

A fast method using polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait

Luigi Palla, Frank Dudbridge

Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.

Correspondence: frank.dudbridge@lshtm.ac.uk

Abstract

Several methods have been proposed to estimate the variance in disease liability explained by large sets of genetic markers. However, current methods do not scale up well to large sample sizes. Linear mixed models require solving high-dimensional matrix equations, while methods using polygenic scores are very computationally intensive. Here we propose a fast analytic method using polygenic scores, based on the formula for the non-centrality parameter of the association test of the score. We estimate model parameters from the results of multiple polygenic score tests based on markers with P -values in different intervals. We estimate parameters by maximum likelihood and use profile likelihood to compute confidence intervals. We compare various options for constructing polygenic scores, based on nested or disjoint intervals of P -values, weighted or unweighted effect sizes and different numbers of intervals, in estimating the variance explained by a set of markers, the proportion of markers with effects, and the genetic covariance between a pair of traits. Our method provides nearly unbiased estimates and confidence intervals with good coverage, although estimation of the variance is less reliable when jointly estimated with the covariance. We find that disjoint P -value intervals perform better than nested intervals, but the weighting did not affect our results. A particular advantage of our method is that it can be applied to summary statistics from single markers, and so can be quickly applied to large consortium datasets. Our methods are implemented in R software, AVENGEME (Additive Variance Explained and Number of Genetic Effects Method of Estimation).

Introduction

Genome-wide association studies have been successful in identifying many variants linked to complex diseases. To date over 6000 have been found in over 500 quantitative traits and common diseases in humans ¹. However when considering the variance explained by the markers associated with any specific disease, there remains a large gap to match the heritability estimates obtained from family studies ². This observation has spurred the development of theories and investigations to explain the missing heritability, including for example copy number variation ³, rare variants ⁴, epigenetics ⁵ and genetic interactions ⁶.

It has become increasingly clear that a large portion of the missing heritability is represented on current genotyping products, but the associated markers are not statistically significant. Several approaches have been developed to estimate the heritability explained by a set of genetic markers that may not be individually associated. In the linear mixed model approach, the genetic value of each individual is treated as a random effect whose sample covariance matrix is derived from the relatedness matrix, which is estimated from the genotype data ⁷. Solving this model gives an estimate of the additive genetic variance explained by the available genotypes, often called the *chip heritability*. Variations of this approach include multiple classes of variant with different effect size distributions ^{8;9}, regression of pair-wise phenotypic correlation on genetic correlation ¹⁰, and multivariate models to estimate genetic correlation between traits ¹¹.

Another approach uses polygenic scores to estimate chip heritability. Here, effect sizes for all markers are estimated in one sample of data, called the training sample. These effects are then used to construct a score for each subject in a second sample, called the target sample, as the weighted sum of genotypes across a set of markers. Originally, association of the score in the target sample was used to demonstrate the presence of missing heritability among an ensemble of markers ¹². More recently, the strength of this association has been used to infer the chip heritability ^{13; 14}.

A further approach uses empirical Bayes methods to estimate the chip heritability from the distribution of z-scores for individual markers¹⁵. This has the advantage of requiring only summary statistics from standard association analysis. Finally, a very recent method of “LD-scoring”¹⁶ estimates the chip heritability from the correlation between the marginal effect size of a marker and a measure of its linkage disequilibrium (LD) with other markers, also using only summary statistics.

In general the methods using linear mixed models are computationally expensive and require individual-level data to calculate the genetic relatedness matrix. Furthermore, many of these methods estimate only the chip heritability, but it is often of interest also to estimate the proportion of markers that affect a trait. This bears on the design of association studies, since it indicates the number and effect sizes of the associated markers remaining to be found. It is also relevant for the debate on the nature of evolution^{17; 18}, as if a large number of variants affect a trait, mechanisms of selection by polygenic adaptation are possible, acting on standing variation without requiring new mutations¹⁹. Methods for the estimation of the number of genes affecting a trait have been proposed since the early 20th century, including complex segregation analysis comparing single and multi-locus models with or without polygenic background²⁰, but only with the recent availability of dense genome-wide data has it become possible to assess the polygenic background itself.

Linear mixed models have been extended to allow for a proportion of variants with effects⁸ but this remains computationally demanding. Polygenic scoring has also been used to estimate this proportion, but again with a computationally demanding procedure using repeated genome-wide simulations within a Bayesian sampling scheme¹³. On the other hand an analytic method for polygenic scores¹⁴ estimates only one parameter among several defined in its model; thus it can estimate the proportion of variants with effects if the chip heritability is assumed to be known, or vice versa. Empirical Bayes methods are also available to estimate the proportion of markers with effects²¹, but have not been adapted to jointly estimate this proportion with the chip heritability.

Here we extend the analytic approach of Dudbridge¹⁴ to develop a fast analytic method based on polygenic scores for the joint estimation of chip heritability and the proportion of variants affecting the trait, and further estimate the genetic covariance between two related traits. A particular advantage is that our method can be applied when only summary data is available for individual markers, and this allows our approach to be readily applied to the increasingly large datasets that are now being made available by study consortia.

Methods

Parameter estimation: AVENGEME

We consider the model presented by Dudbridge¹⁴ in which a pair of standardised traits $\mathbf{Y}=(Y_1, Y_2)'$ is expressed as a linear combination of m genetic effects and an error term $\mathbf{E}=(E_1, E_2)'$

$$\mathbf{Y} = \boldsymbol{\beta}'\mathbf{G} + \mathbf{E} = (\sum_{i=1}^m \beta_{i1}G_i + E_1, \sum_{i=1}^m \beta_{i2}G_i + E_2)' \quad (\text{equation 1})$$

where \mathbf{G} is an m -vector of coded genetic markers and $\boldsymbol{\beta}$ an $m \times 2$ matrix of coefficients, with \mathbf{E} independent of \mathbf{G} . Assuming that in two independent samples the estimates of the genetic effects are given respectively by $\hat{\beta}_{i1}$ and $\hat{\beta}_{i2}$, $i=1, \dots, m$, either set of estimates can then be used to create polygenic scores $\hat{S}_1 = \sum_{i=1}^m \hat{\beta}_{i2}G_i$ and $\hat{S}_2 = \sum_{i=1}^m \hat{\beta}_{i1}G_i$ to be tested for association respectively with Y_1 and Y_2 . Focusing without loss of generality on \hat{S}_2 , the statistical properties of the test of association have been described¹⁴. In particular the coefficient of determination between \hat{S}_2 and Y_2 , i.e. the variance explained by the polygenic score in the regression of Y_2 on \hat{S}_2 , is given by

$$R^2_{\hat{S}_2, Y_2} = \frac{m \text{cov}(\hat{\beta}_{i1}, \beta_{i2})^2}{\text{var}(\hat{\beta}_{i1}) \text{var}(Y_2)}$$

where the terms on the right-hand side are expressed analytically in terms of the following parameters:

Study design: sample sizes of the two samples, (n_1, n_2) ; number of variants in the marker panel, m , assumed to be uncorrelated; P -value thresholds for selecting a marker into the score from the training sample, (p_L, p_U) ; for binary traits, population prevalences (K_1, K_2) and case sampling fractions (P_1, P_2) .

Genetic model: additive genetic variance in the training sample, σ_1^2 ; genetic covariance between training and testing samples, σ_{12} ; proportion of null markers with no effect on the trait in the training sample, π_{01} . The variance and covariance are marginal over all markers, so include the null markers with $\beta_{i1} = 0$ or $\beta_{i2} = 0$.

The asymptotic non-centrality parameter of the $\chi^2_{(1)}$ test of association between Y_2 and \hat{S}_2 is given

$$\text{by } \lambda = \frac{n_2 R_{\hat{S}_2, Y_2}^2}{1 - R_{\hat{S}_2, Y_2}^2}; \text{ equivalently the expectation of the z- (or t-) test is } \mu = \sqrt{\frac{n_2 R_{\hat{S}_2, Y_2}^2}{1 - R_{\hat{S}_2, Y_2}^2}} \text{ with the sign}$$

taken from the correlation between Y_2 and \hat{S}_2 .

Binary traits are assumed to arise from a liability threshold model, in which each subject has an unobserved trait, called the liability, which is normally distributed in the population. Subjects with liability greater than a fixed threshold have the trait. The same theory then holds when either Y_1 or Y_2 is binary as for when it is quantitative, assuming linear transformations between effects on the liability scale to effects on the observed (0/1) scale, and accounting for ascertainment in case/control studies. Specifically, each effect β_{ij} on the liability scale corresponds to an effect

$$\beta_{ij} \phi(\tau_j) \frac{P_j(1-P_j)}{K_j(1-K_j)} \text{ on the observed binary scale}^{14}, \text{ where } \tau_j = \Phi^{-1}(1-K_j) \text{ with } \phi \text{ and } \Phi \text{ the}$$

standard normal density and cumulative distribution functions.

We aim to estimate the genetic model parameters σ_1^2 , σ_{12} , π_{01} from the association test between \hat{S}_2 and Y_2 . Previously it was shown¹⁴ that one parameter could be estimated by solving for the value at which λ equals the observed χ^2 statistic. To estimate multiple parameters, we now propose using association tests of Y_2 with multiple polygenic scores constructed by selecting markers with different P -value thresholds in the training data. We then fit parameters to the observed association tests using maximum likelihood.

Specifically, let d_1, \dots, d_k denote a set of k intervals within the unit interval, where k is equal to or greater than the number of parameters to be estimated. For each $i = 1, \dots, k$ we select markers with P -values falling in d_i , construct the corresponding polygenic score and obtain its (signed) z -score Z_i for association with Y_2 . The log-likelihood for σ_1^2 , σ_{12} , π_{01} is then

$$\ell(\sigma_1^2, \sigma_{12}, \pi_{01}) = \sum_{i=1}^k \log \phi(Z_i - \mu(\sigma_1^2, \sigma_{12}, \pi_{01}; d_i))$$

where $\mu(\sigma_1^2, \sigma_{12}, \pi_{01}; d_i)$ is the expectation of the z -test as described above, expressed explicitly as a function of the model parameters given selection interval d_i . Maximisation of this log-likelihood yields estimates of the model parameters. Note that any of σ_1^2 , σ_{12} , π_{01} could be held fixed while the other parameter(s) are estimated.

An equivalent procedure estimates (using obvious notation) σ_2^2 , σ_{12} , π_{02} by reversing the roles of the training and target samples. Furthermore, a bidirectional procedure can be used to simultaneously estimate up to five parameters σ_1^2 , σ_2^2 , σ_{12} , π_{01} , π_{02} by fitting to the z -scores for association of both \hat{S}_2 with Y_2 and \hat{S}_1 with Y_1 .

The number of estimated parameters can be reduced by assuming that the genetic architectures are identical in the training and testing samples. This would occur if two samples are drawn from the

same population with the same trait definitions, or if one sample is randomly split into training and target subsets. Then we may assume $\sigma_1^2 = \sigma_2^2 = \sigma_{12}$ and $\pi_{01} = \pi_{02}$, estimating just two parameters in either unidirectional or bidirectional analysis.

Ours is not a proper likelihood because the z-scores Z_i corresponding to the marker selection intervals are not independent. The presence of a marker in one interval determines its presence or absence in all other intervals, creating dependence between the corresponding scores, but this is not reflected in our likelihood. Furthermore the bidirectional likelihood does not account for dependence between the scores calculated in each direction. We are therefore using a quasi-likelihood and will later use simulations to investigate its sensitivity to the assumption of independent likelihood contributions.

Maximisation of the log-likelihood is complicated by constraints on the range of σ_{12} . Because the absolute correlation between β_{i1} and β_{i2} must be no greater than 1, $|\sigma_{12}| \leq \sigma_1 \sigma_2$. In the unidirectional estimation, σ_2^2 is not identified and we need only respect that $\sigma_2^2 \leq 1$, giving the constraint $|\sigma_{12}| \leq \sigma_1$. In the bidirectional estimation we must also consider that the absolute correlation is no greater than 1 for the markers that have non-null effects in both training and testing samples. Denoting this correlation as ρ^* , the correlation over all markers as ρ and the proportion of markers with non-null effects in both samples as $\gamma \leq 1 - \max(\pi_{01}, \pi_{02})$, we have

$$\rho^* = \frac{\sigma_{12} \gamma^{-1}}{\sqrt{\sigma_1^2 (1 - \pi_{01})^{-1} \sigma_2^2 (1 - \pi_{02})^{-1}}} = \frac{\rho \sqrt{(1 - \pi_{01})(1 - \pi_{02})}}{\gamma}$$

$$\sigma_{12} = \rho \sigma_1 \sigma_2 = \frac{\rho^* \gamma \sigma_1 \sigma_2}{\sqrt{(1 - \pi_{01})(1 - \pi_{02})}}$$

$$|\sigma_{12}| \leq \frac{(1 - \max(\pi_{01}, \pi_{02})) \sigma_1 \sigma_2}{\sqrt{(1 - \pi_{01})(1 - \pi_{02})}}$$

We maximise the likelihood numerically by nesting the maximisation for σ_{12} within that for the other parameters: for each proposed value of $\sigma_1^2, \sigma_2^2, \pi_{01}, \pi_{02}$ we perform a univariate maximisation for σ_{12} subject to the constraint imposed by the proposed values.

To obtain analytic confidence intervals we use profile likelihood²². For a general scalar parameter θ , its profile log-likelihood function is $\ell_p(\theta) = \ell(\theta, \hat{\theta}(\theta))$ where $\hat{\theta}(\theta)$ is the maximum likelihood estimate of the remaining parameters in the model given θ . Since for a regular model

$2(\ell(\hat{\theta}, \hat{\theta}(\hat{\theta})) - \ell(\theta, \hat{\theta}(\theta))) \xrightarrow{D} \chi_1^2$, for the estimated value $\hat{\theta}$ we obtain a $(1-\alpha)$ confidence interval as the set $\{\theta : \ell_p(\theta) \geq \ell_p(\hat{\theta}) - \frac{1}{2} \chi_1^2(1-\alpha)\}$ where $\chi_1^2(1-\alpha)$ is the $1-\alpha$ quantile point of the χ_1^2 distribution. This procedure is used to obtain confidence intervals for each of $\sigma_1^2, \sigma_2^2, \sigma_{12}, \pi_{01}, \pi_{02}$.

Often it is the genetic correlation rather than the covariance between two traits that is of interest. Because the unidirectional estimation does not identify σ_2^2 , the correlation cannot be estimated unless a value is assumed for σ_2^2 . In the bidirectional estimation, the correlation and its confidence interval can be obtained using previously derived formulas²³.

Association tests of polygenic scores can be calculated from summary data alone²⁴. The regression of Y_2 on \hat{S}_2 has coefficient

$$\frac{\text{cov}(Y_2, \hat{S}_2)}{\text{var}(\hat{S}_2)} = \frac{\sum \text{cov}(Y_2, \hat{\beta}_{1j} G_j)}{\sum \text{var}(\hat{\beta}_{1j} G_j)} = \frac{\sum \hat{\beta}_{1j} \hat{\beta}_{2j} \text{var}(G_j)}{\sum \hat{\beta}_{1j}^2 \text{var}(G_j)} \approx \frac{\sum \hat{\beta}_{1j} \hat{\beta}_{2j} s_{2j}^{-2}}{\sum \hat{\beta}_{1j}^2 s_{2j}^{-2}}$$

where s_{2j}^2 is the sampling variance of $\hat{\beta}_{2j}$, assuming markers are uncorrelated. This is the inverse-variance weighted mean of $\hat{\beta}_{2j} / \hat{\beta}_{1j}$ and hence has sampling variance $\frac{1}{\sum \hat{\beta}_{1j}^2 s_{2j}^{-2}}$. The Wald

$$\text{statistic} \frac{\sum \hat{\beta}_{1j} \hat{\beta}_{2j} s_{2j}^{-2}}{\sqrt{\sum \hat{\beta}_{1j}^2 s_{2j}^{-2}}} \text{ (equation 2)}$$

is then calculated from summary effect sizes and standard errors for the individual markers. These data are frequently available from research consortia even when access to individual-level data is impractical^{25; 26}.

Our methods are implemented in R software, AVENGEME (Additive Variance Explained and Number of Genetic Effects Method of Estimation), which is available from the authors (see Web Resources).

Method evaluation

To study the statistical and operating characteristics of AVENGEME, we simulated genome-wide marker data under various genetic models. We based our simulations on four complex diseases studied by Stahl *et al*¹³, allowing direct comparisons with their ABPA method, which is conceptually similar to ours. We also performed simulations based on three successively larger studies of schizophrenia²⁷. The study design parameters and the genetic models used for our simulations are given in Table 1.

For each genetic model we simulated estimated effect sizes $\hat{\beta}_{1j}$, $\hat{\beta}_{2j}$ independently for each marker, by drawing the true effects from the bivariate normal distribution in equation 1 and adding independent sampling error to each effect. We then selected markers according to their P -values in the training sample and used the summary statistic formula in equation 2 to obtain tests of association for each polygenic score. We verified this approach for sample sizes up to 10K by explicitly simulating genotypes in cases and controls as previously described¹⁴. Briefly, independent biallelic markers were defined with population minor allele frequencies uniformly distributed on

(0.01,0.5). Their effect sizes were drawn from the bivariate normal distribution such that the desired variances and covariances were attained. Allele frequencies in cases and controls were then derived for cases and controls and genotypes simulated in each. Allelic odds ratios were then computed from the genotype counts. The results from the genotype simulations were indistinguishable from those from summary statistics, so we adopted the summary statistic method, which is much faster and easily scales up to very large sample sizes. Note that in our simulations, markers were assumed to be independent, *ie* in linkage equilibrium, as assumed by AVENGEME. We will later consider the effect of LD on our method.

For the models in table 1, we simulated 1000 sets of polygenic score results and estimated the genetic model parameters using the unidirectional AVENGEME. This was done both when assuming $\sigma_1^2 = \sigma_{12}$ (which reflects the assumption that the two samples have the same genetic model), in which case AVENGEME estimates the two free parameters σ_1^2 and π_{01} , and when allowing $\sigma_1^2 \neq \sigma_{12}$ in which case AVENGEME estimates three free parameters. We evaluated the accuracy from the mean and standard deviation of the parameter estimates and the coverage of the 95% confidence intervals.

We then considered different options for constructing polygenic scores, simulating under the design of the largest schizophrenia study (rightmost column of table 1) (hereafter termed *SCZ simulation*). We fixed ten thresholds (Table S1, right half) and compared the use of disjoint to nested *P*-value intervals using those thresholds, with the nested intervals each having a lower limit of 0. We compared weighted scores to unweighted scores in which all markers were given an equal weight in the direction of disease risk. We performed 1000 simulations and evaluated bias, precision and coverage as before.

We considered the effect of increasing the number of selection intervals and the sample size. Here we simulated different heritabilities in the two samples: $\sigma_1^2 = 0.3$, $\sigma_2^2 = 0.45$, $\sigma_{12} = 0.294$ (giving

genetic correlation of 0.8), and different proportions of null markers: $\pi_{01} = 0.95$, $\pi_{02} = 0.94$ (hereafter termed *bivariate simulation*). We compared the use of 3, 5, 10, 20 and 40 selection intervals in sample sizes of 10K, 20K, 40K and 80K subjects with case sampling fractions $P_1 = 0.425$, $P_2 = 0.515$ and disease prevalences $K_1 = K_2 = 0.01$. This reflected the SCZ PGC2 study design, although as that was a meta-analysis of case/control studies the overall sampling fraction should be adjusted to reflect the different fractions in each study. We did not do this here but have found that such adjustments have very little effect on the estimated model.

We evaluated the bidirectional AVENGEME for the simultaneous estimation of all five parameters. We then returned to the SCZ simulation and applied bidirectional AVENGEME under the constraints $\sigma_1^2 = \sigma_2^2 = \sigma_{12}$, $\pi_{01} = \pi_{02}$, to compare the precision of the bidirectional and unidirectional AVENGEME when estimating only two free parameters.

Finally we compared AVENGEME to the genomic restricted maximum likelihood (GREML) solution of the linear mixed model, as implemented in the popular GCTA program²⁸. We performed the bivariate simulation with a total sample size of 10K. GREML was applied on the entire sample, whereas for AVENGEME it was split into training and testing samples each of 5K subjects. We also compared AVENGEME to the method of So *et al*¹⁵, which also uses summary statistics for estimation of σ_1^2 only, under the SCZ simulation for a total sample size of 10K.

Linkage disequilibrium

The theory underlying AVENGEME assumes that markers are uncorrelated¹⁴. This is approximately ensured in practice by pre-filtering markers using “LD-pruning” algorithms that select markers with limited pairwise correlation. While this practice is common for many methods that estimate chip heritability, it may lead to under-estimation of the true chip heritability since the selected markers may not fully tag the causal variation. Conversely, in our approach the residual LD among the

pruned markers may lead to over-estimation of the explained variance and under-estimation of the proportion of null markers, since marker effects will be biased by LD with other markers²⁹.

We therefore performed simulations on real genotype data to assess the effect of LD-pruning. We combined genotype data from all seven case and both control samples in phase 1 of the Wellcome Trust Case-Control Consortium(WTCCC)³⁰, giving genotypes for 384,845 markers on 15769 subjects after basic quality control (Table S2). We allocated a chip heritability of $\sigma_1^2 = \sigma_2^2 = \sigma_{12} = 0.3$ among a random 5% of the markers ($\pi_{01} = \pi_{02} = 0.95$). We simulated a normally distributed quantitative trait under this model, split the sample into equally sized training and target samples, and estimated the model using AVENGEME on a reduced marker set. We considered both a “pruning” algorithm, which does not take association results into account (“indep-pairwise” option in PLINK³¹, window size 100, step 10) and a “clumping” algorithm which greedily retains the most associated markers in the reduced set (“clump” option in PLINK with index and clumped P-value thresholds of 1 and 100 marker radius). Both algorithms were applied with r^2 thresholds of 0.1 and 0.2, giving reduced sets of approximately 77,000 and 102,000 markers respectively on average. The simulation was repeated 1000 times.

Results

Bias and precision

We simulated data based on the estimates for additive genetic variance and proportion of null markers obtained by Stahl *et al*¹³ for four common diseases (Table 1). We compared the performance of AVENGEME for these four models using the same P -value intervals as those authors (Table S1). Results are shown in Table 2. For the estimation of two parameters only, assuming the same genetic model in the training and target samples, our method yielded nearly unbiased results for both σ_1^2 and π_{01} with small variance, suggesting that it is expected to work very well in practice. However, the coverage was lower than 95%, suggesting that the analytic confidence intervals are

too narrow. This may result from our assumption that the selection intervals make independent contributions to the likelihood. To confirm this, we directly simulated χ^2 statistics from the analytic non-central distributions, independently for each selection interval, and repeated the estimation. The confidence intervals then indeed had appropriate coverage (Supplementary Table 3), confirming that the assumption of independent contributions from each selection interval leads to confidence intervals that are too narrow. Nevertheless this effect appears to be fairly small.

In the estimation of 3 parameters, the estimate of σ_1^2 had some upward bias and much larger variance; π_{01} had greater variance compared to the 2 parameter estimation, but coverage close to 95%. Inspection of individual simulations revealed that the estimated σ_1^2 is often close to 0 or to 1, pulling the mean estimate towards 0.5. Generally this suggests that the variability is too large to allow reliable estimation of σ_1^2 when estimating π_{01} and σ_{12} as well, at least at these sample sizes. The estimates for σ_{12} however showed nearly unbiased estimates and small variance, suggesting that our method is reliable for estimating the genetic covariance when it is not assumed to equal the variance. Coverage was slightly less accurate in the estimation of 3 parameters, but generally close to the nominal level.

We conclude that for the estimation of σ_1^2 and π_{01} it is preferable for the training and target samples to be from the same trait population and to apply AVENGEME under the constraint $\sigma_1^2 = \sigma_{12}$, whereas if the interest lies in the estimation of the genetic covariance between traits then the unconstrained version of AVENGEME is more appropriate.

Nested intervals and unweighted scores

We wondered whether the sample sizes could be a reason for the poorer performance of the 3 parameter estimation; in addition we considered the effect of the score weighting versus an unweighted score, and whether the P -value selection intervals were disjoint or nested . We therefore simulated under a scenario with parameters derived from a large meta-analysis of schizophrenia ²⁷ (Methods; Table 1 rightmost column). The results are shown in Table 3. For the 2 parameter estimation, disjoint intervals had the least bias and most accurate coverage, although its variance was slightly greater than for nested intervals. The reduced coverage of the confidence intervals for nested intervals can be ascribed to the dependence between intervals, which is greater for nested intervals. The bias is possibly due to the imbalance in the sample size between training and test set (reversing the direction of estimation led to a reduction in bias, for example for disjoint intervals, weighted score, mean $\hat{\sigma}_1^2 = 0.291$). Similar patterns were observed when estimating three parameters, with the disjoint intervals generally showing less bias and more accurate coverage than the nested intervals, but with slightly increased variance. The choice of weights seems to be generally neutral although a slight increase in variance was observed for unweighted scores. Taken together these results suggest that the weighted score with disjoint selection intervals is the most reliable and accurate approach for use with AVENGEME.

Sample size and number of selection intervals

We then performed bivariate simulations (see Methods) to consider the effect of varying the sample size and the number of selection intervals. In Tables S4-S7 we show the performance of AVENGEME in each direction. The results confirm the poor ability to estimate σ_1^2 or σ_2^2 , with mean values mostly around 0.5 and high variance reflecting the frequent estimates of 0 or 1. This applies across all numbers of selection intervals, but there is a reduction in variance as the number of intervals increases, and a substantial reduction in bias and variance as the sample size increases from 10K to

80K, while more bias persists for the lower genetic variance (mean $\hat{\sigma}_1^2=0.362$ and $\hat{\sigma}_2^2=0.444$ with 40 intervals and 80K total sample size). A similar pattern was observed for π_{01} and π_{02} , although there was much less bias in general.

For the covariance σ_{12} the estimation again worked well, being nearly unbiased and with low variance regardless of sample size and number of selection intervals. We again observed a general trend of improved bias and precision with more selection intervals and greater sample size.

Bidirectional estimation

We applied the bidirectional method to the same bivariate simulation data for total sample size of 80k. The results (Table S8) showed consistently lower variance for each parameter compared to the unidirectional estimators, but with a similar level of bias resulting in lower coverage of the confidence intervals. The information gain from analysing the bidirectional data together is offset to some degree by the increased number of parameters in the model. Furthermore, this analysis was considerably more time-consuming than the unidirectional analyses.

Similarly, when applying the bidirectional estimation to data simulated under the SCZ model (Table 1 rightmost column), and constraining $\sigma_1^2 = \sigma_2^2 = \sigma_{12}$ and $\pi_{01} = \pi_{02}$ in the estimation, we obtained lower bias for σ_1^2 (mean $\hat{\sigma}_1^2=0.286$, $\hat{\pi}_{01}=0.95$), similar variance ($SD(\hat{\sigma}_1^2)=0.011$, $SD(\hat{\pi}_{01})=0.004$), and greater coverage for σ_1^2 and lower for π_{01} ($=0.498$ for σ_1^2 , $=0.760$ for π_{01}) compared to the unidirectional analyses (first column of Table 3), although the differences were very small.

We performed a sensitivity analysis to compare the performance of the bidirectional estimation with different initial parameter values for the numerical optimization and the results were virtually unchanged, with just a slight change in bias, variance and coverage. A similar analysis conducted for

the complex diseases in Table 1 also revealed that the estimate of covariance was robust to the choice of initial parameter values.

Linkage disequilibrium

We simulated a normally distributed trait on 15769 subjects in the WTCCC (see Methods). Using reduced marker sets with pairwise r^2 constrained to <0.1 and <0.2 , we estimated $\sigma_1^2 = \sigma_2^2 = \sigma_{12}$ and $\pi_{01} = \pi_{02}$ when a) the markers were pruned without regard to their association; b) the markers were clumped by greedily retaining the most strongly associated markers. Table 4 shows that for $r^2 < 0.1$, AVENGEME is unbiased in estimating σ_1^2 when clumping is used, but has a small downward bias in π_{01} . Pruning, however, incurs a strong downward bias in both σ_1^2 and π_{01} . For $r^2 < 0.2$, clumping over-estimates σ_1^2 and under-estimates π_{01} owing to the residual LD. Pruning reduces, but does not eliminate, these biases. These results suggest in practice using a clumping algorithm with pairwise $r^2 < 0.1$ as the least biased approach with AVENGEME.

Comparison with related methods

We analysed our bivariate simulations for total sample size 10K using the bivariate GREML implemented in GCTA²⁸. The mean $\hat{\sigma}_{12}$ was 0.265 with standard deviation 0.032, which compared to the results in Table S4 shows that in this case the GREML estimate has greater bias but less variance than AVENGEME.

We also applied the method by So *et al*¹⁵ to the SCZ simulation (Table 1 rightmost column).

Although their method appeared unbiased in the simulation they performed in which

$\pi_{01} = \pi_{02} = 0.995$, in our setting it yielded seriously biased results for σ_1^2 with a mean estimate of 0.189 compared to the true value of 0.3.

Having established the good operating characteristics of AVENGEME, we applied our method to some published association results for polygenic scores. For the four diseases from Stahl *et al*, our estimates were systematically lower than the ones obtained by their ABPA method (Table 5), and for σ_1^2 our confidence intervals excluded their estimates. These results were surprising since the two methods are conceptually similar, and our simulations had shown that under the models inferred by ABPA, AVENGEME achieved nearly unbiased estimation. LD is unlikely to affect these results as the markers were clumped to $r^2 < 0.1$. We speculate that the differences may arise from ABPA's use of prior distributions, and return to this point in the Discussion. Compared to results from GREML, our estimates for σ_1^2 were lower, with non-overlapping confidence intervals, for Rheumatoid Arthritis and Type-2 Diabetes, whereas the results were similar for Celiac Disease and Myocardial Infarction.

We applied AVENGEME to three waves of SCZ meta-analyses (Table 5). The genetic variance σ_1^2 was similar in the ISC and PGC1 data, but decreased in the PGC2 data. The proportion of null markers decreased in PGC1 and PGC2 compared to ISC. This may reflect increased heterogeneity: as more studies contribute to the meta-analyses, increased genetic heterogeneity could decrease the proportion of null markers, whereas increased environmental heterogeneity could decrease the genetic variance, which on the liability scale is expressed relative to the total variance. GREML has been applied to the ISC and PGC1 data ³²; for the former the estimate is similar to ours, whereas it is significantly lower in the latter. ABPA has been applied to an expanded PGC1 analysis ³³, yielding a significantly higher estimate of σ_1^2 than ours.

We finally applied AVENGEME to estimate genetic covariance between psychiatric traits using published summary data ³⁴. These data included 5 pairs from four disorders: schizophrenia, bipolar

disorder, major depressive disorder and autistic spectrum disorder (other combinations, for which only two selection intervals were reported, were excluded as our method requires at least three). The method of Dudbridge ¹⁴ has previously been shown to agree well with GREML for these data ³⁵, but in estimating the genetic covariance it assumes that σ_1^2 and π_{01} are known exactly. Here we estimated all three parameters simultaneously. The results are presented in Table 6 and show that the estimates from AVENGEME are of similar magnitude to those from GREML, but are consistently larger and have narrower confidence intervals. This difference may arise from LD, as here the markers were clumped to $r^2 < 0.25$, which according to table 5 may create an upward bias in AVENGEME.

Discussion

The method we have proposed allows simultaneous estimation of the additive variance explained by a set of genetic markers, the proportion of markers affecting the trait of interest and the genetic covariance between two traits. It does so by solving analytic expressions to obtain maximum likelihood estimates and profile likelihood confidence intervals, and is consequently very fast. Furthermore the polygenic score tests required by our method can be rapidly calculated from summary statistics for individual markers, allowing application to very large data sets and results from published literature. Our simulations show that our method enjoys good bias and coverage properties in spite of its assumption that the tests from different selection intervals are independent. Although we presented results only for case/control designs here, they represent the most challenging scenarios for polygenic modelling and we have observed results of comparable or greater accuracy for quantitative traits (data not shown).

AVENGEME has a number of advantages compared to currently available methods. In comparison with GREML it can deal with very large sample sizes and obtain estimates much more rapidly, and it

additionally estimates the proportion of null markers. Compared to ABPA, it does not require Monte Carlo sampling nor simulation of genome-wide data and is therefore much faster; AVENGEME also extends to estimate the covariance between related traits. Compared to the method of So *et al* and other empirical Bayes methods, it appears to be less biased and can simultaneously estimate up to five model parameters. Compared to the LD-scoring approach, it can estimate the proportion of null markers and does not require calculation of LD between pairs of markers.

The limitations of our approach include the need for two independent datasets, which is often not the case when common controls are used, whereas GREML can estimate a bivariate model from a single sample and LD-scoring is robust to overlapping samples. We assume that population structure has been entirely adjusted for in the target sample, and may over-estimate chip heritability if this is not the case, whereas GREML and LD-scoring adjust for structure explicitly in their calculations. Our method also assumes that markers are uncorrelated. In practice this is approximately ensured by a “LD-pruning” step that is also commonly conducted for other methods. We have shown that if the residual LD between pruned markers is not too high, say $r^2 < 0.1$, then AVENGEME retains its unbiased properties if a “clumping” algorithm is used, but can otherwise overestimate the genetic variance. In contrast, LD-scoring explicitly uses LD to estimate the variance explained. The similarity of estimates obtained by that approach to those of ours and other current methods suggests that this problem is currently not too severe, but as marker densities increase towards whole genome coverage, it will become more important to include all markers and account for LD. Our methods can be extended to allow correlation between markers, and this will be pursued in a subsequent paper.

A limitation is that unless very large sample sizes are used, estimation of the chip heritability in the training sample is unstable if it is jointly estimated with the covariance with the testing sample. Therefore if the variance is of particular interest, we recommend analysing the same trait in both samples, either by splitting a single sample in two, or by drawing two samples from the same trait

population. Then good performance in estimating the variance can be achieved by constraining it to equal the covariance.

The unidirectional estimation provides good estimates in all situations we considered. The bidirectional estimation can also be applied, providing a less variable estimate than the unidirectional estimators, with a similar degree of bias. However, the bidirectional analysis is more time-consuming than the unidirectional, and as its reduction in variance is rather small we do not find a compelling reason to prefer it to the unidirectional.

We recommend using disjoint selection intervals, whereas the influence of the weighting seemed limited in the situations we considered. However the use of nested intervals still provides good estimates if the number of intervals is sufficiently large (say 10) and appears to work well for the covariance across sample sizes, number and type of intervals. Nested intervals seem more appealing for obtaining significant tests of association between polygenic scores and a trait of interest, and to date have been reported more often than disjoint intervals. However for the estimation of the underlying genetic model we suggest that results for disjoint intervals should also be made available. The current fashion for using around 10 intervals appears to be sufficient for obtaining accurate estimates; while precision increases as more intervals are used, the gains diminish rapidly beyond that number.

Our method was generally found to produce under-coverage of confidence intervals. This is due both to some bias in the estimation, though this was generally small, and the assumption of independent tests from each interval. We have observed that our profile likelihood intervals closely match the empirical distribution of parameter estimates in our simulations. The under-coverage is therefore more likely to arise from the slight bias in our estimator rather than from the calculation of its variance. Our experience is that, in this application, an approximately valid confidence interval is generally sufficient for practitioners.

AVENGEME requires numerical optimisation to estimate parameters, and this can be sensitive to the algorithm used and the initial estimates provided. We have used the default settings of the `optim()` function in R (Nelder-Mead non-linear optimization), and in the simulations provided the true parameter values as the initial estimates. This was to obtain, as far as possible, the ideal results from truly maximising the likelihood. We found that slight variations can result from different starting values (our default values are 0.5 for all parameters) but the conclusions remain the same. In practice we suggest using a range of plausible starting values to identify the solution with the maximum likelihood.

AVENGEME is conceptually similar to ABPA¹³, both methods seeking the genetic model that best fits the observed results of polygenic score tests using multiple selection intervals. The main difference is that AVENGEME uses analytic formulae to construct an explicit likelihood, whereas ABPA uses approximate Bayesian computation with Monte Carlo sampling. In the application to the complex diseases in Table 1, we obtained lower estimates for all parameters and the reason for this may be the effect of the prior distributions used by ABPA. Their prior for π_{01} is uniform on the log scale and therefore heavily favours values of π_{01} close to 1. On the other hand, their prior for σ_1^2 is beta distributed on a relative scale and does not have a natural correspondence to maximum likelihood. Furthermore if the true distribution of effects departs from the assumed model (for example, as a mixture of normal distributions^{8;9}) then the two methods may diverge further. Our approach may benefit from imposing prior distributions on the parameters and performing Bayesian estimation, particularly for improving the precision of estimating σ_1^2 jointly with σ_{12} . This is a promising subject for future work.

Our approach provides a fast and accurate method for estimating the genetic model parameters underlying large scale association studies. It is particularly applicable to summary statistics for

individual markers, often made freely available on-line by research consortia. Therefore it will greatly facilitate the estimation of genetic covariance, especially between traits which have been studied by different consortia and for which combined analysis of individual-level data is logistically challenging. The rapid estimation of genetic models at arbitrarily large sample sizes suggests that our approach will prove useful as the sizes of consortium and biobank studies begin to approach millions of subjects.

Supplemental Data

Supplemental Data include 8 tables.

Acknowledgements

We thank Eli Stahl, Stephan Ripke and Dominika Sieradzka for discussions. This work was funded by the MRC (K006215). The authors declare no conflicts of interest.

Web Resources

AVENGEME software, <http://sites.google.com/site/fdudbridge/software/>

References

1. Robinson, M.R., Wray, N.R., and Visscher, P.M. (2014). Explaining additional genetic variation in complex traits. *Trends Genet* 30, 124-132.
2. Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature* 456, 18-21.
3. Gamazon, E.R., Cox, N.J., and Davis, L.K. (2014). Structural architecture of SNP effects on complex traits. *Am J Hum Genet* 95, 477-489.
4. Zuk, O., Schaffner, S.F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M.J., Neale, B.M., Sunyaev, S.R., and Lander, E.S. (2014). Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci U S A* 111, E455-464.
5. Furrow, R.E., Christiansen, F.B., and Feldman, M.W. (2011). Environment-sensitive epigenetics and the heritability of complex diseases. *Genetics* 189, 1377-1387.
6. Zuk, O., Hechter, E., Sunyaev, S.R., and Lander, E.S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* 109, 1193-1198.
7. 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.
8. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44, 821-824.
9. Speed, D., and Balding, D.J. (2014). MultiBLUP: improved SNP-based prediction for complex traits. *Genome Res* 24, 1550-1557.
10. Golan, D., Lander, E.S., and Rosset, S. (2014). Measuring missing heritability: inferring the contribution of common variants. *Proc Natl Acad Sci U S A* 111, E5272-5281.
11. Maier, R., Moser, G., Chen, G.B., Ripke, S., Coryell, W., Potash, J.B., Scheftner, W.A., Shi, J., Weissman, M.M., Hultman, C.M., et al. (2015). Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *Am J Hum Genet* 96, 283-294.
12. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748-752.
13. Stahl, E.A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B.F., Kraft, P., Chen, R., Kallberg, H.J., Kurreeman, F.A., et al. (2012). Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* 44, 483-489.
14. Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 9, e1003348.
15. So, H.C., Li, M., and Sham, P.C. (2011). Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. *Genet Epidemiol* 35, 447-456.
16. Bulik-Sullivan, B., Finucane, H., Anttila, V., Gusev, A., Day, F.R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa, Perry, J.R.B., Patterson, N., et al. (2015). An Atlas of Genetic Correlations across Human Diseases and Traits. *bioRxiv* doi:10.1101/014498
17. Orr, H.A. (2005). The genetic theory of adaptation: a brief history. *Nat Rev Genet* 6, 119-127.
18. Pritchard, J.K., and Di Rienzo, A. (2010). Adaptation - not by sweeps alone. *Nat Rev Genet* 11, 665-667.
19. Pritchard, J.K., Pickrell, J.K., and Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol* 20, R208-215.
20. Lynch, M., and Walsh, B. (1998). *Genetics and analysis of quantitative traits*. (Sinauer Associates).
21. Efron, B., Tibshirani, R., Storey, J.D., and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *J Am Stat Assoc* 96, 1151-1160.
22. Davison, A.R. (2003). *Statistical Models*. (Cambridge: Cambridge University Press).

23. Visscher, P.M., Hemani, G., Vinkhuyzen, A.A., Chen, G.B., Lee, S.H., Wray, N.R., Goddard, M.E., and Yang, J. (2014). Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet* 10, e1004269.
24. Johnson, T. (2012). Efficient Calculation for Multi-SNP Genetic Risk Scores. <http://cran.r-project.org/web/packages/gtx/vignettes/ashg2012.pdf>.
25. Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A.D., Chasman, D.I., Smith, A.V., Tobin, M.D., Verwoert, G.C., Hwang, S.J., et al. (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 478, 103-109.
26. Dastani, Z., Hivert, M.F., Timpson, N., Perry, J.R., Yuan, X., Scott, R.A., Henneman, P., Heid, I.M., Kizer, J.R., Lyytikäinen, L.P., et al. (2012). Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet* 8, e1002607.
27. Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421-427.
28. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M., and Wray, N.R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* 28, 2540-2542.
29. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., 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.
30. WTCCC. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661-678.
31. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559-575.
32. Lee, S.H., DeCandia, T.R., Ripke, S., Yang, J., Sullivan, P.F., Goddard, M.E., Keller, M.C., Visscher, P.M., and Wray, N.R. (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* 44, 247-250.
33. Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kahler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., et al. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet* 45, 1150-1159.
34. Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 381, 1371-1379.
35. Wray, N.R., Lee, S.H., Mehta, D., Vinkhuyzen, A.A., Dudbridge, F., and Middeldorp, C.M. (2014). Research Review: Polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry* 55, 1068-1087.
36. 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., Witte, J.S., et al. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 45, 984-994.

	RA	CD	MI	T2D	SCZ ISC	SCZ PGC1	SCZ PGC2
n_1	16016	5309	6042	14919	5953	19548	77195
n_2	12078	6785	4861	4862	5120	5120	5120
m	82390	91388	89808	75912	84882	93093	103125
σ_1^2	.18	.44	.48	.49	-	-	.30
π_{01}	.973	.972	.980	.962	-	-	.95
p_1	0.248	0.394	0.491	0.416	0.423	0.477	0.425
p_2	0.126	0.273	0.396	0.396	0.515	0.515	0.515
K_1	0.01	0.01	0.06	0.08	0.01	0.01	0.01
K_2	0.01	0.01	0.06	0.08	0.01	0.01	0.01

Table 1: Parameter values for studies of four diseases¹³ and three studies of schizophrenia²⁷. RA,

rheumatoid arthritis; CD, celiac disease; MI, myocardial infarction; T2D, type II diabetes; SCZ,

schizophrenia. ISC, International Schizophrenia Consortium; PGC, Psychiatric Genomics Consortium.

Values of σ_1^2 and π_{01} for RA, CD, MI and T2D were estimated by Stahl *et al*¹³ and subsequently used in our simulations. Those for SCZ are an approximation based on estimates from several studies and methods (table 5).

	Estimation of σ_1^2, π_{01}				Estimation of $\sigma_1^2, \pi_{01}, \sigma_{12}$			
	RA	CD	MI	T2D	RA	CD	MI	T2D
True σ_1^2	0.180	0.440	0.480	0.490	0.180	0.440	0.480	0.490
Mean $\hat{\sigma}_1^2$	0.180	0.438	0.486	0.483	0.270	0.467	0.522	0.581
SD $\hat{\sigma}_1^2$	0.019	0.035	0.050	0.034	0.312	0.325	0.335	0.332
Coverage	0.95	0.89	0.91	0.93	0.97	0.95	0.99	0.99
True π_{01}	0.973	0.972	0.980	0.962	0.958	0.972	0.979	0.961
Mean $\hat{\pi}_{01}$	0.972	0.972	0.979	0.961	0.968	0.972	0.979	0.957
SD $\hat{\pi}_{01}$	0.0054	0.0046	0.0040	0.0052	0.028	0.016	0.011	0.018
Coverage	0.94	0.85	0.88	0.90	0.98	0.95	0.98	0.98
True σ_{12}	-	-	-	-	0.180	0.440	0.480	0.490
Mean $\hat{\sigma}_{12}$	-	-	-	-	0.190	0.442	0.491	0.509
SD $\hat{\sigma}_{12}$	-	-	-	-	0.034	0.048	0.061	0.072
Coverage	-	-	-	-	0.98	0.93	0.94	0.993

Table 2: Application of AVENGEME to simulated data for 4 genetic models shown in Table 1. Mean and standard deviation of parameter estimates and coverage of 95% confidence interval are shown over 1000 simulations. Monte Carlo error for the mean is $SD/\sqrt{1000}$ and for coverage of 0.95 is 0.007.

	Estimation of σ_1^2, π_{01}				Estimation of $\sigma_1^2, \pi_{01}, \sigma_{12}$			
	Disjoint		Nested		Disjoint		Nested	
	W	U	W	U	W	U	W	U
Mean $\hat{\sigma}_1^2$	0.274	0.274	0.254	0.258	0.299	0.298	0.422	0.471
SD $\hat{\sigma}_1^2$	0.011	0.011	0.008	0.009	0.105	0.106	0.045	0.081
Coverage	0.36	0.37	0	0	0.94	0.93	0.01	0.20
Mean $\hat{\pi}_{01}$	0.950	0.950	0.951	0.950	0.946	0.946	0.941	0.933
SD $\hat{\pi}_{01}$	0.004	0.004	0.003	0.003	0.016	0.017	0.006	0.008
Coverage	0.93	0.93	0.80	0.78	0.95	0.94	0.37	0.14
Mean $\hat{\sigma}_{12}$	-	-	-	-	0.281	0.280	0.289	0.309
SD $\hat{\sigma}_{12}$	-	-	-	-	0.042	0.043	0.013	0.021
Coverage	-	-	-	-	0.91	0.91	0.69	0.85

Table 3: Comparison of AVENGEME performance when using weighted (W) or unweighted (U)

score with nested or disjoint intervals. The SCZ simulation model with $\sigma_1^2 = \sigma_{12} = 0.3$,

$\pi_{01} = 0.95$ was used (see main text for full details). Mean and standard deviation of parameter

estimates and coverage of 95% confidence interval are shown over 1000 simulations. Monte Carlo

error for the mean is $SD/\sqrt{1000}$ and for coverage of 0.95 is 0.007.

	Pruned		Clumped		Independent	
r^2	0.1	0.2	0.1	0.2	0.1	0.2
Mean $\hat{\sigma}_1^2$	0.173	0.281	0.297	0.389	0.297	0.300
SD $\hat{\sigma}_1^2$	0.041	0.053	0.042	0.05	0.039	0.046
Mean $\hat{\pi}_{01}$	0.559	0.579	0.900	0.879	0.949	0.931
SD $\hat{\pi}_{01}$	0.428	0.400	0.066	0.080	0.02	0.096

Table 4. Application of AVENGEME to normally distributed traits simulated on real genotypes.

Pruned, markers are randomly retained in the reduced set. Clumped, most strongly associated markers are greedily retained in the reduced set. r^2 , threshold on residual pairwise LD within the reduced set. Independent, results for simulated markers with no LD between any pair. True

$$\sigma_1^2 = 0.3, \pi_{01} = 0.95.$$

	RA	CD	MI	T2D	SCZ ISC	SCZ PGC1	SCZ PGC2
AVENGEME $\hat{\sigma}_1^2$.13 (.09-.17)	.28 (.21-.35)	.34 (.24-.45)	.30 (.23-.37)	.31 (.28-.34)	.31 (.29-.33)	.24 (.24-.25)
ABPA $\hat{\sigma}_1^2$.18 (.11-.25)	.44 (.34-.54)	.48 (.32-.64)	.49 (.39-.59)	-	.50 (.45-.54)*	-
GREML $\hat{\sigma}_1^2$.32 (.25-.39)	.33 (.25-.41)	.41 (.28-.54)	.51 (.38-.64)	.33 (.27-.39)	.23 (.21-.25)	-
AVENGEME $\hat{\pi}_{01}$.946 (.887-.975)	.969 (.950-.982)	.965 (.933-.982)	.954 (.929-.971)	.953(.940-.963)	.867 (.841-.887)	.852(.835-.867)
ABPA $\hat{\pi}_{01}$.973 (.953-.993)	.972 (.954-.990)	.980 (.965-.995)	.962 (.941-.983)	-	.936 (.922-.952)*	-

Table 5. Genetic model parameters estimated by AVENGEME, ABPA^{13; 33}, and GREML^{13; 32}. *

includes an additional Swedish case/control study. 95% confidence intervals given in parentheses, those for ABPA converted from the reported 50% credible intervals by assuming normally distributed posteriors and those for GREML from the reported standard error by assuming normally distributed estimators.

	AVENGEME $\hat{\sigma}_{12}$	GREML $\hat{\sigma}_{12}$
BPD-SCZ	0.199 (0.186-0.209)	0.151 (0.131-0.171)
MDD-BPD	0.134 (0.120-0.148)	0.102 (0.077-0.127)
SCZ-MDD	0.165 (0.153-0.177)	0.087 (0.065-0.110)
SCZ-ASD	0.050 (0.038-0.059)	0.03 (0.008-0.052)
ASD-BPD	0.042 (0.030-0.055)	0.008 (-0.017-0.033)

Table 6: Genetic covariance estimates for 5 pairs of 4 psychiatric traits. BPD, Bipolar Disorder; SCZ, Schizophrenia; MDD, Major Depressive Disorder; ASD, Autistic Spectrum Disorder. AVENGEME estimates are from bidirectional analysis. GREML confidence intervals derived from published standard errors ³⁶ assuming normally distributed estimators.