Supplementary Information

Pervasive Downward Bias in Estimates of Liability-Scale Heritability in Genome-Wide Association Study Meta-Analysis: A Simple Solution

S1. Properties of the Standard GWAS Model

The standard additive GWAS model estimates the marginal effect size, *b*, from the linear regression phenotype *y* on genetic variant *x*, can be written as:

$$y = bx + e \quad .$$

The sampling variance σ_b^2 (i.e. the squared standard error, *SE*) of the *b* estimate is given as:

$$\sigma_b^2 = (SE_b)^2 = \frac{\sigma_e^2}{\sigma_x^2 n} = \frac{\sigma_y^2 - \sigma_x^2(b^2)}{\sigma_x^2 n} \quad ,$$

where σ_e^2 is the residual variance of *y*, *n* is the sample size, σ_y^2 is the variance of *y*, and σ_x^2 is the variance of the variant. Without loss of generality, we assume that *y* and *x* have each been standardized to mean 0, and standard deviation 1, such that σ_b^2 reduces to:

$$\sigma_b^2 = (SE_b)^2 = \frac{\sigma_e^2}{n} = \frac{1 - b^2}{n} \quad .$$

Because GWAS effect sizes for complex traits are extremely small, $1 - b^2 \approx 1$ such that:

$$\sigma_b^2 = (SE_b)^2 \approx \frac{1}{n} \quad ,$$

$$Z = \frac{b}{SE} = \frac{b}{\frac{1}{\sqrt{n}}} = b\sqrt{n} \quad ,$$

$$Z^2 = \chi^2 = nb^2 \quad .$$

S2. Estimating Heritability and Genetic Covariance from GWAS Summary Statistics

We can model *K* phenotypes and *M* SNPs measured in *N* individuals according to the equation:

$$y_{i,k} = x_{i,j}\beta_{j,k} + \epsilon_{i,k}$$

where $y_{i,k}$ is the score for person *i* on the standardized phenotype *k,* $x_{i,j}$ is the standardized genotype for person *i* on SNP *j,* $\beta_{j,k}$ is the true standardized effect size for SNP *j* on phenotype *k,* and $\epsilon_{i,k}$ is the residual for person *i* on phenotype *k*. This model can be written in matrix form as:

$$Y = XB + E$$

where Y is an $N\times K$ matrix of standardized scores for person i on phenotype k, X is an $N\times M$ matrix of standardized genotypes for person i on SNP j, B is an $M\times K$ matrix of true standardized genotype effect sizes for SNP j on phenotype k, and E is an $N\times K$ matrix of residuals for person i on phenotype k.

Assuming independence of genotypes $x_{i,j}$ from one another, but incorporating each genotype's average linkage disequilibrium with the others for model estimation, LD Score regression[1–3] (LDSC) is used to model $\beta_{j,k}$ as phenotype-specific random effects, varying over SNPs, with $\mathbb{E}[\mathrm{B}] = 0$ and $cov[\mathrm{B}] = \frac{1}{M}S$. The diagonal elements of S contain the SNP-based heritability (h^2_{SNP}), and the off-diagonal elements of S contain the genetic covariances (σ_g) between phenotypes, i.e. the genetic correlations between phenotypes scaled relative to the SNP heritabilities of the respective phenotypes. The elements of S are estimated from GWAS summary statistics by regressing the product of Z statistics for the linear regression of phenotypes 1 and 2 on SNP j on the LD score of SNP j and solving for σ_g as follows:

$$\mathbb{E}\left[Z_{1j}Z_{2j}\right] = \sqrt{N_1N_2}\frac{\sigma_g}{M}\ell(j) + \frac{\rho N_s}{\sqrt{N_1N_2}} + a \quad ,$$

where N_1 and N_2 are the sample sizes for phenotypes 1 and phenotypes 2, M is the number of SNPs, $\ell(j)$ is the LD score of SNP j (that is, the sum of squared correlations between SNP j and all other SNPs in the reference panel), N_s is the number of individuals included in both GWAS samples, ρ is the phenotypic correlation within the overlapping samples, and a is a term representing unmeasured sources of confounding such as shared population stratification across GWASs. When the Z statistics for the same phenotype are double entered into the left-hand side of the above equation, such that $\mathbb{E}\left[Z_{1j}Z\right]$ becomes $\mathbb{E}\left[Z_j^2\right] = \mathbb{E}\left[\chi_j^2\right]$, the equation reduces to the univariate LDSC model, and σ_g becomes an estimate of h^2_{SNP}, as follows

$$\mathbb{E}\left[Z_j^2\right] = \mathbb{E}\left[\chi_j^2\right] = N\frac{h^2_{SNP}}{M}\ell(j) + a + 1 \quad .$$

Without loss of generality, we focus here on the univariate LDSC model.

The h^2_{SNP} estimate in LDSC is determined from the LDSC Slope via the coefficient $N\frac{h^2_{SNP}}{M}$. Noting that $\frac{h^2_{SNP}}{M}$ is the average variance explained per SNP, it follows that:

$$LDSC\ Slope = N\ var\left(\beta_{j,k}\right) = N\overline{\beta^2_{j,k}}$$

Noting further that LD Slope relates the LD score to the squared Z statistic, $\mathbb{E}\left[Z_j^2\right]$, we can see that the foundation of this association is the simple relation between Z and the standardized linear regression coefficient, derived earlier as:

$$Z^2 = nb^2 \quad .$$

For GWAS of binary traits, the heritability estimate from the standard LDSC equation produces an estimate of "observed scale" SNP-based heritability that depends on the sample prevalence, which for

ascertained samples, will not correspond to the population prevalence and may vary considerably across studies. This observed scale SNP-based heritability can be transformed into a more interpretable liability-scale SNP-based heritability estimate; i.e. the heritability of the continuous underlying liability toward the binary outcome that does not depend on sample prevalence. Liability-scale heritability (h_l^2) can be computed from observed scale heritability (h_o^2) as:[4–6]

$$h_l^2 = h_o^2 \frac{P^2(1-P)^2}{\phi^2 v(1-v)} \ ,$$

or more directly computed in the reduced form LDSC equation for binary traits as follows:

$$\mathbb{E}[\chi_j^2] = \frac{\phi^2}{P^2(1-P)^2} v(1-v)N \ \frac{h_l^2}{M} \ell(j) + a + 1 \ ,$$

where v is the sample prevalence, P is the population prevalence, and ϕ is the height of the standard normal density of at the threshold corresponding to P. We compare the LDSC slope to the χ^2 statistic for the transformed regression coefficient appropriate for binary traits in section **S4**.

S3. GWAS Meta-Analysis of Quantitative Traits

An inverse variance weighted meta-analysis can be used to combine effect size estimates, b_k, across k independent sets of GWAS summary statistics using weights $w_k = \frac{1}{\sigma^2_{b_k}}$. (de la Fuente et al.[7] formally demonstrate the equivalence of inverse-variance weighted meta-analysis of effect sizes and sample-size weighted meta-analysis of Z statistics.) For continuously distributed phenotypes (quantitative traits), effect sizes from linear regression are appropriate to combine directly. In this circumstance, $\frac{1}{\sigma^2_{b_k}} \approx n$ such that:

$$b_{inv\ var} = \frac{\sum w_k b_k}{\sum w_k} = \frac{\sum \frac{1}{\sigma^2_{b_k}} b_k}{\sum \frac{1}{\sigma^2_{b_k}}} = \frac{\sum n_k b_k}{\sum n_k} \ ,$$

$$SE_{b_{inv\ var}} = \frac{1}{\sqrt{\sum w_k}} = \frac{1}{\sqrt{\sum \frac{1}{\sigma^2_{b_k}}}} = \frac{1}{\sqrt{\sum n_k}} \ ,$$

$$Z_{inv\ var} = \frac{b_{inv\ var}}{SE_{b_{inv\ var}}} = \frac{b_{inv\ var}}{\frac{1}{\sqrt{\sum n_k}}} = b_{inv\ var}\sqrt{\sum n_k} \ ,$$

where $b_{inv\ var}$ is the fixed effects meta-analytic estimate of b. It follows that:

$$\chi^2_{inv\ var} = Z^2_{inv\ var} = b^2_{inv\ var} \sum n_k \ ,$$

which can be compared to the form of the LDSC slope for continuous traits:

$$LDSC\ Slope = \overline{\beta_{J,k}^2}\, N \ ,$$

thus indicating that when using meta-analytic summary statistics for traits analyzed via linear GWAS, $\sum n_k$ (i.e., the total sample size) is appropriate to enter for *N* in LDSC, as is standard practice.

S4. GWAS Meta-Analysis of Binary Traits

When the GWAS phenotype is a binary trait, the effect sizes, b_k, from linear regression must be transformed such that they are comparable metrics across GWAS before they can be combined via meta-analysis. This is of particular relevance for ascertained samples (where cases are typically oversampled) in which the degree of ascertainment varies across contributing GWAS. This is because, under a standard polygenic model of disease liability, the linear slope relating the allele count to the binary phenotype (the slope of the so-called linear probability function) depends on the sample prevalence. In contrast, because odds ratios are not dependent on sample prevalence, the logistic regression coefficient is an appropriate effect for placing GWAS on the same metric. In the context of GWAS of complex traits where individual SNP effects are extremely small, the regression coefficient for the logistic regression of the binary phenotype on the standardized variant ($b_{logit\ STDX}$) can be closely approximated from the coefficient from the linear regression of the standardized binary phenotype on the standardized variant *X* ($b_{linear\ STD}$) as[8,9]

$$b_{logit\ STDX} \approx \frac{b_{linear\ STD}}{\sqrt{v(1-v)}} \ ,$$

$$\sigma^2_{b_{logit\ STDX}} = (SE_{b_{logit\ STDX}})^2 \approx \frac{\sigma^2_{b_{linear\ STD}}}{v(1-v)} = \frac{1}{v(1-v)n} \ ,$$

$$Z = \frac{b_{logit\ STDX}}{SE_{b_{logit\ STDX}}} = \frac{b_{logit\ STDX}}{\frac{1}{\sqrt{v(1-v)n}}} = b_{logit\ STDX}\sqrt{v(1-v)n} \ ,$$

where *v* is the proportion of cases, such that *v(1-v)* is the observed variance of the binary phenotype. We use this relation to derive a corrected linear effect size, b^*, that would have been obtained had the case-control GWAS been on a balanced sample (i.e. 50% cases, 50% controls). We compute b^* by substituting .5 for *v* into the equation approximating logistic regression from linear regression:

$$b_{logit\ STDX} \approx \frac{b_{linear\ STD}}{\sqrt{v(1-v)}} = \frac{b^*}{\sqrt{.5(1-.5)}} \ ,$$

$$\sqrt{.5(1-.5)}\,\frac{b_{linear\ STD}}{\sqrt{v(1-v)}} = b^* \ ,$$

$$b^* = .5\,\frac{b_{linear\ STD}}{\sqrt{v(1-v)}} \ ,$$

$$Z = b_{logit\ STDX}\sqrt{v(1-v)n} = \frac{b^*}{\sqrt{.5(1-.5)}}\sqrt{v(1-v)n} \;=\; 2b^*\sqrt{v(1-v)n}\ ,$$

therefore,

$$Z^2 \;=\; (b^*)^2 v(1-v)4n \quad .$$

The SE of b^* is derived as:

$$Z = 2b^*\sqrt{v(1-v)n} = \frac{b^*}{SE_{b*}},$$

$$SE_{b*} = \frac{1}{2\sqrt{v(1-v)n}}\,,$$

$$\sigma^2_{b*} = (SE_{b*})^2 = \frac{1}{4v(1-v)n}\,.$$

Noting that the standard term for SE in OLS is $\frac{1}{n}$, it is sensible to refer to $4v(1-v)n$ as the *effective sample size (EffN)* for a balanced design.

Because b^* represents the effect size that would have been obtained had all studies had balanced ascertainment, they are on the same scale as one another and can be meta-analyzed. We can specify the inverse weighted meta-analysis of b^*, its *SE*, and its *Z* and χ^2 statics as:

$$b^*_{inv\ var} = \frac{\sum w_k b^*_k}{\sum w_k} = \frac{\sum 4v_k(1-v_k)n_k b^*_k}{\sum 4v_k(1-v_k)n_k} \;= \frac{\sum 4v_k(1-v_k)n_k .5\,\frac{b_{linear\ STD}}{\sqrt{v(1-v)}}}{\sum 4v_k(1-v_k)n_k}$$

$$= \;\frac{\sum \sqrt{v(1-v)}n_k b_{linear\ STD}}{2\sum v_k(1-v_k)n_k},$$

$$SE_{b^*_{inv\ var}} = \frac{1}{\sqrt{\sum w_k}} = \;\frac{1}{\sqrt{\sum 4v_k(1-v_k)n_k}}\ ,$$

$$Z^*_{inv\ var} = \frac{b^*_{inv\ var}}{SE_{b^*_{inv\ var}}} = \frac{b^*_{inv\ var}}{\frac{1}{\sqrt{\sum 4v_k(1-v_k)n_k}}} = \;b^*_{inv\ var}\sqrt{\sum 4v_k(1-v_k)n_k}\ \ ,$$

where $b^*_{inv\ var}$ is the fixed effects meta-analytic estimate of b^*. Note that, apart from a difference in scaling constants, this is equivalent to an inverse-variance weighted meta-analysis of the logistic regression coefficients themselves.

We can see that the relation between squared *Z* statistics and the squared meta-analytic effect size is governed by:

$$\chi^{*2}_{inv\ var} = Z^{*2}_{inv\ var} = \;b^{*2}_{inv\ var}\sum 4v_k(1-v_k)n_k$$

$$= b^{*2}_{inv\ var}\sum EffN_k\,,$$

which can be compared to the form of the LDSC slope

$$LDSC\ Slope = \overline{\beta^2_{J,k}}\,N\ \ ,$$

thus indicating that $\sum EffN_k$ is appropriate to use for LDSC of binary traits to produce an unbiased estimate of observed scale heritability for a balanced design, $h^2_{o^*}$.liability-scale heritability can then be obtained from the following, which now only includes terms that are constant across samples:

$$h^2_l = h^2_{o^*}\frac{P^2(1-P)^2}{\phi^2.5(1-.5)}$$

Liability scale heritability can similarly be derived for the reduced form equation by substituting $\sum EffN_k$ for *N* and setting v to .5 to reflect the fact that *EffN* corresponds to a balanced design, yielding the following for the LDSC slope:

$$LDSC\ Slope = \frac{\phi^2}{P^2(1-P)^2}v(1-v)\left(\sum EffN_k\right)\frac{h^2_l}{M} = \frac{\phi^2}{P^2(1-P)^2}.5(1-.5)\frac{h^2_l}{M}\sum 4v_k(1-v_k)n_k$$
$$= \frac{\phi^2}{P^2(1-P)^2}\frac{h^2_l}{M}\left(\sum v_k(1-v_k)n_k\right) \ .$$

Of particular relevance here is that the term in the LDSC equation that involves the product of sample size, case proportions, and control proportions must be derived via an *n*-weighted summation of individual case proportions for each set of summary statistics in the meta-analysis, not as the product of total sample size and the proportion of total cases and controls in the full, meta-analytic sample. In other words:

$$\sum v_k(1-v_k)n_k \neq \frac{(\sum v_k\, n_k\,)}{\sum n_k}\frac{(\sum(1-v_k)\, n_k)}{\sum n_k}\left(\sum n_k\right) = \left(\frac{N_{Cases\ Total}}{N_{Total}}\right)\left(\frac{N_{Controls\ Total}}{N_{Total}}\right)N_{Total},$$

or similarly put:

$$\sum EffN_k = \sum 4v_k(1-v_k)n_k \neq 4\frac{(\sum v_k\, n_k\,)}{\sum n_k}\frac{(\sum(1-v_k)\, n_k)}{\sum n_k}\left(\sum n_k\right)$$
$$= 4\left(\frac{N_{Cases\ Total}}{N_{Total}}\right)\left(\frac{N_{Controls\ Total}}{N_{Total}}\right)N_{Total}.$$

The equations on the left of the inequalities represent the correct terms, whereas the equations on right of the inequalities represent how the sample size and proportion of cases in meta-analysis are conventionally entered into LDSC when calculating h^2_l from GWAS meta-analysis of binary phenotypes. In other words, the inequality indicates that the conventional approaches are incorrect.

S5. Combined GWAS Meta-Analysis of Continuous and Binary Traits

Theoretical Justification

Researchers may seek to meta-analyze summary statistics from a GWAS (or GWAS meta-analysis) of a binary variable with summary statistics from a GWAS (or GWAS meta-analysis) of a continuous variable, under the assumption that the binary and continuous variables represent different approaches to measuring the same trait. For example, a researcher may be interested in meta-analyzing case-control

GWAS of major depressive disorder with a continuous measure of depression symptomology assuming that the diagnosis of major depressive disorder represents individuals who exceed a diagnostic threshold along the depression symptomology distribution. Here we formally describe a principled method for obtaining unbiased SNP-based heritability estimates.

To integrate continuous and binary GWAS at the level of the heritability model, we rely on the assumption that the binary and continuous variables represent different approaches to measuring the same trait, such that liability-scale heritability of the binary variable is on a comparable scale to the heritability of the continuous variable. For the binary summary statistics, the liability-scale heritability relates to the LDSC slope as:

$$LDSC\ Slope_{bin} = \frac{h^2_{liab}}{M}\frac{\phi^2}{P^2(1-P)^2}\left(\sum v_{kbin}(1-v_{kbin})n_{kbin}\right) = \frac{h^2_{liab}}{M}\frac{\phi^2}{4P^2(1-P)^2}\left(\sum EffN_{kbin}\right).$$

and for the continuous summary statistics, the heritability relates to the LDSC slope as:

$$LDSC\ Slope_{cont} = \frac{h^2_{cont}}{M}\left(\sum n_{kcont}\right) \quad ,$$

where the subscript *bin* refers to the terms corresponding to the binary variables for which a liability threshold model is used, and the subscript *cont* refers to the continuous variable.

When the summary data have been appropriately meta-analyzed across (non-overlapping) binary and continuous GWAS at the SNP level, the quantities by which the respective $\frac{h^2}{M}$ terms are multiplied in the LDSC Slope for continuous and binary variables can be summed, as in **S4**, such that a single interpretable heritability estimate, $h^2_{combined}$ (corresponding to both the continuous and liability-scales, which are expected to be equivalent), can be obtained from the meta-analytic summary statics, i.e.:

$$LDSC\ Slope_{bin+cont} = \frac{h^2_{combined}}{M}\left(\left(\sum n_{kcont}\right) + \frac{\phi^2}{4P^2(1-P)^2}\left(\sum EffN_{kbin}\right)\right) \quad .$$

The ϕ^2 term can be computed directly, by first calculating the threshold, *t*, corresponding to *P* (e.g. using the *qnorm* function in R) and then by calculating the density of the cumulative normal distribution corresponding to *t* (e.g. using the *dnorm* function in R). Thus, in practice, the entire quantity by which $\frac{h^2}{M}$ is multiplied can be computed and input as the *N* directly in LDSC, treating the summary statistics as having been derived from a continuous trait. This will produce a meta-analytic estimate of the heritability that is on the continuous scale and need not be further transformed.

Simulations of GWAS Meta-Analysis of Continuous and Binary Traits

In a separate set of simulations, we examined the proposed approach described directly above for estimating SNP-based heritability for meta-analyses of binary and continuous measures of the same trait. All simulation conditions began by simulating GWAS summary statistics for one binary trait and one continuous trait following the general formula outlined in *Eqs. 8* and *9* of the main text. These binary and

continuous outcomes were set to have a genetic correlation of 1, and the liability-scale heritability for the binary trait set to be equal to the continuous scale heritability at 15% (i.e., meta-analyzed heritability = 15%). In addition, there was no sample overlap across the cohorts (i.e., bivariate LDSC intercept = 0), and the LDSC univariate intercept was set to 1.04. These two GWAS summary statistics were subsequently meta-analyzed using a sample size weighted approach wherein the binary trat was weighted by the liability-scale corrected sample size $\frac{\phi^2}{4P^2(1-P)^2}(\sum EffN_k)$ and the continuous trait weighted by the continuous cohort sample size. Note that when sample size weighted meta-analysis was conducted using observed or effective sample sizes for the binary trait, the resulting Z statistics were extremely similar, as were the results of analyses pertaining to heritability bias. In other words, in these simulations, the choice of what form of *N* was used to compute the meta-analytic summary statistics had little bearing on the bias associated with using different forms *of N* in the LDSC equation for estimating heritability.

The combined sample size used as input for LDSC was calculated as: $(\sum n_{kcont}) + \frac{\phi^2}{4P^2(1-P)^2}(\sum EffN_{kbin})$, and the meta-analyzed trait was treated as continuous for SNP-based heritability estimation. We examined 24 population generating conditions that specified different combinations of sample prevalence (0.1, 0.3, or 0.5), population prevalence (.01, .05, .15), and relative contributions of case/control and continuous participants (50,000 case/control and 50,000 continuous *or* 150,000 case/control and 50,000 continuous). For each of the 24 conditions, we simulated 100 sets of GWAS summary statistics (i.e., 100 binary and 100 continuous summary statistics per condition, for a total of 4,800 simulated summary statistics).

We compared the SNP-based heritability results produced using our proposed approach for obtaining the meta-analytic sample size to the SNP-based heritability that would be obtained by inputting the same meta-analyzed summary statistics but providing the sample size as either: (i) the sum of total sample sizes or (ii) the sum of the continuous sample size and the effective sample size for the binary trait. Note that the latter approach is distinguishable from our proposed approach in that it does not apply the $\frac{\phi^2}{4P^2(1-P)^2}$ component of the liability threshold model. Simulation results show that for a range of population generating conditions that our proposed approach produces an unbiased estimate of SNP-based heritability for the meta-analytic summary statistics Supplementary Table 3). Inputting the total sample size produced average biases in estimates ranging from 60.81% to -56.44% depending on the condition. Observable trends for using total sample size included bias shifting upward for higher sample prevalences and greater downward bias as population prevalence increased. Inputting the sum of the effective sample size and the continuous sample size produced average biases in estimates ranging from 60.81% to -23.62%. Observable trends in this case again included greater downward bias as population prevalence

increased, whereas the trends for sample prevalence were contingent on population prevalence. More specifically, there was greater upward bias at increasing sample prevalence when population prevalence as 1% or 5%, and greater downward bias at increasing sample prevalence for population prevalences of 15% or 30%. In both cases, there was greater overall bias when the case/control cohort was larger than the continuous cohort.

Extended Data Simulations.

The distinction between the different approaches we examined for estimating SNP-based heritability for meta-analysis of continuous and binary traits is how the sample size is calculated. The expectation for the degree of bias in SNP-based heritability estimates relative to our proposed approach can be quantified using the ratio of our proposed approach for calculating sample size over the two alternatives we examine for calculating sample size. More specifically, the degree of bias when using the total sample size for the binary and continuous traits can be expressed as:

$$\frac{\sum n_{kcont} + \sum n_{k\mathrm{bin}}}{\left((\sum n_{kcont}) + \frac{\phi^2}{4P^2(1-P)^2}(\sum EffN_{kbin})\right)} - 1 \,,$$

and the bias when calculating sample size as the sum of the continuous and effective sample size expressed as :

$$\frac{\sum n_{kcont} + \sum EffN_{kbin}}{\left((\sum n_{kcont}) + \frac{\phi^2}{4P^2(1-P)^2}(\sum EffN_{kbin})\right)} - 1$$

With this in mind, we performed an additional series of simulations that did not simulate GWAS summary statistics, but rather directly simulated a range of sample and population prevalences and examined the corresponding degree of bias with respect to the different approaches for calculating sample size. This involved running 10,000 simulations that specified a sample size of 50,000 for the continuous trait and randomly generated a proportion of cases for the binary trait between 5% and 50%, a total sample size for the binary trait between 50,000 and 200,000, and a population prevalence between 1% and 30%,

The % bias ranged from -68.9% to 62.6% when using the sum of the binary and continuous sample sizes and from -25.1% to 64.4% when using the sum of effective sample size and continuous sample size. The trends mirrored what was observed for the simulations that generated GWAS summary statistics. More specifically, both approaches showed downward shifts in bias at increasing population prevalence (Supplementary Figure 2), increasing overall levels of bias as the ratio of binary to continuous sample size increased (Supplementary Figure 3), increasing bias at greater sample prevalence for sample size calculated as the sum of binary and continuous, and bias trends with respect to sample prevalence

sand sample size calculated as the sum of effective sample size and continuous sample size that were contingent on additional population generating parameters.

S6. Approximating $\sum N$ and $\sum EffN_k$

Researchers may encounter instances in which the summary data have been appropriately meta-analyzed at the SNP level, but the needed information to compute the appropriate entry of *N* is unavailable (e.g. cohort level $EffN_k$ isn't available for the binary data, or the data were meta-analyzed using a multivariate method such as Genomic SEM that corrects for sample overlap but precludes summing sample sizes). In these cases, the sample size can be estimated from the data using the meta-analytic betas, the *SE*s, and SNP *MAF*s, so long as the scale of the betas is known.

As described by Privé et al.[10] (also see Mallard et al.[11]), when the effects of individual variants are very small and coefficients are on a continuous scale (i.e. for GWAS meta-analysis of continuous traits or combined GWAS meta-analysis of continuous and binary traits using the approach described in **S5.**), $\sum N$ can be estimated from the data as:

$$\widehat{\sum N} \approx \frac{\sigma_y^2}{\sigma_x^2\, \sigma_{b_{cont}}^2},$$

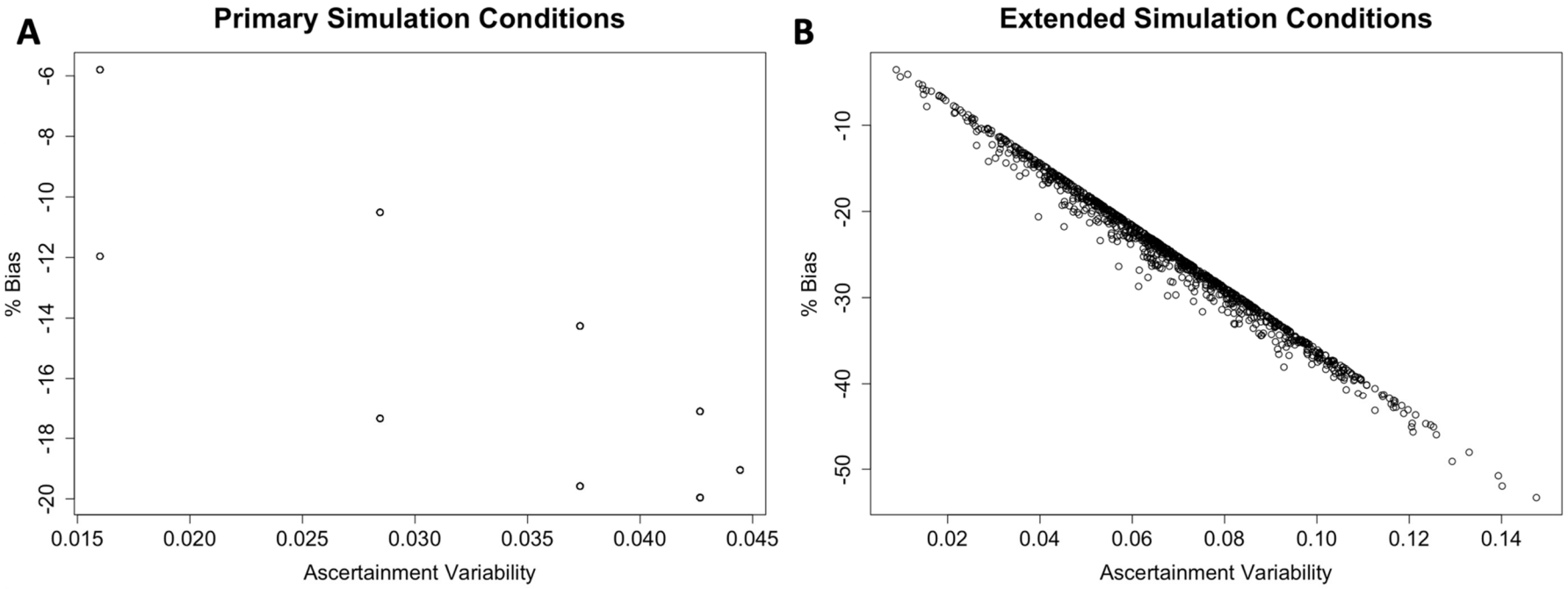
where σ_y^2 is the variance of the continuous outcome (typically 1.0), σ_x^2 is the variance of the genotype (estimated as *2pq*, assuming Hardy-Weinberg equilibrium, where *p* = minor allele frequency and *q* = 1-*p*), b_{cont} is the continuous scale GWAS coefficient, and $\sigma_{b_{cont}}^2$ is the sampling variance (i.e. the squared standard error) of the continuous scale GWAS coefficient.

When the effects of individual variants are very small, and the scale of the betas is on the unstandardized logistic scale, $\sum EffN$ can be estimated from the data as:

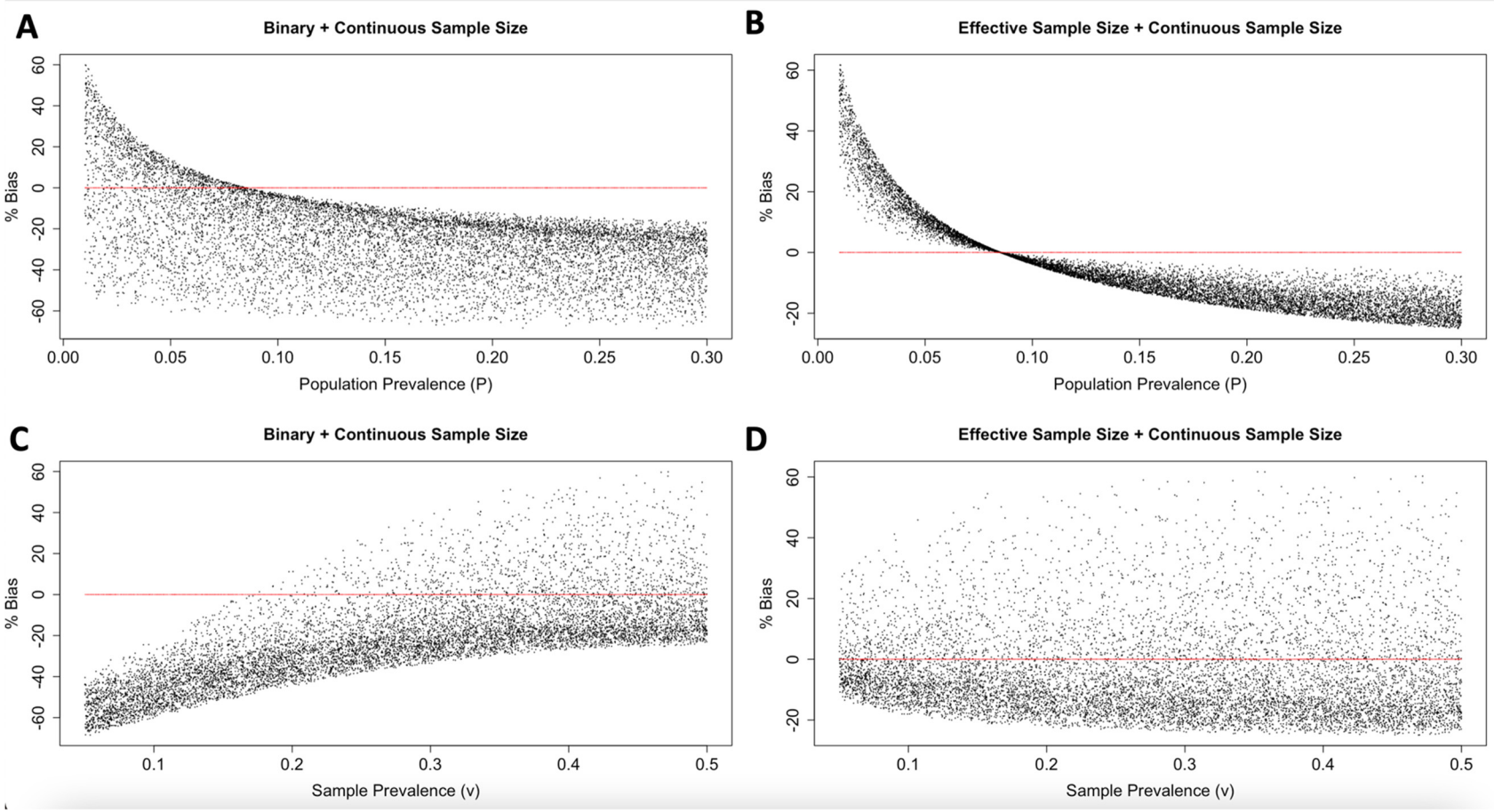
$$\widehat{\sum EffN} \approx \frac{4}{\sigma_x^2\, \sigma_{b_{logit}}^2}$$

where $\sigma_{b_{logit}}^2$ is the sampling variance (squared standard error) of the untransformed logistic regression coefficient. Following the recommendation from Privé et al.[10], these estimates can be capped at the lower end of 0.5**EffN* and 1.1**EffN,* where *EffN* denotes the total effective sample size calculated using the total number of cases and controls. Note that in practice, the SNP variance (σ_x^2) is computed using minor allele frequencies, which may not be present in the summary statistics file. In this instance, minor allele frequency from a reference panel can be used, though mismatch between the reference panel and participant sample will introduce error, and this mismatch is likely to be largest for lower frequency variants. We therefore recommend directly computing $\sum N$ and $\sum EffN_k$ when possible, particularly when in-sample MAF is unavailable. In our analysis of real GWAS summary data for 12 major psychiatric and neurological traits (Supplementary Table 2), we find that estimated $\sum EffN_k$ was

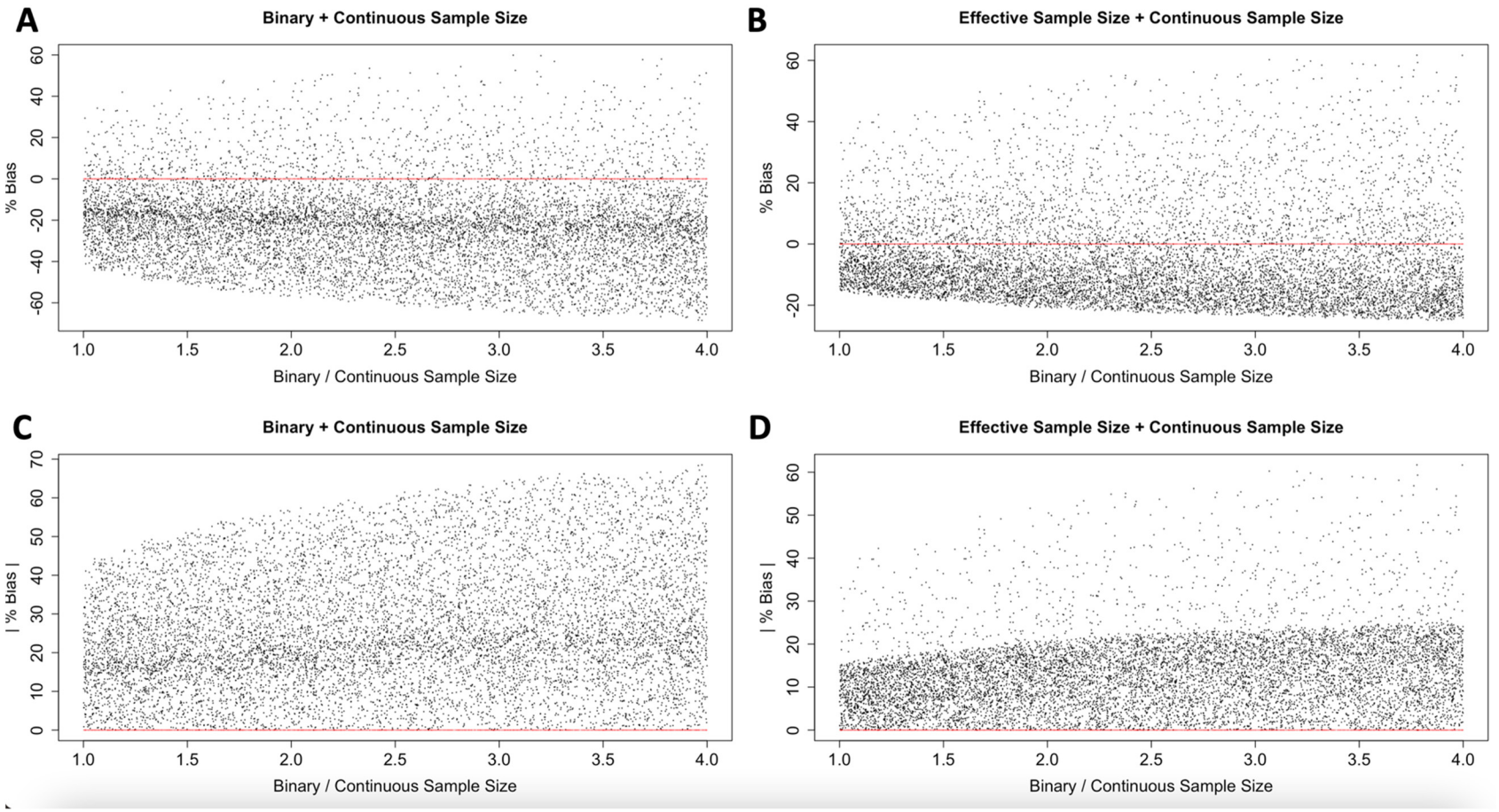
generally very similar to observed $\sum EffN_k$, with the exception of CUD and PTSD, for which estimated and observed values for $\sum EffN_k$ were highly discrepant. A variety of factors may contribute to such discrepancies, including but not limited to: (1) the noted minor allele frequency mismatch, (2) cohorts with high levels of uncontrolled for population stratification wherein a corresponding correction is applied to standard errors (e.g., multiplying standard errors by the univariate LDSC intercept) such that $\sum \widehat{EffN}$ decreases relative to observed $\sum EffN_k$ (3) inaccurate reporting of sample size from the original meta-analysis, (4) incorrect reporting of the effect size metric contained in the summary datafile, (5) sample overlap across cohorts contributing to the meta-analysis, (6) unreported pooled analysis of multiple raw data cohorts at the GWAS stage, and (7) if the GWAS is a stage 2 meta-analysis of prior stage 1 meta-analyses such that SNP-specific, sum of effective sample sizes calculated by meta-analytic software may not capture ascertainment variability in the cohorts comprising the stage 1 meta-analysis (i.e., if the total number of cases and controls across cohorts comprising the stage 1 meta-analysis are used for stage 2 meta-analysis). Such factors are important to attend to during quality control of the summary statistics and analytic pipeline.



Supplementary Figure S1. Ascertainment Variability Simulations. Both panels depict the % bias of the field standard approach on the *y*-axis, as a function of ascertainment variability across contributing cohorts on the x axis. Bias represents $\frac{h_l^2\ Estimate\ for\ v_{Total}}{h_l^2\ \ estimate\ for\ \sum EffN_k} - 1$, which is calculated as $\frac{\sum v_k(1-v_k)n_k}{\frac{(\sum v_k\, n_k)(\sum(1-v_k)\, n_k)}{\sum n_k}(\sum n_k)} - 1$. Ascertainment variability is calculated as the variance in sample prevalence across the 10 cohorts that contributed to each simulation datapoint. **Panel A** depicts simulation results when using the same generating conditions as were used for the direct simulations of GWAS summary statistics (Table 1). The overall levels of bias can be seen to mirror the bias observed for liability-scale heritability from the GWAS summary statistics simulations (Table 1). **Panel B** depicts simulation results for a broader range of simulating conditions, where 1,000 simulations were conducted that randomly varied the proportion of cases within each contributing cohort from 5% to 95%.



Supplementary Figure S2. Binary and Continuous Trait Extended Data Simulations: Sample and Population Prevalence. Panels depict the % bias for using different forms of calculating sample size for meta-analysis of binary and continuous traits. **Panels A and B** depicts simulation results as a function of population prevalence when calculating sample size as the sum of the binary and continuous sample size or the sum of the continuous and effective sample size, respectively. **Panels C and D** depicts simulation results as a function of sample prevalence when calculating sample size as the sum of the binary and continuous sample size or the sum of the continuous and effective sample size, respectively. 10,000 simulations were conducted that randomly varied the sample prevalence, population prevalence, and ratio of case/control to continuous sample sizes. The red horizontal line depicted at the y-axis of 0 corresponds to the results of our proposed solution, denoting no bias.



Supplementary Figure S3. Binary and Continuous Trait Extended Data Simulations: Binary / Continuous Sample Size. Panels depict the % bias for using different forms of calculating sample size for meta-analysis of binary and continuous traits as a function of the ratio of binary trait to continuous trait sample sizes. **Panels A and B** depicts simulation results as a function of the ratio of binary to continuous sample size when calculating sample size as the sum of the binary and continuous sample size or the sum of the continuous and effective sample size, respectively. **Panels C and D** depict the same results with the exception that the y-axis depicts the absolute value of the % bias. 10,000 simulations were conducted that randomly varied the sample prevalence, population prevalence, and ratio of case/control to continuous sample sizes. The red horizontal line is depicted at the y-axis of 0 corresponds to the results of our proposed solution, denoting no bias.

Supplemental References

1. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, 291 (2015).

2. Consortium, R. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature genetics* **47**, 1236–1241 (2015).

3. Yengo, L., Yang, J. & Visscher, P. M. Expectation of the intercept from bivariate LD score regression in the presence of population stratification. *bioRxiv* 310565 (2018).

4. Dempster, E. R. & Lerner, I. M. Heritability of Threshold Characters. *Genetics* **35**, 212–36 (1950).

5. Peyrot, W. J., Boomsma, D. I., Penninx, B. W. J. H. & Wray, N. R. Disease and Polygenic Architecture: Avoid Trio Design and Appropriately Account for Unscreened Control Subjects for Common Disease. *Am J Hum Genetics* **98**, 382–391 (2016).

6. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating Missing Heritability for Disease from Genome-wide Association Studies. *Am J Hum Genetics* **88**, 294–305 (2011).

7. Fuente, J. de la, Grotzinger, A. D., Marioni, R. E., Nivard, M. G. & Tucker-Drob, E. M. Multivariate Modeling of Direct and Proxy GWAS Indicates Substantial Common Variant Heritability of Alzheimer's Disease. *Medrxiv* 2021.05.06.21256747 (2021) doi:10.1101/2021.05.06.21256747.

8. Cook, J. P., Mahajan, A. & Morris, A. P. Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *Eur J Hum Genet* **25**, 240–245 (2017).

9. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature human behaviour* **3**, 513 (2019).

10. Privé, F., Arbel, J., Aschard, H. & Vilhjálmsson, B. J. Identifying and correcting for misspecifications in GWAS summary statistics and polygenic scores. *Biorxiv* 2021.03.29.437510 (2022) doi:10.1101/2021.03.29.437510.

11. Mallard, T. T. *et al.* Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *Biorxiv* 603134 (2020) doi:10.1101/603134.