



Published in final edited form as:

*Nat Genet.* 2020 July ; 52(7): 740–747. doi:10.1038/s41588-020-0631-4.

## Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics

Jean Morrison<sup>1</sup>, Nicholas Knoblauch<sup>1</sup>, Joseph H. Marcus<sup>1</sup>, Matthew Stephens<sup>1,2</sup>, Xin He<sup>1</sup>

<sup>1</sup>Department of Human Genetics, University of Chicago, Chicago, Illinois, USA.

<sup>2</sup>Department of Statistics, University of Chicago, Chicago, Illinois, USA.

### Abstract

Mendelian randomization (MR) is a valuable tool for detecting causal effects using genetic variant associations. Opportunities to apply MR are growing rapidly with the number of genome-wide association studies (GWAS). However, existing MR methods rely on strong assumptions that are often violated, leading to false positives. Correlated horizontal pleiotropy, which arises when variants affect both traits through a heritable shared factor, remains a particularly challenging problem. We propose a new MR method, Causal Analysis Using Summary Effect Estimates (CAUSE), that accounts for correlated and uncorrelated horizontal pleiotropic effects. We demonstrate in simulations that CAUSE avoids more false positives induced by correlated horizontal pleiotropy than other methods. Applied to traits studied in recent GWAS, we find that CAUSE detects causal relationships with strong literature support and avoids identifying most unlikely relationships. Our results suggest that shared heritable factors are common and may lead to many false positives using alternative methods.

---

Inferring causal relationships between traits is important for understanding the etiology of disease and designing new treatments. Randomized trials are considered the gold standard for inferring causality but are expensive and sometimes impossible. Observational studies are often more feasible, but observational associations may be biased by confounding and reverse causality.

Mendelian randomization (MR) is an approach to studying causal relationships using trait associations with genetic variants. MR can be performed using only GWAS summary statistics, making it potentially an extremely valuable technique with wide applicability. The key idea of MR is to treat genotypes as naturally occurring “randomizations”<sup>1,2,3,4</sup>. Suppose

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\* xinhe@uchicago.edu.

#### Author Contributions

J.M. and X.H. conceived and designed the model. J.M. designed the algorithm, implemented the software, conducted analyses, and performed simulations. J.M., X.H., and M.S. contributed to writing the manuscript. J.H.M. contributed to preparing GWAS data. N.K. contributed to software development and data preparation and computed LD.

#### Competing Interests Statement

The authors declare no competing interests.

we are interested in the causal effect,  $\gamma$ , of  $M$  (for “Mediator”) on  $Y$ . If a variant  $G_j$  affects  $M$  and does not affect  $Y$  through pathways not mediated by  $M$  (see Figure 1a), then  $G_j$  can be viewed as an unconfounded proxy for  $M$ . Under these assumptions,

$$\beta_{Y,j} = \gamma\beta_{M,j} \quad \#(1)$$

where  $\beta_{Y,j}$  is the association of  $G_j$  with  $Y$ ,  $\beta_{M,j}$  is the association of  $G_j$  with  $M$ . This relationship is the core of many MR methods, including the commonly used inverse variance weighted (IVW) regression, which regresses estimates of  $\beta_{Y,j}$  on  $\beta_{M,j}$  for selected variants strongly associated with  $M$ .

A variant that affects both  $M$  and  $Y$  through a pathway not mediated by a causal effect is termed *horizontally pleiotropic*. Horizontal pleiotropy leads to false positives in methods that assume the relationship in Equation (1) and is unfortunately common<sup>6,7</sup>. Accounting for horizontal pleiotropy has therefore been a major focus of ongoing MR research<sup>4</sup>. It is helpful to distinguish between two types of horizontal pleiotropy: *uncorrelated pleiotropy*, where horizontal effects on  $Y$  are uncorrelated with effects on  $M$ , and *correlated pleiotropy*, where horizontal effects on  $Y$  are correlated with effects on  $M$ . Uncorrelated pleiotropy occurs when a variant affects  $Y$  and  $M$  through separate mechanisms, whereas correlated pleiotropy occurs when a variant affects  $Y$  and  $M$  through a shared heritable factor,  $U$ , such as a shared process or pathway (Fig. 1a). Both types of pleiotropy may occur for any pair of traits.

Of the two types, uncorrelated pleiotropy is easier to account for. If it is mean zero, it only adds noise to the relationship in Equation (1). Multiple methods have been proposed to deal with this, including Egger regression<sup>8,9</sup>, which allows for directional (non-mean zero) uncorrelated pleiotropy, and methods that rely on outlier removal, including GSMR<sup>10</sup> and MR-PRESSO<sup>7</sup>. Correlated pleiotropy results in correlation between  $\beta_{M,j}$  and  $\beta_{Y,j}$  for a subset of variants. This leads to false positives if it is not accounted for. Genetic correlation has been found between many pairs of traits including some that are unlikely to be causally linked, suggesting that correlated pleiotropy is common and may be an important source of MR false positives<sup>11–14</sup>.

If a suspected shared factor is measured, false positives can be avoided by eliminating variants associated with the shared factor or adjusting for it using multivariable MR<sup>15</sup>. Two methods, the weighted median<sup>16</sup> and the weighted mode<sup>17</sup>, allow for some correlated pleiotropy. In their unweighted versions, the median and mode estimators assume that fewer than half of variants display horizontal pleiotropy and that the modal horizontal pleiotropic effect is zero, respectively. Another method, LCV<sup>14</sup>, does not directly test for a causal effect, but estimates the “genetic causality proportion”, with higher magnitude estimates suggesting a causal effect and lower magnitude estimates suggesting correlated pleiotropy.

We present a new MR method that accounts for both uncorrelated and correlated pleiotropy, Causal Analysis Using Summary Effect Estimates (CAUSE), using genome-wide summary statistics. The intuition behind CAUSE is that a causal effect of  $M$  on  $Y$  leads to correlation between  $\beta_{M,j}$  and  $\beta_{Y,j}$  for all variants with non-zero effect on  $M$ , while a shared factor induces correlation for only a subset of  $M$  effect variants (Fig. 1b). CAUSE uses this

distinction to differentiate causal effects from correlated pleiotropy. Unlike many existing methods, CAUSE incorporates information from all variants, rather than only those most strongly associated with  $M$ . This can improve power when the GWAS for trait  $M$  is underpowered. CAUSE provides a test-statistic, an estimate of the causal effect, and a variant-level summary indicating how each variant contributes to the overall test, and which variants are likely to be acting through a shared factor.

We demonstrate in simulations that CAUSE makes fewer false detections in the presence of correlated pleiotropy than existing methods. In applications to GWAS data, CAUSE identifies plausible causal relationships while avoiding many likely false positives. CAUSE fills a gap in existing methodology by accounting for correlated pleiotropy created by unknown or unmeasured shared factors.

## Results

### CAUSE models uncorrelated and correlated pleiotropy

CAUSE assesses whether genome-wide effect estimates for  $M$  and  $Y$  are consistent with a causal effect. In a simple MR analysis, variants with strong associations with  $M$  are selected and assumed to follow the causal diagram in Figure 1a. In our proposal, illustrated in Figure 1c, we use all variants but assume that most have no effect on either trait. We assume that the majority of variants that do affect  $M$  follow the left-hand causal diagram in Figure 1c, which is the same as the diagram in Figure 1a with the addition of an uncorrelated pleiotropic effect. We allow a small proportion,  $q$ , of  $M$  effect variants to follow the right-hand causal diagram in Figure 1c, acting on  $M$  and  $Y$  through an unobserved heritable shared factor,  $U$ . Under this model, the relationship between  $\beta_{Y,j}$  and  $\beta_{M,j}$  is:

$$\beta_{Y,j} = \underbrace{\gamma\beta_{M,j}}_{\text{causal effect}} + \underbrace{Z_j\eta\beta_{M,j}}_{\text{correlated pleiotropy}} + \underbrace{\theta_j}_{\text{uncorrelated pleiotropy}}, \quad \#(2)$$

where  $Z_j$  is an indicator that is 1 if  $G_j$  acts on  $M$  through  $U$  and 0 otherwise, and  $\eta$  is the effect of  $U$  on  $Y$ . This relationship is an extension of Equation (1) that includes terms allowing for both types of horizontal pleiotropic effects. We assume that  $q$  is small, so  $Z_j$  is equal to 0 for most variants (see Methods).

Equation (2) captures the patterns in Figure 1b. If  $\gamma = 0$  and  $q = 0$  (no causal effect and no shared factor), then  $\beta_{Y,j}$  and  $\beta_{M,j}$  are uncorrelated for all  $j$ . If there is a shared factor and no causal effect ( $\gamma = 0$ ,  $q$  and  $\eta$  non-zero), then  $\beta_{Y,j}$  and  $\beta_{M,j}$  are correlated for variants with  $Z_j = 1$  (Fig. 1b, left). If there is a causal effect ( $\gamma \neq 0$ ),  $\beta_{Y,j}$  and  $\beta_{M,j}$  are correlated for all variants (Fig. 1b, right). Including all variants allows us to model uncertainty about variant effects on  $M$  and gain information from weakly associated variants (see Methods). For computational simplicity, we use a likelihood for independent variants and prune variants for LD before estimating posterior distributions and computing test statistics (see Methods and Supplementary Note, Section SN4).

We assess whether GWAS summary statistics for the two traits are consistent with a causal effect by comparing two nested models. The *sharing model* has  $\gamma$  fixed at 0, allowing for

horizontal pleiotropic effects but no causal effect, and the *causal model* in which  $\gamma$  is a free parameter. We compare the models using the expected log pointwise posterior density<sup>18</sup> (ELPD), a Bayesian model comparison approach that estimates how well the posterior distributions of a particular model are expected to predict a new set data (see Methods). We produce a one-sided  $p$ -value testing that the sharing model fits the data at least as well as the causal model. If this hypothesis is rejected, we conclude that the data are consistent with a causal effect.

The right-hand causal diagram in Figure 1c is related to the model used in the latent causal variable (LCV) method<sup>14</sup> (see Supplementary Note, Section SN5). Unlike other MR methods, LCV does not provide a test for a causal effect. LCV considers the proportion of heritability of each trait that is mediated by a shared factor, summarized as the “genetic causality proportion” (GCP), which ranges from  $-1$  to  $1$ . A causal model has a GCP of  $1$  or  $-1$ , depending on the direction of effect. O’Connor and Price<sup>14</sup> consider a large magnitude estimated GCP more likely to be causal and use a threshold of  $0.6$  to identify suggestive pairs.

### CAUSE can distinguish causality from correlated pleiotropy in simulations

We simulate summary statistics with realistic LD patterns to assess CAUSE in a variety of scenarios and compare performance with other MR methods. Both traits are simulated with a polygenic trait architecture—an average of  $1,000$  variants with normally distributed effects contribute to a total heritability of  $0.25$ . The power of trait  $M$  and  $Y$  GWAS can influence performance of all methods, so we consider low power and high power (median  $8$  and  $107$  genome-wide significant loci) settings for both traits. Power is controlled by adjusting the simulated GWAS sample size and genome-wide significance is defined as  $p < 5 \cdot 10^{-8}$ . We generate effects on trait  $Y$  from the relationship in Equation (2) and then simulate effect estimates from true effects and LD structure using results of Zhu and Stephens<sup>19</sup> (see Methods).

We compare CAUSE to six MR methods: IVW regression using multiplicative random effects<sup>20</sup>, Egger regression with random effects<sup>8</sup>, GSMR<sup>10</sup>, MR-PRESSO<sup>7</sup>, the weighted median<sup>16</sup>, and the weighted mode<sup>17</sup>. IVW regression, Egger regression, GSMR and MR-PRESSO assume that no variants exhibit correlated pleiotropy. The weighted median and weighted mode methods each allow some horizontal pleiotropy of any kind. All six alternative methods require selection of variants with strong evidence of association with trait  $M$ . We follow common practice selecting variants with  $p < 5 \cdot 10^{-8}$  for association with  $M$  and pruning for LD so that no pair of variants has pairwise  $r^2 > 0.1$ .

We first evaluate the robustness of each method to correlated pleiotropy by simulating data with no causal effect and a proportion  $q = 0$  to  $0.5$  of variants acting through a shared factor, considering weak and strong shared factor effects (Fig. 2a). CAUSE controls the false positive rate in the presence of correlated pleiotropy ( $q > 0$ ) better than other methods except for Egger regression and the weighted mode. CAUSE makes more false detections when  $q$  is large, the confounder effect is large, and both GWAS have high power. This is expected because the data pattern resulting from a high proportion of shared variants is similar to the pattern that results from a causal effect. Although in these settings CAUSE has a higher false

positive rate when the trait  $MGWAS$  has higher power, as sample size and GWAS power continue to increase, CAUSE's false positive rate eventually drops to zero, indicating that, asymptotically, CAUSE is able to determine the correct model (Supplementary Note, Section SN6.4).

We next compare the power of each method when there is a true causal effect of  $M$  on  $Y$  and no shared factor (Fig. 2b). CAUSE has substantially better power than Egger regression and the modal estimator, the two methods that control the false positive rate for high levels of correlated pleiotropy. CAUSE has somewhat lower power than other methods in most settings. However, when the trait  $MGWAS$  has low power and the trait  $YGWAS$  has high power, CAUSE can achieve better power than other methods for larger causal effects. This is a result of using all variants, rather than only those reaching genome-wide significance. The posterior median of  $\gamma$  can be taken as a point estimate of the causal effect. This estimate tends to be shrunk slightly towards zero compared to alternative methods but has a substantially lower mean squared error than estimates obtained using the weighted mode or Egger regression (Supplementary Note, Section SN6.2). In simulations with both a causal effect and a shared factor, CAUSE maintains similar power when  $\eta$  has a smaller magnitude than  $\gamma$  but loses power when  $\eta$  is large and has opposite sign to  $\gamma$  (Supplementary Note, Section SN6.3).

Although in many settings CAUSE has both lower power and lower false positive rate than other methods, the difference between CAUSE and other methods is not simply calibration. In Figure 2c, we compare the trade-off between power and false positive rate for CAUSE and other methods, additionally including the LCV GCP estimate. CAUSE is better able to distinguish a causal scenario from a non-causal scenario with 30% correlated pleiotropy. This pattern is consistent for levels of correlated pleiotropy between 0% and 50% (Extended Data Fig. 1 and Supplementary Table 1). Additionally, CAUSE is better or equal to other methods at discriminating scenarios with a causal effect and a shared factor from those with only a shared factor (Supplementary Note, Section SN6.3).

### CAUSE reduces false positives due to reverse causality

Many MR methods obtain false positives when there is a causal effect of  $Y$  on  $M$ , but a test is performed for an effect of  $M$  on  $Y$  (reverse causal effects). This occurs because if  $Y$  affects  $M$ , some variants will have correlated effects on both traits. However, for most polygenic traits, there will be some variants that affect  $M$  through pathways independent of  $Y$  and therefore do not exhibit this correlation. By modeling a subset of correlated pleiotropic variants, CAUSE can avoid false positives due to reverse causality as long as only a proportion of  $M$  effect variants act through  $Y$ .

To verify this expectation, we simulate data with a causal effect of  $Y$  on  $M$  and no shared factor and test for an effect of  $M$  on  $Y$  using CAUSE and other methods. In all scenarios, the expected heritability of both  $M$  and  $Y$  is 0.25. For trait  $Y$ , 1000 variants with normally distributed effects explain this heritability. The causal effect of  $Y$  on  $M$  explains 20% of the trait  $M$  heritability while the rest comes from a set of 250, 500, or 1,000 variants (see Methods). CAUSE and the modal estimator both avoid false positives in all settings while the other methods do not (Fig. 3). More false positives from other methods occur when a

higher proportion of  $M$  effect variants act through  $Y$  and therefore have correlated effects. Some false positives from alternative methods could be avoided by carefully filtering variants, for example using Steiger filtering<sup>21</sup>.

Although CAUSE is often able to avoid false positives from reverse causality, it may still find evidence of causality in both directions for some pairs of traits. This can occur for several reasons. First, there may be a causal effect in one direction that accounts for nearly all of the heritability of the downstream/child trait. Second, there may be causal effects in both directions. Third, the two traits may be very closely related and share nearly all genetic variants despite having no causal relationship.

### Identifying causal risk factors of common diseases

We use CAUSE and other methods to test for causal effects of twelve possible risk factors for cardio-metabolic diseases on coronary artery disease (CAD), stroke, and type 2 diabetes (T2D) (Supplementary Table 2). In order to increase the number of tested relationships that are unlikely to be causal, we also test for effects of each risk factor on asthma, though not all of these relationships are negative controls. Focusing on well-studied risk factors and diseases allows us to compare results of each method with evidence from scientific literature.

Table 1 summarizes  $p$ -values and estimated effect direction for each method as well as GCP estimates from LCV and genetic correlation estimates from LD-score regression (complete results in Extended Data Fig. 2 and Supplementary Table 3). Full lists of variants used for each MR method are provided in Supplementary Table 4. We classified trait pairs in Table 1 by evidence of causality from existing literature (see Supplementary Note, Section SN7) and ordered them by genetic correlation  $p$ -value. CAUSE obtains a  $p$ -value less than 0.05 for 8 of 9 relationships in the considered causal category, and for 4 of 10 relationships with literature support. Of the remaining pairs in these two categories, 4 are not identified as causal by any method. The remaining negative results from CAUSE are discussed in Supplementary Note, Section SN8. Using LCV, very few pairs of traits have estimated GCP larger than 60%, the threshold used by O'Connor and Price<sup>14</sup> as suggestive of a causal relationship. Variant estimates for smoking and CAD, a strong causal effect detected by all methods except LCV, are shown in Figure 4a colored by individual variant contribution to the CAUSE test statistic. Although there is some heterogeneity in effect size correlation, the large number of variants with correlated effects provide enough evidence for CAUSE to reject the sharing model in favor of the causal model.

To assess false positives, we consider traits in the non-causal and implausible categories, and those in the unknown category that have no significant genetic correlation. CAUSE identifies fewer of these pairs as causal than any other method. All methods except Egger regression detect an effect of HDL cholesterol on CAD risk, which is likely to be an error<sup>22</sup>. This example represents a limitation of CAUSE, as it is possible to obtain false positives when a high proportion of variants are shared. In this case, CAUSE estimates that 54% of variants act through a shared factor.

LDL cholesterol and T2D is an example of a pair of traits identified by other methods but not by CAUSE (Fig. 4b). The true relationship between this pair is unknown. Levels of LDL cholesterol are often elevated in individuals with T2D, an effect that is typically considered a consequence of insulin insensitivity<sup>23</sup> (see Supplementary Note, Section SN7). Clinical trials of statins, which reduce LDL cholesterol, show an increased risk of T2D with statin use<sup>24,25</sup>. However, statins and their target enzyme affect multiple systems, and it remains unknown whether statin effects on T2D are mediated by LDL cholesterol or other mechanisms<sup>26</sup> (see Supplementary Note, Section SN7). The effects estimated by IVW regression and other methods are negative, opposite of observational associations, but consistent with previous findings of Fall et al.<sup>27</sup>. CAUSE does not reject the sharing model ( $p = 0.32$ ) and estimates that 13% of LDL variants act through a shared factor.

We expect that most effects detected in the opposite direction, of diseases on potential mediators are false positives. Testing opposite direction effects therefore gives us another opportunity to evaluate the robustness of methods to false positives. Additionally, as discussed previously, a reciprocal result for CAUSE can arise from both causal and non-causal scenarios, so it is important to consider tests in both directions. CAUSE identifies five reverse pairs at a  $p$ -value threshold of 0.05, fewer than any other method. IVW, Egger regression, the weighted median, the weighted mode, and MR-PRESSO identify 15, 6, 11, 7, and 17 reverse effects respectively (Supplementary Table 5 and Extended Data Fig. 3). The case of CAD and LDL cholesterol provides an example of how CAUSE is able to avoid false positives from reverse effects (Fig. 4c).

All of the mediators in this analysis have high-powered GWAS with many genome-wide significant variants. Such studies are not always available, and simulations suggest that CAUSE may have a power advantage over other methods if the trait  $MGWAS$  is underpowered. In order to explore the effect of GWAS power, we compared results using an earlier GWAS for blood pressure with a smaller sample size<sup>28</sup>. With these data, CAUSE is able to detect effects of blood pressure on CAD that are missed by IVW and Egger regression (Supplementary Note, Section SN9). This suggests that CAUSE may be able to boost power to detect causal effects of mediators with underpowered GWAS.

### Links between immune-mediated disease and blood cell counts

There are a growing number of GWAS for intermediate or molecular traits. Relationships between these traits and clinical outcomes are appealing targets for MR studies because molecular traits may be accessible as drug targets or suggest physiological pathways contributing to disease. We use CAUSE and other MR methods to search for evidence of causal relationships between 13 measures of blood cell composition and five immune-mediated diseases (Supplementary Table 6). Results are summarized in Figure 5 and Supplementary Table 7 (full results are shown in Supplementary Tables 8 and 9, and variants used are shown in in Supplementary Table 10).

Blood cell composition traits can be divided into red blood cell, white blood cell, and platelet traits. Although there is less literature information to guide us, we expect that effects of white blood cell traits on immune-mediated disorders are more plausible than effects of red blood cell or platelet traits.

CAUSE identifies only 4 pairs of traits as consistent with a causal effect at  $p < 0.05$ . The two most significant effects are positive effects of eosinophil count on asthma and allergy risk ( $p = 1.9 \cdot 10^{-5}$  and  $4.0 \cdot 10^{-11}$ , respectively). Eosinophils have proven to be promising drug targets for treating asthma and allergic disease<sup>29</sup>. All other methods, including Egger regression and the weighted mode, obtain many more positive results than CAUSE, including likely false positive effects of red blood cell and platelet traits. Some of these results suggest that there are cases when methods that down-weight or remove outliers may be less robust than IVW regression. For example, an effect of immature reticulocyte fraction (IRF, a red blood cell trait) and inflammatory bowel disease (IBD) is found only by the weighted median ( $p = 1.2 \cdot 10^{-5}$ ), the weighted mode ( $p = 2.6 \cdot 10^{-5}$ ), and MR-PRESSO ( $p = 0.0015$ ) (Fig. 4d). There is no visible correlation in variant effect sizes, and CAUSE as well as the simpler IVW regression obtain non-significant results ( $p = 0.32$  and  $0.16$ , respectively). By removing or down-weighting the contribution of some variants, putatively robust methods can obtain false positives by discounting conflicting evidence.

## Discussion

We have introduced CAUSE, a new approach to MR analysis that accounts for uncorrelated and correlated pleiotropy. Compared to existing methods, CAUSE is able to reduce false positives and, in some circumstances, increase power. Previous authors have identified uncorrelated horizontal pleiotropy as a pervasive phenomenon adversely impacting MR analyses<sup>7</sup>. Our results demonstrate that correlated horizontal pleiotropy is also common and may explain a substantial fraction of positive MR. Alternative MR methods perform best when information is available to guide removal of pleiotropic variants. For example, in analyzing HDL cholesterol and heart disease, Voight et al.<sup>22</sup> avoid a false positive by discarding variants associated with other lipid and metabolic traits. CAUSE does not require such selection and performs well in an automated manner, making it well suited for applications where shared factors are unknown or unmeasured.

Caution must be used when interpreting results of CAUSE, as well as any other MR method. CAUSE tests that the GWAS summary statistics for  $M$  and  $Y$  are consistent with a model where every  $M$  effect variant has a correlated effect on  $Y$ . This pattern occurs when  $M$  has a causal effect on  $Y$  but can also occur in other circumstances. Notably, if most of the heritable variation of two traits is mediated by the same unobserved process, we expect to observe this pattern in both directions. It is therefore good practice to test trait pairs in both directions<sup>13</sup>. We also note that, while CAUSE is more robust than alternatives, it still has a high false positive rate when a shared factor accounts for a large proportion of the trait  $M$  variants. A low  $p$ -value from CAUSE (or any MR method) should not be regarded as proof of a causal effect. Instead, it is an indicator that the summary statistics for the two traits are consistent with a causal effect. The power and false positive rate of CAUSE can be affected by the prior on  $q$  (Supplementary Note, Section SN6.1). We have chosen a default prior that gives good performance over a range of settings. However, it may also be good practice to conduct sensitivity analyses using more and less stringent priors.

CAUSE has several limitations that provide interesting future directions. First, it is not possible to account for known shared factors when they are measured. Second, CAUSE



models only a single unobserved shared factor, so it may not fully account for shared genetic components between two traits of interest. This problem is partially alleviated by the flexibility of the empirical effect size prior distribution (see Methods). Finally, CAUSE prunes variants for LD rather than explicitly modeling variant correlation. This ensures that problem is computationally tractable, but may lead to loss of information.

## Methods

### CAUSE model for GWAS summary statistics

We use effect estimates and standard errors measured in GWAS of  $M$  and  $Y$  (summary statistics) to evaluate evidence of a causal effect of  $M$  on  $Y$ . Let  $(\hat{\beta}_{M,j}, \hat{s}_{M,j})$  and  $(\hat{\beta}_{Y,j}, \hat{s}_{Y,j})$  be effect estimates and standard errors at variant  $G_j$  ( $j = 1, \dots, p$ ) for traits  $M$  and  $Y$ , respectively. Let  $\beta_{M,j}$  and  $\beta_{Y,j}$  be the true marginal associations of  $G_j$  with  $M$  and  $Y$ . We model effect estimates as normally distributed given the true effects, allowing for global correlation that can result from overlapping GWAS samples. We model:

$$p(\hat{\beta}_{M,j}, \hat{\beta}_{Y,j} | \beta_{M,j}, \beta_{Y,j}, \rho, \hat{s}_{M,j}, \hat{s}_{Y,j}) = N_2\left(\begin{pmatrix} \hat{\beta}_{M,j} \\ \hat{\beta}_{Y,j} \end{pmatrix}; \begin{pmatrix} \beta_{M,j} \\ \beta_{Y,j} \end{pmatrix}; S_j(\rho)\right), \quad \#(3)$$

where  $S_j(\rho) = \begin{pmatrix} \hat{s}_{M,j}^2 & \rho \hat{s}_{M,j} \hat{s}_{Y,j} \\ \rho \hat{s}_{M,j} \hat{s}_{Y,j} & \hat{s}_{Y,j}^2 \end{pmatrix}$  and  $N_2(x; \mu, \Sigma)$  is the bivariate normal density with

mean  $\mu$  and variance  $\Sigma$  evaluated at  $x$ . This model implicitly assumes that  $\hat{s}_{M,j}$  and  $\hat{s}_{Y,j}$  are measured without error. The correlation term,  $\rho$ , which accounts for sample overlap, is estimated empirically (Supplementary Note, Section SN1).

In the CAUSE model, a proportion,  $q$ , of variants exhibit correlated pleiotropy, modeled as an effect on a shared factor,  $U$  (Fig. 1c, right). The remaining proportion,  $1 - q$ , are independent of  $U$  (Fig. 1c, left). All variants may have pleiotropic effects,  $\theta_j$  on  $Y$  that are uncorrelated with their effects on  $M$ . Let  $Z_j$  be an indicator that variant  $j$  effects  $U$ . Then,

$$\beta_{Y,j} | Z_j = \underbrace{\gamma \beta_{M,j}}_{\text{causal effect}} + \underbrace{Z_j \eta \beta_{M,j}}_{\text{correlated pleiotropy}} + \underbrace{\theta_j}_{\text{uncorrelated pleiotropy}}, \quad \#(4)$$

$$Z_j \sim \text{Bernoulli}(q),$$

where  $\eta$  is the effect of  $U$  on  $Y$  and  $U$  is scaled so that the effect of  $U$  on  $M$  is 1. Note that if there are no uncorrelated or correlated pleiotropic effects, then Equation (4) reduces to  $\beta_{Y,j} = \gamma \beta_{M,j}$ , the relationship assumed by simple MR approaches such as IVW regression. We substitute the right side of Equation (4) into Equation (3) and integrate out  $Z_j$  to obtain:

$$\begin{aligned}
& p(\hat{\beta}_{M,j}, \hat{\beta}_{Y,j} | \beta_{M,j}, \theta_j, \rho, \hat{s}_{M,j}, \hat{s}_{Y,j}, \gamma, \eta, q) \\
& = q N_2 \left( \begin{pmatrix} \hat{\beta}_{M,j} \\ \hat{\beta}_{Y,j} \end{pmatrix}; \begin{pmatrix} \beta_{M,j} \\ (\gamma + \eta)\beta_{M,j} + \theta_j \end{pmatrix}; S_j(\rho) \right) \\
& + (1 - q) N_2 \left( \begin{pmatrix} \hat{\beta}_{M,j} \\ \hat{\beta}_{Y,j} \end{pmatrix}; \begin{pmatrix} \beta_{M,j} \\ \gamma\beta_{M,j} + \theta_j \end{pmatrix}; S_j(\rho) \right).
\end{aligned} \tag{5}$$

The parameters of interest in the CAUSE model are  $\gamma$ ,  $\eta$ , and  $q$ . Rather than estimating individual variant effects  $\beta_{M,j}$  and  $\theta_j$ , we model their joint distribution empirically and integrate them out to obtain a marginal density for  $\hat{\beta}_{M,j}$  and  $\hat{\beta}_{Y,j}$ . We model  $\beta_{M,j}$  and  $\theta_j$  as draws from a mixture of bivariate normal distributions. This strategy is based on Adaptive Shrinkage (ASH) approach of Stephens<sup>30</sup> for modeling univariate distributions and provides a flexible unimodal distribution with mode at (0,0). We model:

$$p(\beta_{M,j}, \theta_j | \pi_0, \dots, \pi_K, \Sigma_0, \dots, \Sigma_K) = \sum_{k=0}^K \pi_k N_2 \left( \begin{pmatrix} \beta_{M,j} \\ \theta_j \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \Sigma_k \right), \tag{6}$$

where  $\Sigma_0, \Sigma_1, \dots, \Sigma_K$  are pre-specified covariance matrices of the form  $\Sigma_k = \begin{pmatrix} \sigma_{k,1}^2 & 0 \\ 0 & \sigma_{k,2}^2 \end{pmatrix}$ , and

$\pi_0, \pi_1, \dots, \pi_K$  are mixing proportions that sum to 1. The set of parameters  $\Omega = \{ \pi_0, \dots, \pi_K, \Sigma_0, \dots, \Sigma_K \}$  are estimated from the data along with  $\rho$  in a single pre-processing step (Supplementary Note, Section SN1). Integrating  $\beta_{M,j}$  and  $\theta_j$  out of Equation (5), we obtain:

$$\begin{aligned}
& p(\hat{\beta}_{M,j}, \hat{\beta}_{Y,j} | \gamma, \eta, q, \hat{s}_{M,j}, \hat{s}_{Y,j}, \Omega, \rho) = q \sum_{k=0}^K \pi_k N_2 \\
& \left( \begin{pmatrix} \hat{\beta}_{M,j} \\ \hat{\beta}_{Y,j} \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}; A(\gamma + \eta) \Sigma_k A(\gamma + \eta)^\top + S_j(\rho) \right) \\
& + (1 - q) \sum_{k=0}^K \pi_k N_2 \left( \begin{pmatrix} \hat{\beta}_{M,j} \\ \hat{\beta}_{Y,j} \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}; A(\gamma) \Sigma_k A(\gamma)^\top + S_j(\rho) \right),
\end{aligned} \tag{7}$$

where  $A(x) = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}$ . Treating variants as independent, we obtain a joint density for the entire set of summary statistics as a product over variants. We place independent prior distributions on  $\gamma$ ,  $\eta$ , and  $q$ , described below, and estimate posterior distributions via adaptive grid approximation (Supplementary Note, Section SN3). A flow-chart illustrating a CAUSE analysis including parameter estimation is shown in Extended Data Figure 4. CAUSE is implemented in an open source R package and we also provide pipelines for running CAUSE on large-scale applications.

### Prior distributions for $\gamma$ , $\eta$ , and $q$

We place normal prior distributions with mean zero on  $\gamma$  and  $\eta$ . We find in simulations that results are robust to different choices of prior variance for  $\gamma$  and  $\eta$ , as long the same value is

used for both parameters (Supplementary Note, Section SN2). Since the magnitude of a possible causal effect is difficult to know a priori due to differences in trait scaling and covariate adjustment across GWAS, we use the data to suggest a prior variance (Supplementary Note, Section SN2).

By default, and in all analyses presented here, we use a Beta(1,10) prior distribution for  $q$ . This distribution gives a prior probability of 0.001 that  $q > 0.5$  and 0.056 that  $q > 0.25$ . The parameters of the Beta distribution can be adjusted by the user to reflect different prior beliefs about the size of  $q$ . If the prior on  $q$  places more weight on values close to zero, CAUSE will behave more similarly to standard MR, obtaining high false positive rates when the true value of  $q$  is large. With a more permissive prior on  $q$ , CAUSE is more robust to larger amounts of correlated pleiotropy but has reduced power. We set Beta(1,10) as the default choice because it performed acceptably over a range of settings in simulations (Supplementary Note, Section SN6.1).

### Accounting for linkage disequilibrium

We treat variants as independent when we compute the joint density of summary statistics. In reality, variants are correlated due to linkage disequilibrium (LD). We define the *LD-transformed effects*  $\beta_M^* = S_M R S_M^{-1} \beta_M$  and  $\beta_Y^* = S_Y R S_Y^{-1} \beta_Y$ , where  $R$  is the variant correlation matrix and  $S_M$  and  $S_Y$  are diagonal matrices with diagonal elements  $(s_{M,j})$  and  $(s_{Y,j})$ <sup>19</sup>. We assume that  $R$  is the same in both GWAS. The pair of estimates  $(\hat{\beta}_{M,j}, \hat{\beta}_{Y,j})$  can be approximated as normally distributed with mean  $(\beta_{M,j}^*, \beta_{Y,j}^*)$  and variance  $S(\rho)$ . If the relationship in Equation (4) holds for true effects,  $\beta_{M,j}$  and  $\beta_{Y,j}$ , and either  $M$  effects are sparse relative to LD structure (most  $M$  effect variants are independent) or  $q$  is small, then the relationship between the LD-transformed effects is:

$$\beta_{Y,j}^* = \gamma \beta_{M,j}^* + Z_j \eta \beta_{M,j}^* + \theta_j^*, \quad \#(8)$$

where  $\theta_j^* = S_Y R S_Y^{-1} \theta_j$  (Supplementary Note, Section SN4). In this case, the mean relationship for summary statistics at a single locus is the same with or without LD and Equation (5) and (7) remain valid for variants in LD if  $\beta_{M,j}$  and  $\theta_j$  are replaced with  $\beta_{M,j}^*$  and  $\theta_j^*$ .

Correlations among variants can affect the joint density of all summary statistics, which we compute as a product of the densities at each variant. To account for this, we use a subset of variants with low mutual LD ( $r^2 < 0.1$  by default), prioritizing variants with low trait  $M$ - $p$ -values to improve power.

If  $M$  effect variants are dense relative to LD structure and  $q$  is large, LD can induce a positive correlation between  $Y$  and  $M$  effect estimates for all variants, even when there is no causal effect (Supplementary Note, Section SN4). In this case, CAUSE may have an inflated false positive rate.

## Model comparison using ELPD

To determine whether GWAS summary statistics are consistent with a causal effect of  $M$  on  $Y$  we compare a model in which the causal effect is fixed at zero (the *sharing model*) to a model that allows a non-zero causal effect (the *causal model*). To compare the fit of these models, we estimate the difference in the *expected log pointwise posterior density* (ELPD)<sup>18</sup>. The ELPD measures how well the posterior distributions estimated under a given model are expected to predict a hypothetical new set of summary statistics obtained from GWAS of  $M$  and  $Y$  in different samples.

Let  $\Theta$  be the set of parameters  $(\gamma, \eta, q)$ . Let  $p_C(\Theta|\text{Data})$  and  $p_S(\Theta|\text{Data})$  be the posterior density of  $\Theta$  given the observed summary statistics under the causal model and sharing model respectively. Let  $y'_j$  denote a new observation of  $(\hat{\beta}_{M,j}, \hat{\beta}_{Y,j}, \hat{s}_{M,j}, \hat{s}_{Y,j})$ . The ELPD for model  $m \in \{C, S\}$  (for causal and sharing) is:

$$\text{ELPD}_m = \sum_{j=1}^p \left( \int \left( \int p(y'_j | \theta, \Omega, \rho, \hat{s}_{M,j}, \hat{s}_{Y,j}) p_m(\theta | \text{Data}) d\theta \right) p_{\text{true}}(y'_j) dy'_j \right). \quad \#(9)$$

where  $p(y'_j | \theta, \Omega, \rho, \hat{s}_{M,j}, \hat{s}_{Y,j})$  is given in Equation (7) and  $p_{\text{true}}$  is the probability density under the true data generating mechanism. If  $\text{ELPD} = \text{ELPD}_C - \text{ELPD}_S$  is positive, then the posteriors from the causal model predict the data better so the causal model is a better fit. If  $\text{ELPD} = 0$ , then the sharing model fits at least as well, indicating that the data are not consistent with a causal effect.

We estimate  $\text{ELPD}$  and a standard error of the estimator using the Pareto-smoothed importance sampling method described by Vehtari, Gelman, and Gabry<sup>18</sup> and implemented in the R package `loo`. We then compute a  $z$ -score,  $z_{\text{elpd}} = \frac{\Delta \text{ELPD}}{\widehat{\text{se}}(\Delta \text{ELPD})}$ , that is larger when posteriors estimated under the causal model fit the data better than the posteriors estimated under the sharing model. We compute a one-sided  $p$ -value by comparing the  $z$ -score to a standard normal distribution. The  $p$ -value estimates the probability of obtaining a  $z$ -score larger than the one observed if the true value of  $\text{ELPD}$  were less than or equal to zero.

## Generating simulated summary statistics

To create data with a realistic correlation structure, we estimate LD for 19,490 HapMap variants on chromosome 19 in the CEU 1,000 Genomes population using LDShrink<sup>31</sup> (<https://github.com/stephenslab/ldshrink>) and replicate this pattern 30 times to create a genome sized data set of  $p = 584,700$  variants. We generate effect estimates from the CAUSE model in Equation (4), given setting specific values of  $\gamma$ ,  $\eta$ , and  $q$ , and parameters defining the trait architecture and power of the GWAS (heritability, number of effect variants, and GWAS sample size).

We simulate an effect estimate and standard error for each variant using the following procedure. First, standardized effects  $\tilde{\beta}_{M,j} = \beta_{M,j} \sqrt{2f_j(1-f_j)}$  and  $\tilde{\theta}_j = \theta_j \sqrt{2f_j(1-f_j)}$  are drawn from a mixture distribution:

$$\begin{aligned} \tilde{\beta}_{M,j} &\sim (1 - \pi_M)\delta_0 + \pi_M N(0, \sigma_M^2) \\ \tilde{\theta}_j &\sim (1 - \pi_\theta)\delta_0 + \pi_\theta N(0, \sigma_\theta^2), \end{aligned} \quad \#(10)$$

where  $f_j$  is the allele frequency of variant  $j$  in the 1000 Genomes CEU population. Variances  $\sigma_M^2$  and  $\sigma_\theta^2$  are chosen to give the desired expected heritability as  $\sigma_M^2 = \frac{h_M^2}{m_M}$  and  $\sigma_\theta^2 = \frac{h_Y^2 - (\gamma^2 + q\eta^2)h_M^2}{m_\theta}$ . In these simulations,  $h_M^2 = h_Y^2 = 0.25$ . Parameters  $\pi_M = m_M/p$  and  $\pi_\theta = m_\theta/p$  define the expected number of variants with non-zero values of  $\beta_{M,j}$  and  $\theta_j$  respectively. In simulations in Figure 2,  $m_M = m_\theta = 1,000$ . Note that  $\gamma^2 h_M^2 / h_Y^2$  is the proportion of trait  $Y$  heritability mediated by the causal effect and  $q\eta^2 h_M^2 / h_Y^2$  is the proportion of trait  $Y$  heritability mediated by  $U$ .

Second, standardized effects are converted to non-standardized effects and standard errors are computed as  $s_{\cdot j} = \sqrt{\frac{1}{2N \cdot f_j(1 - f_j)}}$ , where  $\cdot$  may be  $M$  or  $Y$  and  $N_M$  and  $N_Y$  are GWAS sample sizes for traits  $M$  and  $Y$ , respectively. Given the other parameters, we selected sample sizes of 40,000 for high powered GWAS settings and 12,000 for low powered GWAS settings to provide desired numbers of genome-wide significant variants (about 107 and 8, respectively).

Third,  $Z_j$  are drawn from a Bernoulli( $q$ ) distribution and true effects  $\beta_{Y,j}$  are computed using Equation (4). In Figure 2a, a weak shared factor effects correspond to  $\eta = \sqrt{0.02}$ , while large shared factor effects correspond to  $\eta = \sqrt{0.05}$ . Finally, effect estimates are simulated from true effects as:

$$\hat{\beta}_{\cdot} \sim N_p(S \cdot RS \cdot \beta_{\cdot}, S \cdot RS \cdot^{-1}), \quad \#(11)$$

where  $S$  is the diagonal matrix of standard errors and  $R$  is the variant correlation matrix<sup>19</sup>. Simulations can be replicated using the online tutorial (<https://jean997.github.io/cause/simulations.html>).

### Existing MR methods

We compare the performance of CAUSE in simulated data with six other MR methods and LCV<sup>14</sup>. These are implemented as follows.

- Random effect IVW regression, random effect Egger regression, the weighted median, and weighted mode methods are implemented in the MendelianRandomization R package. IVW regression is run using model = “random” and weights = “delta” options. The weighted mode estimator is computed using the phi = 1 and stderror=“delta” options. All other methods are run with default options.
- MR-PRESSO is performed using the MRPRESSO R package (<https://github.com/rondolab/MR-PRESSO>) with outlier and distortion tests.

- GSMR is performed using the `gsmr` R package (<http://cnsgenomics.com/software/gsmr/>) with the Heidi outlier test, default threshold 0.01, and minimum number of instruments lowered to 1.
- LCV is performed using R scripts available from the author (<https://github.com/lukejconnor/LCV>) using default parameters.

### Reporting summary

Additional details on study design are provided in the **Life Sciences Reporting Summary**.

### Code availability

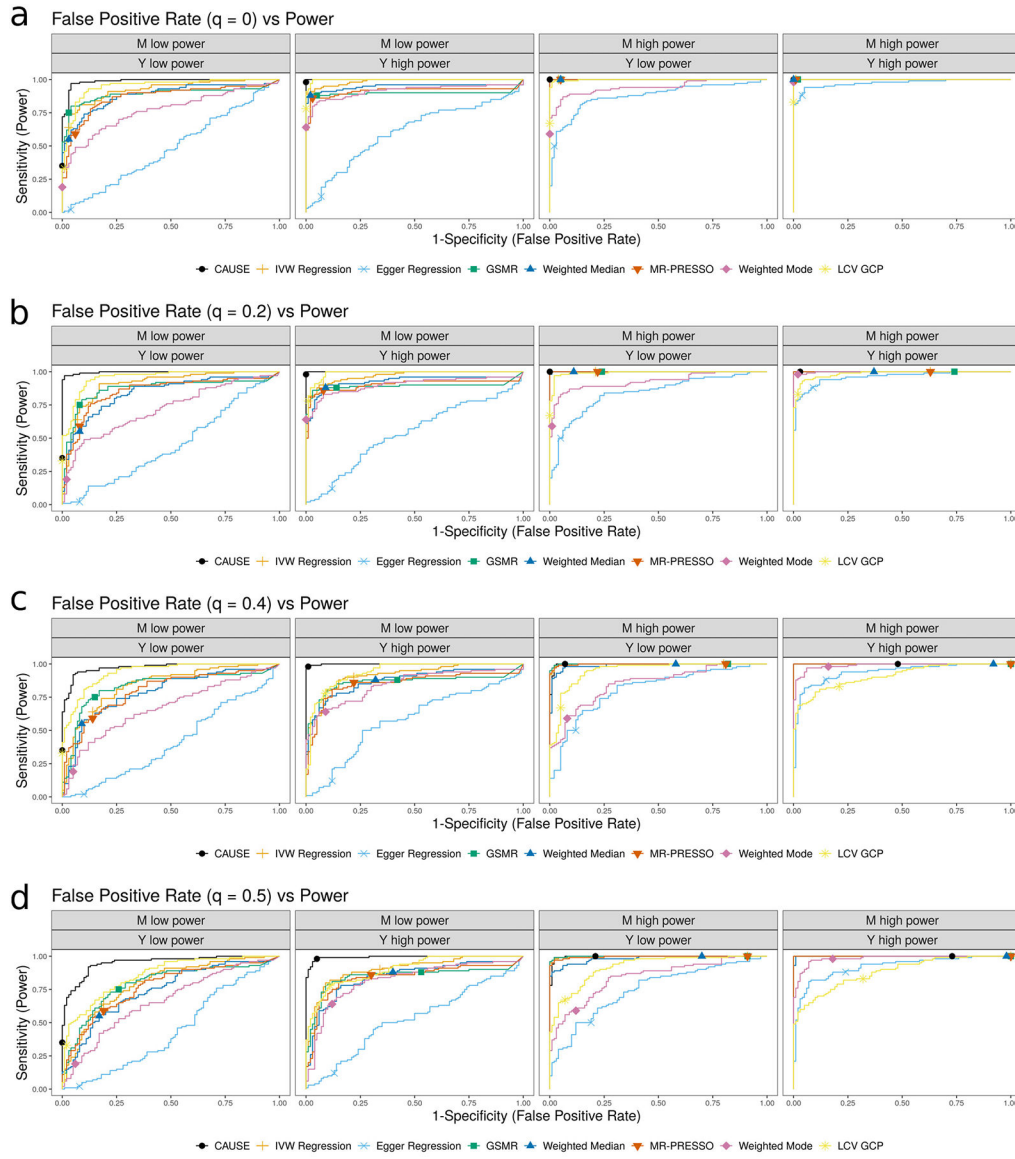
All software and analysis code is publicly available. The CAUSE method is implemented in an R package available through GitHub.

- Website: <https://jean997.github.io/cause/>  
Includes pipelines and instructions for replicating all results presented in this paper.
- CAUSE software (R package): <https://github.com/jean997/cause>
- Simulations software (R package): <https://github.com/jean997/causeSims>

### Data availability

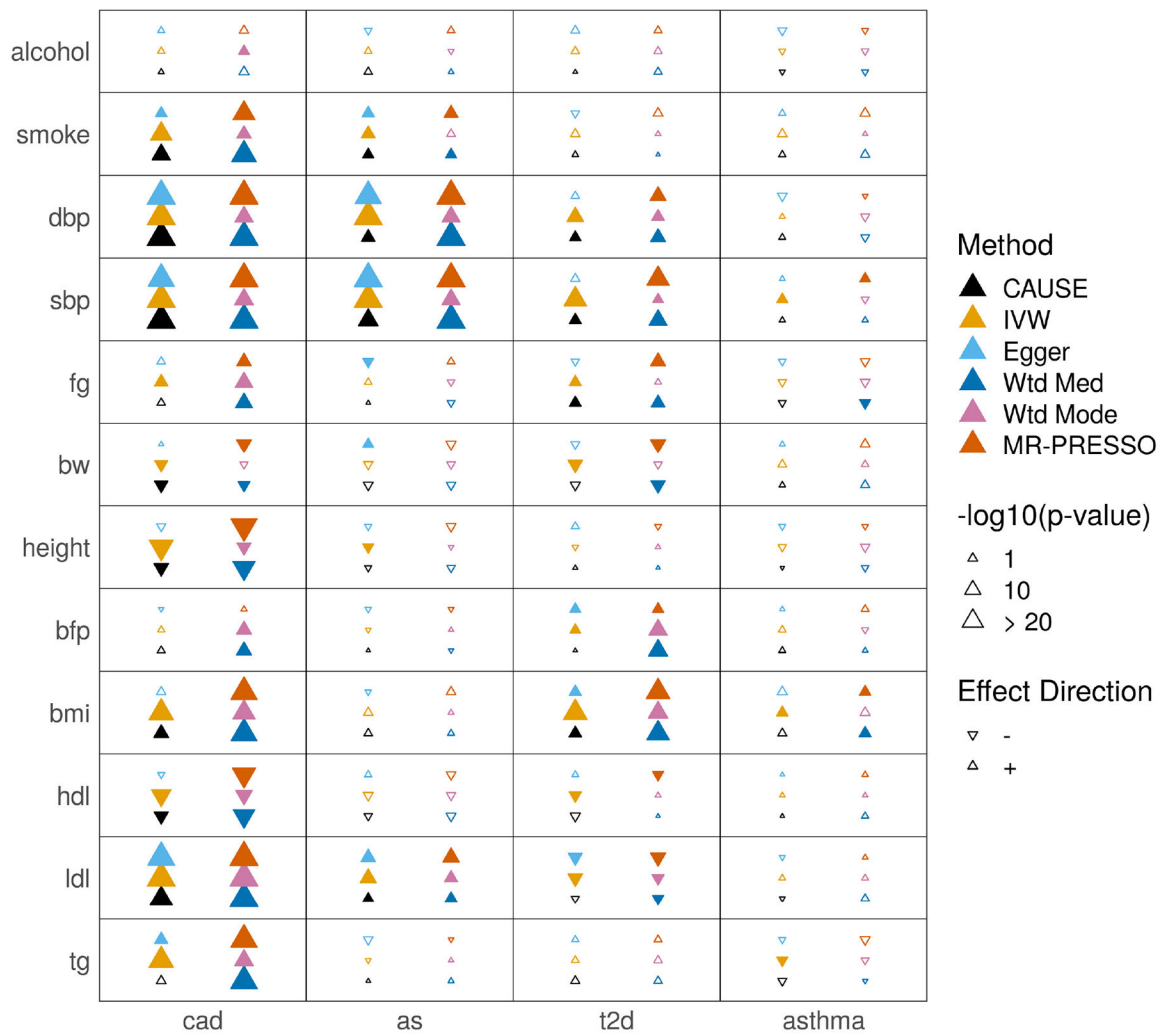
All of the data analyzed are publicly available with the exception of blood pressure summary statistics from Ehret et al.<sup>28</sup>. These are available through dbGaP Accession phs000585.v2.p1. Download links for all other data sets are available in Supplementary Table 11. Instructions and code for formatting and processing data and reproducing CAUSE analysis results can be found on the website <https://jean997.github.io/cause/>.

### Extended Data



**Extended Data Fig. 1. False positive-power trade-offs for different proportions of correlated pleiotropic variants.**

We compare the power when  $\gamma = \sqrt{0.05}$  and  $q = 0$  to the false positive rate when  $\gamma = 0$ ,  $q$  varies from 0 to 0.5 and  $\eta = \sqrt{0.05}$ . There are 100 simulations each in the causal and non-causal scenarios. Curves are created by varying the significance threshold. Points indicate the power and false positive rate achieved at a threshold of  $p = 0.05$ .

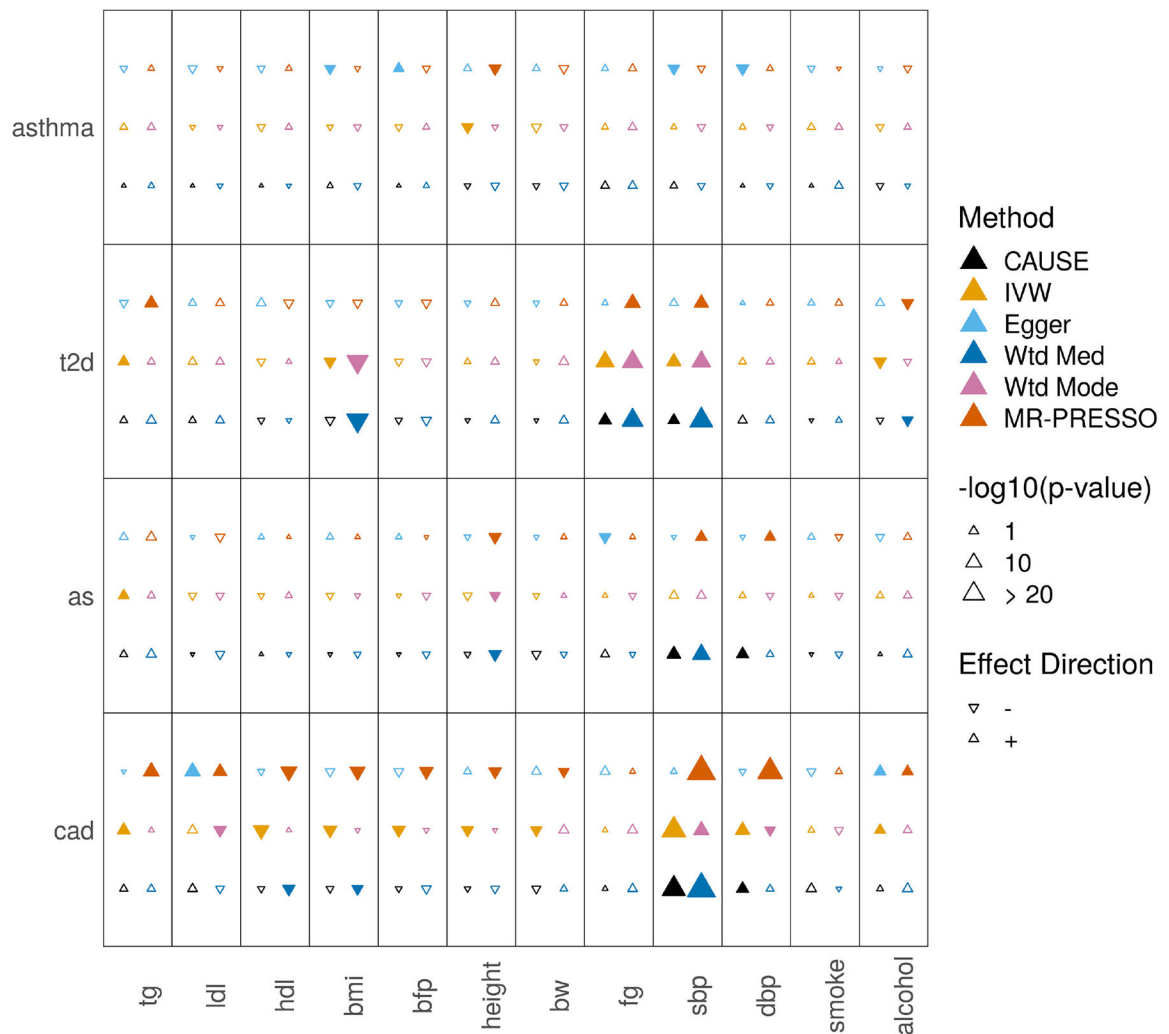


**Extended Data Fig. 2. Tests for casual effects of risk factors on diseases.**

Each cell summarizes the results of six methods for a pair of traits. In the left column of the cell, methods from bottom to top are CAUSE, IVW regression, and Egger regression. In the right column, methods from bottom to top are weighted median, weighted mode, and MR-PRESSO.

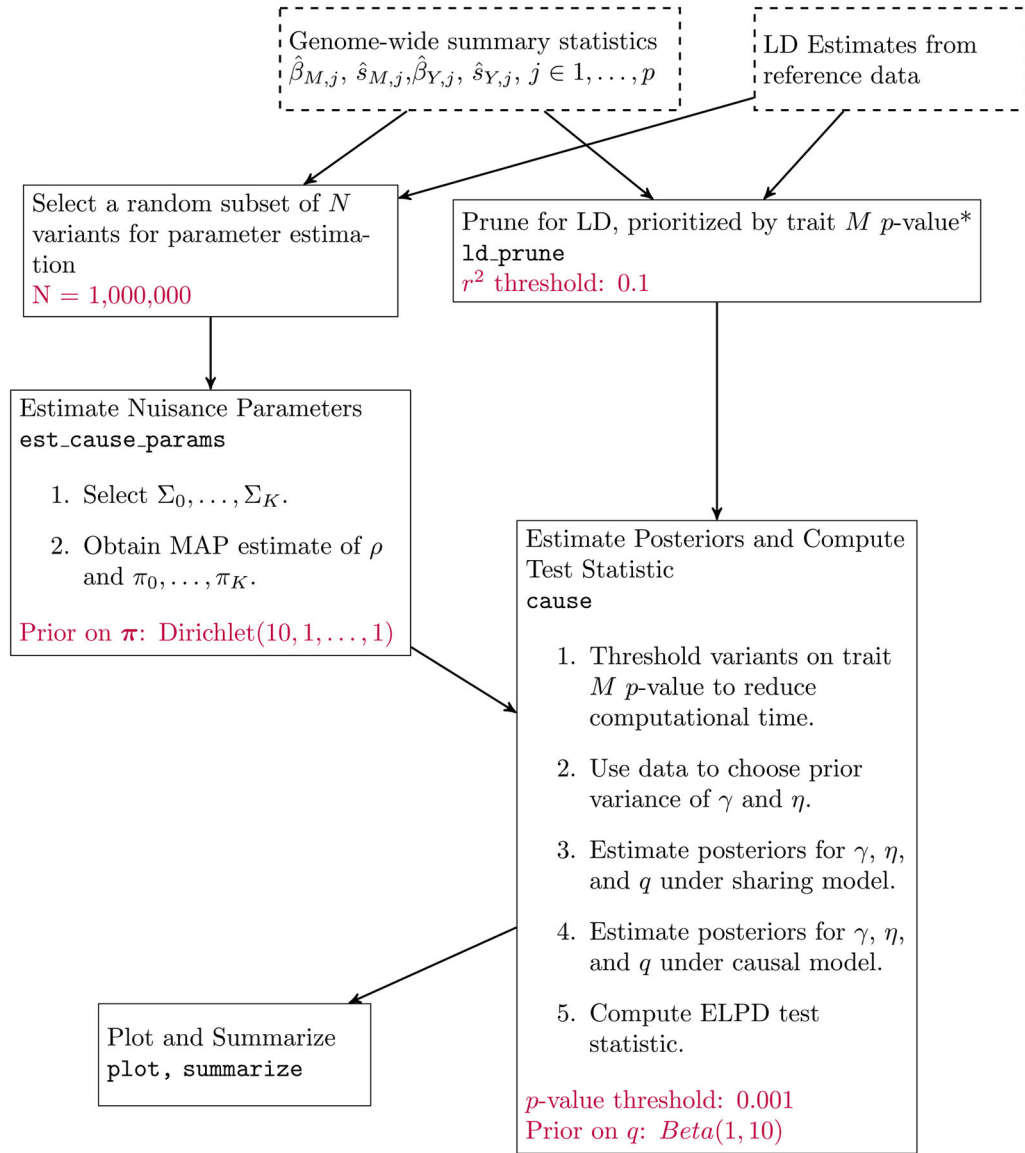
Filled symbols indicate a nominally significant  $p = 0.05$ .





**Extended Data Fig. 3. Tests for casual effects of disease outcomes on risk factors.**

Tests for casual effects of disease outcomes on mediators. Each cell summarizes the results of six methods for a pair of traits. In the left column of the cell, methods from bottom to top are CAUSE, IVW regression, and Egger regression. In the right column, methods from bottom to top are weighted median, weighted mode, and MR-PRESSO. Filled symbols indicate a nominally significant  $p = 0.05$ .



**Extended Data Fig. 4. Workflow of a CAUSE analysis.**

Dashed boxes represent input data. Each solid box is an analysis step completed by the given function in the cause R package. LD pruning can be parallelized over chromosomes. Text at the bottom of boxes indicates user provided parameters and their default values. All analyses presented are run with default parameters.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

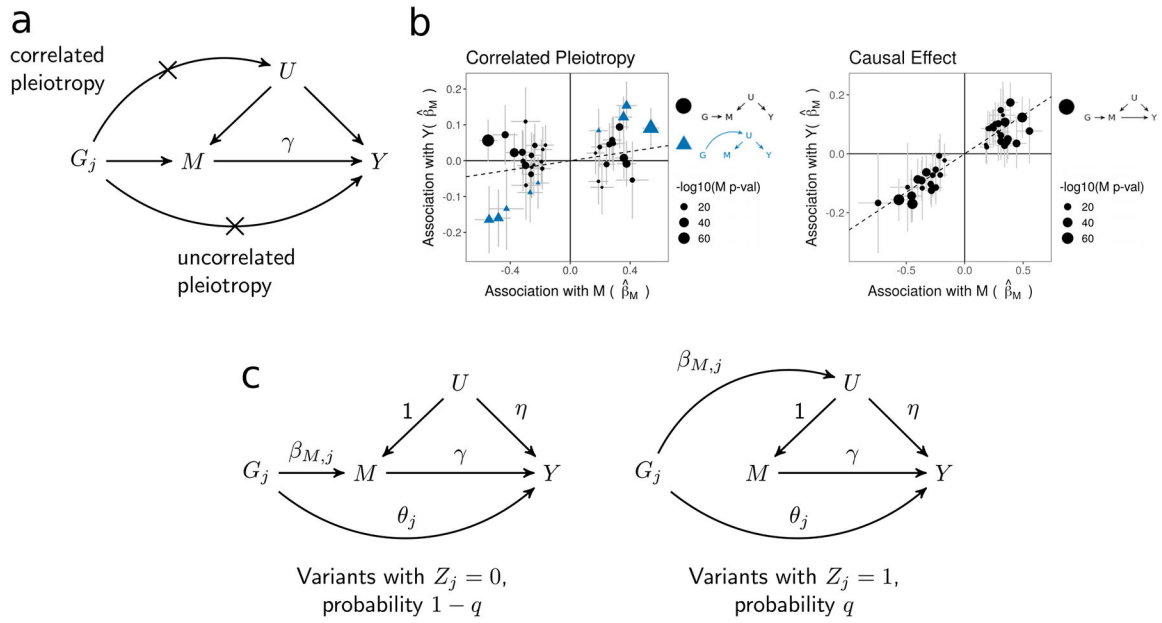
**Acknowledgements**

This work was supported by National Institutes of Health (NIH) grants MH110531 (to X.H.) and HG002585 (to M.S.), and a Research Grant from the March of Dimes (X.H.).

## References

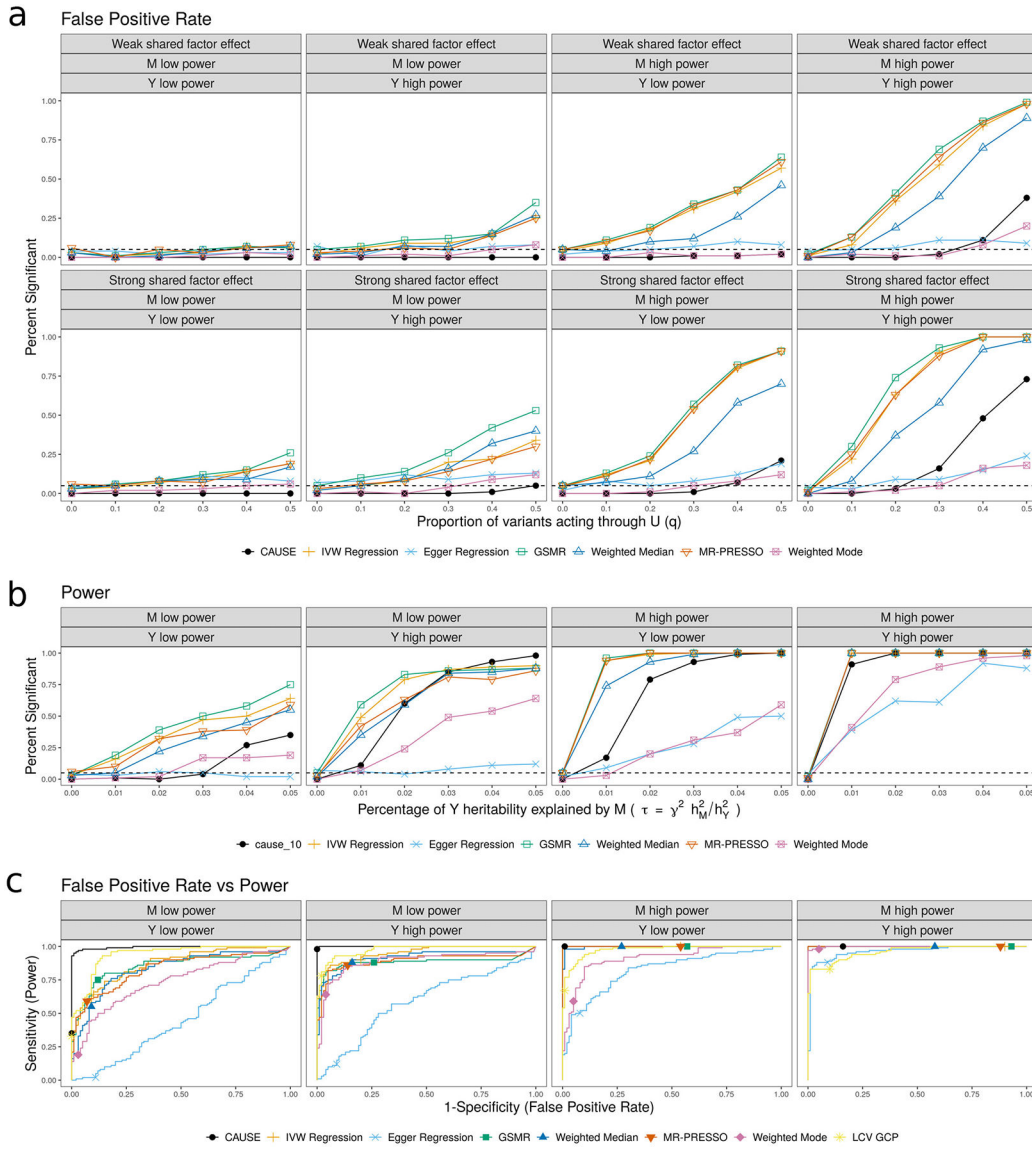
1. Smith GD & Ebrahim S 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol* 32, 1–22 (2003). [PubMed: 12689998]
2. Smith GD & Hemani G Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet* 23, 89–98 (2014).
3. Boef AGC, Dekkers OM & Le Cessie S Mendelian randomization studies: a review of the approaches used and the quality of reporting. *Int. J. Epidemiol* 44, 496–511 (2015). [PubMed: 25953784]
4. Zhang G et al. Genetic associations with gestational duration and spontaneous preterm birth. *N. Engl. J. Med* 377, 1156–1167 (2017). [PubMed: 28877031]
5. Burgess S, Dudbridge F & Thompson SG Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods. *Stat. Med* 35, 1880–1906 (2016). [PubMed: 26661904]
6. Hemani G, Bowden J & Davey Smith G Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet* 27, R195–R208 (2018). [PubMed: 29771313]
7. Verbanck M, Chen C, Neale B & Do R Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet* 50, 693–698 (2018). [PubMed: 29686387]
8. Bowden J, Smith GD & Burgess S Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol* 44, 512–525 (2015). [PubMed: 26050253]
9. Barfield R, Feng H, Gusev A, Wu L & Zheng W Transcriptome-wide association studies accounting for co-localization using Egger regression. *Genet. Epidemiol* 42, 418–433 (2018). [PubMed: 29808603]
10. Zhu Z et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun* 9, 224 (2018). [PubMed: 29335400]
11. Bulik-Sullivan B et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet* 47, 1236–1241 (2015). [PubMed: 26414676]
12. Anttila V et al. Analysis of shared heritability in common disorders of the brain. *Science* 360, eaap8757 (2018). [PubMed: 29930110]
13. Pickrell JK et al. Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet* 48, 709–717 (2016). [PubMed: 27182965]
14. O'Connor LJ & Price AL Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat. Genet* 50, 1728–1734 (2018). [PubMed: 30374074]
15. Burgess S & Thompson SG Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol* 181, 251–260 (2015). [PubMed: 25632051]
16. Bowden J, Davey Smith G, Haycock PC & Burgess S Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol* 40, 304–314 (2016). [PubMed: 27061298]
17. Hartwig FP, Smith GD & Bowden J Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol* 46, 1985–1998 (2017). [PubMed: 29040600]
18. Vehtari A, Gelman A & Gabry J Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput* 27, 1413–1432 (2016).
19. Zhu X & Stephens M Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. *Ann. Appl. Stat* 11, 1561–1592 (2017). [PubMed: 29399241]
20. Burgess S, Butterworth A & Thompson SG Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol* 37, 658–665 (2013). [PubMed: 24114802]

21. Hemani G, Tilling K & Davey Smith G Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* 13, e1007081 (2017). [PubMed: 29149188]
22. Voight BF et al. Plasma HDL cholesterol and risk of myocardial infarction: a Mendelian randomisation study. *Lancet* 380, 572–580 (2012). [PubMed: 22607825]
23. Moordian AD Dyslipidemia in type 2 diabetes mellitus. *Nat. Clin. Pract* 5, 150–159 (2009).
24. Sattar N et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet* 375, 735–742 (2010). [PubMed: 20167359]
25. Crandall JP et al. Statin use and risk of developing diabetes: results from the diabetes prevention program. *BMJ Open Diabetes Res. Care* 5, e000438 (2017).
26. Swerdlow DI et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: Evidence from genetic analysis and randomised trials. *Lancet* 385, 351–361 (2015). [PubMed: 25262344]
27. Fall T et al. Using genetic variants to assess the relationship between circulating lipids and type 2 diabetes. *Diabetes* 64, 2676–2684 (2015). [PubMed: 25948681]
28. Ehret GB et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 478, 103–109 (2011). [PubMed: 21909115]
29. Fulkerson PC & Rothenberg ME Targeting eosinophils in allergy, inflammation and beyond. *Nat. Rev. Drug Discov* 12, 117–129 (2013). [PubMed: 23334207]
30. Stephens M False discovery rates: a new deal. *Biostatistics* 18, 275–295 (2017). [PubMed: 27756721]
31. Wen X & Stephens M Using linear predictors to impute allele frequencies from summary or pooled genotype data. *Ann. Appl. Stat* 4, 1158–1182 (2010). [PubMed: 21479081]



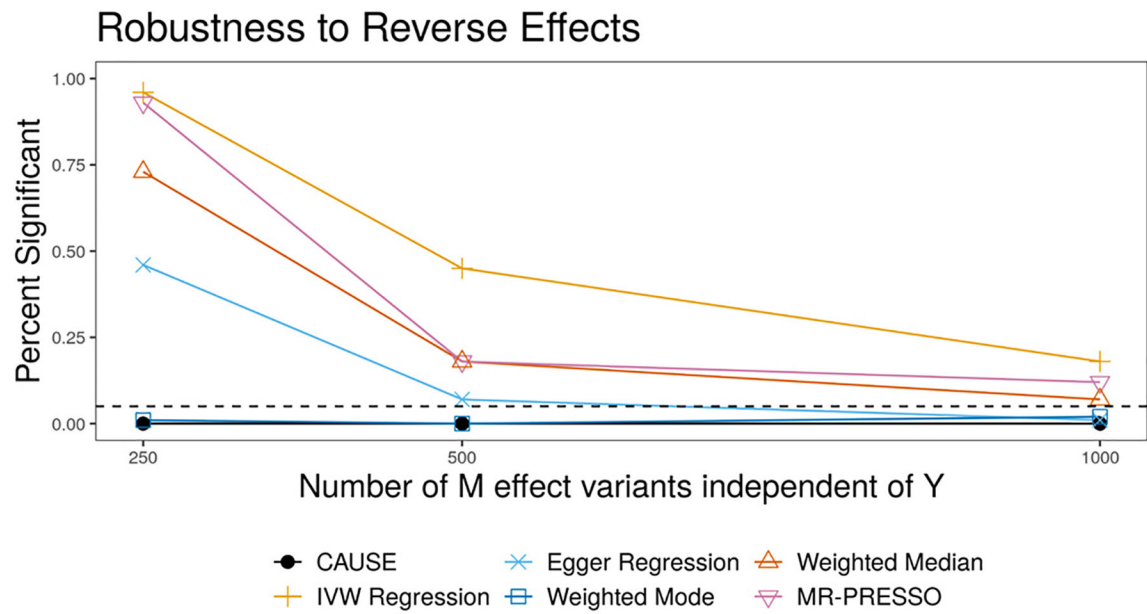
**Figure 1 |. Assumptions of traditional MR and CAUSE model.**

**a**, Causal diagram assumed by traditional MR approaches. The causal effect of trait  $M$  on trait  $Y$ ,  $\gamma$ , is the target of inference. Crosses mark horizontal pleiotropic effects that are assumed absent by traditional MR. **b**, Simulated effect estimates illustrating the pattern induced by a shared factor (correlated pleiotropy) with no causal effect (left), and the pattern induced by a causal effect (right). In both plots, effect size estimates for  $M$  and  $Y$  are indicated by points. Error bars around points have length 1.96 times the simulated standard error of the estimate on each side. SE's are simulated with a sample size of 5,000. Only variants that are strongly associated with  $M$  ( $p < 5 \cdot 10^{-8}$ ) are shown. **c**, CAUSE assumes that variants affect trait  $M$  through one of two mechanisms. A proportion  $1 - q$  of variants have the left causal diagram, while the remaining proportion,  $q$ , have the right causal diagram.



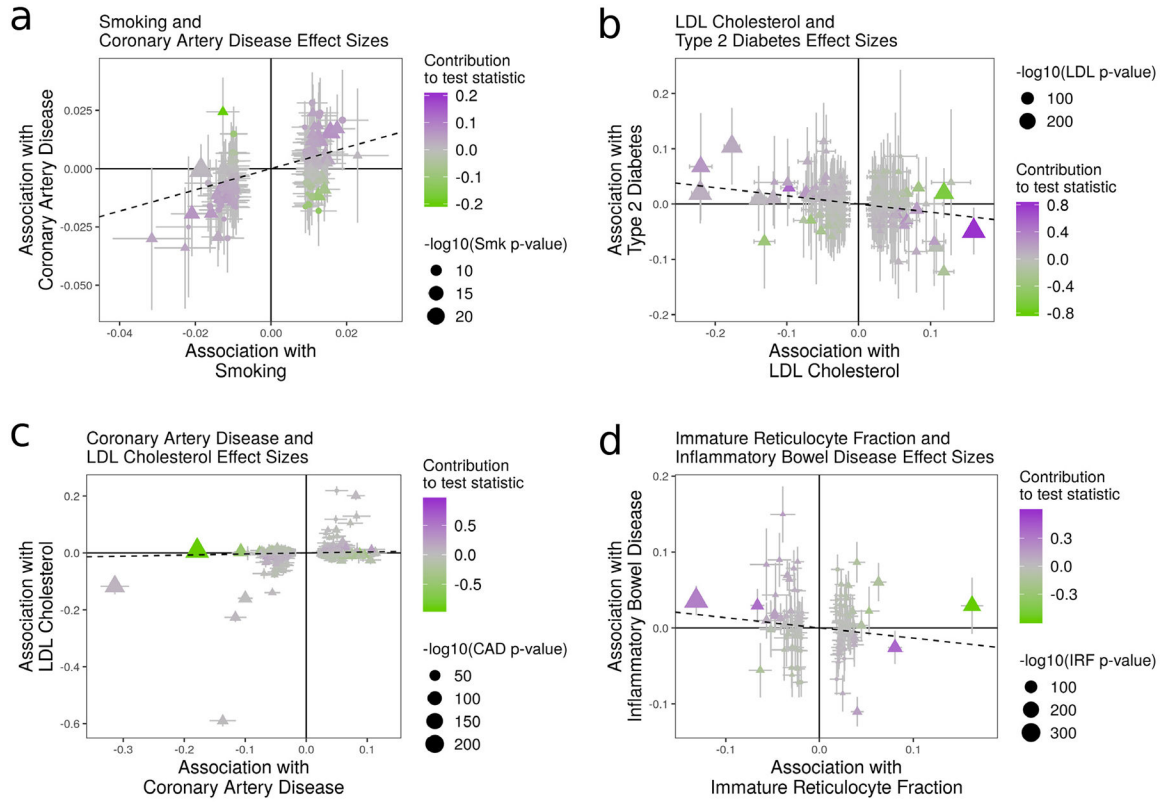
**Figure 2 | Performance of CAUSE and other MR methods in simulated data.**

**a**, False positive rate averaged over 100 simulated data sets in settings with no causal effect and a proportion of correlated pleiotropic variants ranging from 0 to 50%. **b**, Power averaged over 100 simulated data sets in settings with a causal effect and no shared factor. **c**, Comparison of false positive-power trade-off. We compare the power when  $\gamma = \sqrt{0.05}$  and  $q = 0$  to the false positive rate  $\gamma = 0$ ,  $q = 0.3$  and  $\eta = \sqrt{0.05}$ . There are 100 simulations each in the causal and non-causal scenarios. Curves are created by varying the significance threshold. Points indicate the power and false positive rate achieved at a threshold of  $p = 0.05$  or  $\widehat{GCP} > 0.6$  for LCV.



**Figure 3 |. False positives resulting from reverse causal effects.**

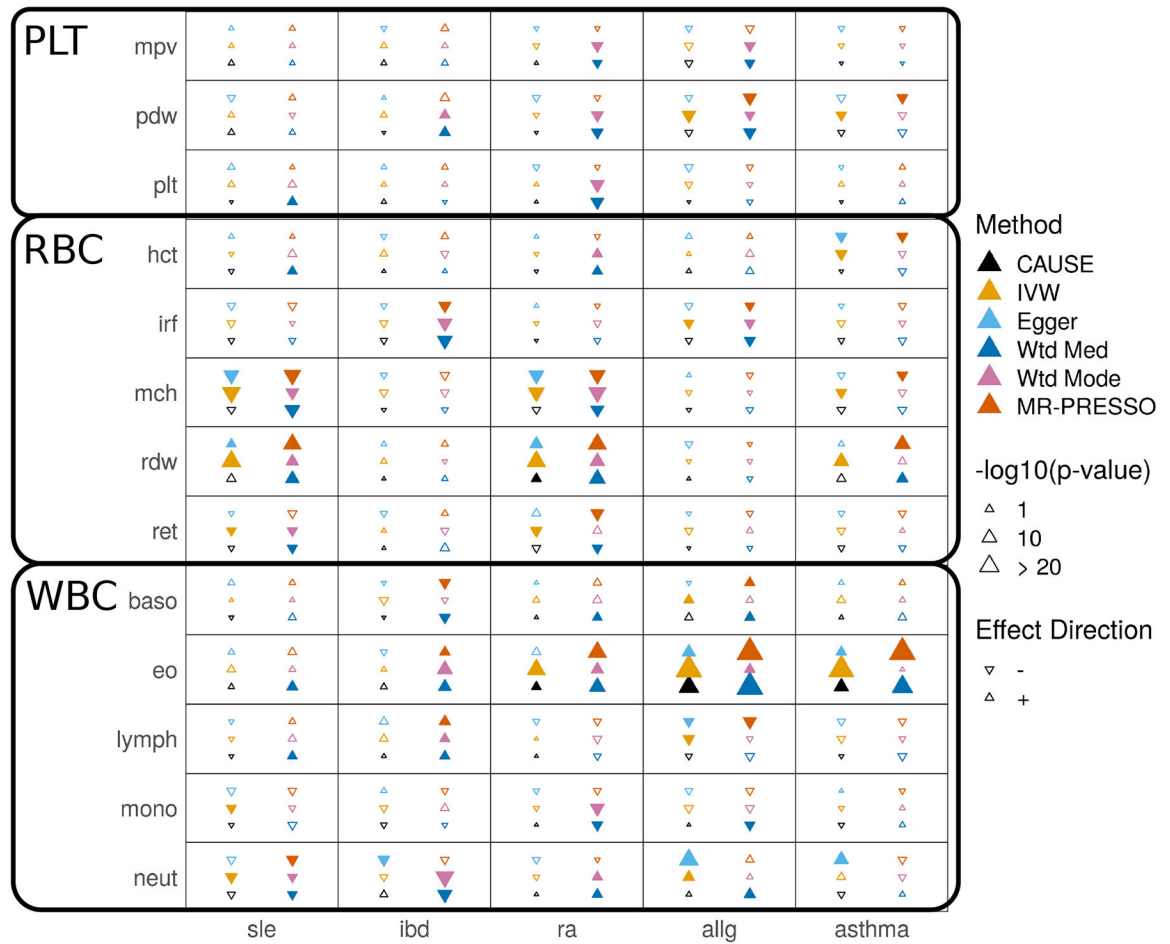
Data are simulated with a true effect of  $Y$  on  $M$ , but tests are performed for an effect of  $M$  on  $Y$ . Each point shows the average over 100 simulations. Only CAUSE and the weighted mode control the false positive rate.



**Figure 4 | Effect size estimates and variant level contribution to CAUSE test statistics for four trait pairs.**

Effect estimates for trait  $M$  (horizontal axis) are plotted against estimates for trait  $Y$  (vertical axis). Error bars have length 1.96 times the standard error of the estimate. Triangles indicate variants reaching genome-wide significance for trait  $M$  ( $p < 5 \cdot 10^{-8}$ ). Variants with trait  $M$   $p$ -value  $< 5 \cdot 10^{-6}$  are shown. Dotted lines show the IVW estimate obtained using only genome-wide significant variants. **a**, Smoking ( $M$ ) and CAD ( $Y$ ). All methods detect evidence of a causal effect. **b**, LDL ( $M$ ) and T2D ( $Y$ ). Only CAUSE does not detect a causal effect. Under the CAUSE model, these data can be explained by a shared factor accounting for 13% of LDL cholesterol effect variants. **c**, CAD ( $M$ ) and LDL cholesterol ( $Y$ ). CAUSE avoids a likely false positive obtained by other methods as a result of a reverse direction effect. Egger regression, the weighted mode, and MR-PRESSO all find a significant effect (see Supplementary Table 5). **d**, IRF ( $M$ ) and IBD ( $Y$ ). MR-PRESSO, the weighted median, and modal estimators obtain a positive result by downweighting or removing variants supplying conflicting evidence.





**Figure 5 | Tests for causal effects of blood cell composition on immune mediated traits.** Each cell summarizes the results of six methods for a pair of traits. Filled symbols indicate  $p$ -value < 0.05. Blood cell traits are grouped into platelet traits (PLT), red blood cell traits (RBC), and white blood cell traits (WBC).

**Table 1 |**  
**Summary of results for pairs of GWAS traits categorized by prior information about causality.**

Columns 2–7 give the  $p$ -value for each MR method. Values are bold if  $p < 0.05$ . Arrows indicate the sign of the corresponding effect estimate. LCV GCP and LCV pval give estimated GCP from LCV and  $p$ -value testing that GCP = 0. Values are bold if estimated GCP > 0.6. The Cause  $q$  column gives the posterior median of  $q$  in the CAUSE sharing model. GC and GC pval give the genetic correlation and  $p$ -value testing that genetic correlation is zero estimated by LD score regression. In each section, pairs are ordered by increasing genetic correlation  $p$ -value. BF, body fat; BMI, body mass index; BW, body weight; CAD, coronary artery disease; DBP, diastolic blood pressure; FG, fasting glucose; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SBP, systolic blood pressure; TG, triglycerides; T2D, type 2 diabetes.

Traits	CAUSE	IVW	Egger	Wtd Med	Wtd Mode	MR-PRESSO	LCV GCP	LCV pval	CAUSE $q$	GC	GC pval
Considered causal											
SBP → CAD	<b>6.6</b> · 10 <sup>-31</sup> ↑	<b>1.8</b> · 10 <sup>-148</sup> ↑	<b>7</b> · 10 <sup>-18</sup> ↑	<b>1.8</b> · 10 <sup>-157</sup> ↑	<b>3.8</b> · 10 <sup>-8</sup> ↑	<b>1.3</b> · 10 <sup>-167</sup> ↑	0.02	0.95	0.91	0.35	2.1 · 10 <sup>-49</sup>
DBP → CAD	<b>1.3</b> · 10 <sup>-26</sup> ↑	<b>8.1</b> · 10 <sup>-111</sup> ↑	<b>5.9</b> · 10 <sup>-21</sup> ↑	<b>6.4</b> · 10 <sup>-138</sup> ↑	<b>1.6</b> · 10 <sup>-7</sup> ↑	<b>1.7</b> · 10 <sup>-132</sup> ↑	0.54	0.069	0.88	0.27	1.3 · 10 <sup>-27</sup>
Smoking → CAD	<b>9.2</b> · 10 <sup>-8</sup> ↑	<b>5.4</b> · 10 <sup>-11</sup> ↑	<b>0.014</b> ↑	<b>1.7</b> · 10 <sup>-15</sup> ↑	<b>0.00034</b> ↑	<b>1.6</b> · 10 <sup>-11</sup> ↑	0.44	0.21	0.65	0.22	2.3 · 10 <sup>-27</sup>
SBP → Stroke	<b>4.1</b> · 10 <sup>-9</sup> ↑	<b>9.3</b> · 10 <sup>-93</sup> ↑	<b>9.2</b> · 10 <sup>-25</sup> ↑	<b>5.1</b> · 10 <sup>-56</sup> ↑	<b>2.2</b> · 10 <sup>-7</sup> ↑	<b>3.4</b> · 10 <sup>-82</sup> ↑	0.14	0.16	0.77	0.33	1.8 · 10 <sup>-19</sup>
DBP → Stroke	<b>0.0011</b> ↑	<b>1.2</b> · 10 <sup>-77</sup> ↑	<b>1.1</b> · 10 <sup>-17</sup> ↑	<b>1.1</b> · 10 <sup>-54</sup> ↑	<b>2.6</b> · 10 <sup>-7</sup> ↑	<b>1.2</b> · 10 <sup>-73</sup> ↑	0.17	0.054	0.64	0.28	1.9 · 10 <sup>-11</sup>
Smoking → Stroke	<b>0.023</b> ↑	<b>0.00068</b> ↑	<b>0.0062</b> ↑	<b>0.026</b> ↑	0.21 ↑	<b>0.00073</b> ↑	0.47	0.28	0.27	0.22	5.8 · 10 <sup>-11</sup>
LDL → CAD	<b>6.3</b> · 10 <sup>-12</sup> ↑	<b>1.9</b> · 10 <sup>-79</sup> ↑	<b>5.6</b> · 10 <sup>-19</sup> ↑	<b>1.5</b> · 10 <sup>-50</sup> ↑	<b>4.2</b> · 10 <sup>-51</sup> ↑	<b>2.7</b> · 10 <sup>-52</sup> ↑	<b>0.87</b>	3.6 · 10 <sup>-57</sup>	0.79	0.19	2.9 · 10 <sup>-5</sup>
Smoking → T2D	0.67 ↑	0.11 ↑	0.23 ↓	1 ↓	0.86 ↑	0.077 ↑	0.05	0.69	0.05	0.04	0.19
LDL → Stroke	<b>0.046</b> ↑	<b>1.2</b> · 10 <sup>-5</sup> ↑	<b>0.00032</b> ↑	<b>0.0097</b> ↑	<b>0.002</b> ↑	<b>5.2</b> · 10 <sup>-6</sup> ↑	0.39	0.62	0.09	0.03	0.5
Supported by literature											
BMI → CAD	<b>0.00012</b> ↑	<b>6.9</b> · 10 <sup>-15</sup> ↑	0.12 ↑	<b>3.5</b> · 10 <sup>-17</sup> ↑	<b>4.9</b> · 10 <sup>-12</sup> ↑	<b>8.7</b> · 10 <sup>-18</sup> ↑	0.49	1.2 · 10 <sup>-9</sup>	0.56	0.23	3.1 · 10 <sup>-24</sup>
TG → CAD	0.085 ↑	<b>4.7</b> · 10 <sup>-15</sup> ↑	<b>0.0023</b> ↑	<b>2.4</b> · 10 <sup>-18</sup> ↑	<b>6.4</b> · 10 <sup>-8</sup> ↑	<b>7.3</b> · 10 <sup>-18</sup> ↑	<b>0.94</b>	7.8 · 10 <sup>-70</sup>	0.59	0.28	2.6 · 10 <sup>-20</sup>
BMI → T2D	<b>0.0048</b> ↑	<b>9.1</b> · 10 <sup>-14</sup> ↑	<b>0.0089</b> ↑	<b>2.5</b> · 10 <sup>-12</sup> ↑	<b>3.9</b> · 10 <sup>-9</sup> ↑	<b>3.9</b> · 10 <sup>-14</sup> ↑	0.49	0.002	0.54	0.34	4.3 · 10 <sup>-15</sup>
BF → CAD	0.33 ↑	0.52 ↑	0.93 ↓	<b>9.8</b> · 10 <sup>-5</sup> ↑	<b>8.6</b> · 10 <sup>-5</sup> ↑	0.78 ↑	0.09	0.053	0.05	0.26	5.2 · 10 <sup>-12</sup>
BF → T2D	1 ↑	<b>0.046</b> ↑	<b>0.026</b> ↑	<b>1.1</b> · 10 <sup>-8</sup> ↑	<b>1.1</b> · 10 <sup>-7</sup> ↑	<b>0.019</b> ↑	-0.46	0.0073	0.04	0.38	4.3 · 10 <sup>-9</sup>
FG → T2D	<b>0.013</b> ↑	<b>0.01</b> ↑	0.29 ↓	<b>0.0011</b> ↑	0.62 ↑	<b>1</b> · 10 <sup>-4</sup> ↑	0.07	0.85	0.29	0.62	7.8 · 10 <sup>-9</sup>
Height → CAD	<b>0.00018</b> ↓	<b>2.4</b> · 10 <sup>-14</sup> ↓	0.14 ↓	<b>2.8</b> · 10 <sup>-12</sup> ↓	<b>0.00052</b> ↓	<b>2</b> · 10 <sup>-17</sup> ↓	0.07	0.53	0.44	-0.1	1.1 · 10 <sup>-7</sup>

Traits	CAUSE	IVW	Egger	Wtd Med	Wtd Mode	MR-PRESSO	LCV GCP	LCV pval	CAUSE $q$	GC	GC pval
BMI → Stroke	0.23 ↑	0.12 ↑	0.73 ↓	0.59 ↑	0.81 ↑	0.1 ↑	0.48	0.19	0.07	0.13	$1 \cdot 10^{-4}$
BF → Stroke	1 ↑	0.88 ↓	0.7 ↓	0.93 ↓	0.94 ↑	0.83 ↓	-0.18	0.8	0.04	0.2	$7 \cdot 10^{-4}$
Smoking → Asthma	0.5 ↑	0.089 ↑	0.49 ↑	0.12 ↑	0.94 ↑	0.053 ↑	0.06	0.5	0.04	0.1	0.0036
Unknown or conflicting evidence											
SBP → T2D	<b>0.01</b> ↑	<b>1.2 · 10<sup>-12</sup></b> ↑	0.13 ↑	<b>2.5 · 10<sup>-7</sup></b> ↑	<b>0.044</b> ↑	<b>4.5 · 10<sup>-12</sup></b> ↑	-0.55	0.038	0.34	0.25	$2.8 \cdot 10^{-15}$
HDL → T2D	0.056 ↓	<b>0.0071</b> ↓	0.6 ↑	1 ↓	0.84 ↑	<b>0.025</b> ↓	<b>0.69</b>	$4 \cdot 10^{-5}$	0.1	-0.31	$2.1 \cdot 10^{-12}$
DBP → T2D	<b>0.02</b> ↑	<b>5.4 · 10<sup>-6</sup></b> ↑	0.25 ↑	<b>0.00011</b> ↑	<b>0.0041</b> ↑	<b>1.2 · 10<sup>-5</sup></b> ↑	-0.13	0.48	0.27	0.19	$2.7 \cdot 10^{-10}$
TG → T2D	0.15 ↑	0.38 ↑	0.52 ↑	0.23 ↑	0.22 ↑	0.32 ↑	0.44	0.34	0.13	0.28	$5.2 \cdot 10^{-10}$
BW → CAD	<b>0.0014</b> ↓	<b>0.0021</b> ↓	0.94 ↑	<b>0.0099</b> ↓	0.29 ↓	<b>0.00013</b> ↓	0.5	0.23	0.31	-0.16	$2.2 \cdot 10^{-8}$
BW → T2D	0.064 ↓	<b>0.00019</b> ↓	0.16 ↓	<b>0.00043</b> ↓	0.25 ↓	<b>6.8 · 10<sup>-5</sup></b> ↓	0.44	0.44	0.12	-0.26	$2.1 \cdot 10^{-7}$
TG → Stroke	0.99 ↑	0.84 ↓	0.1 ↓	0.8 ↑	0.95 ↑	0.9 ↓	-0.31	0.36	0.03	0.15	$3 \cdot 10^{-4}$
BMI → Asthma	0.16 ↑	<b>0.018</b> ↑	0.064 ↑	<b>0.0091</b> ↑	0.065 ↑	<b>0.017</b> ↑	0.28	0.21	0.1	0.09	0.0031
FG → CAD	0.25 ↑	<b>0.0011</b> ↑	0.24 ↑	<b>2.2 · 10<sup>-6</sup></b> ↑	<b>9.9 · 10<sup>-7</sup></b> ↑	<b>0.00025</b> ↑	0.48	0.18	0.04	0.12	0.0049
FG → Stroke	0.99 ↑	0.42 ↑	<b>0.031</b> ↓	0.37 ↓	0.41 ↓	0.37 ↑	-0.67	0.0093	0.04	0.22	0.0082
BF → Asthma	0.57 ↑	0.4 ↑	0.93 ↑	0.85 ↑	0.53 ↓	0.36 ↑	0.08	0.63	0.04	0.12	0.048
BW → Stroke	0.088 ↓	0.12 ↓	<b>0.039</b> ↑	0.11 ↓	0.21 ↓	0.052 ↓	-0.02	0.97	0.15	-0.08	0.11
Alcohol → Stroke	0.25 ↑	0.41 ↑	0.39 ↓	0.87 ↑	0.77 ↓	0.37 ↑	0.1	0.71	0.07	0.06	0.13
Height → Stroke	0.55 ↓	<b>0.034</b> ↓	0.53 ↓	0.24 ↓	0.91 ↓	0.077 ↓	-0.18	0.58	0.03	-0.05	0.15
LDL → T2D	0.32 ↓	<b>0.00017</b> ↓	<b>0.00068</b> ↓	<b>0.022</b> ↓	<b>0.0064</b> ↓	<b>5.5 · 10<sup>-5</sup></b> ↓	0.57	$4 \cdot 10^{-13}$	0.13	0.05	0.24
Alcohol → T2D	0.99 ↑	0.27 ↑	0.15 ↑	0.26 ↑	0.29 ↑	0.29 ↑	-0.24	0.51	0.03	-0.03	0.41
Alcohol → CAD	0.82 ↑	0.47 ↑	0.54 ↑	0.06 ↑	<b>0.036</b> ↑	0.12 ↑	-0.08	0.85	0.03	0	0.99
Implausible or unsupported											
Alcohol → Asthma	0.81 ↓	0.57 ↓	0.14 ↓	0.57 ↓	0.43 ↓	0.56 ↓	-0.08	0.43	0.03	-0.08	0.031
Height → Asthma	1 ↓	0.27 ↓	0.56 ↓	0.42 ↓	0.11 ↓	0.69 ↓	-0.07	0.62	0.03	-0.04	0.29
SBP → Asthma	0.76 ↑	<b>0.026</b> ↑	0.85 ↑	0.73 ↑	0.35 ↓	<b>0.035</b> ↑	-0.02	0.93	0.04	0.03	0.4

Traits	CAUSE	IVW	Egger	Wtd Med	Wtd Mode	MR-PRESSO	LCV GCP	LCV pval	CAUSE $q$	GC	GC pval
DBP → Asthma	0.59 ↑	0.79 ↑	0.086 ↓	0.2 ↓	0.15 ↓	0.88 ↓	-0.17	0.5	0.04	0.03	0.43
FG → Asthma	0.35 ↓	0.24 ↓	0.39 ↓	<b>0.025</b> ↓	0.082 ↓	0.091 ↓	0.2	0.099	0.06	-0.05	0.56
TG → Asthma	0.17 ↓	<b>0.03</b> ↓	0.51 ↓	0.93 ↓	0.34 ↓	0.065 ↓	0.02	0.88	0.06	0.03	0.58
Height → T2D	0.96 ↑	0.68 ↓	0.3 ↑	1 ↓	0.9 ↑	0.62 ↓	0	0.75	0.03	-0.01	0.7
LDL → Asthma	0.87 ↓	0.65 ↑	0.82 ↓	0.25 ↑	0.61 ↑	0.87 ↑	0.15	0.87	0.03	-0.01	0.75
BW → Asthma	0.73 ↑	0.18 ↑	0.85 ↑	0.14 ↑	0.42 ↑	0.082 ↑	-0.02	0.5	0.04	-0.01	0.83
HDL → Asthma	1 ↑	0.91 ↑	0.99 ↑	0.52 ↑	0.93 ↑	0.74 ↑	-0.02	0.76	0.03	-0.01	0.84
Considered non-causal											
HDL → CAD	<b>0.00044</b> ↓	<b>4.9 · 10<sup>-9</sup></b> ↓	0.55 ↓	<b>5.0 · 10<sup>-11</sup></b> ↓	<b>5.6 · 10<sup>-6</sup></b> ↓	<b>1.3 · 10<sup>-13</sup></b> ↓	<b>0.67</b>	0.0012	0.54	-0.26	4.8 · 10 <sup>-18</sup>
HDL → Stroke	0.27 ↓	0.076 ↓	0.48 ↑	0.07 ↓	0.13 ↓	0.096 ↓	<b>0.67</b>	3.5 · 10 <sup>-8</sup>	0.04	-0.15	2.3 · 10 <sup>-5</sup>