LDSC++: Improving linkage disequilibrium score regression estimation of heritability and genetic correlation for multivariate GWAS analysis

Johan Zvrskovec[1,2(0000-0002-8862-0874)], Alexandra Gillett[1,2(0000-0002-5069-3197)], Michel Nivard[5,6,7], Jonathan R I Coleman[1,2(0000-0002-6759-0944)], Christopher Hübel[1,3,4(0000-0002-1267-8287)], Raquel Iniesta[1,2], Gerome Breen[1,2(0000-0003-2053-1792)]

1. Institute of Psychiatry, Psychology and Neuroscience, King's College London, Denmark Hill, Camberwell, London, UK
2. National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, UK
3. Clinic for Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, German Red Cross Hospitals Berlin Westend, Berlin, Germany
4. National Centre for Register-based Research, Aarhus University, Aarhus, Denmark
5. Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK
6. Population Health Sciences, University of Bristol, Bristol, UK
7. Dept of biological psychology, Vrije Universiteit Amsterdam, the Netherlands

Abstract

Introduction: Linkage disequilibrium (LD) score regression is widely used for estimating common variant heritability and genetic correlations from genome-wide association study (GWAS) summary statistics. We hypothesise that segmented regression (also known as piecewise regression) improves on previous LD score regression implementations, when estimating both genetic covariance and its standard error.

Methods: We present novel extensions to LD score regression (LDSC++) improving I.) handling of varying numbers of shared genetic variants across trait pairs and reference panels, II.) estimation of genetic covariance and its variance, and III.) handling of imputation quality. We propose supporting statistical tests that use our novel extensions to improve sensitivity, and are further aimed at comparing parameter estimates that are highly correlated, such as those obtained from the same trait but from different methods. We validate LDSC++ first on real-world individual level data from the Genetic Links to Anxiety and Depression study and the United

Kingdom National Institute of Health and Social Care Research BioResource (N: 14,190 - 20,144), second on simulated data with different degrees of shared QTL, and third on a battery of publicly available GWASs of ten diverse traits of varying statistical power and heritability.

Results: Using variance-component method (GCTA-GREML) estimates for reference, LDSC++ extensions were found to yield heritability estimates with a bias of about -10% to -20% while standard LD score regression yielded a bias of -30%, and heritability variability estimates with a bias of -1% to -7% while standard LD score regression yielded a bias of 8%. For ten external trait GWASs, LDSC++ was shown to recover 5% to 8% larger heritabilities with 4% smaller variability on average compared to standard LD score regression. Weighting by imputation quality in the model, rather than excluding genetic variants of low imputation quality, contributed to retaining information. Our supporting statistical tests enabled us to detect statistically significant differences in genetic covariance and its standard error while considering the varying number of shared genetic variants across bivariate trait pairs.

Conclusion: LDSC++ was confirmed to produce less biassed estimates of genetic covariance and its variability in our GLAD+ sample compared to standard LD score regression, using GCTA-REML as reference. This performance was supported by results from external trait GWASs of varying character, also implying an important performance of our extended weighting schemes. Our proposed extensions to LD score regression, among which genome-wide parameters are constructed as aggregates of heterogeneous local parameters, may prove important for large-scale multivariate studies such as genomic structural equation models or local genetic covariance analyses.

1 | INTRODUCTION

Genome-wide association study (GWAS) summary statistics are readily available to researchers worldwide compared to individual level genotyped or sequenced data, which typically require considerable cost and time to access due to data sharing restrictions[1,2]. This prevents the use of optimal methods that use individual level genotype data to estimate components of variance, such as GCTA-GREML[3,4], BOLT-REML[5,6] and REML in LDAK[7]. Popular alternatives[8] are Linkage Disequilibrium (LD) score regression[9] (LDSR) and SumHer[10], which instead use GWAS summary statistics and an LD reference panel to obtain common genetic variant covariance (covG) and heritability (h^2), which are standardised to estimate genetic correlations[11] (r_g). LDSR was later re-implemented for calculation of large-scale bivariate trait pair analyses to obtain covG matrices and variance estimates of the elements of these for genomic structural equation modelling in Genomic SEM[12].

However, available implementations of LDSR do not: 1.) explicitly account for varying numbers of genetic variants between association statistics from contributing GWASs, 2.) consider local association and LD patterns in GWASs, or 3.) adjust for imputation quality differences. In bivariate LDSR, genetic variant overlap between two traits and the LD score panel is usually drastically smaller compared to the regressions performed on single traits (supplementary tab. S6), due to differing patterns of missingness across datasets. This limits the power to carry out many downstream multivariate GWAS analyses, such as in Genomic SEM[12], and related methods such as mtCOJO[13,4], MTAG[14], and METAL[15]. These methods rely on the covG estimates from LDSR, and/or use the LDSR intercept to adjust associations (genomic control) for spurious associations due to confounding from population stratification or sample overlap. In comparison, the variance-component REML method in LDAK[7] is both assuming a variable contribution to heritability across the genome and includes an information score weighting[16] to differentially treat variants of varying genotyping or imputation quality. It may be of interest to bring this functionality

over to an LDSR setting to be applicable to summary level data rather than individual level data. Standard LDSR further assumes there to exist a heteroskedastic relationship between genome-wide covG and its standard error, mainly due to the expected larger variability of variants in high LD regions. Despite a weighting correction to account for genome-wide covG heteroskedasticity, it also served as justification[9] for employing resampling to estimate standard errors that are robust to bias from heteroskedasticity. It may therefore be of interest to investigate whether this assumption can be relaxed when estimating covG standard errors, to allow for sampling across the genome.

Here we present an enhanced version of LDSR, LDSC++, extended to improve on 1.) the handling of varying numbers of shared genetic variants across trait pairs and reference panels, 2.) the estimation of genetic covariance and its standard error, and 3.) the handling of imputation quality of imputed genotypes or genetic associations. Finally, through our extensions we aim to allow for tractable statistical tests of covG, its standard error, or differences between them, also accounting for high correlations between compared variables. We validate the performance of our extensions to LDSR, first using simulations, second on real-world data in the form of GWASs performed on selected measurements in the Genetic Links to Anxiety and Depression study and the United Kingdom National Institute of Health and Social Care Research BioResource, and third on a selection of ten GWAS summary statistics representing a wide variety of traits. We hypothesise the accuracy of estimating both covG and its standard error may be improved and the average measured variability of covG estimates across a sample of multiple diverse traits will be reduced when applying our extensions in LDSC++ individually or in combination, as compared to standard LDSR.

2 | MATERIALS AND METHODS

To enable more precise covG estimation from GWAS summary statistics, we extend the LDSR method based on the multivariate implementation included in Genomic SEM. We have named this extension LDSC++ (supplementary info. 1.8). The extensions to LDSR are:

- **Extension 1 - Variable block-count:** Allowing the number of blocks to vary when estimating covG and covG standard error, in contrast to a fixed number of blocks.
- **Extension 2 - Extended block definitions to allow for variable block-count:** Block definitions to be used while allowing the number of blocks to vary with Extension 1. These are based on either a set number of variants or a recombination distance in cM.
- **Extension 3 - Variable block-count sampling:** A novel approach to sample across genome blocks to estimate genome-wide covG and covG variance, rather than performing block jackknife resampling.
- **Extension 4 - Extended weighting scheme for imputation quality:** An addition to the standard LDSR weighting scheme to also control for imputation quality.
- **Extension 5 - Adjusted weighting scheme to correct for the correlation between LD score and association statistic:** We adjusted the standard weighting scheme in LDSR to correct for correlation between the regression variables, to be less influenced by extreme LD values due to noise.

LDSC++ alongside standard LDSR (emulated in our LDSC++ code) were validated in both real-world data and simulated traits, using REML covG estimates for reference.

An individual level sample from cohorts part of the National Institute of Health Research (NIHR) BioResource was used for assessing genetic associations with phenotypes and for simulation. This selection included participants of European ancestry from the Genetic Links to Anxiety and Depression study (GLAD)[17], the United Kingdom Eating Disorders Genetics Initiative

123 (EDGI)[18,19], the NIHR BioResource Inflammatory Bowel Disease BioResource panel and the
124 NIHR BioResource General Population panel, which together form a collection of cohorts
125 colloquially referred to as GLAD+ (N=36,746).

126 We selected the traits height, weight, neuroticism measured by the Eysenck Personality
127 Questionnaire-Revised Short form (EPQR-S)[20], and negative affect measured by the PID-5 Brief
128 Form[21,22], in the GLAD+ cohort. Additionally, we simulated two phenotypes (further referred to as
129 SIM1 and SIM2 respectively) created to have 20% heritability and a genetic correlation of 0.4 with
130 each other, in four pairs with fractions of 100%, 75%, 50%, and 25% shared quantitative trait loci
131 (QTL) respectively. A simulation was performed in GCTA[4] for each phenotype and fraction
132 combination, using 30,000 QTL and GLAD+ as an individual level sample.

133 We used GCTA-GREML[3] implemented in GCTA[4] to estimate variance component covG
134 for our individual level GLAD+ samples. Association tests were performed in GCTA fastGWA[23].
135 The covariates included in the association analysis were: genotyping array type/version, tissue
136 source, genotyping batch, and the first seven principal component scores of participant genotypes
137 as computed by the genetic principal component eigenvalues of the 1000 Genome Project[24]
138 Phase 3 (1kG3) reference panel projected on the genetic variants of GLAD+. Additionally, ten
139 external GWAS summary statistics covering a range of traits of various statistical power, h^2, and
140 genome coverage in terms of the number of variants included were selected for validating
141 methods (Tab. 1). All summary statistics were documented to be based on samples with
142 predominantly European ancestry.

143 To match GCTA-GREML, we used GCTA[4,25] to generate the GLAD+ LD scores, using a
144 window of 1 Mb and a prior filter on minor allele frequency (MAF) > 0.01. For our external GWAS
145 summary statistics, we instead created LD scores based on the European ancestry subset of the
146 1kG3[26], not including Finnish ancestry. For this second LD score library, we used a 1 cM window
147 and 250 blocks setting in LDSC[9,27].

148 GWAS summary statistics were first harmonised using our *Supermunge* routine part of
149 the *shru*[28] package for the R language [https://github.com/tnggroup/tng_fork_shru_jz2024],
150 based on the original LDSC munge routine[9], the Genomic SEM routines for munging and
151 preparation for latent factor GWAS[12], and the MungeSumstats routine[29] (supplementary info.
152 1.4.4).

153 To account for heteroskedastic relationships for comparisons of covG inferential
154 uncertainty (sampling variance and thus standard errors) across traits of varying heritabilities, we
155 applied the coefficient of variation (CV) of covG rather than comparing direct estimates
156 (supplementary info. 1.5.1).

157 We formulated simple parametric statistical tests specifically accounting for the novel
158 varying number of blocks in the LDSR sampling and resampling methods and non-independent
159 variables (supplementary info. 1.9), supported by paired samples Wilcoxon signed-rank tests[30]
160 (supplementary info. 1.9.12). Test results were corrected for multiple comparisons using the
161 Benjamini & Hochberg[31] method (q-values, also called false discovery rate; FDR), additionally
162 and unconventionally taking the effective number of comparisons from correlated variables[32–34]
163 into account (supplementary info. 1.9.13).

164

165

166

167

168

169

170

171

172

173

174 **Table 1. Selected external GWAS summary statistics**

175 A selection of ten diverse publicly available GWAS summary statistics. The table describes the dataset
176 reference, the number of cases or total number of participants, and for binary traits; the number of
177 controls, the sample prevalence estimate, and the population prevalence estimate used throughout the
178 study. 'INFO range' describes the range of the provided imputation quality information score.

Trait	Reference	Year	N case	N control	Sample prev.	Population prev.	INFO range
Anxiety disorder (ANXD)	**Purves et al.[35]**	**2019**	**25,453**	**58,113**	**0.305**	**0.16[36]**	**0.8-1.0**
Bipolar disorder (BIPD)	**Stahl et al.[37]**	**2019**	**20,352**	**31,358**	**0.394**	**0.025[38]**	**0.1-1.6**
BMI (BMI)	**Pulit et al.[39]**	**2019**	**806,834**				**0.3-1.0**
MDD (MDD)	**Wray et al.[40]**	**2018**	**16,823**	**25,632**	**0.396**	**0.146[41]**	**0.6-1.9**
Broad depression (MDDB)	**Howard et al.[42]**	**2019**	**170,756**	**329,443**	**0.341**	**0.302[42]**	**-**
MDD symptoms (MDDS)	**Wray et al.[40]**	**2018**	**14,260**	**15,480**	**0.479**	**0.146[41]**	**0.9-1.0**
Educational attainment (EDUA)	**Lee et al.[43]**	**2018**	**766,345**				**-**
Insomnia (INSO)	**Jansen et al.[44]**	**2019**	**109,402**	**277,131**	**0.283**	**0.69[45]**	**0.9-1.0**
Neuroticism (NEUR)	**Nagel et al.[46]**	**2018**	**449,484**				**0.9-1.0**
Schizophrenia (SCHI)	**Pardiñas et al.[47]**	**2018**	**40,675**	**64,643**	**0.386**	**0.0072[48]**	**-**

179

180

181

182

183

184

185

186 # 3 | RESULTS

187 We estimated REML h^2 and r_g of our four selected GLAD+ traits and their bivariate combinations.
188 The REML h^2 of the simulated traits SIM1 and SIM2 were estimated in the range from 13% to
189 15%, and the REML r_g between traits in a pair to 0.43, 0.24, 0.19, and 0.2 for the QTL sharing
190 fractions 100%, 75%, 50%, and 25% respectively. Deviations of LDSR method results from
191 GCTA-REML estimates were obtained, run on both the selected GLAD+ traits (Tab. 2,
192 supplementary tab. S7A, S7B) and the simulated traits (supplementary tab. S8A, S8B). Most
193 deviations between covG and covG variance estimates by our LDSR method extensions
194 compared to standard LDSR were seen to be statistically significant, supporting our extensions
195 to yield empirically distinct results. As results were quite different comparing results from diagonal
196 and off-diagonal elements of covG matrices, we decided to stratify them accordingly.

197 Only considering h^2 estimates (diagonal elements in the covG or covG SE matrix), the top
198 performing LDSR method combination was variable block-count sampling using the
199 recombination distance block-definition, which yielded average relative h^2 differences as small as
200 -11% and average relative h^2 CV differences of -1% compared to GCTA-GREML in GLAD+. The
201 same pattern was seen for the simulated traits only using the 100% QTL sharing fraction, while
202 for lesser sharing fractions the performance among methods was mixed with the standard LDSR
203 block jackknife or LDSC++ variable block-count sampling with recombination distance blocks
204 yielding the on average most accurate and less biassed estimates. In the GLAD+ data, all LDSR
205 methods under-estimated diagonal covG with about -30% for block jackknife resampling and -
206 10% to -20% for variable block-count sampling. Block jackknife resampling was also consistently
207 seen to over-estimate diagonal variability estimates from 5% to 9%, while variable block-count
208 sampling seemed to under-estimate diagonal variability estimates with about -1% to -7%. The
209 influence of our weighting extensions were seen to only have tiny effects on the differences in the

210 preferred direction for the GLAD+ traits, primarily detectable for h^2 estimates and when using
211 block jackknife resampling and imputation quality weighting.

212 Off-diagonal deviations were markedly larger by about a factor of two to four for average
213 relative covG estimates and a factor of ten to sixty for average relative covG CV estimates,
214 compared to the corresponding diagonal matrix element results for the GLAD+ traits. A similar
215 but attenuated pattern was seen for the simulated traits. As for the diagonal elements, the LDSR
216 methods under-estimated off-diagonal covG estimates with about -60% for block jackknife
217 resampling and -40% to -50% for variable block-count sampling in the GLAD+ traits. Off-diagonal
218 covG CVs were consistently and grossly over-estimated by all LDSR methods. For block jackknife
219 resampling, the recombination distance block definition brought the off-diagonal covG CV
220 difference to about 100% compared to 120% when using number-of-variants defined blocks.
221 Variable block count sampling produced off-diagonal covG CV estimates with deviations from
222 about 60% with the recombination distance block definition to 80% with the number-of-variants
223 block definition.

224 We applied the different LDSR method combinations in our selection of ten external trait
225 GWASs (Tab. 1) for which samples we did not have individual level data (supplementary tab.
226 S9A, S9B). Default block-size settings resulted in block-counts in the range of 249 - 405 for the
227 number-of-variants block definition and 338 - 364 for the recombination distance block definition,
228 without INFO filter. Estimates from method combinations using our extensions were first assessed
229 independently in reference to suitable baselines (supplementary tab. S9A). Compared to standard
230 LDSR, applying our block definitions and sampling extensions yielded 4% to 8% larger h^2
231 estimates, 3% smaller h^2 CV estimates using number-of-variants blocks, but 7% larger h^2 CV
232 estimates using recombination distance blocks. The weighting extensions made a relatively small
233 contribution to average differences in estimates compared to the block and sampling extensions,
234 but specific traits showed statistically significant results. Applying imputation score weighting
235 resulted in an approximate gain of 4% to 5% ($q<0.05$) in BIPD h^2 and a 1% to 7% ($q<0.02$)

236 reduction in BIPD bivariate covG CV with specific traits across block and sampling methods.
237 Similarly, applying alternative correlation correction weighting yielded an approximate gain of 3%
238 to 4% ($q<0.05$) in BMI h^2 and multiple significant bivariate BMI covG CV reductions in the range
239 of 1% to 10% ($q<0.05$). Applying all extensions in concert and comparing to standard LDSR
240 yielded 5% to 8% larger h^2 estimates, 4% smaller h^2 CV estimates using number-of-variants
241 blocks, but 6% larger h^2 CV estimates using recombination distance blocks.

Table 2. LDSR estimate deviation from a REML ideal measured over four GLAD+ traits

The table shows deviations of LD score regression (LDSR) method estimates of genetic covariance (covG) and covG SE (converted to coefficient of variation, CV) from the chosen ideal; estimates from GCTA-GREML using individual level data. We used the LDSC++ implementation of LDSR applied to our selection of four real-world traits from the GLAD+ sample: height, weight, neuroticism measured by the Eysenck Personality Questionnaire-Revised Short form (EPQR-S), and negative affect measured by the PID-5 Brief Form. Deviation was measured as average relative difference (Avg. Δ) and root-mean-square error (RMSE). LDSR method combinations were defined by their configuration in regard to block definition, (re)sampling method, alternative correlation correction weighting, and imputation quality weighting. Estimates are stratified in diagonal (heritability) and off-diagonal (bivariate genetic covariance) estimates from a genetic covariance matrix. Only significant covG values (q<0.05) were considered.

Block definition	(Re)sampling method	Alternative correlation correction weighting	Imputation quality weighting	covG				covG CV			
				Diagonal		Off-diagonal		Diagonal		Off-diagonal	
				Avg. Δ (%)	RMSE	Avg. Δ (%)	RMSE	Avg. Δ (%)	RMSE	Avg. Δ (%)	RMSE
Fixed 200 blocks	Block jackknife resampling	☐	☐	-32.7	0.0754	-61.0	0.0994	7.6	0.0756	118.0	0.1283
Number-of-variants blocks	Block jackknife resampling	☐	☐	-32.7	0.0754	-61.0	0.0994	7.9	0.0788	120.0	0.1300
		☑	☐	-33.0	0.0759	-61.3	0.0998	8.7	0.0769	122.4	0.1327
		☐	☑	-33.0	0.0756	-61.3	0.0998	6.4	0.0769	116.5	0.1231
		☑	☑	-33.2	0.0760	-61.5	0.1002	7.3	0.0750	118.8	0.1256
	Variable block-count sampling	☐	☐	-21.7	0.0542	-50.1	0.0836	-5.8	0.1182	79.4	0.0783
		☑	☐	-22.0	0.0549	-50.4	0.0841	-5.6	0.1173	80.5	0.0795
		☐	☑	-21.0	0.0523	-52.3	0.0861	-6.7	0.1185	81.4	0.0807
		☑	☑	-21.2	0.0529	-52.6	0.0867	-6.5	0.1175	82.8	0.0821
Recombination distance blocks	Block jackknife resampling	☐	☐	-31.7	0.0722	-59.8	0.0972	5.1	0.0727	101.8	0.1155
		☑	☐	-31.9	0.0726	-60.0	0.0976	5.9	0.0707	103.6	0.1176
		☐	☑	-31.9	0.0721	-60.0	0.0975	5.7	0.0703	103.1	0.1177
		☑	☑	-32.1	0.0726	-60.2	0.0979	6.5	0.0683	105.0	0.1198
	Variable block-count sampling	☐	☐	-11.1	0.0475	-38.5	0.0638	-1.6	0.1132	59.7	0.0925
		☑	☐	-11.4	0.0480	-38.5	0.0643	-1.1	0.1127	60.9	0.0944
		☐	☑	-11.2	0.0482	-46.4	0.0784	-1.4	0.1134	99.0	0.0962
		☑	☑	-11.5	0.0487	-39.3	0.0647	-0.8	0.1129	62.9	0.0910

4 | DISCUSSION

We have adapted the LDSR routine from Genomic SEM into a modified implementation called LDSC++, which performs LDSR with extensions specifically aimed at improving covG estimation in large-scale multivariate analyses with varying genomic overlaps among trait GWAS dataset and reference panels. For this purpose, we created a variable block-count sampling routine in LDSC++ aimed to be applied with a variable number of blocks to discriminate between situations with a varying number of participating genetic variants, which further makes it possible to use this information in statistical tests. Variable block-count sampling performs segmented (also called piecewise) regression using one of our two block definitions, which putatively makes the aggregate genome-wide parameter estimates (averages of segmented regression parameters) robust to heteroskedasticity and outliers. The most obvious and important assumption for both the variable block-count sampling method and the statistical tests is that the estimated genome-wide covG can be assumed to have the same variance across the genome, or rather can be estimated as an aggregate of piecewise estimates. While resampling techniques[49,50] traditionally have been used to obtain robust covG standard errors in LDSR, a sufficiently better model from the segmented regression is in our project theorised to make robust standard errors superfluous[51]. LDSC++ further takes advantage of imputation quality information to weight the LDSR, instead of relying on filters which otherwise reduce the overall information content in the analysis. Variable block-count sampling may be interpreted as applying a slightly different h^2 model, allowing for a different contribution to h^2 across the genome similar to the LDAK h^2 model[7,16,52], in contrast to the standard LDSR heritability model[9,16,52]. We additionally proposed statistical tests in which the varying number of blocks from our extensions can be applied to translate statistical power from a large number of genetic variants in a GWAS into statistical power for LDSR covG and covG

275 variance, and thus making these tests more sensitive and differentiating in collections of bivariate
276 trait pairs proportional to the variability in genetic variant overlap.

277 Our LDSR extensions were shown to yield more accurate results as validated by GCTA-
278 GREML than standard LDSR settings, both when using each extension independently and when
279 applied in concert, yielding average relative deviations from about -10% to -20% for h^2 and -1%
280 to -6% for h^2 CV compared to the corresponding deviations of -30% and 8% respectively for
281 standard LDSR. This performance was seen in estimates made in the individual level GLAD+
282 sample and replicated to a lesser extent in simulated data using GCTA-GREML. The performance
283 increase when using the combined extensions could be attributed to allowing the number of
284 blocks to vary and using our new block definitions or using our variable block-count sampling, and
285 only a small fraction could be attributed to the weighting extensions in this setting. Applying our
286 extensions to a set of ten trait GWASs external to our individual level sample, supported the above
287 validations through convergent results primarily for h^2 estimates rather than off-diagonal covG
288 estimates and when using number-of-variants block definitions. In this diverse selection of
289 GWASs of varying power and coverage our combined LDSC++ extensions yielded on average
290 up to 8% larger and 4% less variable h^2 estimates compared to standard LDSR. Larger h^2
291 estimates were convergent with the previous individual level data validation validations, but only
292 the number-of-variants block definition were seen to produce smaller average h^2 estimates
293 compared to standard LDSR, while using the recombination distance block definition instead
294 yielded larger average h^2 estimates - contrary to the previous validation results. This may indicate
295 a sensitivity of the recombination distance block definition to missingness patterns or other factors
296 that were not represented by our four GLAD+ trait GWASs. The external GWASs also allowed us
297 to explore other imputation score ranges and highly polygenic trait GWASs of high power, which
298 implicated putative improved estimates when applying the imputation score weighting to the BIPD
299 GWAS and when applying the alternate correlation correction weighting to the BMI GWAS. Our
300 proposed explanation for these results is that the BIPD GWAS has a wide distribution of

301 imputation quality INFO scores in contrast to other included datasets which the corresponding
302 weighting extension would be able to leverage, and the BMI GWAS is relatively well-powered and
303 represents a highly polygenic trait for which the alternative correlation correction weighting may
304 be able to correct bias due to tiny LD scores to a higher degree.

305 The major strength in our simulations came from using a relatively large sample size of
306 36,746 individuals from the GLAD+ cohort, in contrast to the Bulik-Sullivan study which simulated
307 associations based on a sample of 2,062 individuals. We additionally simulated whole genome
308 associations, where the Bulik-Sullivan study only simulated chromosome 2. Unfortunately, we did
309 not manage to perform sufficiently many repetitions in our simulations for them to appear
310 trustworthy, with an effective number of GCTA simulations of 2.5 for any randomly sampled QTL
311 compared to 100 repetitions done in a previous study by Bulik-Sullivan[53]. Our effort to investigate
312 the influence of effect QTL overlap on REML and LDSR estimates was interesting, but proved
313 practically challenging enough to be one of the reasons for why we did not manage to repeat the
314 simulations to a higher degree.

315 The discrepancy between validation results from diagonal and off-diagonal covG elements
316 was identified as a cause for concern about the LDSR method in general. Because of this we
317 were inclined to trust off-diagonal validation results less. Off-diagonal covG estimates from
318 bivariate LDSR are however the foundation for genetic correlation analyses, and also provide the
319 majority and arguably the more important information for analyses performed on covG such as
320 Genomic SEM, underscoring the importance of accurate off-diagonal covG estimation and the
321 need to critically investigate bivariate LDSR further.

322 Our results provide support for the accuracy of statistical models where genome-wide h^2
323 and covG parameters are constructed as aggregates of heterogeneous local parameters. The
324 recombination distance block definition was part of the method combination displaying the
325 strongest performance in our individual level real-world and simulated GCTA-GREML validations.
326 In the external GWASs, the recombination distance block definition also recovered some of the

largest covG estimates on average compared to the number-of-variants block definition and standard LDSR. We would suggest that these properties are promising and may prompt further research into block definitions for LDSR that are consistent across traits and bivariate trait pairs in the way our recombination distance definition was made or how local genetic correlation methods such as LAVA[54] define consistent genomic blocks. The arbitrary definition of blocks is a general limitation of LDSR. Creating more optimised block definitions is indeed an active research area[55,56]. Reconciling the recombination based block definition approach with the utility of the number-of-variants block definition may be important to advance LDSR and similar methods further.

5 | CONFLICT OF INTEREST

Prof Breen has received honoraria, research or conference grants and consulting fees from Illumina, Otsuka, and COMPASS Pathfinder Ltd.

6 | ACKNOWLEDGEMENTS

This study/research is funded by the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Christopher Hübel acknowledges funding by Lundbeckfonden (R276-2018-4581).

Alexandra Gillett is supported by the Medical Research Council (MR/X009815/1).

The authors acknowledge use of the research computing facility at King’s College London, King's Computational Research, Engineering and Technology Environment (CREATE) (https://doi.org/10.18742/rnvf-m076)[57].

7 | DATA AVAILABILITY

All data produced in the present study are available upon reasonable request to the authors. While some data is contained in the manuscript, an attempt has been made to additionally present results in table format in the accompanying supplementary tables. GWAS summary statistics of our traits used for validation are available upon reasonable request. LD scores based on the GLAD+ sample are available upon reasonable request. The analysis code will be made available at https://github.com/tnggroup/jz-phd-c2c3-2024. The software package 'shru' will be made available on https://github.com/johanzvrskovec/shru.

8 | ETHICS APPROVAL STATEMENT

The London—Fulham Research Ethics Committee approved the GLAD Study on 21st August 2018 (REC reference: 18/LO/1218). The London—Fulham Research Ethics Committee approved the EDGI UK Study on 29th July 2019 (REC reference: 19/LO/1254). The East of England—Cambridge Central Committee approved the NIHR BioResource as a Research Tissue Bank (REC reference: 17/EE/0025). The South West—Central Bristol Research Ethics Committee approved the COVID-19 Psychiatry and Neurological Genetics study on 27th April 2020 (REC reference: 20/SW/0078).

9 | REFERENCES

1. Abdellaoui, A. & Verweij, K. J. H. Dissecting polygenic signals from genome-wide association studies on human behaviour. *Nat. Hum. Behav.* 1–9 (2021) doi:10.1038/s41562-021-01110-y.
2. Pasaniuc, B. & Price, A. L. Dissecting the genetics of complex traits using summary association statistics. *Nat. Rev. Genet.* **18**, 117–127 (2017).
3. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
4. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
5. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).
6. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
7. Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved Heritability Estimation from Genome-wide SNPs. *Am. J. Hum. Genet.* **91**, 1011–1021 (2012).
8. Barry, C.-J. S. *et al.* How to estimate heritability: a guide for genetic epidemiologists. *Int. J. Epidemiol.* **52**, 624–632 (2023).
9. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
10. Speed, D. & Balding, D. J. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat. Genet.* **51**, 277–284 (2019).
11. van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H. & Wray, N. R. Genetic correlations of polygenic disease traits: from theory to practice. *Nat. Rev. Genet.* **20**, 567–581 (2019).
12. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the

393 multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
394 13. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from
395 GWAS summary data. *Nat. Commun.* **9**, 224 (2018).
396 14. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using
397 MTAG. *Nat. Genet.* **50**, 229–237 (2018).
398 15. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of
399 genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
400 16. Speed, D., Cai, N., Johnson, M. R., Nejentsev, S. & Balding, D. J. Reevaluation of SNP
401 heritability in complex human traits. *Nat. Genet.* **49**, 986–992 (2017).
402 17. Davies, M. R. *et al.* The Genetic Links to Anxiety and Depression (GLAD) Study: Online
403 recruitment into the largest recontactable study of depression and anxiety. *Behav. Res.*
404 *Ther.* (2019) doi:10.1016/j.brat.2019.103503.
405 18. Monssen, D. *et al.* The United Kingdom Eating Disorders Genetics Initiative. *Int. J. Eat.*
406 *Disord.* **57**, 1145–1159 (2024).
407 19. Bulik, C. M. *et al.* The Eating Disorders Genetics Initiative (EDGI): study protocol. *BMC*
408 *Psychiatry* **21**, 234 (2021).
409 20. Eysenck, H. J. & Eysenck, M. W. Personality and Individual Differences. *Plenum Press N.*
410 *Y.* (1985) doi:10.1007/978-1-4613-2413-3.
411 21. Krueger, R. F., Derringer, J., Markon, K. E., Watson, D. & Skodol, A. E. Initial construction
412 of a maladaptive personality trait model and inventory for DSM-5. *Psychol. Med.* **42**, 1879–
413 1890 (2012).
414 22. Maples, J. L. *et al.* Supplemental Material for Testing Whether the DSM-5 Personality
415 Disorder Trait Model Can Be Measured With a Reduced Set of Items: An Item Response
416 Theory Investigation of the Personality Inventory for DSM-5. *Psychol. Assess.* **27**, 1195–
417 1210 (2015).
418 23. Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-scale

419 data. *Nat. Genet.* (2019) doi:10.1038/s41588-019-0530-8.

420 24. Altshuler, D. M. *et al.* An integrated map of genetic variation from 1,092 human genomes.
421 *Nat. 2012 4917422* **491**, 56–65 (2012).

422 25. Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing
423 heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120 (2015).

424 26. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

425 27. Bulik-Sullivan, B. K. LDSC (LD SCore). (2015).

426 28. Zvrskovec, J. shru. https://github.com/johanzvrskovec/shru (2023).

427 29. Murphy, A. E. & Skene, N. G. MungeSumstats: A Bioconductor package for the
428 standardisation and quality control of many GWAS summary statistics. *bioRxiv*
429 2021.06.21.449239 (2021) doi:10.1101/2021.06.21.449239.

430 30. Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biom. Bull.* **1**, 80–83 (1945).

431 31. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and
432 Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300
433 (1995).

434 32. Gao, X., Starmer, J. & Martin, E. R. A multiple testing correction method for genetic
435 association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* **32**,
436 361–369 (2008).

437 33. Nyholt, D. R. A Simple Correction for Multiple Testing for Single-Nucleotide Polymorphisms
438 in Linkage Disequilibrium with Each Other. *Am. J. Hum. Genet.* **74**, 765–769 (2004).

439 34. Cheverud, J. M. A simple correction for multiple comparisons in interval mapping genome
440 scans. *Heredity* **87**, 52–58 (2001).

441 35. Purves, K. L. *et al.* A major role for common genetic variation in anxiety disorders. *Mol.*
442 *Psychiatry* (2019) doi:10.1038/s41380-019-0559-1.

443 36. Kessler, R. C. *et al.* The global burden of mental disorders: An update from the WHO World
444 Mental Health (WMH) Surveys. *Epidemiol. Psichiatr. Soc.* **18**, 23–33 (2009).

445 37. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar
446 disorder. *Nat. Genet.* **51**, 793–803 (2019).
447 38. Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M. & Wittchen, H.-U. Twelve-
448 month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the
449 United States. *Int. J. Methods Psychiatr. Res.* **21**, 169–184 (2012).
450 39. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution
451 in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).
452 40. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the
453 genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
454 41. Kessler, R. C. & Bromet, E. J. The Epidemiology of Depression Across Cultures. *Annu. Rev.*
455 *Public Health* **34**, 119–138 (2013).
456 42. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent
457 variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**,
458 343–352 (2019).
459 43. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association
460 study of educational attainment in 1.1 million individuals. *Nat. Genet. 2018 508* **50**, 1112–
461 1121 (2018).
462 44. Jansen, P. R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies
463 new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).
464 45. Morphy, H., Dunn, K. M., Lewis, M., Boardman, H. F. & Croft, P. R. Epidemiology of
465 Insomnia: a Longitudinal Study in a UK Population. *Sleep* **30**, 274–280 (2007).
466 46. Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in
467 449,484 individuals identifies novel genetic loci and pathways. *Nat. Genet. 2018 507* **50**,
468 920–927 (2018).
469 47. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant
470 genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).

48. McGrath, J., Saha, S., Chant, D. & Welham, J. Schizophrenia: A Concise Overview of Incidence, Prevalence, and Mortality. *Epidemiol. Rev.* **30**, 67–76 (2008).
49. Friedl, H. & Stampfer, E. Jackknife Resampling. *Encycl. Environmetrics* **2**, (2001).
50. Efron, B. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika* **68**, 589–599 (1981).
51. King, G. & Roberts, M. E. How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It. *Polit. Anal.* **23**, 159–179 (2015).
52. Speed, D., Holmes, J. & Balding, D. J. Evaluating and improving heritability models using summary statistics. *Nat. Genet.* **52**, 458–462 (2020).
53. Bulik-Sullivan, B. Relationship between LD Score and Haseman-Elston Regression. 018283 Preprint at https://doi.org/10.1101/018283 (2015).
54. Werme, J., Sluis, S. van der, Posthuma, D. & Leeuw, C. A. de. LAVA: An integrated framework for local genetic correlation analysis. *bioRxiv* (2021) doi:10.1101/2020.12.31.424652.
55. Privé, F. Optimal linkage disequilibrium splitting. *Bioinformatics* **38**, 255–256 (2022).
56. Muggeo, V. M. R. Estimating regression models with unknown break-points. *Stat. Med.* **22**, 3055–3071 (2003).
57. King's College London. King's Computational Research, Engineering and Technology Environment (CREATE). (2022) doi:https://doi.org/10.18742/rnvf-m076.