

**Schwartzman, Armin; Schork, Andrew J.; Zablocki, Rong; Thompson, Wesley K.**  
**A simple, consistent estimator of SNP heritability from genome-wide association studies.**  
(English) [Zbl 1435.62399](#)  
*Ann. Appl. Stat.* 13, No. 4, 2509-2538 (2019).

Summary: Analysis of genome-wide association studies (GWAS) is characterized by a large number of univariate regressions where a quantitative trait is regressed on hundreds of thousands to millions of single-nucleotide polymorphism (SNP) allele counts, one at a time. This article proposes an estimator of the SNP heritability of the trait, defined here as the fraction of the variance of the trait explained by the SNPs in the study. The proposed GWAS heritability (GWASH) estimator is easy to compute, highly interpretable and is consistent as the number of SNPs and the sample size increase. More importantly, it can be computed from summary statistics typically reported in GWAS, not requiring access to the original data. The estimator takes full account of the linkage disequilibrium (LD) or correlation between the SNPs in the study through moments of the LD matrix, estimable from auxiliary datasets. Unlike other proposed estimators in the literature, we establish the theoretical properties of the GWASH estimator and obtain analytical estimates of the precision, allowing for power and sample size calculations for SNP heritability estimates and forming a firm foundation for future methodological development.

**MSC:**

- [62P10](#) Applications of statistics to biology and medical sciences; meta analysis
- [62R07](#) Statistical aspects of big data and data science
- [92D20](#) Protein sequences, DNA sequences

**Keywords:**

[high dimensional data](#); [massively univariate regression](#); [summary statistics](#); [single nucleotide polymorphism](#)

**Software:**

[copula](#); [copula](#)

**Full Text:** [DOI](#) [Euclid](#)

**References:**

- [1] The 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S. et al. (2015). A global reference for human genetic variation. *Nature* 526 68-74.
- [2] Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* 36 199-227. · [Zbl 1132.62040](#)
- [3] Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M. J., Price, A. L. et al. (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47 291-295.
- [4] Cai, T. T., Zhang, C.-H. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* 38 2118-2144. · [Zbl 1202.62073](#)
- [5] Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M. and Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4 7.
- [6] de Leeuw, C. A., Mooij, J. M., Heskes, T. and Posthuma, D. (2015). MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* 11 e1004219.
- [7] Dicker, L. H. (2014). Variance estimation in high-dimensional linear models. *Biometrika* 101 269-284. · [Zbl 1452.62495](#)
- [8] Falconer, D. S. and Mackay, T. F. C. (1996). *Introduction to quantitative genetics*, 4th ed. Longman, Harlow.
- [9] Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52 399-433.
- [10] Gazal, S., Finucane, H. K., Furlotte, N. A., Loh, P.-R., Palamara, P. F., Liu, X., Schoech, A., Bulik-Sullivan, B., Neale, B. M. et al. (2017). Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection.

Nat. Genet. 49 1421.

- [11] Hemani, G., Shakhbazov, K., Westra, H.-J., Esko, T., Henders, A. K., McRae, A. F., Yang, J., Gibson, G., Martin, N. G. et al. (2014a). Detection and replication of epistasis influencing transcription in humans. *Nature* 508 249.
- [12] Hemani, G., Shakhbazov, K., Westra, H.-J., Esko, T., Henders, A. K., McRae, A. F., Yang, J., Gibson, G., Martin, N. G. et al. (2014b). Another explanation for apparent epistasis. *Nature* 514 E5.
- [13] Hill, W. G., Goddard, M. E. and Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* 4 e1000008.
- [14] Hofert, M., Kojadinovic, I., Maechler, M. and Yan, J. (2014). *copula: Multivariate dependence with copulas*. R package version 0.999-9.
- [15] Kojadinovic, I., Jun Yan, J. Y. et al. (2010). Modeling multivariate distributions with continuous margins using the copula R package. *J. Stat. Softw.* 34 1-20.
- [16] Li, Y., Willer, C., Sanna, S. and Abecasis, G. (2009). The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Genotype Imputation* 10 387-406.
- [17] Locke, A. E. et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518 197-206.
- [18] Lynch, M. and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Vol. 1. Sinauer Sunderland, MA.
- [19] MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A. et al. (2017). The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Res.* 45 D896-D901.
- [20] Okbay, A. et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 533 539-542.
- [21] Pasaniuc, B. and Price, A. L. (2017). Dissecting the genetics of complex traits using summary association statistics. *Nat. Rev. Genet.* 18 117-127.
- [22] Schork, A. J., Thompson, W. K., Pham, P., Torkamani, A., Roddey, J. C., Sullivan, P. F., Kelsoe, J. R., O'Donovan, M. C., Furberg, H. et al. (2013). All SNPs are not created equal: Genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet.* 9 e1003449.
- [23] Schwartzman, A., Schork, A. J., Zablocki, R. and Thompson, W. K (2019). Supplement to "A simple, consistent estimator of SNP heritability from genome-wide association studies." DOI:10.1214/19-AOAS1291SUPPA, DOI:10.1214/19-AOAS1291SUPPB.
- [24] Sniekers, S., Stringer, S., Watanabe, K., Jansen, P. R., Coleman, J. R. I., Krapohl, E., Taskesen, E., Hammerschlag, A. R., Okbay, A. et al. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* 49 1107-1112.
- [25] Spain, S. L. and Barrett, J. C. (2015). Strategies for fine-mapping complex traits. *Hum. Mol. Genet.* 24 R111-R119.
- [26] Speed, D., Cai, N., Consortium, U., Johnson, M. R., Nejentsev, S. and Balding, D. J. (2017). Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* 49 986-992.
- [27] Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A. and Yang, J. (2017). 10 years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* 101 5-22.
- [28] Wood, A. R., Tuke, M. A., Nalls, M. A., Hernandez, D. G., Bandinelli, S., Singleton, A. B., Melzer, D., Ferrucci, L., Frayling, T. M. et al. (2014). Another explanation for apparent epistasis. *Nature* 514 E3-E5.
- [29] Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G. et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42 565-569.
- [30] Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A., Lee, S. H., Robinson, M. R., Perry, J. R., Nolte, I. M. et al. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* 47 1114.
- [31] Zhou, X., Carbonetto, P. and Stephens, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet.* 9 e1003264.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.