysis used to generate the PGS, and the left-out target cohort where PGSs are calculated); the y axis is the out-of-sample prediction R^2 on the liability scale calculated in the target cohort. Each dot represents a target cohort (only cohorts where estimated h_t^2 is between 0 and 1 are considered, sizes of dots are in proportion to the effective sample size). The horizontal line denotes h_{ma}^2 (red dotted line) and its 95% C.I. (blue dotted line). (Although we left a different cohort out each time, the h_{ma}^2 and 95% CI remain unchanged because the sample size of the left-out cohort is small when compared with the total sample size of the meta-analysis.) Our derivations show that when the h_{SNP}^2 in the cohort being predicted into is higher than the predicted threshold (vertical black line), out-of-sample prediction R^2 will exceed the h_{ma}^2 in the meta-analysis used to generate the predictor. Small cohorts have estimates (depicted by small dots) with large standard errors (see Table S2).

the left-out sample and the remaining cohorts were estimated using LDSC and HapMap3 SNPs. For the left-out cohort, we retained only those h_{SNP}^2 estimates that are greater than 0 and smaller than 1 (Figure 4) or retained all cohorts whose h_{SNP}^2 are available (Figure S5). Results show that the derived inequality (Equation 6) is consistent with empirical results.

For EA, we obtained cohort-specific SNP-based heritability estimates and between cohort genetic correlations for all pairs of cohorts from the supplemental file of Lee et al.¹⁰ We removed cohorts with heritability estimates smaller than 0 or larger than 1, or not available (supplemental method S3.2). 35 out of 71 cohorts met the criteria. We conducted analyses as for MD, but LOO PGS results were available for only 2 cohorts. The empirical results again agree with our derived expectation (Figure S6).

In conclusion, in this report, we demonstrate that the h_{ma}^2 can be expressed in terms of per-cohort h_{SNP}^2 , between-cohort genetic correlations, and meta-analysis weights (which are a function of the sample sizes under the commonly used fixed effect IVM model). Under idealized conditions where between-cohorts genetic correlations are all equal to 1 and all cohorts have a common

SNP-based heritability, the out-of-sample prediction R^2 will always be smaller than SNP-based heritability (smaller because of error associated with estimates of SNP effect sizes). The difference between h_{SNP}^2 and R^2 will tend to be 0 with an infinitely large sample size. However, when h_{SNP}^2 and r_g heterogeneities exist, h_{SNP}^2 estimates made from GWAS meta-analysis results will decrease as the numbers of meta-analyzed cohorts increases (Equation 2) until reaching a plateau, and the out-of-sample prediction R^2 can be greater than SNP-based heritability (Equation 6). Notably, a key assumption of the fixed-effect meta-analysis is that true underlying effect sizes of SNPs are the same for each cohort, and the experimental settings and measured phenotype are the same.¹⁶ These assumptions do not always hold, especially when population-based databases are used where phenotypes may be proxy phenotypes. With the knowledge of how between-cohort heterogeneity can impact SNP-based heritability estimates, it may be relevant to select cohorts that represent the focal trait (e.g., clinically measured major depressive disorder for MD or years of education [as opposed to the proxy trait of attended college yes/no] for EA) and treat other cohorts as genetically correlated traits (i.e., an MTAG analysis).²⁶

Consortia

Members of the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium include Naomi R. Wray, Stephan Ripke, Manuel Mattheisen, Maciej Trzaskowski, Enda M. Byrne, Abdel Abdellaoui, Mark J. Adams, Esben Agerbo, Tracy M. Air, Till F.M. Andlauer, Silviu-Alin Bacanu, Marie Bækvad-Hansen, Aartjan T.F. Beekman, Tim B. Bigdeli, Elisabeth B. Binder, Julien Bryois, Henriette N. Buttenschøn, Jonas Bybjerg-Grauholm, Na Cai, Enrique Castelao, Jane Hvarregaard Christensen, Toni-Kim Clarke, Jonathan R.I. Coleman, Lucía Colodro-Conde, Baptiste Couvy-Duchesne, Nick Craddock, Gregory E. Crawford, Gail Davies, Franziska Degenhardt, Eske M. Derks, Nese Direk, Conor V. Dolan, Erin C. Dunn, Thalia C. Eley, Valentina Escott-Price, Farnush Farhadi Hassan Kiadeh, Hilary K. Finucane, Jerome C. Foo, Andreas J. Forstner, Josef Frank, Héléna A. Gaspar, Michael Gill, Fernando S. Goes, Scott D. Gordon, Jakob Grove, Lynsey S. Hall, Christine Søholm Hansen, Thomas F. Hansen, Stefan Herms, Ian B. Hickie, Per Hoffmann, Georg Homuth, Carsten Horn, Jouke-Jan Hotenga, David M. Hougaard, David M. Howard, Marcus Ising, Rick Jansen, Ian Jones, Lisa A. Jones, Eric Jorgenson, James A. Knowles, Isaac S. Kohane, Julia Kraft, Warren W. Kretschmar, Zoltán Kutalik, Yihan Li, Penelope A. Lind, Donald J. MacIntyre, Dean F. MacKinnon, Robert M. Maier, Wolfgang Maier, Jonathan Marchini, Hamdi Mbarek, Patrick McGrath, Peter McGuffin, Sarah E. Medland, Divya Mehta, Christel M. Middeldorp, Evelin Mihailov, Yuri Milaneschi, Lili Milani, Francis M. Mondimore, Grant W. Montgomery, Sara Mostafavi, Niamh Mullins, Matthias Nauck, Bernard Ng, Michel G. Nivard, Dale R. Nyholt, Paul F. O'Reilly, Hogni Oskarsson, Michael J. Owen, Jodie N. Painter, Carsten Bøcker Pedersen, Marianne Giørtz Pedersen, Roseann E. Peterson, Wouter J. Peyrot, Giorgio Pistis, Danielle Posthuma, Jorge A. Quiroz, Per Qvist, John P. Rice, Brien P. Riley, Margarita Rivera, Saira Saeed Mirza, Robert Schoevers, Eva C. Schulte, Ling Shen, Jianxin Shi, Stanley I. Shyn, Engilbert Sigurdsson, Grant C.B. Sinnamon, Johannes H. Smit, Daniel J. Smith, Hreinn Stefansson, Stacy Steinberg, Fabian Streit, Jana Strohmaier, Katherine E. Tansey, Henning Teismann, Alexander Teumer, Wesley Thompson, Pippa A. Thomson, Thorgeir E. Thorgeirsson, Matthew Traylor, Jens Treutlein, Vassily Trubetskoy, André G. Uitterlinden, Daniel Umbricht, Sandra Van der Auwera, Albert M. van Hemert, Alexander Viktorin, Peter M. Visscher, Yunpeng Wang, Bradley T. Webb, Shantel Marie Weinsheimer, Jürgen Wellmann, Gonneke Willemssen, Stephanie H. Witt, Yang Wu, Hualin S. Xi, Jian Yang, Futao Zhang, Volker Arolt, Bernhard T. Baune, Klaus Berger, Dorret I. Boomsma, Sven Cichon, Udo Dannlowski, E.J.C. de Geus, J. Raymond DePaulo, Enrico Domenici, Katharina Domschke, Tõnu Esko, Hans J. Grabe, Steven P. Hamilton, Caroline Hayward, Andrew C. Heath,

Kenneth S. Kendler, Stefan Kloiber, Glyn Lewis, Qingqin S. Li, Susanne Lucae, Pamela A.F. Madden, Patrik K. Magnusson, Nicholas G. Martin, Andrew M. McIntosh, Andres Metspalu, Ole Mors, Preben Bo Mortensen, Bertram Müller-Myhsok, Merete Nordentoft, Markus M. Nöthen, Michael C. O'Donovan, Sara A. Paciga, Nancy L. Pedersen, Brenda W.J.H. Penninx, Roy H. Perlis, David J. Porteous, James B. Potash, Martin Preisig, Marcella Rietschel, Catherine Schaefer, Thomas G. Schulze, Jordan W. Smoller, Kari Stefansson, Henning Tiemeier, Rudolf Uher, Henry Völzke, Myrna M. Weissman, Thomas Werge, Cathryn M. Lewis, Douglas F. Levinson, Gerome Breen, Anders D. Børglum, and Patrick F. Sullivan.

Data and code availability

The major depression data used in this study is available from the Psychiatric Genomics Consortium: <https://pgc.unc.edu/researchers/data-access-committee/data-access-information/>.

The meta-analysis summary statistics of all cohorts can be downloaded directly from the PGC website. To access cohort-specific GWAS summary statistics, the researcher must request data access by submitting a research proposal to the PGC.

The educational attainment data used in the study were available in the supplemental files of Lee et al. study.¹⁰

Code for this report can be found on Github: <https://github.com/mark-xiaotong-wang/out-of-sample-prediction-r2>.

Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2023.06.006>.

Acknowledgments

We acknowledge funding from the Australian National Health & Medical Research Council (1173790, 1113400), Australian Research Council (FL180100072), and the National Institute of Mental Health (R01MH124871, R01MH121545). This work would not have been possible without the contributions of the investigators who comprise the PGC-MDD working group. The procedures followed in the PGC-MDD working group were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and that proper informed consent was obtained. Data analysis was conducted under the University of Queensland Human Research Ethics Committee approval HE002938. For a full list of acknowledgments and ethical statements of all individual cohorts, please see the original publications. The PGC has received major funding from the US National Institute of Mental Health and the US National Institute on Drug Abuse (U01 MH109528 and U01 MH1095320). Some statistical analyses were carried out on the NL Genetic Cluster Computer (<http://www.geneticcluster.org/>) hosted by SURFsara who support the PGC through grants to Danielle Posthuma. GWAS summary statistics from 23andMe were included in the meta-analyzed GWAS summary statistics. We thank the customers, research participants, and employees of 23andMe for making this work possible. The study protocol used by 23andMe was

approved by an external AAHRPP-accredited institutional review board. The graphical abstract was created with BioRender.com.

Author contributions

Study motivation, N.R.W., A.M.McI.; theory, X.W., P.M.V., N.R.W.; simulations & analyses, X.W.; data preparation, A.W., J.A.R., G.N., M.J.A.; first draft, X.W., N.R.W., P.M.V.; final draft, all authors read and approved the manuscript.

Declaration of interests

The authors declare no competing interests.

Received: February 6, 2023

Accepted: June 12, 2023

Published: June 27, 2023

References

1. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* *101*, 5–22.
2. Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A.E., Lee, S.H., Robinson, M.R., Perry, J.R.B., Nolte, I.M., van Vliet-Ostaptchouk, J.V., et al. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* *47*, 1114–1120.
3. Visscher, P.M., Yang, J., and Goddard, M.E. (2010). A commentary on ‘common SNPs explain a large proportion of the heritability for human height’ by Yang et al. (2010). *Twin Res. Hum. Genet.* *13*, 517–524.
4. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* *42*, 565–569.
5. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* *88*, 76–82.
6. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics, C., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* *47*, 291–295.
7. Evans, L.M., Tahmasbi, R., Vrieze, S.I., Abecasis, G.R., Das, S., Gazal, S., Bjelland, D.W., De Candia, T.R., Haplotype Reference Consortium, and Goddard, M.E., et al. (2018). Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat. Genet.* *50*, 737–745.
8. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., et al. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* *45*, 984–994.
9. Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* *50*, 668–681.
10. Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghziyan, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* *50*, 1112–1121.
11. Escott-Price, V., and Hardy, J. (2022). Genome-wide association studies for Alzheimer’s disease: bigger is not always better. *Brain Commun.* *4*, fac125.
12. Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S.M., Sidorenko, J., Kweon, H., Goldman, G., Gjorgjieva, T., et al. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat. Genet.* *54*, 437–449.
13. de Vlaming, R., Okbay, A., Rietveld, C.A., Johannesson, M., Magnusson, P.K.E., Uitterlinden, A.G., van Rooij, F.J.A., Hofman, A., Groenen, P.J.F., Thurik, A.R., and Koellinger, P.D. (2017). Meta-GWAS Accuracy and Power (MetaGAP) calculator shows that hiding heritability is partially due to imperfect genetic correlations across studies. *PLoS Genet.* *13*, e1006495.
14. Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* *9*, e1003348.
15. Daetwyler, H.D., Villanueva, B., and Woolliams, J.A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* *3*, e3395.
16. Borenstein, M., Hedges, L.V., Higgins, J.P.T., and Rothstein, H.R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res. Synth. Methods* *1*, 97–111.
17. Cai, N., Revez, J.A., Adams, M.J., Andlauer, T.F.M., Breen, G., Byrne, E.M., Clarke, T.K., Forstner, A.J., Grabe, H.J., Hamilton, S.P., et al. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nat. Genet.* *52*, 437–447.
18. Goddard, M.E., Hayes, B.J., and Meuwissen, T.H.E. (2011). Using the genomic relationship matrix to predict the accuracy of genomic selection. *J. Anim. Breed. Genet.* *128*, 409–421.
19. Hill, W.G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* *38*, 226–231.
20. Wray, N.R., Yang, J., Hayes, B.J., Price, A.L., Goddard, M.E., and Visscher, P.M. (2013). Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* *14*, 507–515.
21. Howard, D.M., Adams, M.J., Clarke, T.-K., Hafferty, J.D., Gibson, J., Shireli, M., Coleman, J.R.I., Hagenaars, S.P., Ward, J., Wigmore, E.M., et al. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* *22*, 343–352.
22. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* *26*, 2190–2191.
23. Sullivan, P.F., Agrawal, A., Bulik, C.M., Andreassen, O.A., Børglum, A.D., Breen, G., Cichon, S., Edenberg, H.J., Faraone, S.V., Gelernter, J., et al. (2018). Psychiatric genomics: an update and an agenda. *Am. J. Psychiatry* *175*, 15–27.

24. Ni, G., Zeng, J., Revez, J.A., Wang, Y., Zheng, Z., Ge, T., Restuadi, R., Kiewa, J., Nyholt, D.R., Coleman, J.R.I., et al. (2021). A comparison of ten polygenic score methods for psychiatric disorders applied across multiple cohorts. *Biol. Psychiatry* *90*, 611–620.
25. Lloyd-Jones, L.R., Zeng, J., Sidorenko, J., Yengo, L., Moser, G., Kemper, K.E., Wang, H., Zheng, Z., Magi, R., Esko, T., et al. (2019). Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* *10*, 5086.
26. Turley, P., Walters, R.K., Maghzian, O., Okbay, A., Lee, J.J., Fontana, M.A., Nguyen-Viet, T.A., Wedow, R., Zacher, M., Furlotte, N.A., et al. (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* *50*, 229–237.