

The American Journal of Human Genetics

Supplemental Data

**A Fast Method that Uses Polygenic Scores to Estimate
the Variance Explained by Genome-wide Marker Panels
and the Proportion of Variants Affecting a Trait**

Luigi Palla and Frank Dudbridge

	Association between score and disease (Stahl et al., 2012)					Association between score and disease (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014)		
<i>P</i> -value interval	RA	CD	MI	T2D	<i>P</i> -value interval	SCZ-ISC	SCZ-PGC1	SCZ-PGC2
[0,10 ⁻⁴]	9×10 ⁻⁶	2×10 ⁻¹⁶	1×10 ⁻⁶	1×10 ⁻¹⁹	[0,5×10 ⁻⁸]	-	2×10 ⁻⁴	1×10 ⁻²³
(10 ⁻⁴ ,10 ⁻³]	0.03	0.001	0.05	1×10 ⁻⁴	[0,10 ⁻⁶]	3×10 ⁻⁴	4×10 ⁻⁵	4×10 ⁻³⁶
(10 ⁻³ ,0.01]	5×10 ⁻⁴	8×10 ⁻⁶	6×10 ⁻⁵	7×10 ⁻¹⁰	[0,10 ⁻⁴]	2×10 ⁻¹⁰	9×10 ⁻¹²	2×10 ⁻⁷¹
(0.01,0.05]	2×10 ⁻⁶	5×10 ⁻⁹	5×10 ⁻⁶	3×10 ⁻⁷	[0,10 ⁻³]	2×10 ⁻¹⁶	2×10 ⁻²⁹	8×10 ⁻¹⁰³
(0.05,0.1]	0.1	0.003	0.005	8×10 ⁻⁵	[0,0.01]	1×10 ⁻¹⁸	3×10 ⁻⁴⁸	2×10 ⁻¹³⁸
(0.1,0.2]	0.2	0.001	0.2	0.02	[0,0.05]	1×10 ⁻²²	1×10 ⁻⁶³	1×10 ⁻¹⁶⁴
(0.2,0.3]	0.5	0.003	0.2	0.1	[0,0.1]	6×10 ⁻²⁶	7×10 ⁻⁷⁰	6×10 ⁻¹⁶⁶
(0.3,0.4]	0.01	0.06	0.01	0.2	[0,0.2]	1×10 ⁻²⁷	8×10 ⁻⁷⁴	4×10 ⁻¹⁶⁴
(0.4,0.5]	0.03	0.08	0.3	0.02	[0,0.5]	1×10 ⁻²⁶	1×10 ⁻⁷³	8×10 ⁻¹⁵⁹
					[0,1]	6×10 ⁻²⁷	2×10 ⁻⁷⁴	2×10 ⁻¹⁵⁷

Table S1: *P*-value selection intervals and *P*-values for association between the corresponding polygenic scores and each of the diseases presented in Table 1. RA, rheumatoid arthritis; CD, celiac disease; MI, myocardial infarction; T2D, type II diabetes; SCZ, schizophrenia. ISC, International Schizophrenia Consortium; PGC, Psychiatric Genomics Consortium.

Criterion	Threshold	PLINK2 flag
Pairwise relatedness	0.025	--rel-cutoff 0.025
Subject missingness	0.03	--mind 0.03
Subject heterozygosity	<0.23 or >0.3	--het
Genotype call rate	0.01	--geno 0.01
Hardy-Weinberg <i>P</i> -value in combined NBS+58C	5e-7	--hwe 5e-7
Minor allele frequency	0.01	--maf 0.01

Table S2: Quality control criteria used in analysis of WTCCC data.

	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 $\hat{\sigma}_1^2$	0.181	0.441	0.485	0.492	0.370	0.607	0.652	0.619
SD $\hat{\sigma}_1^2$	0.018	0.029	0.043	0.030	0.335	0.323	0.327	0.318
Coverage	0.95	0.95	0.95	0.96	0.40	0.91	0.90	0.93
True π_{01}	0.972	0.972	0.979	0.961	0.958	0.972	0.979	0.961
Mean $\hat{\pi}_{01}$	0.972	0.972	0.980	0.962	0.960	0.967	0.976	0.957
SD $\hat{\pi}_{01}$	0.0049	0.0033	0.0031	0.0041	0.028	0.014	0.009	0.017
Coverage	0.95	0.95	0.96	0.96	0.95	0.93	0.93	0.94
True σ_{12}	-	-	-	-	0.180	0.440	0.480	0.490
Mean $\hat{\sigma}_{12}$	-	-	-	-	0.200	0.454	0.500	0.514
SD $\hat{\sigma}_{12}$	-	-	-	-	0.036	0.040	0.050	0.068
Coverage	-	-	-	-	0.96	0.98	0.98	0.998

Table S3: Application of AVENGEME to independently simulated χ^2 statistics for each selection interval, 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.

	3 Intervals	5 Intervals	10 Intervals	20 Intervals	40 Intervals
Mean $\hat{\sigma}_1^2$ (0.3)	0.532	0.459	0.448	0.405	0.302
SD $\hat{\sigma}_1^2$	0.45	0.435	0.430	0.413	0.364
Coverage	1	0.996	0.994	0.997	0.997
Mean $\hat{\pi}_{01}$ (0.95)	0.912	0.926	0.930	0.938	0.953
SD $\hat{\pi}_{01}$	0.085	0.076	0.069	0.062	0.053
Coverage	1	0.991	0.992	0.989	0.991
Mean $\hat{\sigma}_{12}$ (0.294)	0.288	0.286	0.287	0.288	0.296
SD $\hat{\sigma}_{12}$	0.038	0.039	0.038	0.038	0.037
Coverage	0.97	0.959	0.955	0.961	0.964
Mean $\hat{\sigma}_2^2$ (0.45)	0.534	0.484	0.502	0.451	0.351
SD $\hat{\sigma}_2^2$	0.446	0.428	0.429	0.420	0.387
Coverage	1	0.987	0.992	0.991	0.989
Mean $\hat{\pi}_{02}$ (0.94)	0.928	0.935	0.933	0.939	0.952
SD $\hat{\pi}_{02}$	0.062	0.058	0.058	0.056	0.050
Coverage	1	0.986	0.985	0.985	0.988
Mean $\hat{\sigma}_{12}$ (0.294)	0.286	0.285	0.287	0.288	0.295
SD $\hat{\sigma}_{12}$	0.037	0.038	0.038	0.038	0.036
Coverage	0.97	0.957	0.956	0.963	0.968

Table S4 : Unidirectional AVENGEME performance for estimating three parameters (true values in parentheses), under the bivariate simulation with $N_1=N_2=5000$ (see main text). Estimation in each direction is shown. Monte Carlo error for the mean is $SD/\sqrt{1000}$ and for coverage of 0.95 is 0.007. The P -value selection thresholds, for disjoint intervals, are as follows:

3 intervals: 0, 10^{-3} , 0.2, 1;

5 intervals: 0, 10^{-6} , 10^{-3} , 0.05, 0.2, 1;

10 intervals: 0, 5×10^{-8} , 10^{-6} , 10^{-4} , 10^{-3} , 0.01, 0.05, 0.1, 0.2, 0.5, 1;

20 intervals: 0, 10^{-8} , 5×10^{-8} , 10^{-7} , 10^{-6} , 10^{-5} , 10^{-4} , 5×10^{-4} , 10^{-3} , 0.005, 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.3, 0.5, 0.75, 1;

40 intervals: 0, 5×10^{-9} , 10^{-8} , 2.5×10^{-8} , 5×10^{-8} , 7.5×10^{-8} , 10^{-7} , 5×10^{-7} , 10^{-6} , 5×10^{-6} , 10^{-5} , 5×10^{-5} , 10^{-4} , 2.5×10^{-4} , 5×10^{-4} , 7.5×10^{-4} , 10^{-3} , 2.5×10^{-3} , 5×10^{-3} , 7.5×10^{-3} , 0.01, 0.0175, 0.025, 0.0375, 0.05, 0.0625, 0.075, 0.0875, 0.1, 0.125, 0.15, 0.175, 0.2, 0.25, 0.3, 0.4, 0.5, 0.625, 0.75, 0.875, 1.

	3 Intervals	5 Intervals	10 Intervals	20 Intervals	40 Intervals
Mean $\hat{\sigma}_1^2$ (0.3)	0.456	0.357	0.436	0.442	0.432
SD $\hat{\sigma}_1^2$	0.414	0.321	0.378	0.385	0.392
Coverage	1	0.95	0.882	0.852	0.842
Mean $\hat{\pi}_{01}$ (0.95)	0.937	0.946	0.939	0.939	0.941
SD $\hat{\pi}_{01}$	0.051	0.041	0.045	0.045	0.045
Coverage	1	0.971	0.914	0.909	0.904
Mean $\hat{\sigma}_{12}$ (0.294)	0.290	0.286	0.291	0.291	0.292
SD $\hat{\sigma}_{12}$	0.029	0.026	0.029	0.030	0.031
Coverage	0.996	0.960	0.933	0.935	0.938
Mean $\hat{\sigma}_2^2$ (0.45)	0.504	0.521	0.532	0.583	0.584
SD $\hat{\sigma}_2^2$	0.421	0.330	0.353	0.361	0.373
Coverage	1	0.960	0.917	0.917	0.908
Mean $\hat{\pi}_{02}$ (0.94)	0.938	0.932	0.932	0.928	0.929
SD $\hat{\pi}_{02}$	0.046	0.034	0.036	0.037	0.038
Coverage	1	0.963	0.924	0.928	0.926
Mean $\hat{\sigma}_{12}$ (0.294)	0.286	0.289	0.291	0.295	0.296
SD $\hat{\sigma}_{12}$	0.030	0.026	0.028	0.030	0.031
Coverage	0.996	0.979	0.953	0.954	0.955

Table S5 : Unidirectional AVENGEME performance for estimating three parameters (true values in parentheses), under the bivariate simulation with $N_1=N_2=10000$ (see main text). Estimation in each direction is shown. The Monte Carlo errors and P -value selection thresholds are as in Supplementary Table 2.

	3 Intervals	5 Intervals	10 Intervals	20 Intervals	40 Intervals
Mean $\hat{\sigma}_1^2$ (0.3)	0.445	0.493	0.434	0.409	0.435
SD $\hat{\sigma}_1^2$	0.385	0.298	0.274	0.258	0.260
Coverage	0.994	0.625	0.591	0.614	0.629
Mean $\hat{\pi}_{01}$ (0.95)	0.938	0.930	0.935	0.938	0.935
SD $\hat{\pi}_{01}$	0.042	0.032	0.031	0.029	0.029
Coverage	0.993	0.626	0.606	0.627	0.637
Mean $\hat{\sigma}_{12}$ (0.294)	0.294	0.305	0.301	0.298	0.304
SD $\hat{\sigma}_{12}$	0.051	0.042	0.043	0.042	0.043
Coverage	0.980	0.789	0.667	0.691	0.692
Mean $\hat{\sigma}_2^2$ (0.45)	0.516	0.501	0.451	0.431	0.428
SD $\hat{\sigma}_2^2$	0.407	0.269	0.254	0.245	0.238
Coverage	0.989	0.699	0.676	0.664	0.683
Mean $\hat{\pi}_{02}$ (0.94)	0.936	0.933	0.938	0.940	0.940
SD $\hat{\pi}_{02}$	0.040	0.026	0.026	0.026	0.025
Coverage	0.992	0.725	0.691	0.694	0.706
Mean $\hat{\sigma}_{12}$ (0.294)	0.284	0.288	0.283	0.281	0.281
SD $\hat{\sigma}_{12}$	0.054	0.038	0.040	0.040	0.040
Coverage	0.984	0.808	0.694	0.690	0.698

Table S6 : Unidirectional AVENGEME performance for estimating three parameters (true values in parentheses), under the bivariate simulation with $N_1=N_2=20000$ (see main text). Estimation in each direction is shown. The Monte Carlo errors and P -value selection thresholds are as in Supplementary Table 2.

	3 Intervals	5 Intervals	10 Intervals	20 Intervals	40 Intervals
Mean $\hat{\sigma}_1^2$ (0.3)	0.526	0.420	0.386	0.370	0.362
SD $\hat{\sigma}_1^2$	0.410	0.111	0.098	0.094	0.091
Coverage	0.978	0.436	0.505	0.576	0.596
Mean $\hat{\pi}_{01}$ (0.95)	0.927	0.932	0.936	0.939	0.940
SD $\hat{\pi}_{01}$	0.044	0.013	0.012	0.012	0.012
Coverage	0.986	0.352	0.443	0.530	0.552
Mean $\hat{\sigma}_{12}$ (0.294)	0.306	0.307	0.304	0.302	0.302
SD $\hat{\sigma}_{12}$	0.090	0.028	0.027	0.027	0.027
Coverage	0.965	0.768	0.689	0.701	0.697
Mean $\hat{\sigma}_2^2$ (0.45)	0.541	0.513	0.472	0.452	0.444
SD $\hat{\sigma}_2^2$	0.421	0.140	0.131	0.126	0.124
Coverage	0.979	0.711	0.690	0.701	0.685
Mean $\hat{\pi}_{02}$ (0.94)	0.932	0.929	0.934	0.936	0.937
SD $\hat{\pi}_{02}$	0.042	0.015	0.014	0.014	0.014
Coverage	0.984	0.679	0.706	0.728	0.731
Mean $\hat{\sigma}_{12}$ (0.294)	0.276	0.293	0.288	0.286	0.285
SD $\hat{\sigma}_{12}$	0.087	0.030	0.030	0.030	0.030
Coverage	0.971	0.848	0.731	0.691	0.685

Table S7 : Unidirectional AVENGEME performance for estimating three parameters (true values in parentheses), under the bivariate simulation with $N_1=N_2=40000$ (see main text). Estimation in each direction is shown. The Monte Carlo errors and P -value selection thresholds are as in Supplementary Table 2.

	5 Intervals	10 Intervals	20 Intervals	40 Intervals
Mean $\hat{\sigma}_1^2$ (0.3)	0.398	0.364	0.347	0.340
SD $\hat{\sigma}_1^2$	0.083	0.075	0.072	0.069
Coverage	0.385	0.522	0.579	0.606
Mean $\hat{\pi}_{01}$ (0.95)	0.935	0.939	0.942	0.942
SD $\hat{\pi}_{01}$	0.010	0.010	0.009	0.009
Coverage	0.295	0.450	0.544	0.574
Mean $\hat{\sigma}_2^2$ (0.45)	0.563	0.526	0.507	0.499
SD $\hat{\sigma}_2^2$	0.104	0.097	0.094	0.092
Coverage	0.472	0.569	0.622	0.64
Mean $\hat{\pi}_{02}$ (0.94)	0.926	0.930	0.932	0.932
SD $\hat{\pi}_{02}$	0.011	0.010	0.010	0.010
Coverage	0.386	0.512	0.585	0.608
Mean $\hat{\sigma}_{12}$ (0.294)	0.307	0.304	0.302	0.301
SD $\hat{\sigma}_{12}$	0.022	0.022	0.022	0.022
Coverage	0.771	0.686	0.695	0.701

Table S8 : Bidirectional AVENGEME performance for estimating five parameters (true values in parentheses), under the bivariate simulation with $N_1=N_2=40000$ (see main text). The Monte Carlo errors and P -value selection thresholds are as in Supplementary Table 2. No results for 3 intervals are shown as the number of intervals must be at least equal to the number of estimated parameters, in this case 5.

References

- SCHIZOPHRENIA WORKING GROUP OF THE PSYCHIATRIC GENOMICS CONSORTIUM 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511, 421-7.
- 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., KATHIRESAN, S., WIJMENGA, C., GREGERSEN, P. K., ALFREDSSON, L., SIMINOVITCH, K. A., WORTHINGTON, J., DE BAKKER, P. I., RAYCHAUDHURI, S. & PLENGE, R. M. 2012. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet*, 44, 483-9.