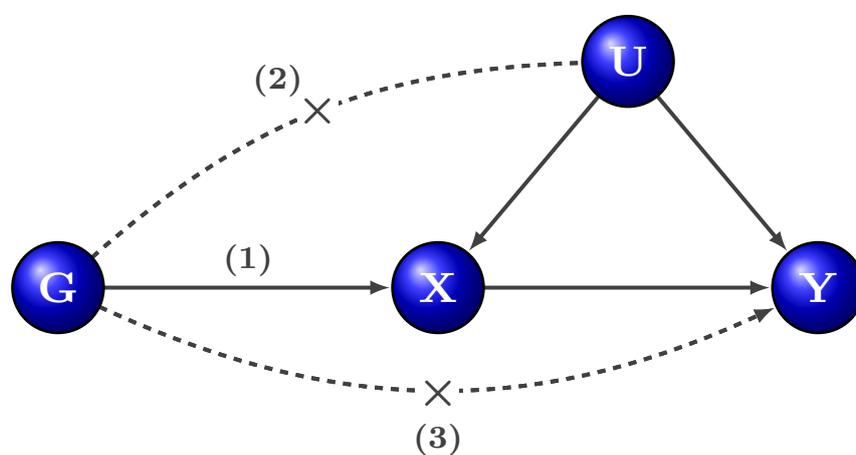
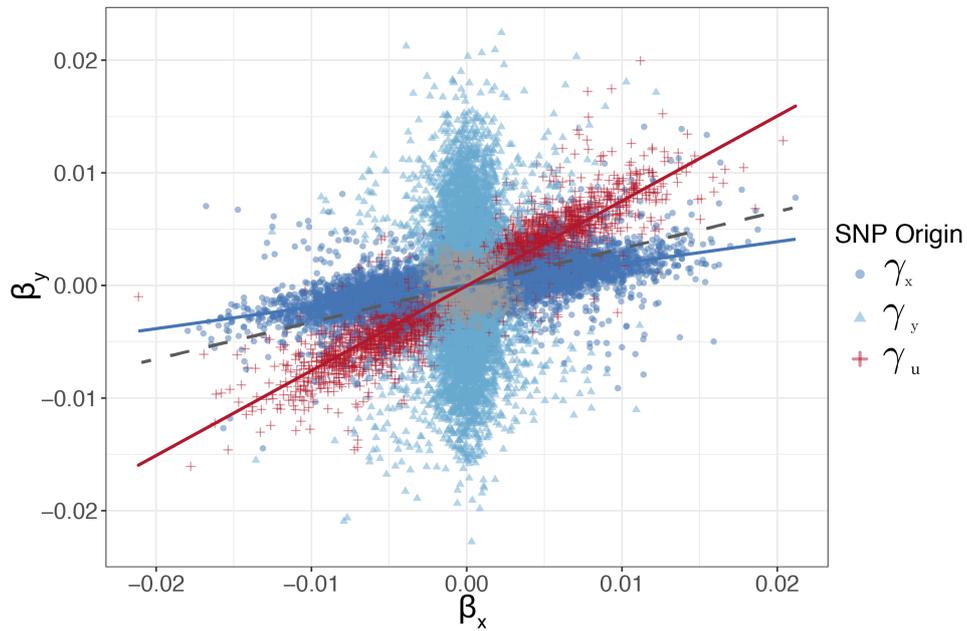


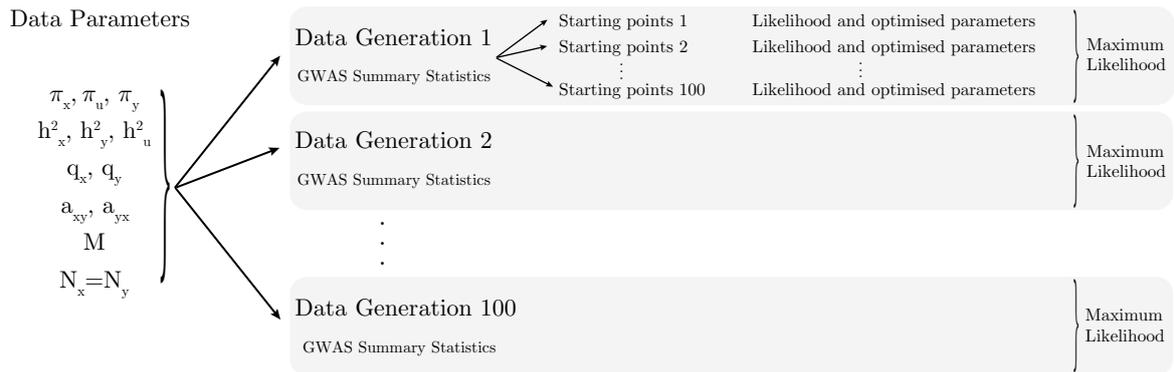
## Supplementary Figures



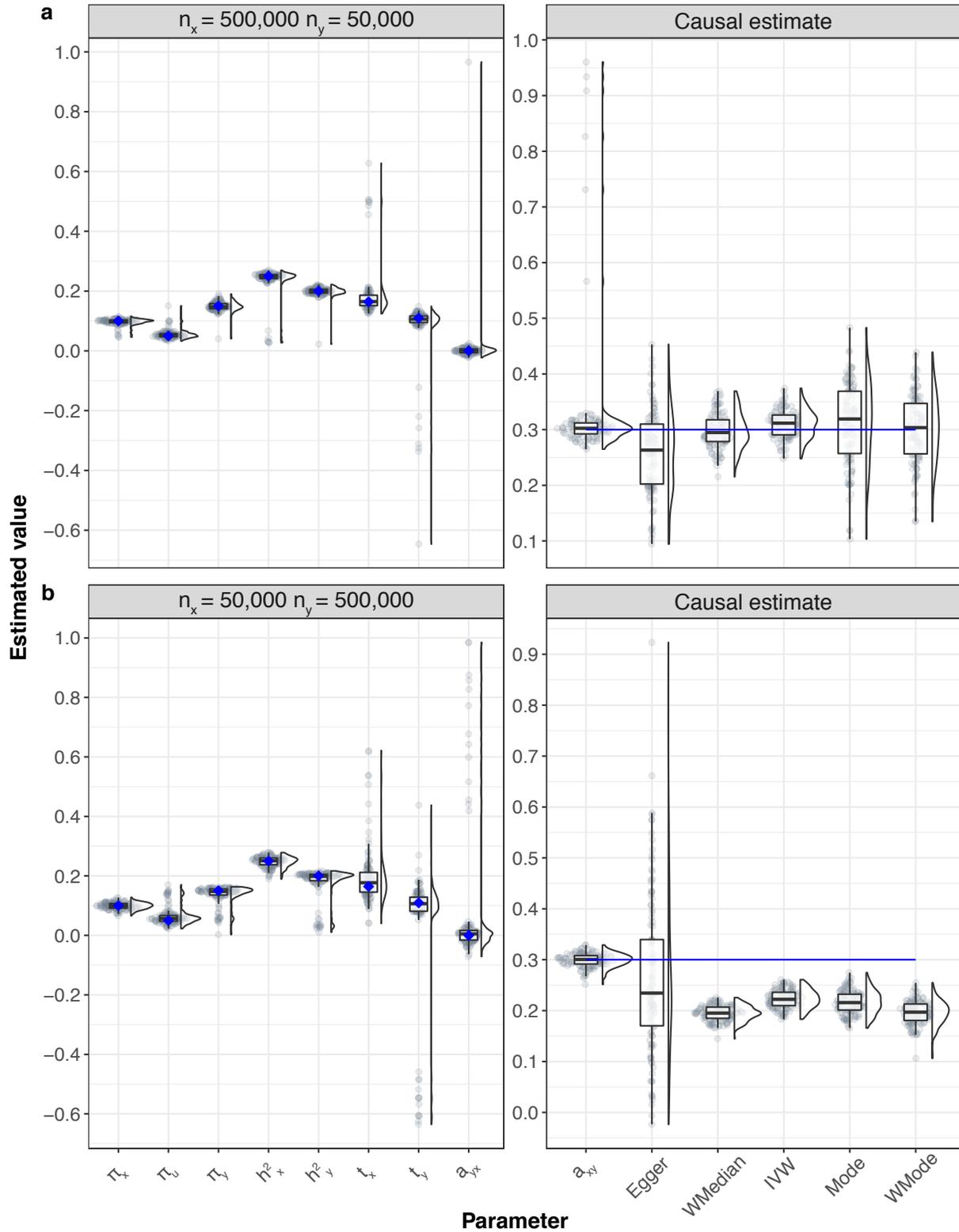
**Figure S1:** Basic assumptions of Mendelian randomisation. (1) Relevance – genetic data, denoted by  $G$ , is robustly associated with the exposure. (2) Exchangeability –  $G$  is not associated with any confounder of the exposure-outcome relationship. (3) Exclusion restriction –  $G$  is independent of the outcome conditional on the exposure and all confounders of the exposure-outcome relationship (i.e. the only path between the instrument and the outcome is via the exposure).



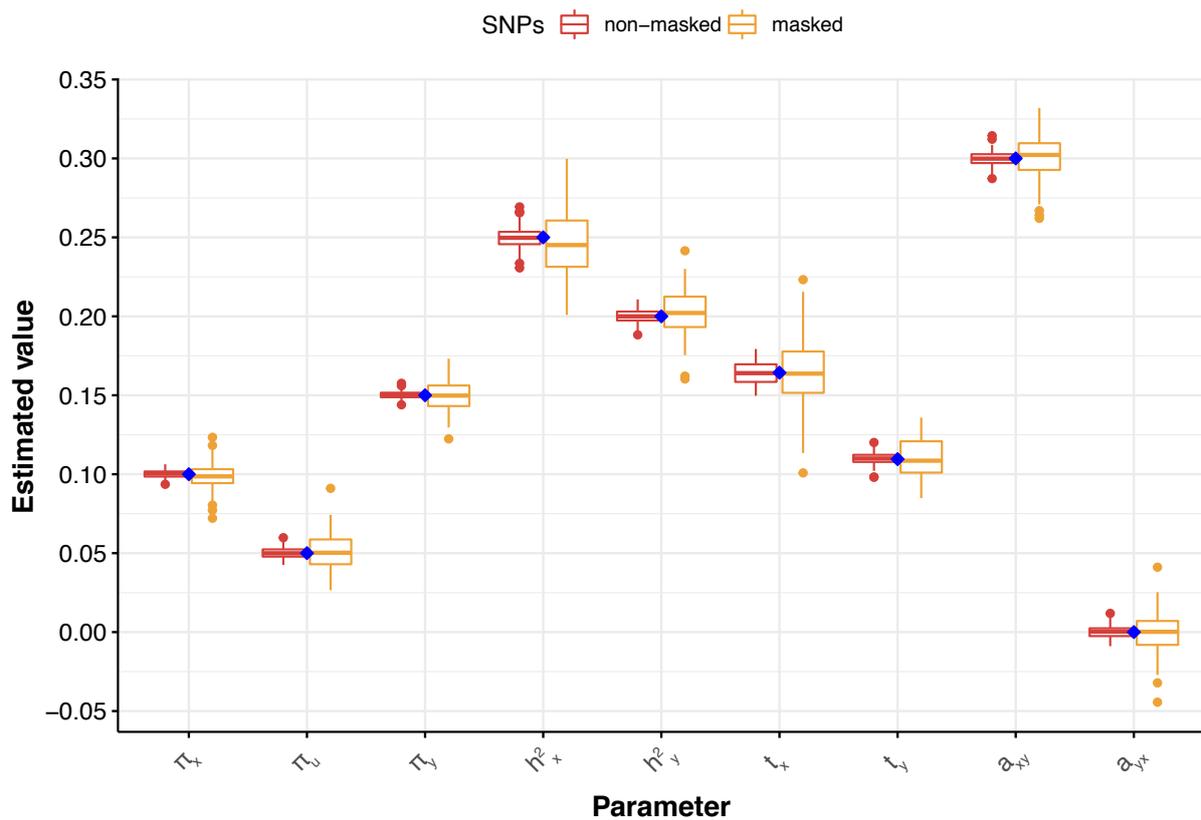
**Figure S2:** A plot of the observed SNP effects on traits  $X$  and  $Y$ , coloured by the strongest effect between the three vectors  $\gamma_x, \gamma_y, \gamma_u$ . SNPs in grey are those with no effect on any of the traits. This illustration shows the distinct clusters that could arise in the presence of a confounder. The dark blue cluster of SNPs represents those that are not in violation of any of the MR assumption, and hence its slope reflects the true causal effect of  $X$  on  $Y$ , while the red cluster of SNPs are those associated with the confounder. The steeper slope of the red cluster of SNPs causes a typical regression line - shown in grey - that represents the causal effect (estimated using conventional MR methods) to be overestimated.



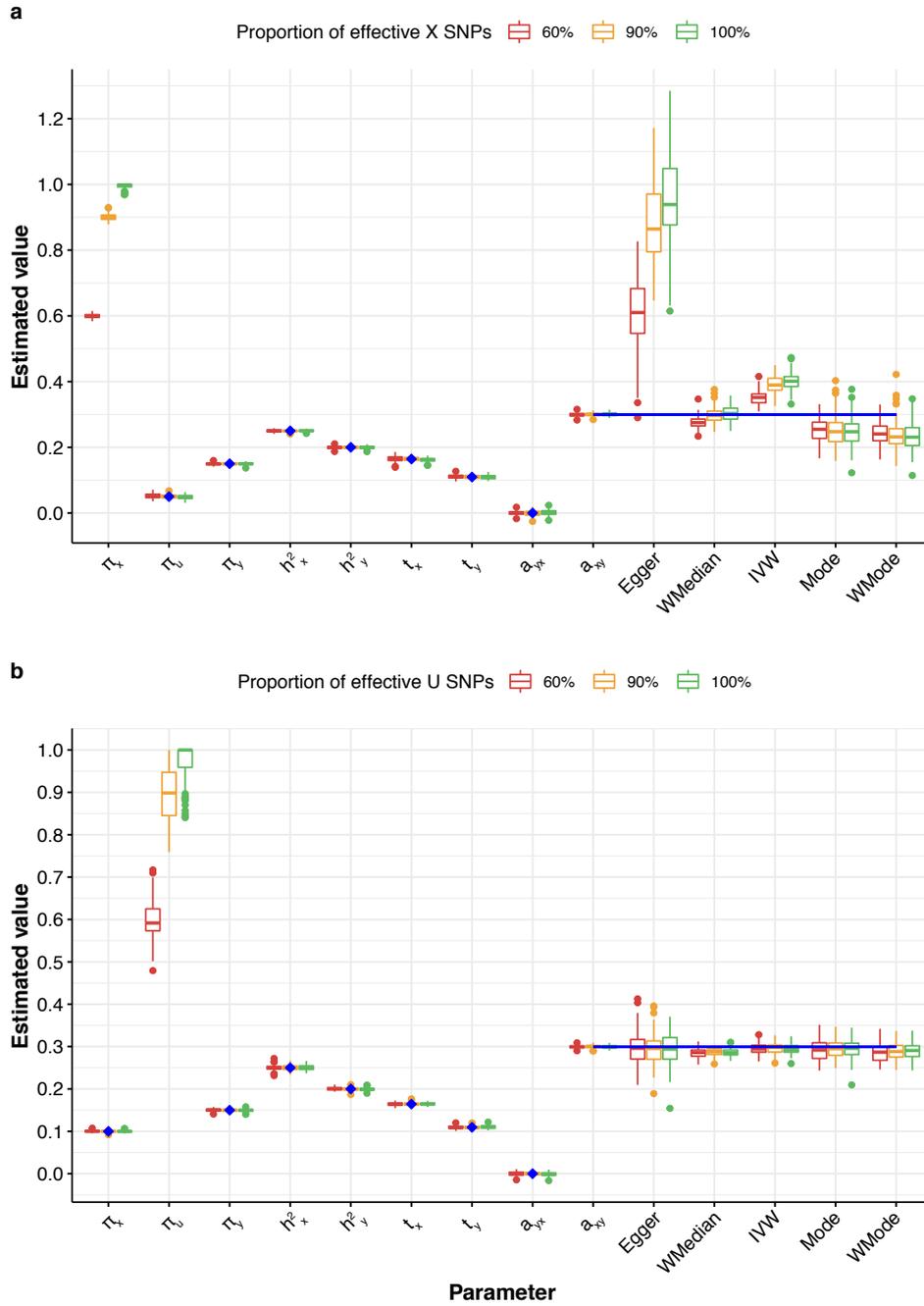
**Figure S3:** A schema showing the workflow of the simulation results. For a single set of parameter settings, 100 different data generations of GWAS summary statistics are created for trait  $X$  and  $Y$ . The summary statistics of a single data generation, as well as the sample size and SNP number, are used in the likelihood optimisation function that is run with 100 different random starting points in order to explore the likelihood surface. A single maximum likelihood and its corresponding estimated parameters are selected to represent the estimates of that data generation. And this is repeated for the other generations. The results for several data generation are often represented in boxplots throughout the paper.



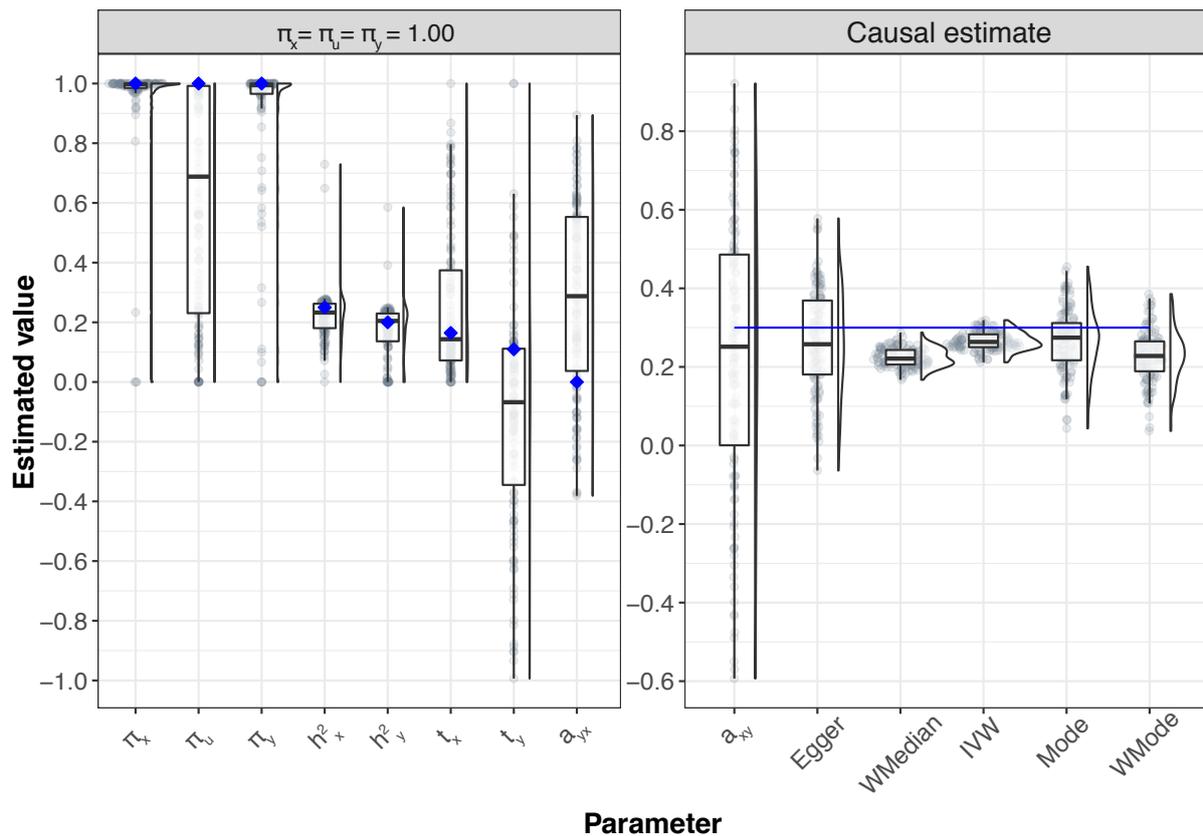
**Figure S4:** Raincloud boxplot representing the distribution of parameter estimates from 100 different data generations. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. The true values of the parameters used in the data generations are represented by the blue dots/lines. In this figure, samples sizes for the two traits differ as such  $n_x = 500,000$  and  $n_y = 50,000$  for panel a, and  $n_x = 50,000$  and  $n_y = 500,000$  for panel b.



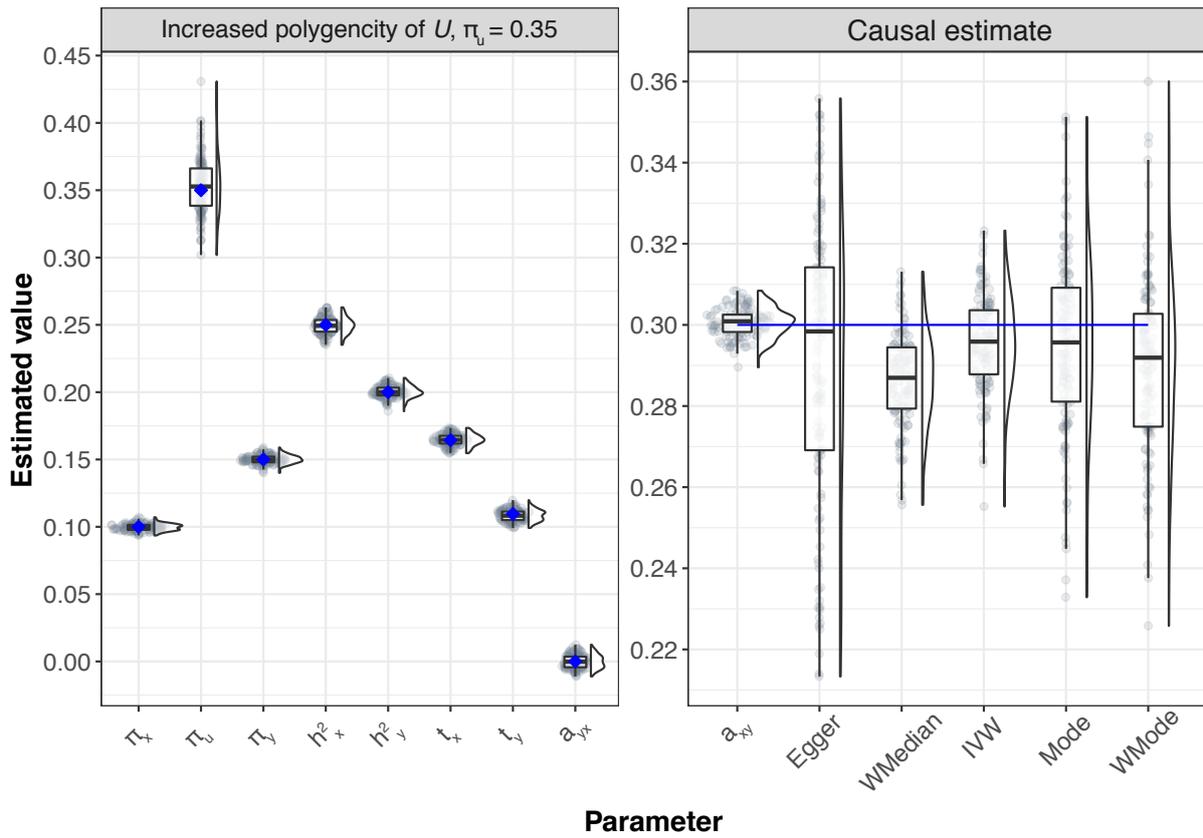
**Figure S5:** Boxplots representing the distribution of parameter estimates from 100 different data generations. The true values of the parameters used in the data generations are represented by the blue dots. The different coloured boxplots represent the masking of the SNPs during the likelihood optimisation process. Red boxplots are the runs that used the full set of SNPs in the optimisation, whereas orange boxplots had a tenfold masking with only 5,000 SNPs being used in the optimisation.



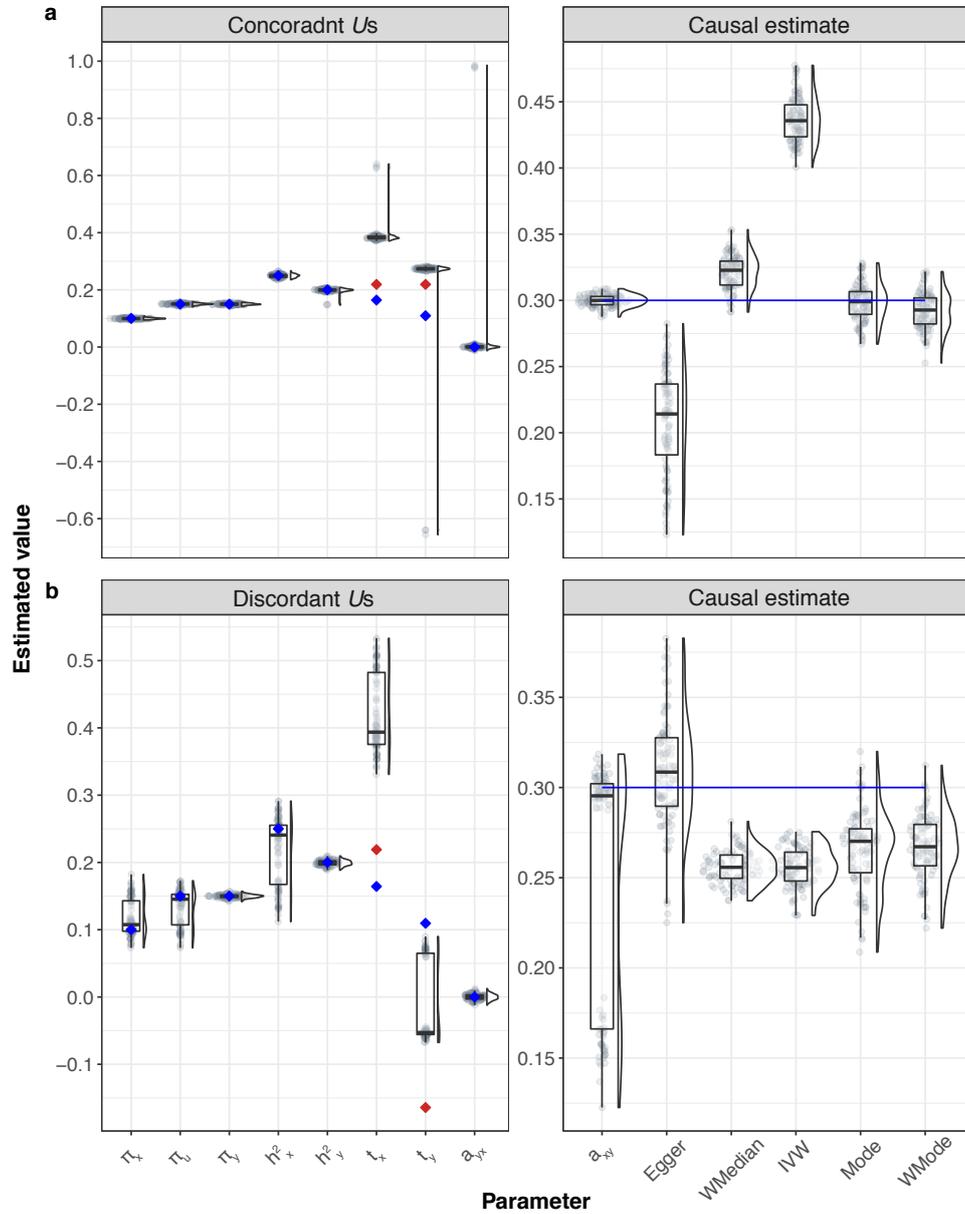
**Figure S6:** Boxplots representing the distribution of parameter estimates from 100 different data generations. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. The true values of the parameters used in the data generations are represented by the blue dots/lines. The different coloured boxplots represent the different proportion of effective SNPs that make up the spike-and-slab distributions of the  $\gamma$  vectors. The red boxplots are those when trait  $X$  (panel  $a$ ) or  $U$  (panel  $b$ ) has 60% effective SNPs, the orange ones are when the proportion of effective SNPs for those two is 90%, and the green boxplots is when they have a 100% effective SNPs.



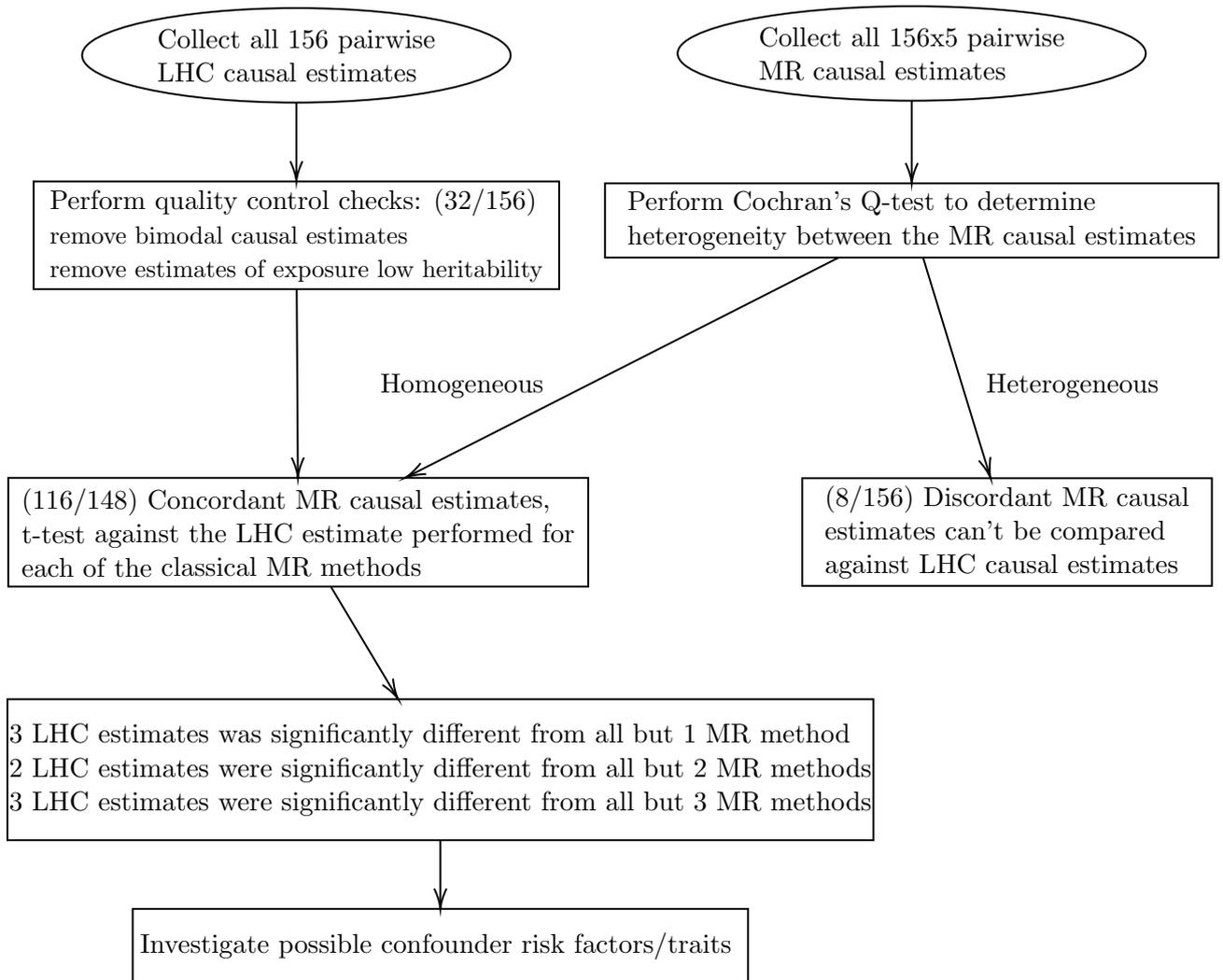
**Figure S7:** Raincloud Boxplot representing the distribution of parameter estimates from 100 different data generations. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. The true values of the parameters used in the data generations are represented by the blue dots/lines. The proportion of effective SNPs that make up the spike-and-slab distributions of the  $\gamma$  vectors in this setting is 100% for all traits  $X, Y$  and  $U$ .



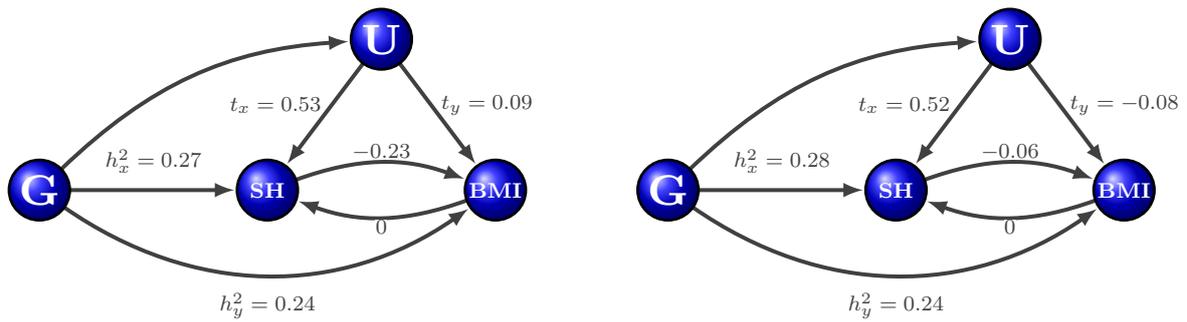
**Figure S8:** Raincloud boxplot representing the distribution of parameter estimates from 100 different data generations. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. The true values of the parameters used in the data generations are represented by the blue dots/lines. In this figure, the polygenicity for  $U$  is increased in the form of higher  $\pi_u = 0.35$ .



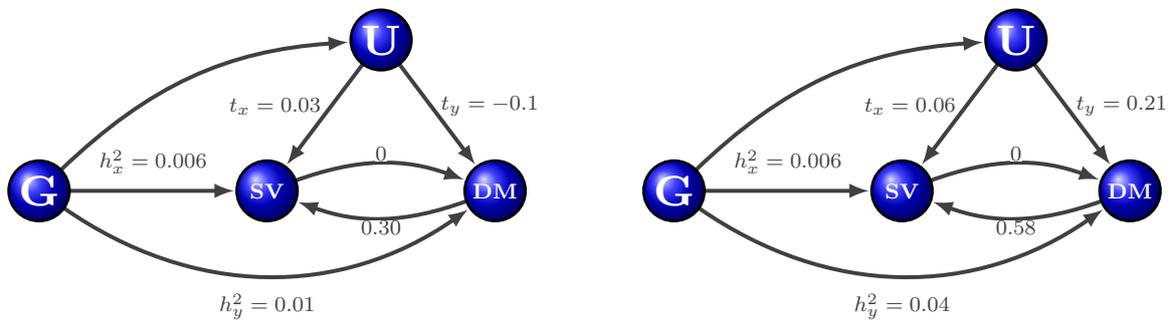
**Figure S9:** Boxplots representing the distribution of parameter estimates from 100 different data generations. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect.. The true values of the parameters used in the data generations are represented by the blue dots/lines. Figure A represents the underlying data generation with two concordant heritable confounders  $U_1$  and  $U_2$  with positive effects on traits  $X$  and  $Y$ , whereas figure B shows data generations with two discordant heritable confounders with  $t_x^{(1)} = 0.16$ ,  $t_y^{(1)} = 0.11$ ,  $t_x^{(2)} = 0.22$ ,  $t_y^{(2)} = -0.16$ .



**Figure S10:** The decision tree used to compare the different causal estimates from standard MR methods to those of LHC.

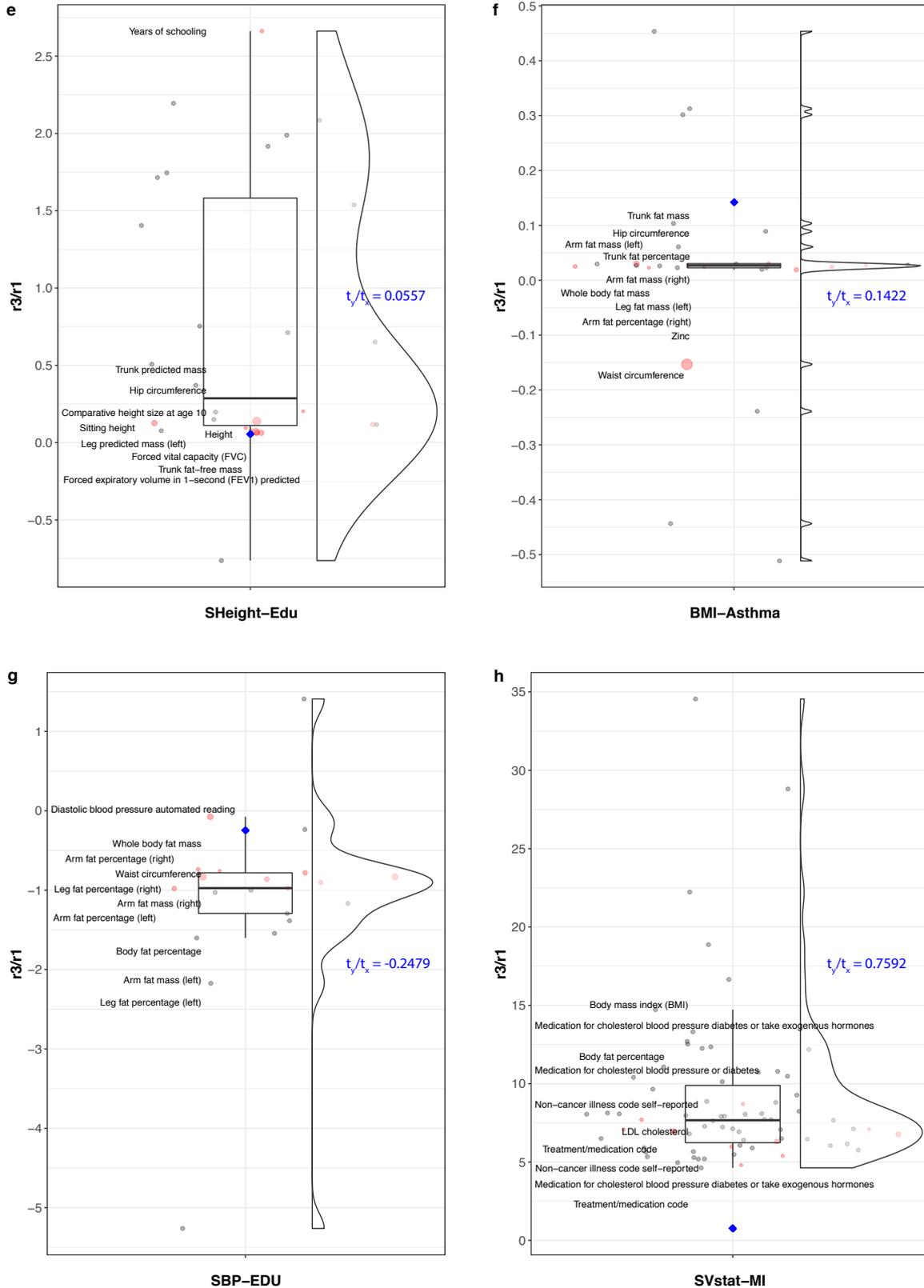


**Figure S11:** The two equally likely models underlying the observed summary statistics in the case of standing height (SH) and its effect on BMI.



**Figure S12:** The two equally likely models underlying the observed summary statistics in the case simvastatin (SV/SVstat) on DM.





**Figure S13:** Confounders' effects obtained from EpiGraphDB plotted as a raincloud with the  $r_3/r_1$  ratio on the y-axis. The blue diamonds represent the  $t_y/t_x$  ratio derived from the LHC model for that trait pair, also reported in blue text. Labelled dots in red and their varying size show the ten largest confounder traits in terms of their absolute effect product on the two traits, whereas grey dots represent the rest of the confounder traits found by EpiGraphDB.

<b>a</b>		MR-Egger			<b>b</b>		Weighted median			<b>c</b>		IVW		
LHC-MR		Sig+	Sig-	nonSig		Sig+	Sig-	nonSig		Sig+	Sig-	nonSig		
	Sig+	4	0	13	Sig+	14	0	3	Sig+	15	0	2		
	Sig-	0	1	9	Sig-	0	5	5	Sig-	1	5	4		
	nonSig	2	1	93	nonSig	8	8	80	nonSig	9	9	78		

<b>d</b>		Simple mode			<b>e</b>		Weighted mode			<b>f</b>		IVW		
LHC-MR		Sig+	Sig-	nonSig		Sig+	Sig-	nonSig	Weighted mode		Sig+	Sig-	nonSig	
	Sig+	8	0	9	Sig+	9	0	8		Sig+	14	0	1	
	Sig-	0	2	8	Sig-	0	2	8		Sig-	0	4	3	
	nonSig	4	1	91	nonSig	6	5	85		nonSig	11	10	80	

**Figure S14:** Cross tables between LHC-MR and various standard MR methods comparing the significance and sign of each respective causal estimate. Panel *f* shows a cross table between the two-least correlated MR methods in terms of their estimates.

## Supplementary Tables

Sample Size	LHC	MR Egger	IVW	Mode	Weighted median	Weighted mode
10,000	0.4	4.44	0.27	0.29	0.27	0.28
50,000	0.11	0.26	0.08	0.1	0.1	0.11
100,000	0.07	0.12	0.05	0.07	0.07	0.08
200,000	0.04	0.07	0.02	0.04	0.04	0.05
300,000	0.01	0.06	0.01	0.03	0.03	0.04
400,000	0.01	0.06	0.01	0.03	0.02	0.03
500,000	0.00	0.06	0.02	0.02	0.02	0.02

**Table S1:** Table reporting the RMSE in estimating the causal effect using different MR methods and LHC-MR with varying sample sizes.

UKBB ID / Data Origin	Trait Name	Abbreviation	Sample Size	PMID
845	Age completed full time education	Edu	240,547	25826379
21001_irnt	Body mass index (BMI)	BMI	359,983	25826379
2443	Diabetes diagnosed by doctor	DM	360,192	25826379
20002_1075	Non-cancer illness code, self-reported: heart attack/myocardial infarction	MI	361,141	25826379
20002_1111	Non-cancer illness code, self-reported: asthma	Asthma	361,141	25826379
2887	Number of cigarettes previously smoked daily	PSmoke	84,456	25826379
20022_irnt	Birth weight	BWeight	205,475	25826379
50_irnt	Standing height	SHeight	360,388	25826379
4080	Systolic blood pressure, automated reading	SBP	340,159	25826379
20003_1140861958	Treatment/medication code: simvastatin	SVstat	361,141	25826379
GLGC	LDL Cholesterol	LDL	89,138	24097068
GLGC	HDL Cholesterol	HDL	93,561	24097068
UKBB + CARDIoGRAMplusC4D	Coronary Artery Disease	CAD	380,831	29212778

**Table S2:** Details of the origin study of each trait, its abbreviation used in this paper, the sample size of the study for that trait, as well as the PubMed article ID.

Exposure	Outcome	Lower CI ratio	mean ratio	Upper CI ratio
Edu	BMI	0.849	0.909	1.088
SHeight	BMI	0.641	0.832	1.189
SBP	SHeight	0.612	0.685	0.757
SHeight	Edu	-1.023	0.528	2.141
SVstat	DM	0.156	0.191	0.233
SBP	DM	0.025	0.070	0.110

**Table S3:** Table reporting the ratio of the genetic covariance attributed to the bi-directional causal effect, to the genetic covariance attributed to heritable confounding. For each trait pair, parameter estimates from LHC-MR were used to calculate those two values from Eq. 2, by sampling 1,000 times from distributions provided by LHC-MR (parameter estimates and covariance matrices). The mean of the 1,000 computed ratios, as well as the lower and upper bounds of the 95% confidence interval (CI) are reported above.