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

Jean Morrison *[1], Nicholas Knoblauch[1], Joseph Marcus[1], Matthew Stephens †[1,2], and Xin He ‡[1]

[1]Department of Human Genetics, University of Chicago
[2]Department of Statistics, University of Chicago

January 28, 2020

Abstract

Mendelian randomization (MR) is a valuable tool for detecting evidence of causal relationships using genetic variant associations. Opportunities to apply MR are growing rapidly with the number of genome-wide association studies (GWAS) with publicly available results. However, existing MR methods rely on strong assumptions that are often violated, leading to false positives. Many methods have been proposed loosening these assumptions. However, it has remained challenging to account for correlated pleiotropy, which arises when variants affect both traits through a heritable shared factor. 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 is more robust to correlated 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 many pairs of traits identified as causal using alternative methods may be false positives due to horizontal pleiotropy.

1 Introduction

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 a feasible alternative, but observational associations may be biased by confounding and reverse causality.

Mendelian randomization (MR) is a potentially powerful approach to studying causal relationships using data from observational genetic studies. The key idea of MR is to treat genotypes as naturally occurring "randomizations" [1, 2, 3, 4]. Suppose we are interested in the causal effect, γ, of M (for "Mediator") on Y. If a variant G_j affects M and doesn't affect Y through pathways not mediated by M (see Figure 1), then G_j can be viewed as an un-confounded proxy for M. Under these assumptions

$$\beta_{Y,j} = \gamma \beta_{M,j}, \tag{1}$$

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 simple MR methods. Many methods based on Equation (1), including the commonly used inverse variance weighted (IVW) regression, first obtain estimates of $\beta_{Y,j}$ and $\beta_{M,j}$ for several variants strongly associated with M, and then estimate γ by regressing the estimates of $\beta_{Y,j}$ on the estimates of $\beta_{M,j}$ [5].

A variant that affects M and also affects Y through a pathway not mediated by a causal effect of M on Y is termed *horizontally pleiotropic*. Historically, horizontal pleiotropy was assumed to be rare because it was thought that each variant had few biological effects and that most confounders (U in Figures 1)

*jeanm@uchicago.edu
†mstephens@uchicago.edu
‡xinhe@uchicago.edu

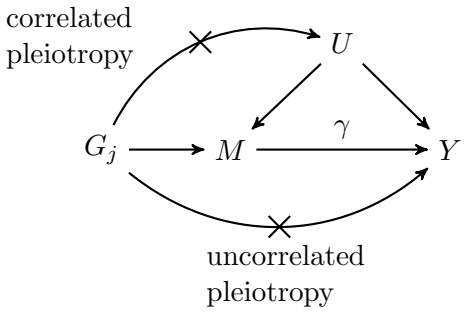


Figure 1: Causal diagram assumed by simple MR. Arrows indicate causal effects. Crosses mark causal effects that are assumed absent by simple MR methods. The causal effect of trait M on trait Y, γ, is the target of inference. Variant G_j affects trait M but is assumed to have no effects on Y that are not mediated through M. Pleiotropic effects on Y that are independent of the effect of G_j on M are referred to as uncorrelated pleiotropy. Effects that are mediated by a shared factor, U, are referred to as correlated pleiotropy.

were non-heritable environmental risks. It has subsequently become clear that horizontal pleioptropy is common and leads to MR false positives if it is not accounted for [6, 7]. Accounting for these effects has been a major focus of ongoing MR research [4]. It is helpful to distinguish between two types of horizontal pleiotropy: *uncorrelated pleiotropy*, where horizontal pleiotropic effects on Y are uncorrelated with effects on M, and *correlated pleiotropy*, where horizontal pleiotropic effects on Y are correlated with effects on M. Uncorrelated pleiotropy occurs when a genetic variant affects Y and M through separate mechanisms, whereas correlated pleiotropy occurs when a genetic variant affects Y and M through a shared heritable factor, U (Figure 1). U may represent a shared biological process or pathway. Both types of pleiotropy may occur for any pair of traits.

Of the two types, uncorrelated horizontal pleiotropy is easier to account for. If it is mean zero, it only adds noise to the relationship in Equation (1) rather than inducing a systematic correlation between $\beta_{Y,j}$ and $\beta_{M,j}$. Several methods have recently been developed to deal with this, including Egger regression [8, 9] which allows for directional, or non-mean zero uncorrelated pleiotropy, and several methods that rely on outlier removal including GSMR [10] and MR-PRESSO [7]. However, because these methods do not model correlated pleiotropy, they remain vulnerable to false positives in its presence. Genetic correlation has been found between many pairs of traits including some that are unlikely to be causally linked, which suggests that correlated pleiotropy is common and may be an important source of false positives in MR analyses [11, 12, 13, 14]. If a potential shared factor is suspected and measured, false positives can be reduced by eliminating variants associated with the shared factor or adjusting for it using multivariable MR [15]. Two methods, the weighted median approach [16] and the weighted mode approach [17], allow for some correlated pleiotropy. These methods rely respectively on the assumption that fewer than 50% of variants display any form of horizontal pleiotropy and that mode of pleiotropic effects is zero. Another method, LCV [14], does not directly test for a causal effect, but estimates the "genetic causality proportion", with higher magnitude estimates suggestive of a causal effect and lower magnitude estimates suggestive of correlated pleiotropy.

We present a new MR method that accounts for both uncorrelated and correlated pleiotropy. Causal Analysis Using Summary Effect Estimates (CAUSE) uses genome-wide summary statistics to identify pairs of traits that are consistent with a causal effect. 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 (Figure 2b). On the other hand, if M has no causal effect on Y, but both traits are affected by a shared factor, then only the subset of M effect variants acting on the shared factor will have correlated estimates (Figure 2a). CAUSE uses this distinction to differentiate causal effects from correlated pleiotropy. In addition to accounting for correlated pleiotropy, CAUSE differs from other MR methods by using information from all variants, rather than only those most strongly associated with M. This can improve power when the GWAS for trait M is under-powered. 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 are likely to be acting through a shared factor.

We demonstrate in simulations that CAUSE makes fewer false detections in the presence of correlated

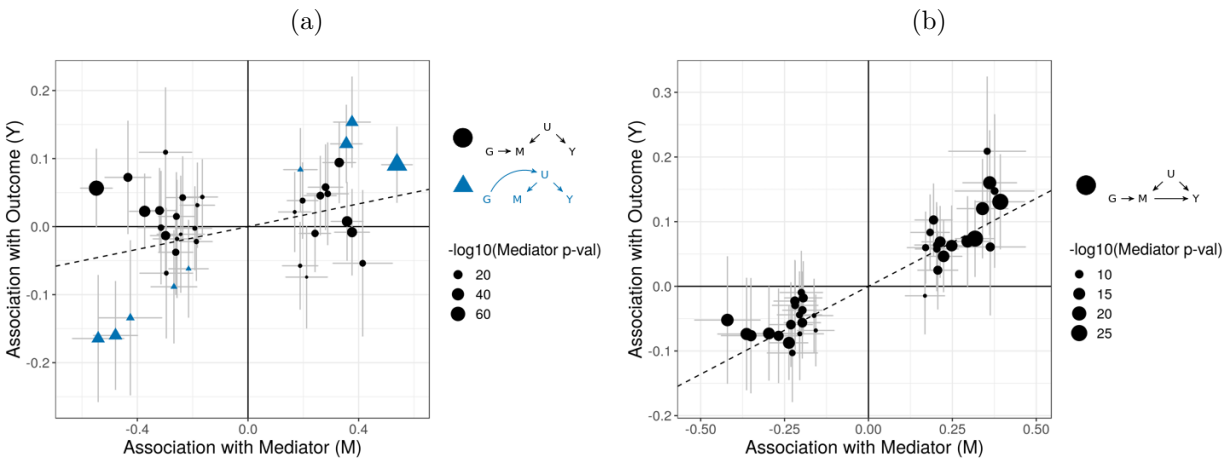


Figure 2: Simulated data illustrating (a) the pattern induced by a shared factor (correlated pleiotropy) without a causal effect; and (b) the pattern induced by a causal effect. In both plots, effect size estimates are indicated by points. Error bars around points have length 1.96 times the standard error on each side. Only variants that are strongly associated with M ($p < 5 \cdot 10^{-8}$) are shown. (a) Summary statistics are simulated for two traits with no causal effect. 15% of the M effect variants act on M through a shared factor that affects both M and Y (triangles). Although there are many variants associated with M and not with Y, the IVW estimator (dotted line) is significantly different from zero ($p = 0.01$), driven by the effect size correlation of the 15% pleiotropic variants (triangles). (b) Summary statistics are simulated under a model where M causally affects Y. Here *every* variant associated with M is also associated with Y and the effect sizes are 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.

2 Results

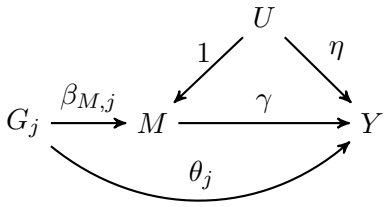
2.1 CAUSE models uncorrelated and correlated pleiotropy

Using CAUSE, we assess whether genome-wide variant effect estimates and standard errors 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 1. In our proposal, illustrated in Figure 3, 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 causal diagram in Figure 3a, which is the same as the diagram in Figure 1 with the addition of an uncorrelated pleiotropic effect. We allow a small proportion, q, of M effect variants to follow the causal diagram in Figure 3b, acting on both M and Y through an unobserved heritable shared factor, U. Under this model, the relationship between $\beta_{Y,j}$ and $\beta_{M,j}$ is

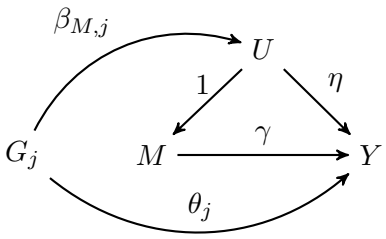
$$\beta_{Y,j} = \underbrace{\gamma\beta_{M,j}}_{\text{causal effect}} + \underbrace{Z_j\eta\beta_{M,j}}_{\substack{\text{correlated}\\ \text{pleiotropy}}} + \underbrace{\theta_j}_{\substack{\text{uncorrelated}\\ \text{pleiotropy}}}, \tag{2}$$

where Z_j is an indicator that is 1 if G_j acts on M through U and 0 otherwise, and η 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) allows us to model the patterns in Figure 2. 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 η non-zero), then $\beta_{Y,j}$ and $\beta_{M,j}$ are correlated for variants with $Z_j = 1$ (Figure 2a). If there is a causal effect ($\gamma \neq 0$), $\beta_{Y,j}$ and $\beta_{M,j}$ are correlated for all variants (Figure 2b). Including all variants allows us to model uncertainty about variant effects on M and gain information from weakly associated



(a) Causal diagram for variants that are independent of the shared factor U.



(b) Causal diagram for variants that act on M through U.

Figure 3: CAUSE assumes that variants affect trait M through one of two mechanisms. A proportion $1-q$ of variants have the causal diagram in (a), while the remaining proportion, q, have the causal diagram in (b). (a) G_j has an effect on M of size $\beta_{M,j}$ and is independent of U. G_j can affect Y through two pathways: through the uncorrelated effect, θ_j, and through the effect of M on Y (if $\gamma \neq 0$). (b) The effect of G_j on M is mediated by U. We have assumed that U is scaled so that its effect on M is 1. This makes the effect of G_j on U equal to the effect of G_j on M, $\beta_{M,j}$. The effect of U on Y is equal to η, an unknown parameter. G_j can affect Y through three pathways: through θ_j, through the causal effect of M on Y, and through the effect of U on Y.

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 γ fixed at 0, allowing for pleiotropic effects but no causal effect, and the *causal model* in which γ is a free parameter. We compare the models using the expected log pointwise posterior density (ELPD; [18]), 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 causal diagram in Figure 3b is related to the model used in the latent causal variable (LCV) method proposed by O'Connor and Price [14] (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 [14] consider a large magnitude estimated GCP more likely to be causal and use a threshold of 0.6 to identify suggestive pairs.

2.2 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 (median 8 genome-wide significant loci) and a high power (median 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 $p < 5 \cdot 10^{-8}$. Variant effects on trait Y are generated from the relationship in Equation (2) and effect estimates are then simulated from true effects and LD structure using results of Zhu and Stephens [19] (see Methods).

We compare CAUSE to six MR methods: IVW regression using multiplicative random effects [15], Egger regression with random effects [8], GSMR [10], MR-PRESSO [7], the weighted median estimator [16], and the modal estimator [17]. IVW regression, Egger regression, GSMR and MR-PRESSO assume that no variants exhibit correlated pleiotropy. The weighted median and modal methods are more permissive, assuming respectively that fewer than half the variants are pleiotropic and that the modal pleiotropic effect is zero. All six alternative methods require selection of variants with strong evidence of association with

trait M. We follow the typical practice of 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 (Figure 4a). We consider weak ($\eta = \sqrt{0.02}$) and strong ($\eta = \sqrt{0.05}$) shared factor effects. CAUSE controls the false positive rate in the presence of correlated pleiotropy ($q > 0$) better than other methods except for Egger regression and the modal estimator. CAUSE makes more false detections when q is large, the confounder effect is large, and the GWAS have higher 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.

We next compare the power of each method when there is a true causal effect of M on Y and no shared factor (Figure 4b). CAUSE has substantially better power than Egger regression and the modal estimator, the two methods that were able to 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 M GWAS has low power and trait Y has high power, CAUSE can achieve better power than other methods for larger causal effects. This is a result of using all variants genome-wide, rather than only those reaching genome-wide significance. The posterior median of γ 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 η has a smaller magnitude than γ but loses power when η is large and has opposite sign to γ (Supplementary Note Section SN6.3).

Although in most 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 4c 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% (Figure S2, Table S3). 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).

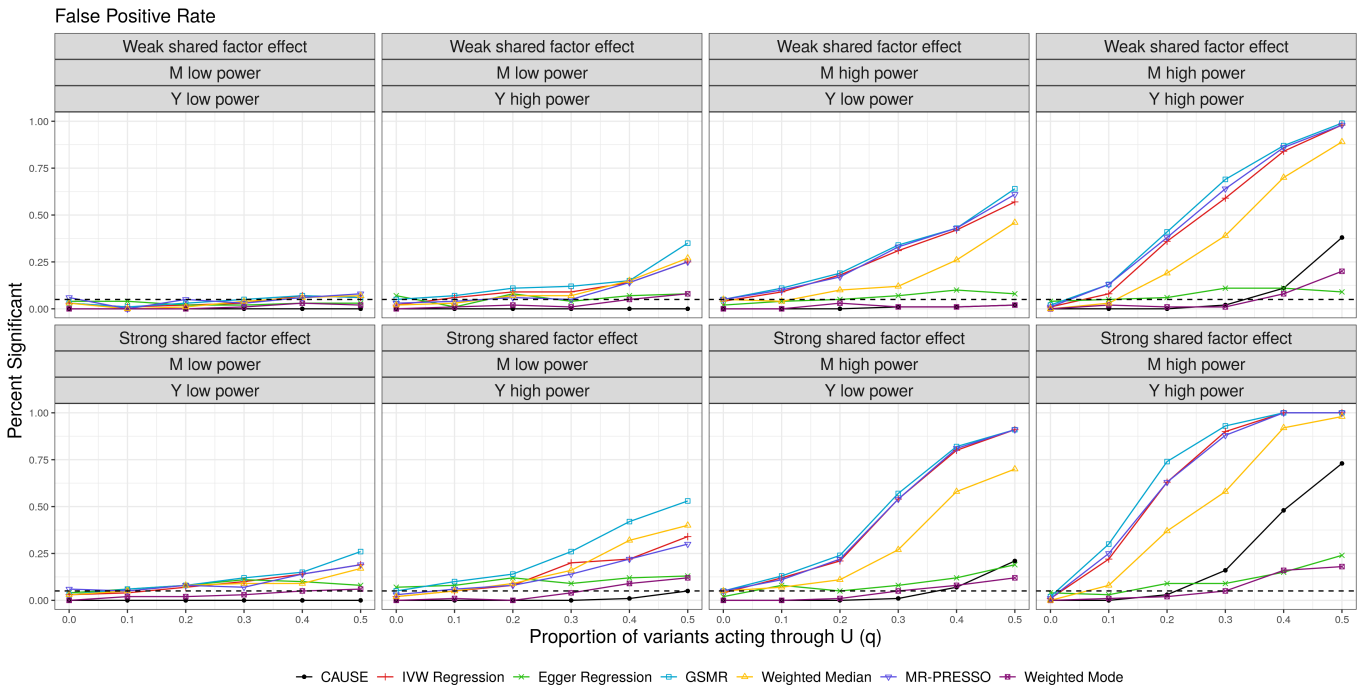
2.3 CAUSE avoids false positives due to reverse causality

Many MR methods obtain false positives when there is a true causal effect of Y on M but a test is performed for an effect of M on Y (reverse causal effects). This occurs because a reverse effect creates a similar pattern of variant associations to the pattern created by a shared factor. 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. Since CAUSE models correlated pleiotropy, it can avoid false positives due to reverse causality as long as there are a sufficient number of M effect variants that are independent of the causal path from Y to M.

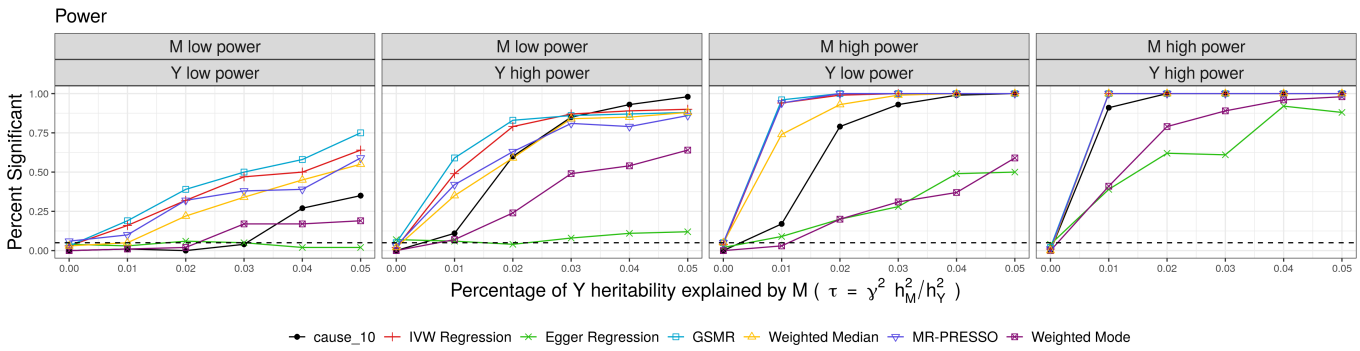
To verify this expectation, we simulate data with a causal effect of Y on M and no shared factor and test for an effect of M on Y using CAUSE and other methods. In all scenarios, the expected heritability of both M and Y is 0.25. For trait Y, 1000 variants with normally distributed effects explain this 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 1000 variants (see Methods). CAUSE and the modal estimator both avoid false positives in all settings while the other methods do not (Figure 5). More false positives from other methods occur when the number of independent M effect variants is low. In these scenarios, many of the variants significantly associated with M act on M through Y and thus have correlated effects. Some false positives from alternative methods could be avoided by carefully filtering variants, for example using Steiger filtering [20]. This analysis demonstrates that CAUSE is able to avoid false positives from reverse effects without filtering, when there are independent variants associated with M independent of Y.

CAUSE may still find evidence of causality in both directions for some pairs of traits 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.

(a) False Positive Rate



(b) Power



(c) False Positive Rate vs Power

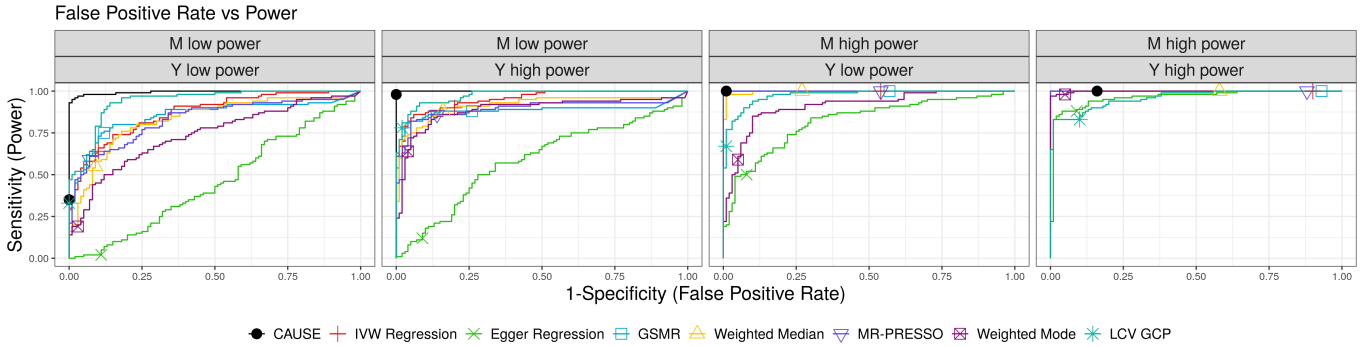


Figure 4: Performance of CAUSE and other MR methods in simulated data. (a) False positive rate averaged over 100 simulated data sets in settings with no causal effect and a proportion of correlated pleiotropic variants (q) ranging from 0 to 50%. (b) Power averaged over 100 simulated data sets in settings 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 there is no shared factor to the false positive rate when there is no causal effect, but a proportion $q = 0.3$ of variants act through a shared factor with effect $\eta = \sqrt{0.05}$ on Y. 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 $p \leq 0.05$ or $G\hat{C}P0.06$ for LCV.

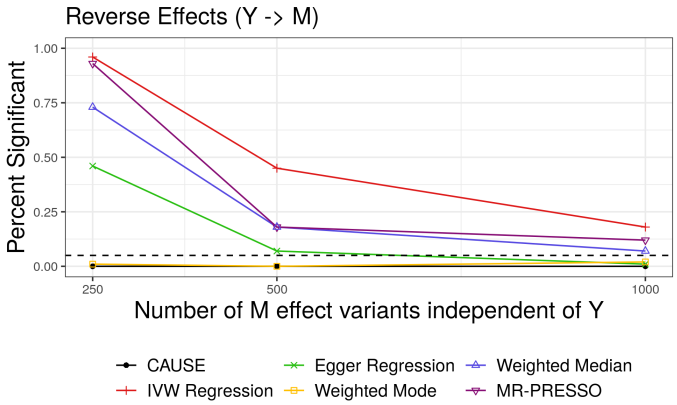


Figure 5: False positives resulting from reverse causality. Only CAUSE and the weighted mode control the false positive rate.

2.4 Identifying causal risk factors of common diseases

We use CAUSE and other methods to test for effects of twelve possible risk factors for cardio-metabolic diseases coronary artery disease (CAD), stroke, and type 2 diabetes (T2D). These include anthropometric traits, lipids, blood pressure, fasting glucose, smoking, and alcohol consumption (Table S1). In order to increase the number of tested relationships that are unlikely to be causal, we also test for effects of each mediator 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 (see also Figure S3 and Table S5). Full lists of variants used for each MR method are provided in Tables S9 and S10. We classified trait pairs in Table 1 by evidence of causality from existing literature (see Supplementary Note Section SN7) and ordered them by strength of evidence for genetic correlation. 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 trait 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 as suggestive of a causal relationship, however many pairs have moderate GCP estimates that are significantly different from zero. Variant estimates for smoking and CAD, a strong causal effect detected by all methods, are shown in Figure 6a 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, suggesting that it is avoiding more false positives. All methods except Egger regression detect an effect of HDL cholesterol on CAD risk, which is likely to be an error[21]. 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 (Figure 6b). The true relationship between this pair of traits is unknown. Levels of LDL cholesterol are often elevated in individuals with T2D, an effect that is sometimes considered a consequence of insulin insensitivity [22] (see Supplementary Note Section SN7). The effects estimated by IVW regression and other methods are negative, opposite of what might be expected from observational associations, but consistent with previous findings of Fall et al. [23]. 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, making reverse direction analyses important to consider. At a p-value threshold of 0.05, CAUSE identifies five reverse pairs. These are effects of CAD and stroke on diastolic and systolic blood pressure and of T2D on fasting glucose. IVW, Egger regression, the weighted median, the weighted mode, and MR-PRESSO identify 15, 6, 11, 7, and 17 reverse effects respectively (Table S6, Figure S4). The case of CAD and LDL cholesterol provides an example of how CAUSE is able to avoid false positives from reverse effects (Figure 6c, full results in Table S6).

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 M GWAS is under-powered. In order to explore the effect of GWAS power, we compared results using an earlier GWAS for blood pressure that has a much smaller sample size [24]. With these data, IVW regression and Egger regression are unable to detect the effect of blood pressure on CAD. CAUSE is able to detect the effect by incorporating information from less significant variants (Supplementary Note Section SN9). This suggests that CAUSE may be able to boost power to detect causal effects of mediators with under-powered GWAS.

2.5 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 they may be accessible as drug targets and may suggest molecular and 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: allergic disease, asthma, rhumatoid arthritis, inflammatory bowel disease, and systemic lupus erythematosus (Table S2). Results are summarized in Figure 7 and Table S4 (full results in Tables S7 and S8).

The blood cell composition traits can be divided into red blood cell traits, white blood cell traits, and platelet traits. There is less literature information about which relationships are expected to be causal than in the previous application. However, we expect that relationships between immune-mediated disorders and immune related white blood cell traits are more plausible than relationships with red blood cell or platelet traits.

CAUSE identifies only 4 pairs of traits as consistent with a causal effect at a p-value threshold of 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 [25]. All other methods, including Egger regression and the weighted mode obtain many more positive results than CAUSE, including many causal relationships of red blood cell and platelet traits which are likely false positives. Figure 6d shows a pair of traits, immature reticulocyte fraction (IRF, a red blood cell trait) and inflammatory bowel disease (IBD), that are only detected by methods designed to be robust to pleiotropy, the weighted median ($p = 1.2 \cdot 10^{-5}$), the weighted mode ($p = 2.6 \cdot 10^{-5}$), and MR-PRESSO ($p = 0.0015$). 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). However, by removing or down-weighting the contribution of some variants, these methods can obtain false positives by discounting conflicting evidence.

3 Discussion

We have introduced CAUSE, a new approach to MR analysis that accounts for uncorrelated and correlated pleiotropy. CAUSE has several advantages over existing methods, most notably reduction of false positives. Previous authors have identified uncorrelated horizontal pleiotropy as a pervasive phenomenon that may adversely impact MR analyses [7]. Our analyses demonstrate that correlated horizontal pleiotropy is also common and may explain a substantial fraction of positive results obtained using other MR methods. Alternative MR methods perform best when information is available to guide removal of pleiotropic variants variants. For example, in analyzing HDL cholesterol and heart disease, Voight et al. avoid a false positive by discarding variants associated with other lipid and metabolic traits. CAUSE, on the other hand, is well suited for applications where it is difficult to predict shared factors, shared factors cannot be measured, or

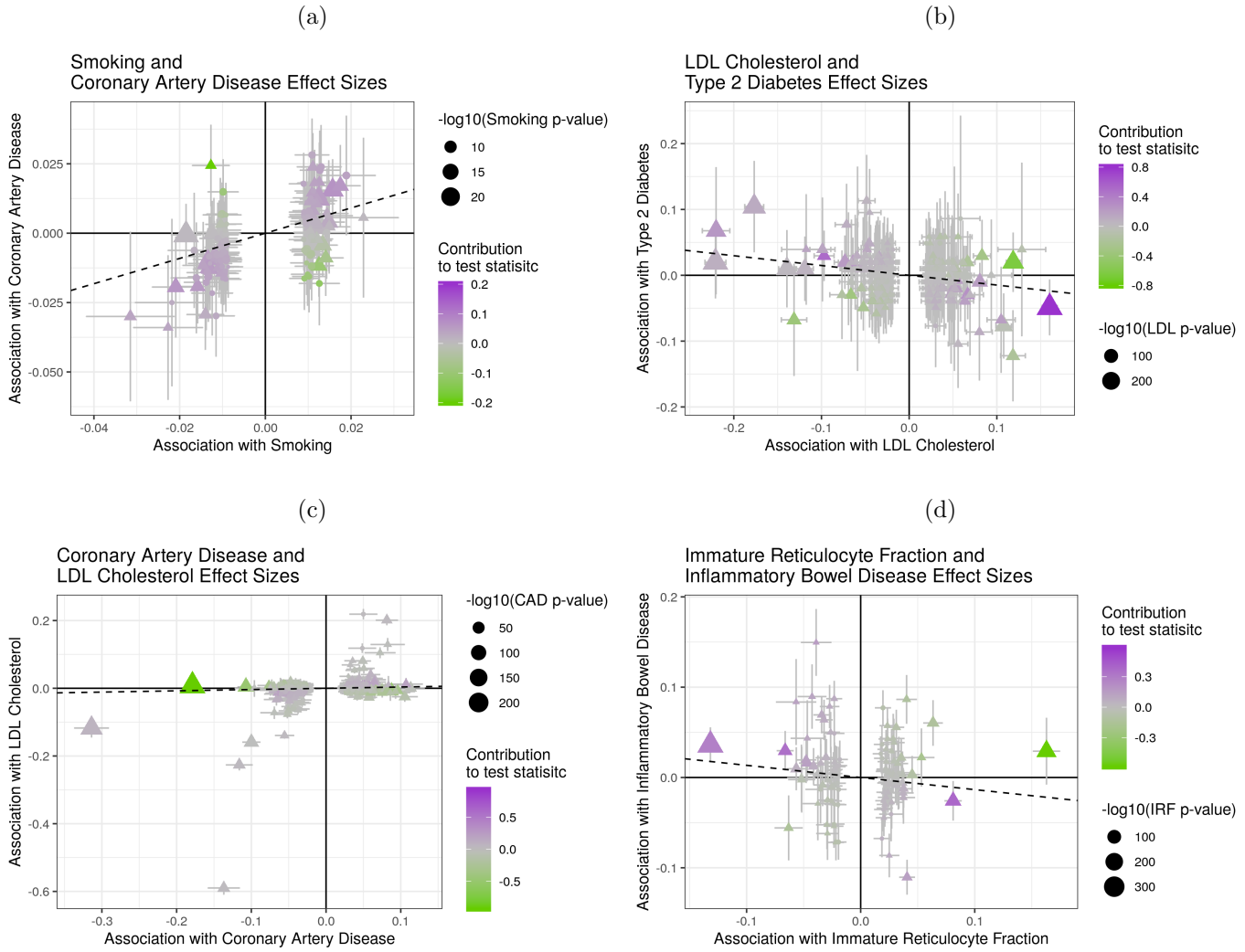


Figure 6: Effect size estimates and variant level contribution to CAUSE test statistics for four trait pairs. Effect estimates for trait M (horizontal axis) are plotted against estimates for trait Y (vertical axis). Error bars have length 1.96 times the standard error of the estimate. Triangles indicate 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). CAUSE does not detect a causal effect. Under the CAUSE model, these data can be explained by a shared factor associated with 13% of LDL effect variants. (c) CAD (M) and LDL cholesterol (Y). Some strong CAD variants have correlated effects on LDL, likely acting on CAD through effects on LDL. These lead to false positives for Egger regression, the weighted mode, and MR-PRESSO ($p = 4.8 \cdot 10^{-5}, 0.0068$ and 0.0013 respectively). CAUSE avoids a false positive ($p = 0.11$) by considering evidence from the many strong CAD variants that are not associated with LDL. IVW regression and the weighted median achieve p-values of 0.065 and 0.22. (d) IRF (M) and IBD (Y). MR-PRESSO, the weighted median, and modal estimators obtain a positive result by down-weighting or removing variants supplying conflicting evidence.

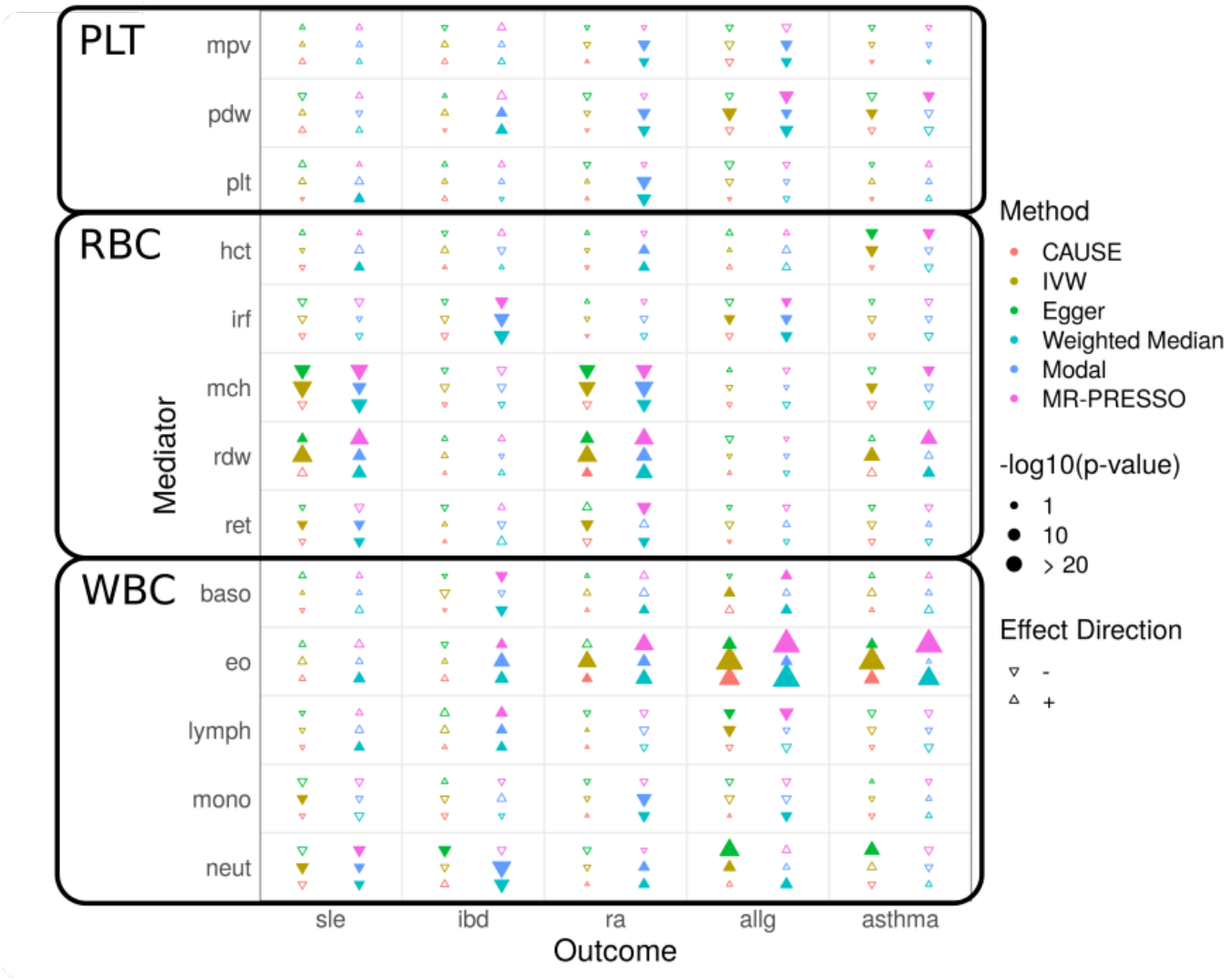


Figure 7: Tests for causal effects of blood cell composition on immune mediated traits. Each cell summarizes the the results of six methods for a pair of traits. Filled symbols indicate p-value < 0.05. Blood cell traits are grouped into platelet traits (PLT), red blood cell traits (RBC), and white blood cell traits (WBC).

the researcher would like to screen a large number of potentially causal variables in an automated manner.

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 in which 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 trait directions. With this in mind, it is good practice to test trait pairs in both directions [13]. We also note that while CAUSE is more robust than alternatives, it still has a high false positive rate when the shared factor has a strong effect and accounts for a large proportion of the trait M variants. Thus, a low p-value from CAUSE (as with other MR methods) 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.

CAUSE has several limitations that provide interesting future research directions. First, it is not currently 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 simply prunes variants for LD rather than explicitly modeling variant correlation. This helps ensure that the problem is computationally tractable, however, using only one variant per LD block may lead to loss of information.

Traits	CAUSE	IVW	Egger	Weighted Median	Weighted Mode	MR-PRESSO	LCV	CAUSE sharing q	Genetic Correlation
Considered Causal									
SBP → CAD	$\mathbf{6.6 \cdot 10^{-31}} \uparrow$	$\mathbf{1.8 \cdot 10^{-148}} \uparrow$	$\mathbf{0} \uparrow$	$\mathbf{1.8 \cdot 10^{-157}} \uparrow$	$\mathbf{3.8 \cdot 10^{-8}} \uparrow$	$\mathbf{1.3 \cdot 10^{-167}} \uparrow$	$0.02(0.95)$	0.91	$0.35(2.1 \cdot 10^{-49})$
DBP → CAD	$\mathbf{1.3 \cdot 10^{-26}} \uparrow$	$\mathbf{8.1 \cdot 10^{-111}} \uparrow$	$\mathbf{0} \uparrow$	$\mathbf{6.4 \cdot 10^{-138}} \uparrow$	$\mathbf{1.6 \cdot 10^{-7}} \uparrow$	$\mathbf{1.7 \cdot 10^{-132}} \uparrow$	$0.54(0.069)$	0.88	$0.27(1.3 \cdot 10^{-27})$
Smoking → CAD	$\mathbf{9.2 \cdot 10^{-8}} \uparrow$	$\mathbf{5.4 \cdot 10^{-11}} \uparrow$	$\mathbf{0.014} \uparrow$	$\mathbf{1.7 \cdot 10^{-15}} \uparrow$	$\mathbf{0.00034} \uparrow$	$\mathbf{1.6 \cdot 10^{-11}} \uparrow$	$0.44(0.21)$	0.65	$0.22(2.3 \cdot 10^{-27})$
SBP → Stroke	$\mathbf{4.1 \cdot 10^{-9}} \uparrow$	$\mathbf{9.3 \cdot 10^{-93}} \uparrow$	$\mathbf{0} \uparrow$	$\mathbf{5.1 \cdot 10^{-56}} \uparrow$	$\mathbf{2.2 \cdot 10^{-7}} \uparrow$	$\mathbf{3.4 \cdot 10^{-82}} \uparrow$	$0.14(0.16)$	0.77	$0.33(1.8 \cdot 10^{-19})$
DBP → Stroke	$\mathbf{0.0011} \uparrow$	$\mathbf{1.2 \cdot 10^{-77}} \uparrow$	$\mathbf{0} \uparrow$	$\mathbf{1.1 \cdot 10^{-54}} \uparrow$	$\mathbf{2.6 \cdot 10^{-7}} \uparrow$	$\mathbf{1.2 \cdot 10^{-73}} \uparrow$	$0.17(0.054)$	0.64	$0.28(1.9 \cdot 10^{-11})$
Smoking → Stroke	$\mathbf{0.023} \uparrow$	$\mathbf{0.00068} \uparrow$	$\mathbf{0.0062} \uparrow$	$\mathbf{0.026} \uparrow$	$0.21 \uparrow$	$\mathbf{0.00073} \uparrow$	$0.47(0.28)$	0.27	$0.22(5.8 \cdot 10^{-11})$
LDL → CAD	$\mathbf{6.3 \cdot 10^{-12}} \uparrow$	$\mathbf{1.9 \cdot 10^{-79}} \uparrow$	$\mathbf{0} \uparrow$	$\mathbf{1.5 \cdot 10^{-50}} \uparrow$	$\mathbf{4.2 \cdot 10^{-51}} \uparrow$	$\mathbf{2.7 \cdot 10^{-52}} \uparrow$	$\mathbf{0.87(3.6 \cdot 10^{-57})}$	0.79	$0.19(2.9 \cdot 10^{-5})$
Smoking → T2D	$0.67 \uparrow$	$0.11 \uparrow$	$0.23 \downarrow$	$1 \downarrow$	$0.86 \uparrow$	$0.077 \uparrow$	$0.05(0.69)$	0.05	$0.04(0.19)$
LDL → Stroke	$\mathbf{0.046} \uparrow$	$\mathbf{1.2 \cdot 10^{-5}} \uparrow$	$\mathbf{0.00032} \uparrow$	$\mathbf{0.0097} \uparrow$	$\mathbf{0.002} \uparrow$	$\mathbf{5.2 \cdot 10^{-6}} \uparrow$	$0.39(0.62)$	0.09	$0.03(0.5)$
Supported by Literature									
BMI → CAD	$\mathbf{0.00012} \uparrow$	$\mathbf{6.9 \cdot 10^{-15}} \uparrow$	$0.12 \uparrow$	$\mathbf{3.5 \cdot 10^{-17}} \uparrow$	$\mathbf{4.9 \cdot 10^{-12}} \uparrow$	$\mathbf{8.7 \cdot 10^{-18}} \uparrow$	$0.49(1.2 \cdot 10^{-9})$	0.56	$0.23(3.1 \cdot 10^{-24})$
TG → CAD	$0.085 \uparrow$	$\mathbf{4.7 \cdot 10^{-15}} \uparrow$	$\mathbf{0.0023} \uparrow$	$\mathbf{2.4 \cdot 10^{-18}} \uparrow$	$\mathbf{6.4 \cdot 10^{-8}} \uparrow$	$\mathbf{7.3 \cdot 10^{-18}} \uparrow$	$\mathbf{0.94(7.8 \cdot 10^{-70})}$	0.59	$0.28(2.6 \cdot 10^{-20})$
BMI → T2D	$\mathbf{0.0048} \uparrow$	$\mathbf{9.1 \cdot 10^{-14}} \uparrow$	$\mathbf{0.0089} \uparrow$	$\mathbf{2.5 \cdot 10^{-12}} \uparrow$	$\mathbf{3.9 \cdot 10^{-9}} \uparrow$	$\mathbf{3.9 \cdot 10^{-14}} \uparrow$	$0.49(0.002)$	0.54	$0.34(4.3 \cdot 10^{-15})$
BF → CAD	$0.33 \uparrow$	$0.52 \uparrow$	$0.93 \downarrow$	$\mathbf{9.8 \cdot 10^{-5}} \uparrow$	$\mathbf{8.6 \cdot 10^{-5}} \uparrow$	$0.78 \uparrow$	$0.09(0.053)$	0.05	$0.26(5.2 \cdot 10^{-12})$
BF → T2D	$1 \uparrow$	$\mathbf{0.046} \uparrow$	$\mathbf{0.026} \uparrow$	$\mathbf{1.1 \cdot 10^{-8}} \uparrow$	$\mathbf{1.1 \cdot 10^{-7}} \uparrow$	$\mathbf{0.019} \uparrow$	$-0.46(0.0073)$	0.04	$0.38(4.3 \cdot 10^{-9})$
FG → T2D	$\mathbf{0.013} \uparrow$	$\mathbf{0.01} \uparrow$	$0.29 \downarrow$	$\mathbf{0.0011} \uparrow$	$0.62 \uparrow$	$\mathbf{1 \cdot 10^{-4}} \uparrow$	$0.07(0.85)$	0.29	$0.62(7.8 \cdot 10^{-9})$
Height → CAD	$\mathbf{0.00018} \downarrow$	$\mathbf{2.4 \cdot 10^{-14}} \downarrow$	$0.14 \downarrow$	$\mathbf{2.8 \cdot 10^{-12}} \downarrow$	$\mathbf{0.00052} \downarrow$	$\mathbf{2 \cdot 10^{-17}} \downarrow$	$0.07(0.53)$	0.44	$-0.1(1.1 \cdot 10^{-7})$
BMI → Stroke	$0.23 \uparrow$	$0.12 \uparrow$	$0.73 \downarrow$	$0.59 \uparrow$	$0.81 \uparrow$	$0.1 \uparrow$	$0.48(0.19)$	0.07	$0.13(1 \cdot 10^{-4})$
BF → Stroke	$1 \uparrow$	$0.88 \downarrow$	$0.7 \downarrow$	$0.93 \downarrow$	$0.94 \uparrow$	$0.83 \downarrow$	$-0.18(0.8)$	0.04	$0.2(7 \cdot 10^{-4})$
Smoking → Asthma	$0.5 \uparrow$	$0.089 \uparrow$	$0.49 \uparrow$	$0.12 \uparrow$	$0.94 \uparrow$	$0.053 \uparrow$	$0.06(0.5)$	0.04	$0.1(0.0036)$
Unknown or Conflicting Evidence									
SBP → T2D	$\mathbf{0.01} \uparrow$	$\mathbf{1.2 \cdot 10^{-12}} \uparrow$	$0.13 \uparrow$	$\mathbf{2.5 \cdot 10^{-7}} \uparrow$	$\mathbf{0.044} \uparrow$	$\mathbf{4.5 \cdot 10^{-12}} \uparrow$	$-0.55(0.038)$	0.34	$0.25(2.8 \cdot 10^{-15})$
HDL → T2D	$0.056 \downarrow$	$\mathbf{0.0071} \downarrow$	$0.6 \uparrow$	$1 \downarrow$	$0.84 \uparrow$	$\mathbf{0.025} \downarrow$	$\mathbf{0.69(4 \cdot 10^{-5})}$	0.1	$-0.31(2.1 \cdot 10^{-12})$
DBP → T2D	$\mathbf{0.02} \uparrow$	$\mathbf{5.4 \cdot 10^{-6}} \uparrow$	$0.25 \uparrow$	$\mathbf{0.00011} \uparrow$	$\mathbf{0.0041} \uparrow$	$\mathbf{1.2 \cdot 10^{-5}} \uparrow$	$-0.13(0.48)$	0.27	$0.19(2.7 \cdot 10^{-10})$
TG → T2D	$0.15 \uparrow$	$0.38 \uparrow$	$0.52 \uparrow$	$0.23 \uparrow$	$0.22 \uparrow$	$0.32 \uparrow$	$0.44(0.34)$	0.13	$0.28(5.2 \cdot 10^{-10})$
BW → CAD	$\mathbf{0.0014} \downarrow$	$\mathbf{0.0021} \downarrow$	$0.94 \uparrow$	$\mathbf{0.0099} \downarrow$	$0.29 \downarrow$	$\mathbf{0.00013} \downarrow$	$0.5(0.23)$	0.31	$-0.16(2.2 \cdot 10^{-8})$
BW → T2D	$0.064 \downarrow$	$\mathbf{0.00019} \downarrow$	$0.16 \downarrow$	$\mathbf{0.00043} \downarrow$	$0.25 \downarrow$	$\mathbf{6.8 \cdot 10^{-5}} \downarrow$	$0.44(0.44)$	0.12	$-0.26(2.1 \cdot 10^{-7})$
TG → Stroke	$0.99 \uparrow$	$0.84 \downarrow$	$0.1 \downarrow$	$0.8 \uparrow$	$0.95 \uparrow$	$0.9 \downarrow$	$-0.31(0.36)$	0.03	$0.15(3 \cdot 10^{-4})$
BMI → Asthma	$0.16 \uparrow$	$\mathbf{0.018} \uparrow$	$0.064 \uparrow$	$\mathbf{0.0091} \uparrow$	$0.065 \uparrow$	$\mathbf{0.017} \uparrow$	$0.28(0.21)$	0.1	$0.09(0.0031)$
FG → CAD	$0.25 \uparrow$	$\mathbf{0.0011} \uparrow$	$0.24 \uparrow$	$\mathbf{2.2 \cdot 10^{-6}} \uparrow$	$\mathbf{9.9 \cdot 10^{-7}} \uparrow$	$\mathbf{0.00025} \uparrow$	$0.48(0.18)$	0.04	$0.12(0.0049)$
FG → Stroke	$0.99 \uparrow$	$0.42 \uparrow$	$\mathbf{0.031} \downarrow$	$0.37 \downarrow$	$0.41 \downarrow$	$0.37 \uparrow$	$-0.67(0.0093)$	0.04	$0.22(0.0082)$
BF → Asthma	$0.57 \uparrow$	$0.4 \uparrow$	$0.93 \uparrow$	$0.85 \uparrow$	$0.53 \downarrow$	$0.36 \uparrow$	$0.08(0.63)$	0.04	$0.12(0.048)$
BW → Stroke	$0.088 \downarrow$	$0.12 \downarrow$	$\mathbf{0.039} \uparrow$	$0.11 \downarrow$	$0.21 \downarrow$	$0.052 \downarrow$	$-0.02(0.97)$	0.15	$-0.08(0.11)$
Alcohol → Stroke	$0.25 \uparrow$	$0.41 \uparrow$	$0.39 \downarrow$	$0.87 \uparrow$	$0.77 \downarrow$	$0.37 \uparrow$	$0.1(0.71)$	0.07	$0.06(0.13)$
Height → Stroke	$0.55 \downarrow$	$\mathbf{0.034} \downarrow$	$0.53 \downarrow$	$0.24 \downarrow$	$0.91 \downarrow$	$0.077 \downarrow$	$-0.18(0.58)$	0.03	$-0.05(0.15)$
LDL → T2D	$0.32 \downarrow$	$\mathbf{0.00017} \downarrow$	$\mathbf{0.00068} \downarrow$	$\mathbf{0.022} \downarrow$	$\mathbf{0.0064} \downarrow$	$\mathbf{5.5 \cdot 10^{-5}} \downarrow$	$0.57(4 \cdot 10^{-13})$	0.13	$0.05(0.24)$
Alcohol → T2D	$0.99 \uparrow$	$0.27 \uparrow$	$0.15 \uparrow$	$0.26 \uparrow$	$0.29 \uparrow$	$0.29 \uparrow$	$-0.24(0.51)$	0.03	$-0.03(0.41)$
Alcohol → CAD	$0.82 \uparrow$	$0.47 \uparrow$	$0.54 \uparrow$	$0.06 \uparrow$	$\mathbf{0.036} \uparrow$	$0.12 \uparrow$	$-0.08(0.85)$	0.03	$0(0.99)$
Implausible and Unsupported									
Alcohol → Asthma	$0.81 \downarrow$	$0.57 \downarrow$	$0.14 \downarrow$	$0.57 \downarrow$	$0.43 \downarrow$	$0.56 \downarrow$	$-0.08(0.43)$	0.03	$-0.08(0.031)$
Height → Asthma	$1 \downarrow$	$0.27 \downarrow$	$0.56 \downarrow$	$0.42 \downarrow$	$0.11 \downarrow$	$0.69 \downarrow$	$-0.07(0.62)$	0.03	$-0.04(0.29)$
SBP → Asthma	$0.76 \uparrow$	$\mathbf{0.026} \uparrow$	$0.85 \uparrow$	$0.73 \uparrow$	$0.35 \downarrow$	$\mathbf{0.035} \uparrow$	$-0.02(0.93)$	0.04	$0.03(0.4)$
DBP → Asthma	$0.59 \uparrow$	$0.79 \uparrow$	$0.086 \downarrow$	$0.2 \downarrow$	$0.15 \downarrow$	$0.88 \downarrow$	$-0.17(0.5)$	0.04	$0.03(0.43)$
FG → Asthma	$0.35 \downarrow$	$0.24 \downarrow$	$0.39 \downarrow$	$\mathbf{0.025} \downarrow$	$0.082 \downarrow$	$0.091 \downarrow$	$0.2(0.099)$	0.06	$-0.05(0.56)$
TG → Asthma	$0.17 \downarrow$	$\mathbf{0.03} \downarrow$	$0.51 \downarrow$	$0.93 \downarrow$	$0.34 \downarrow$	$0.065 \downarrow$	$0.02(0.88)$	0.06	$0.03(0.58)$
Height → T2D	$0.96 \uparrow$	$0.68 \downarrow$	$0.3 \uparrow$	$1 \downarrow$	$0.9 \uparrow$	$0.62 \downarrow$	$0(0.75)$	0.03	$-0.01(0.7)$
LDL → Asthma	$0.87 \downarrow$	$0.65 \uparrow$	$0.82 \downarrow$	$0.25 \uparrow$	$0.61 \uparrow$	$0.87 \uparrow$	$0.15(0.87)$	0.03	$-0.01(0.75)$
BW → Asthma	$0.73 \uparrow$	$0.18 \uparrow$	$0.85 \uparrow$	$0.14 \uparrow$	$0.42 \uparrow$	$0.082 \uparrow$	$-0.02(0.5)$	0.04	$-0.01(0.83)$
HDL → Asthma	$1 \uparrow$	$0.91 \uparrow$	$0.99 \uparrow$	$0.52 \uparrow$	$0.93 \uparrow$	$0.74 \uparrow$	$-0.02(0.76)$	0.03	$-0.01(0.84)$
Considered Non-Causal									
HDL → CAD	$\mathbf{0.00044} \downarrow$	$\mathbf{4.9 \cdot 10^{-9}} \downarrow$	$0.55 \downarrow$	$\mathbf{5 \cdot 10^{-11}} \downarrow$	$\mathbf{5.6 \cdot 10^{-6}} \downarrow$	$\mathbf{1.3 \cdot 10^{-13}} \downarrow$	$\mathbf{0.67(0.0012)}$	0.54	$-0.26(4.8 \cdot 10^{-18})$
HDL → Stroke	$0.27 \downarrow$	$0.076 \downarrow$	$0.48 \uparrow$	$0.07 \downarrow$	$0.13 \downarrow$	$0.096 \downarrow$	$\mathbf{0.67(3.5 \cdot 10^{-8})}$	0.04	$-0.15(2.3 \cdot 10^{-5})$

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. The LCV column gives the estimated GCP from LCV with p-value testing that GCP$= 0$ in parenthesis. Values are bold if estimated GCP> 0.6. The "Cause q" column gives the posterior median of q in the CAUSE sharing model and the GC column gives the genetic correlation estimated by LD score regression with p-value testing that genetic correlation is zero in parenthesis. In each section, pairs are ordered by increasing genetic correlation p-value.

4 Methods

4.1 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 $(j = 1, \ldots, p)$ 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

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

where where $S_j(\rho) = \begin{pmatrix}\hat{s}^2_{M,j} & \rho\hat{s}_{M,j}\hat{s}_{Y,j}\\ \rho\hat{s}_{M,j}\hat{s}_{Y,j} & \hat{s}^2_{Y,j}\end{pmatrix}$ and $N_2(x; \mu, \Sigma)$ is the bivariate normal density with mean μ and variance Σ evaluated at x. This model implicitly assumes that $\hat{s}_{M,j}$ and $\hat{s}_{Y,j}$ are measured without error. The correlation term, ρ, 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 (Figure 3b). The remaining proportion, $1 - q$, are independent of U (Figure 3a). All variants may have pleiotropic effects, θ_j on Y that are uncorrelated with their effects on M. Let Z_j be an indicator that variant j effects U. Then,

$$\begin{aligned}\beta_{Y,j}|Z_j = &\underbrace{\gamma\beta_{M,j}}_{\text{causal effect}} + \underbrace{Z_j\eta\beta_{M,j}}_{\substack{\text{correlated}\\ \text{pleiotropy}}} + \underbrace{\theta_j}_{\substack{\text{uncorrelated}\\ \text{pleiotropy}}}\\ Z_j \sim &\text{Bernoulli}(q),\end{aligned} \tag{4}$$

where η 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 $\beta_{Y,j} = \gamma\beta_{M,j}$, which is 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

$$\begin{aligned}p\left(\hat{\beta}_{M,j}, \hat{\beta}_{Y,j} \middle| \beta_{M,j}, \theta_j, \rho, \hat{s}_{M,j}, \hat{s}_{Y,j}, \gamma, \eta, q\right) =& qN_2\left(\begin{pmatrix}\hat{\beta}_{M,j}\\ \hat{\beta}_{Y,j}\end{pmatrix}; \begin{pmatrix}\beta_{M,j}\\ (\gamma+\eta)\beta_{M,j} + \theta_j\end{pmatrix}, S_j(\rho)\right) +\\ &(1-q)N_2\left(\begin{pmatrix}\hat{\beta}_{M,j}\\ \hat{\beta}_{Y,j}\end{pmatrix}; \begin{pmatrix}\beta_{M,j}\\ \gamma\beta_{M,j} + \theta_j\end{pmatrix}, S_j(\rho)\right).\end{aligned} \tag{5}$$

The parameters of interest in the CAUSE model are γ, η, and q. Rather than estimating individual variant effects $\beta_{M,j}$ and θ_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 θ_j as draws from a mixture of bivariate normal distributions. This strategy is based on Adaptive Shrinkage (ASH) approach of Stephens [26] for modeling univariate distributions and provides a flexible unimodal distribution with mode at $(0, 0)$. We model

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

where $\Sigma_0, \Sigma_1, \ldots, \Sigma_K$ are pre-specified covariance matrices of the form $\Sigma_k = \begin{pmatrix}\sigma^2_{k,1} & 0\\ 0 & \sigma^2_{k,2}\end{pmatrix}$, and $\pi_0, \ldots, \pi_K$ are mixing proportions that sum to 1. The set of parameters $\Omega = \{\pi_0, \ldots, \pi_K, \Sigma_0, \ldots, \Sigma_K\}$ are estimated from the data along with ρ in a single pre-processing step (Supplementary Note Section SN1). Integrating

$\beta_{M,j}$ and θ_j out of Equation (5), we obtain

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

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 γ, η, 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 Figure S1. CAUSE is implemented in an open source R package and we also provide pipelines for running CAUSE on large scale applications.

4.2 Prior Distributions for γ, η, and q

We place normal prior distributions with mean zero on γ and η. We find in simulations that results are robust to different choices of prior variance for γ and η, 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).

4.3 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 matrix and S_M and S_Y are diagonal matrices with diagonal elements $(s_{M,j})$ and $(s_{Y,j})$ [19]. 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 $\theta^* = S_Y R S_Y^{-1} \theta$ (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 θ_j are replaced with $\beta^*_{M,j}$ and θ^*_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, leading to false positives (Supplementary Note Section SN4). In this case, CAUSE may have an inflated false positive rate.

4.4 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* (ΔELPD) [18]. 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 Θ be the set of parameters (γ, η, q). Let $p_C(\Theta|\text{Data})$ and $p_S(\Theta|\text{Data})$ be the posterior density of Θ 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

$$\text{ELPD}_m = \sum_{j=1}^{p} \left(\int \left(\int p(y'_j|\Theta, \Omega, \rho) p_m(\Theta|\text{Data}) d\Theta \right) p_{\text{true}}(y'_j) dy'_j \right). \tag{9}$$

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

We estimate ΔELPD and a standard error of the estimator using the Pareto-smoothed importance sampling method described by Vehtari, Gelman, and Gabry [18] and implemented in the R package `loo`. We then compute a z-score, $z_{elpd} = \frac{\widehat{\Delta\text{elpd}}}{\hat{se}\left(\widehat{\Delta\text{elpd}}\right)}$, 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 ΔELPD were less than or equal to zero.

4.5 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 ([27]; `https://github.com/stephenslab/ldshrink`) and replicate this pattern 30 times to create a genome sized data set of $p = 584,700$ variants. We set the heritability of each trait, h^2_M and h^2_Y, the sample size for each GWAS, N_M and N_Y, and the expected number of effect variants m_M and m_θ. Additionally, we set γ, η, and q and generate effect estimates from the CAUSE model in Equation (4). We note that m_θ is the expected number of variants with $\theta_j \neq 0$, rather than the expected number of variants with non-zero effect on Y.

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:

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

where f_j is the allele frequency of variant j in the 1,000 Genomes CEU population. Variances σ^2_M and σ^2_θ are chosen to give the desired expected heritability as $\sigma^2_M = \frac{h^2_M}{m_M}$ and $\sigma^2_\theta = \frac{h^2_Y - (\gamma^2 + q\eta^2)h^2_M}{m_\theta}$. Mixing parameters are $\pi_M = m_M/p$ and $\pi_\theta = m_\theta/p$. Note that $\gamma^2 h^2_M/h^2_Y$ is the proportion of trait Y heritability mediated by the causal effect and $q\eta^2 h^2_M/h^2_Y$ 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 as $s_{\cdot j} = \sqrt{\frac{1}{(2*N_{\cdot}*f_j*(1-f_j)}}$, where $\cdot$ may be M or Y. Third, Z_j are drawn from a Bernoulli(q) distribution and true effects $\beta_{\gamma,j}$ are computed using Equation (4). Finally, effect estimates are simulated from true effects as

$$\hat{\boldsymbol{\beta}}_{\cdot} \sim N_p(\boldsymbol{S}_{\cdot}RS_{\cdot}\boldsymbol{\beta}_{\cdot}, S_{\cdot}RS_{\cdot}^{-1}) \tag{11}$$

where $S_\cdot$ is the diagonal matrix of standard errors and R is the variant correlation matrix [19]. Simulations can be replicated using the online tutorial (`https://jean997.github.io/cause/simulations.html`).

4.6 Existing MR Methods

We compare the performance of CAUSE in simulated data with six other MR methods and LCV [14]. 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 ([7], `https://github.com/rondolab/MR-PRESSO`) with outlier and distortion tests.
- GSMR is performed using the `gsmr` R package ([10], `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 ([14], `https://github.com/lukejoconnor/LCV`) using default parameters.

4.7 URLs and Software 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`

4.8 Acknowledgements

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

4.9 Data Availability

All of the data analyzed are publicly available with the exception of blood pressure summary statistics from [24]. These are available through dbGaP Accession phs000585.v2.p1. Download links for all other data sets as well as instructions for formatting and processing data and reproducing CAUSE analysis results can be found on the website `https://jean997.github.io/cause/`.

References

[1] George Davey Smith and Shah Ebrahim. "Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease ?" In: *International Journal of Epidemiology* 32 (2003), pp. 1–22.

[2] George Davey Smith and Gibran Hemani. "Mendelian randomization: genetic anchors for causal inference in epidemiological studies". In: *Human Molecular Genetics* 23.1 (2014), pp. 89–98.

[3] Anna G.C. Boef, Olaf M. Dekkers, and Saskia Le Cessie. "Mendelian randomization studies: A review of the approaches used and the quality of reporting". In: *International Journal of Epidemiology* 44.2 (2015), pp. 496–511.

[4] Jie Zheng, Denis Baird, Maria-Carolina Borges, Jack Bowden, Gibran Hemani, et al. "Recent Developments in Mendelian Randomization Studies". In: *Current Epidemiology Reports* 4.4 (2017), pp. 330–345.

[5] Stephen Burgess, Frank Dudbridge, and Simon G. Thompson. "Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods". In: *Statistics in Medicine* 35.11 (2016), pp. 1880–1906.

[6] Gibran Hemani, Jack Bowden, and George Davey Smith. "Evaluating the potential role of pleiotropy in Mendelian randomization studies". In: *Human Molecular Genetics* 27.R2 (2018), R195–R208.

[7] Marie Verbanck, Chia-yen Chen, Benjamin Neale, and Ron Do. "Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases". In: *Nature Genetics* 50.May (2018).

[8] Jack Bowden, George Davey Smith, and Stephen Burgess. "Mendelian Randomization Methodology Mendelian randomization with invalid instruments : effect estimation and bias detection through Egger regression". In: *International Journal of Epidemiology* June (2015), pp. 512–525.

[9] Richard Barfield, Helian Feng, Alexander Gusev, Lang Wu, and Wei Zheng. "Transcriptome-wide association studies accounting for co- localization using Egger regression". In: *Genetic Epidemiology* 42 (2018), pp. 418–433.

[10] Zhihong Zhu, Zhili Zheng, Futao Zhang, Yang Wu, Maciej Trzaskowski, et al. "Causal associations between risk factors and common diseases inferred from GWAS summary data". In: *Nature Communications* 9.1 (2018), p. 224.

[11] Brendan Bulik-Sullivan, Hilary K Finucane, Verneri Anttila, Alexander Gusev, Felix R Day, et al. "An Atlas of Genetic Correlations across Human Diseases and Traits". In: *Nature Genetics* 47.11 (2015), pp. 1236–1241.

[12] Verneri Anttila, Brendan Bulik-Sullivan, Hilary K. Finucane, Raymond K. Walters, Jose Bras, et al. "Analysis of shared heritability in common disorders of the brain". In: *Science* 360.6395 (2018).

[13] Joseph K Pickrell. "Fulfilling the promise of Mendelian randomization". In: *bioRxiv* (2015). doi: `10.1101/018150`.

[14] Luke J O'Connor and Alkes L. Price. "Distinguishing genetic correlation from causation across 52 diseases and complex traits". In: *Nature Genetics* 50 (2018), pp. 1728–1734.

[15] Stephen Burgess and Jack Bowden. "Integrating summarized data from multiple genetic variants in Mendelian randomization: bias and coverage properties of inverse-variance weighted methods". In: *arXiv* (2015). arXiv: `1512.04486`.

[16] Jack Bowden, M. Fabiola Del Greco, Cosetta Minelli, George Davey Smith, Nuala A. Sheehan, et al. "Assessing the suitability of summary data for two-sample mendelian randomization analyses using MR-Egger regression: The role of the I2statistic". In: *International Journal of Epidemiology* 45.6 (2016), pp. 1961–1974.

[17] Fernando Pires Hartwig, George Davey Smith, and Jack Bowden. "Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption". In: *International Journal of Epidemiology* 46.6 (2017), pp. 1985–1998.

[18] Aki Vehtari, Andrew Gelman, and Jonah Gabry. "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC". In: *Statistics and Computing* September (2016), pp. 1–20.

[19] Xiang Zhu and Matthew Stephens. "Bayesian large-scale multiple regression with summary statistics from genome-wide association studies". In: *Annals of Applied Statistics* 11.3 (2017), pp. 1561–1592.

[20] Gibran Hemani, Jack Bowden, Philip Haycock, Jie Zheng, Oliver Davis, et al. "Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome". In: *bioRx* (2017).

[21] Benjamin F. Voight, Gina M. Peloso, Marju Orho-Melander, Ruth Frikke-Schmidt, Maja Barbalic, et al. "Plasma HDL cholesterol and risk of myocardial infarction: A mendelian randomisation study". In: *The Lancet* 380.9841 (2012), pp. 572–580.

[22] Arshag D Moordian. "Dyslipidemia in type 2 diabetes mellitus". In: *Nature Clinical Practice* 5.3 (2009), pp. 150–159.

[23] Tove Fall, Weijia Xie, Wenny Poon, Hanieh Yaghootkar, Reedik Magi, et al. "Using genetic variants to assess the relationship between circulating lipids and type 2 diabetes". In: *Diabetes* 64.7 (2015), pp. 2676–2684.

[24] Georg B. Ehret, Patricia B. Munroe, Kenneth M. Rice, Murielle Bochud, Andrew D. Johnson, et al. "Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk". In: *Nature* 478.7367 (2011), pp. 103–109. arXiv: `NIHMS150003`.

[25] Patricia C. Fulkerson and Marc E. Rothenberg. "Targeting eosinophils in allergy, inflammation and beyond". In: *Nature Reviews Drug Discovery* 12.2 (2013), pp. 117–129.

[26] Matthew Stephens. "False discovery rates: a new deal". In: *Biostatistics* 18.2 (2017), pp. 275–295.

[27] Xiaoquan Wen and Matthew Stephens. "Using linear predictors to impute allele frequencies from summary or pooled genotype data". In: *Annals of Applied Statistics* 4.3 (2010), pp. 1158–1182.

Supplementary Figures

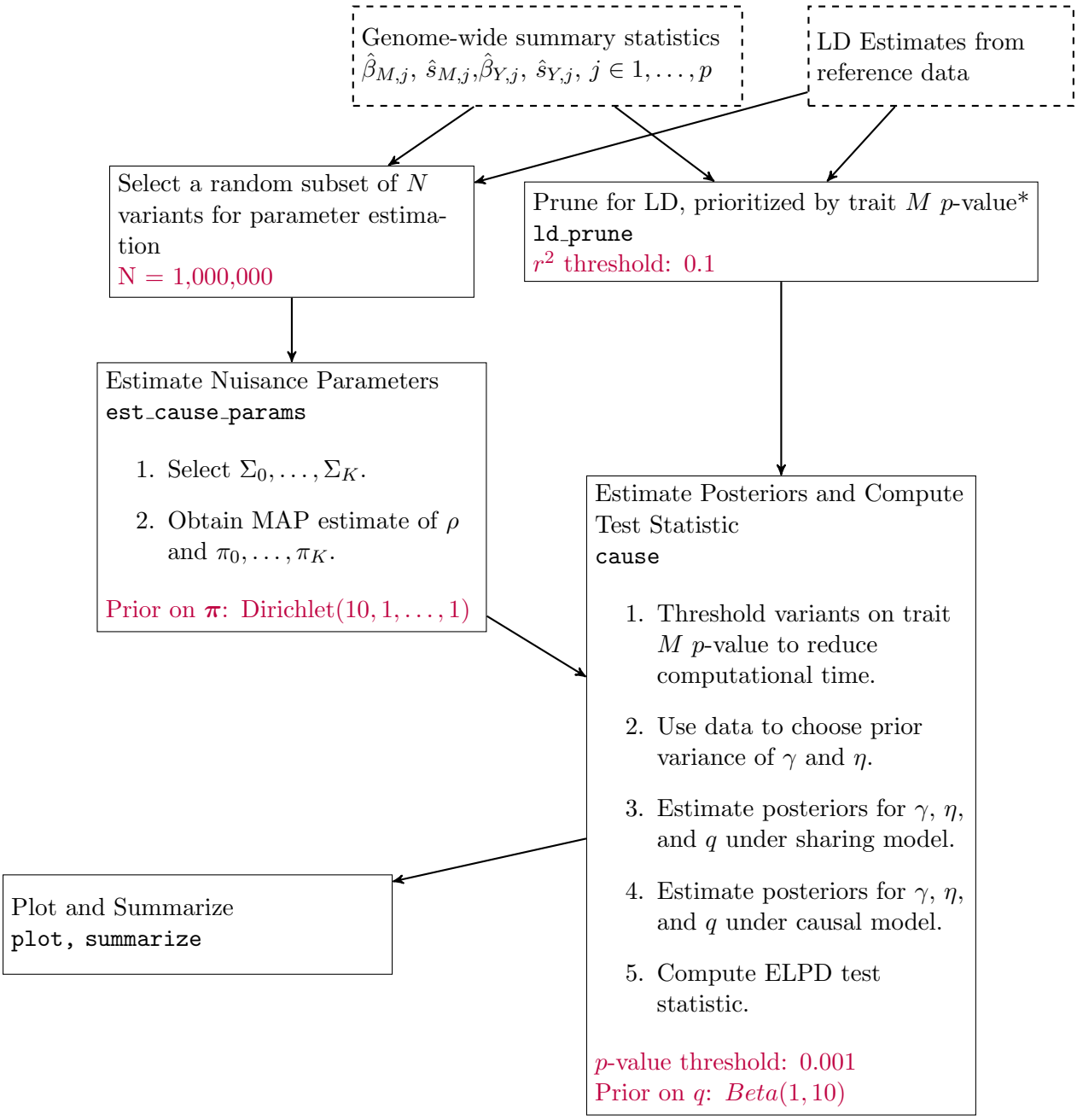
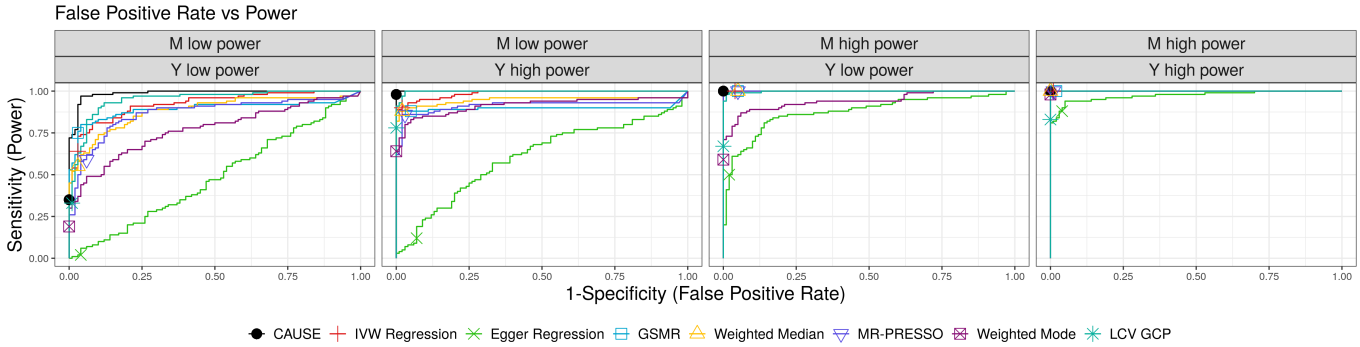
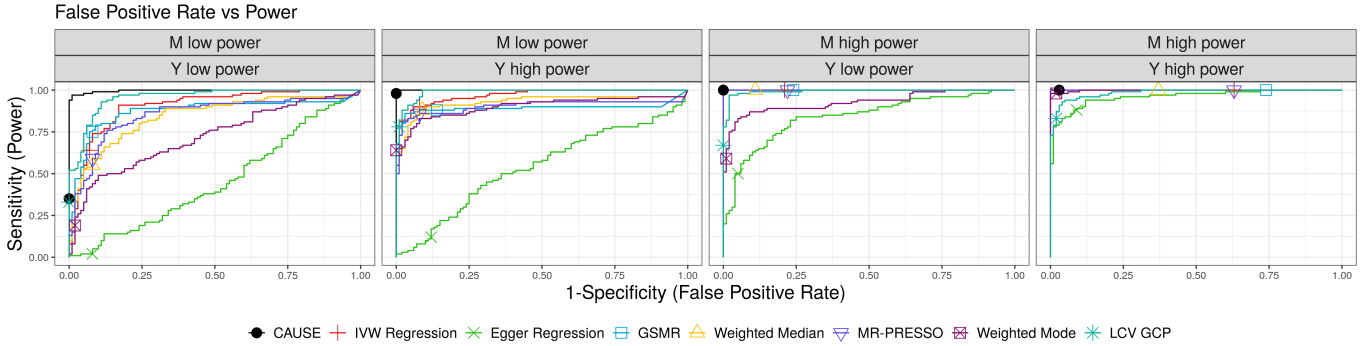


Figure S1: Workflow of a CAUSE analysis. Dashed boxes represent input data. Each solid box is an analysis step completed by the given function in the `cause` R package. LD pruning (*) can be parallelized over chromosomes. Purple text indicates user provided parameters and their default values. All analyses presented are run with default parameters.

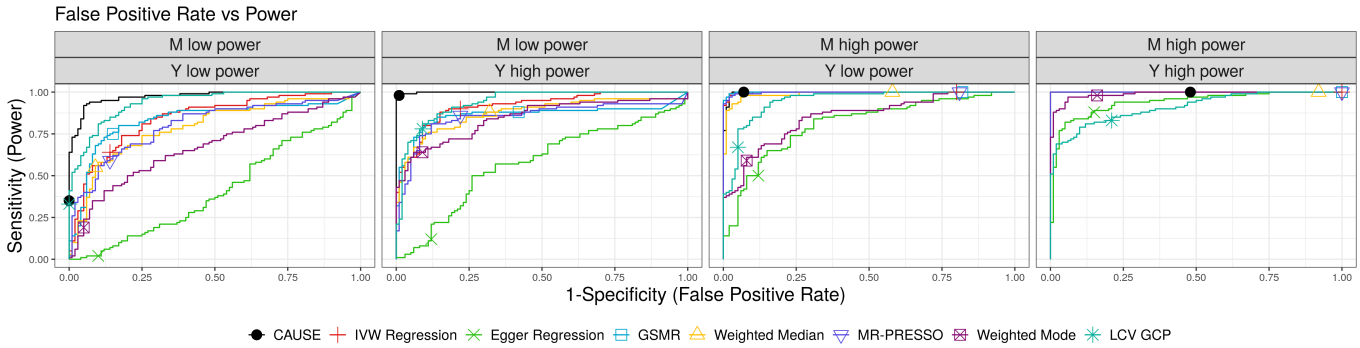
(a) $q = 0$



(b) $q = 0.2$



(c) $q = 0.4$



(d) $q = 0.5$

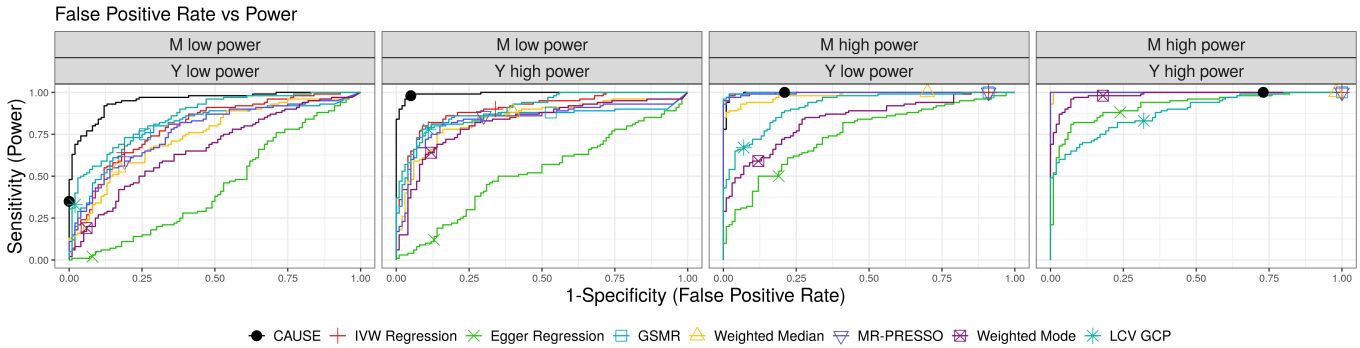


Figure S2: Comparison of false positive-power trade-off. We compare the power when $\gamma = \sqrt{0.05}$ and there is no shared factor to the false positive rate when there is no causal effect, but a proportion of variants between 0 and 50% act through a shared factor with effect $\eta = \sqrt{0.05}$ on Y. 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 $p \leq 0.05$.

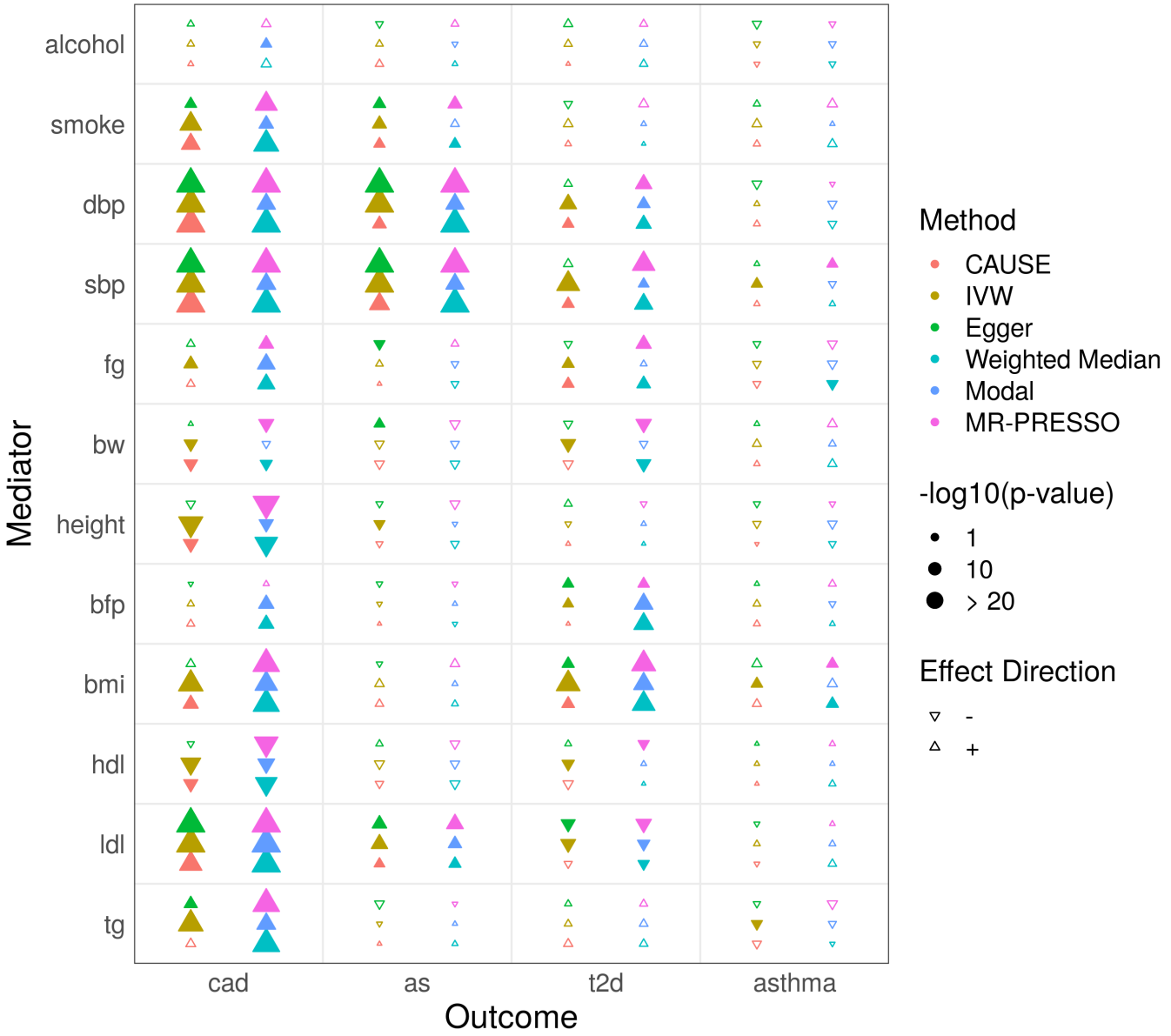


Figure S3: Tests for casual effects of mediators on diseases. Each cell summarizes the the results of six methods for a pair of traits. In the left column of the cell methods from bottom to top are CAUSE, IVW regression, and Egger regression. In the right column methods from bottom to top are weighted median, weighted mode, and MR-PRESSO. Filled symbols indicate a nominally significant p-value < 0.05.

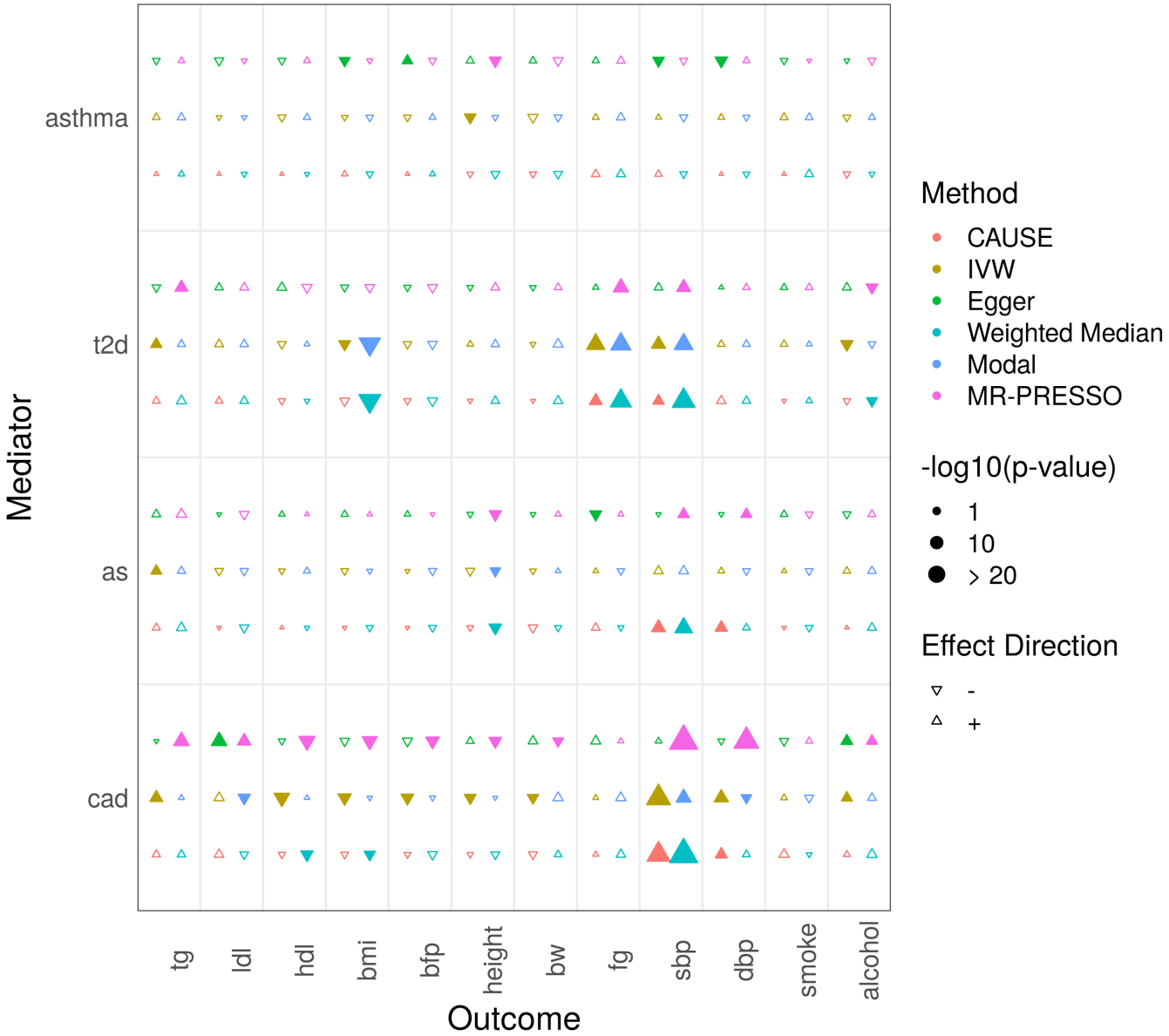


Figure S4: Tests for casual effects of disease outcomes on mediators. Each cell summarizes the the results of six methods for a pair of traits. In the left column of the cell methods from bottom to top are CAUSE, IVW regression, and Egger regression. In the right column methods from bottom to top are weighted median, weighted mode, and MR-PRESSO. Filled symbols indicate a nominally significant p-value < 0.05.

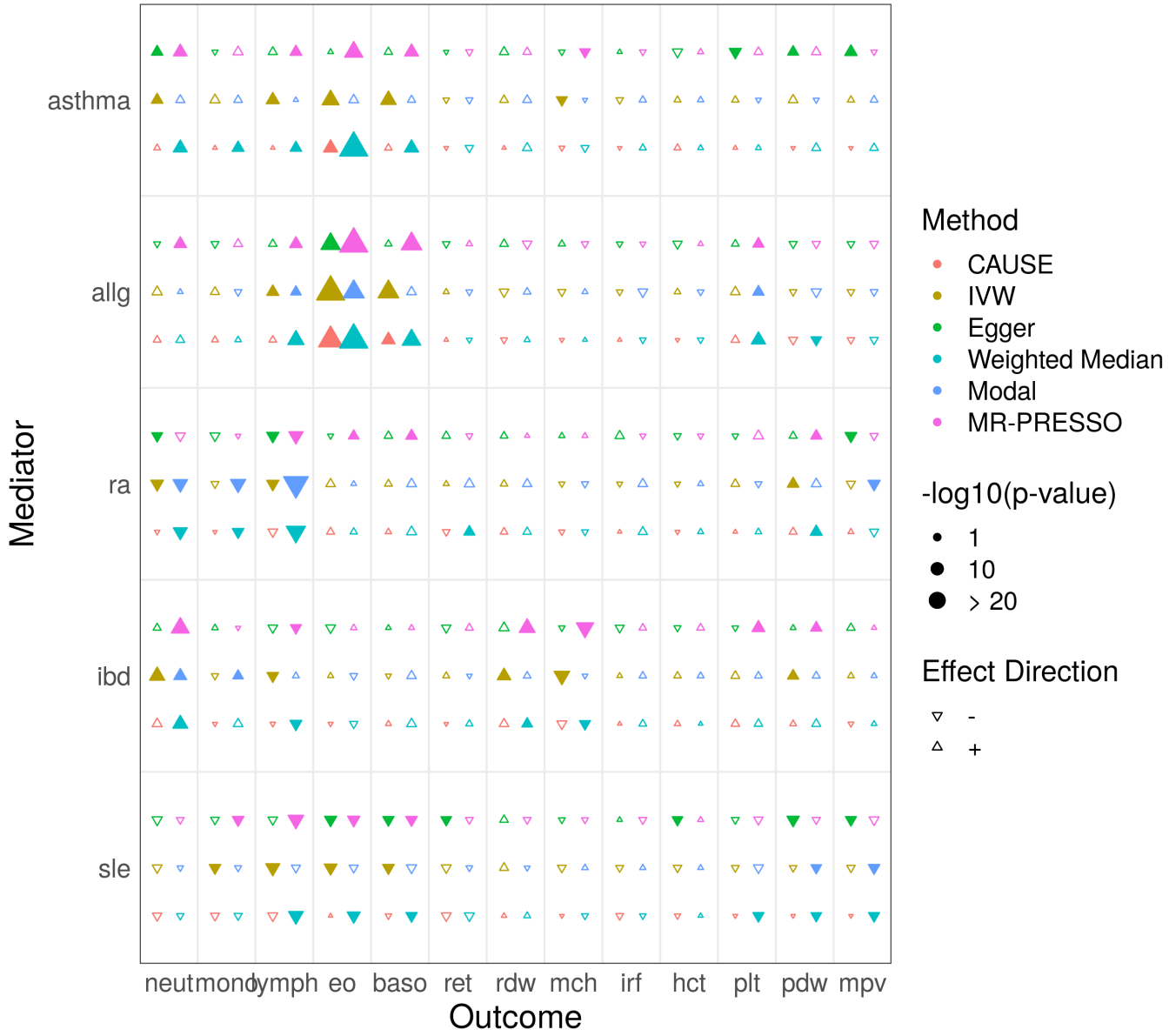


Figure S5: Tests for causal effects of autoimmune traits on blood cell composition. Each cell summarizes the the results of six methods for a pair of traits. In the left column of the cell methods from bottom to top are CAUSE, IVW regression, and Egger regression. In the right column methods from bottom to top are weighted median, weighted mode, and MR-PRESSO. Filled symbols indicate a nominally significant p-value < 0.05.

Supplementary Tables

Abbreviation	Trait	Sample Size	Cases	Controls	PMID	First Author (Year)
tg	triglycerides	188577			24097068	Willer (2013)
ldl	ldl	188577			24097068	Willer (2013)
hdl	hdl	188577			24097068	Willer (2013)
height	height	253288			25282103	Wood (2014)
bmi	body mass index	322154			25673413	Locke (2015)
bf	body fat percentage	100716			26833246	Lu (2016)
bw	birth weight	153781			27680694	Horikoshi (2016)
dbp	diastolic blood pressure	757601			30224653	Evangelou (2018)
sbp	systolic blood pressure	757601			30224653	Evangelou (2018)
fg	fasting glucose	46186			20081858	Dupuis (2010)
smoke	ever regular smoker	1232091			30643251	Liu (2019)
alcohol	drinks per week	941280			30643251	Liu (2019)
cad	coronary artery disease	547261	122733	424528	29212778	van der Harst (2017)
stroke	any stroke	446696	40585	406111	29531354	Malik (2018)
t2d	type 2 diabetes	69033	12171	56862	22885922	Morris (2012)
asthma	asthma	142486	23948	118538	29273806	Demenais (2018)

Table S1: Genome wide association studies for common diseases and risk factors

Abbreviation	Trait	Sample Size	Cases	Controls	PMID	First Author (Year)
baso	basophil count	173480			27863252	Astle (2016)
eo	eosinophil count	173480			27863252	Astle (2016)
hct	hematocrit	173480			27863252	Astle (2016)
irf	immature fraction of reticulocytes	173480			27863252	Astle (2016)
lymph	lymphocyte count	173480			27863252	Astle (2016)
mch	mean corpuscular hemoglobin	173480			27863252	Astle (2016)
mono	monocyte count	173480			27863252	Astle (2016)
mpv	mean platelet volume	173480			27863252	Astle (2016)
neut	neutrophil count	173480			27863252	Astle (2016)
pdw	platelet distribution width	173480			27863252	Astle (2016)
plt	platelet count	173480			27863252	Astle (2016)
rdw	red cell distribution width	173480			27863252	Astle (2016)
ret	reticulocyte count	173480			27863252	Astle (2016)
sle	lupus	23210	7219	15991	26502338	Bentham (2015)
ra	rheumatoid arthritis	103638	29880	73758	24390342	Okada (2014)
ibd	irritable bowel disease	96486	42950	53536	26192919	Liu (2015)
asthma	asthma	142486	23948	118538	29273806	Demenais (2018)
allg	allergic disease	360838	180129	180709	29083406	Ferreira (2017)

Table S2: Genome wide association studies for blood cell traits and immune-mediated disesase.

Method	HH	HL	LH	LL	HH	HL	LH	LL
	$q = 0$				$q = 0.3$			
CAUSE	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.99
Egger Regression	0.97	0.88	0.61	0.48	0.96	0.82	0.59	0.45
GSMR	1.00	1.00	0.90	0.89	1.00	1.00	0.88	0.85
IVW Regression	1.00	1.00	0.99	0.92	1.00	1.00	0.95	0.87
LCV GCP	1.00	1.00	1.00	0.95	0.96	0.97	0.97	0.94
MR-PRESSO	1.00	1.00	0.91	0.87	1.00	1.00	0.89	0.83
Weighted Median	1.00	1.00	0.95	0.88	1.00	1.00	0.91	0.83
Weighted Mode	1.00	0.94	0.91	0.77	1.00	0.91	0.89	0.74
	$q = 0.1$				$q = 0.4$			
CAUSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97
Egger Regression	0.97	0.83	0.60	0.44	0.94	0.81	0.58	0.41
GSMR	1.00	1.00	0.90	0.89	1.00	1.00	0.86	0.83
IVW Regression	1.00	1.00	0.98	0.92	1.00	1.00	0.92	0.84
LCV GCP	1.00	1.00	1.00	0.96	0.91	0.95	0.94	0.94
MR-PRESSO	1.00	1.00	0.91	0.87	1.00	1.00	0.85	0.80
Weighted Median	1.00	1.00	0.94	0.87	1.00	0.99	0.87	0.79
Weighted Mode	1.00	0.94	0.90	0.77	0.99	0.85	0.85	0.67
	$q = 0.2$				$q = 0.5$			
CAUSE	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.95
Egger Regression	0.96	0.84	0.54	0.45	0.92	0.75	0.52	0.44
GSMR	1.00	1.00	0.89	0.87	1.00	1.00	0.85	0.79
IVW Regression	1.00	1.00	0.97	0.90	1.00	1.00	0.90	0.80
LCV GCP	0.98	0.99	0.99	0.96	0.88	0.92	0.90	0.86
MR-PRESSO	1.00	1.00	0.90	0.85	1.00	1.00	0.84	0.77
Weighted Median	1.00	1.00	0.93	0.84	1.00	0.98	0.84	0.74
Weighted Mode	1.00	0.93	0.90	0.72	0.98	0.84	0.82	0.66

Table S3: Area under curve values for CAUSE and other methods for ROC curves shown in Figures 4c and Supplementary Figure S2. Column labels (HH, HL, LH, and LL) indicate the power of trait M and Y GWAS (H = high power, L = low power) with the first letter referring to trait M and the second letter referring to trait Y.

Traits	CAUSE	IVW	Egger	Weighted Median	Weighted Mode	MR-PRESSO	LCV	CAUSE sharing q	Genetic Correlation
White Blood Cell Traits									
eo → allg	$\mathbf{4 \cdot 10^{-11}}$ ↑	$\mathbf{4 \cdot 10^{-38}}$ ↑	**0.00021** ↑	$\mathbf{3.2 \cdot 10^{-46}}$ ↑	**0.014** ↑	$\mathbf{1.1 \cdot 10^{-43}}$ ↑	0.13(0.27)	0.81	$0.39(1.9 \cdot 10^{-18})$
eo → asthma	$\mathbf{1.9 \cdot 10^{-5}}$ ↑	$\mathbf{1.5 \cdot 10^{-19}}$ ↑	**0.015** ↑	$\mathbf{3.1 \cdot 10^{-12}}$ ↑	0.9 ↑	$\mathbf{2.6 \cdot 10^{-21}}$ ↑	0.06(0.76)	0.62	$0.37(9.9 \cdot 10^{-8})$
baso → allg	0.14 ↑	**0.018** ↑	0.84 ↓	**0.021** ↑	0.28 ↑	**0.02** ↑	0.01(0.91)	0.06	0.13(0.0046)
eo → ra	**0.042** ↑	$\mathbf{8.1 \cdot 10^{-9}}$ ↑	0.069 ↑	$\mathbf{1.6 \cdot 10^{-6}}$ ↑	**0.00091** ↑	$\mathbf{4.5 \cdot 10^{-9}}$ ↑	0.2(0.36)	0.31	0.09(0.015)
baso → asthma	0.91 ↑	0.11 ↑	0.68 ↑	0.11 ↑	0.51 ↑	0.54 ↑	−0.18(0.025)	0.03	0.14(0.1)
mono → sle	0.74 ↓	**0.03** ↓	0.077 ↓	0.065 ↓	0.36 ↓	0.12 ↓	0.05(0.62)	0.03	−0.08(0.11)
lymph → ibd	0.9 ↑	0.12 ↑	0.094 ↑	**0.025** ↑	**0.022** ↑	**0.0059** ↑	−0.26(0.7)	0.03	−0.1(0.12)
neut → asthma	0.26 ↓	0.11 ↑	$\mathbf{2.1 \cdot 10^{-5}}$ ↑	0.6 ↑	0.21 ↓	0.099 ↓	−0.34(0.47)	0.05	0.11(0.23)
eo → sle	0.55 ↑	0.12 ↑	0.46 ↑	**0.007** ↑	0.39 ↑	0.098 ↑	−0.41(0.43)	0.05	0.06(0.28)
lymph → allg	0.4 ↓	**0.0032** ↓	**0.014** ↓	0.052 ↓	0.46 ↓	**0.00029** ↓	0(0.54)	0.04	−0.04(0.28)
baso → sle	0.91 ↓	0.97 ↑	0.4 ↑	0.2 ↑	0.81 ↑	0.51 ↑	−0.05(0.65)	0.03	−0.09(0.34)
mono → ibd	0.49 ↓	0.16 ↓	0.59 ↑	0.68 ↓	0.091 ↑	0.35 ↓	−0.04(0.93)	0.05	−0.06(0.43)
baso → ra	0.99 ↑	0.38 ↑	0.95 ↑	**0.05** ↑	0.063 ↑	0.15 ↑	0.09(0.61)	0.03	−0.03(0.47)
mono → allg	1 ↑	0.058 ↓	0.24 ↓	**0.02** ↓	0.078 ↓	0.14 ↓	0.45(0.37)	0.03	0.02(0.5)
neut → sle	0.16 ↓	**0.0036** ↓	0.068 ↓	**0.042** ↓	**0.039** ↓	**0.0042** ↓	0.14(0.48)	0.06	−0.05(0.52)
neut → allg	0.7 ↑	**0.0034** ↑	$\mathbf{7.5 \cdot 10^{-10}}$ ↑	**0.0043** ↑	0.51 ↑	0.14 ↑	−0.07(0.72)	0.03	0.03(0.53)
eo → ibd	0.4 ↑	0.73 ↑	0.43 ↓	**0.00057** ↑	$\mathbf{4.8 \cdot 10^{-6}}$ ↑	**0.017** ↑	−0.01(0.91)	0.05	0.04(0.54)
neut → ibd	0.21 ↑	0.18 ↓	**0.003** ↓	$\mathbf{3.2 \cdot 10^{-5}}$ ↓	$\mathbf{4.6 \cdot 10^{-9}}$ ↓	0.11 ↓	−0.39(0.3)	0.06	0.05(0.71)
lymph → ra	1 ↑	1 ↑	0.42 ↓	0.24 ↓	0.063 ↓	0.18 ↓	−0.15(0.3)	0.03	−0.01(0.78)
lymph → sle	0.9 ↓	0.68 ↓	0.77 ↓	**0.035** ↑	0.14 ↑	0.42 ↑	−0.52(0.23)	0.03	−0.01(0.83)
mono → ra	1 ↑	0.64 ↓	0.55 ↓	**0.0098** ↓	$\mathbf{6.6 \cdot 10^{-5}}$ ↓	0.26 ↓	0.02(0.88)	0.03	0.01(0.87)
baso → ibd	1 ↓	0.051 ↓	0.87 ↓	**0.0082** ↓	0.41 ↓	**0.0051** ↓	0.36(0.18)	0.03	−0.01(0.88)
mono → asthma	0.61 ↓	0.68 ↓	0.96 ↑	0.59 ↑	0.7 ↑	0.51 ↓	−0.26(0.69)	0.03	0.01(0.88)
neut → ra	0.97 ↑	0.46 ↓	0.21 ↓	**0.019** ↑	**0.017** ↑	0.73 ↓	−0.02(0.6)	0.03	−0.01(0.89)
lymph → asthma	0.85 ↓	0.13 ↓	0.17 ↓	0.076 ↓	0.42 ↓	0.18 ↓	−0.3(0.69)	0.03	0(1)
Red Blood Cell Traits									
hct → asthma	0.95 ↓	**0.0024** ↓	**0.0055** ↓	0.13 ↓	0.21 ↓	**0.0065** ↓	0.52(0.11)	0.03	−0.1(0.025)
hct → ibd	0.99 ↑	0.25 ↑	0.6 ↓	0.9 ↑	0.15 ↓	0.26 ↑	−0.04(0.9)	0.03	0.12(0.033)
rdw → ra	**0.03** ↑	$\mathbf{1.5 \cdot 10^{-9}}$ ↑	**0.00052** ↑	$\mathbf{1.1 \cdot 10^{-6}}$ ↑	$\mathbf{1.5 \cdot 10^{-5}}$ ↑	$\mathbf{5.7 \cdot 10^{-9}}$ ↑	0.24(0.47)	0.27	0.06(0.039)
rdw → asthma	0.064 ↑	$\mathbf{5.1 \cdot 10^{-6}}$ ↑	0.55 ↑	**0.0042** ↑	0.15 ↑	$\mathbf{1.2 \cdot 10^{-6}}$ ↑	0.51(0.023)	0.13	0.1(0.046)
ret → sle	0.53 ↓	**0.049** ↓	0.67 ↓	**0.024** ↓	**0.023** ↓	0.06 ↓	−0.31(0.7)	0.04	−0.1(0.063)
irf → sle	0.5 ↓	0.12 ↓	0.11 ↓	0.38 ↓	0.72 ↓	0.073 ↓	−0.37(0.85)	0.07	−0.1(0.076)
mch → ibd	0.87 ↓	0.096 ↓	0.42 ↓	0.47 ↓	0.1 ↓	0.052 ↓	0.07(0.61)	0.04	−0.07(0.16)
rdw → sle	0.056 ↑	$\mathbf{2.4 \cdot 10^{-10}}$ ↑	**0.039** ↑	**0.00015** ↑	**0.0012** ↑	$\mathbf{1.7 \cdot 10^{-8}}$ ↑	0.48(0.28)	0.35	0.11(0.18)
mch → sle	0.11 ↓	$\mathbf{4.5 \cdot 10^{-9}}$ ↓	$\mathbf{4.4 \cdot 10^{-6}}$ ↓	$\mathbf{6.4 \cdot 10^{-6}}$ ↓	$\mathbf{4 \cdot 10^{-4}}$ ↓	$\mathbf{1.5 \cdot 10^{-7}}$ ↓	$\mathbf{0.6(5.2 \cdot 10^{-11})}$	0.16	−0.06(0.19)
ret → asthma	0.4 ↓	0.11 ↓	0.45 ↓	0.36 ↓	0.72 ↑	0.2 ↓	−0.01(0.67)	0.06	−0.04(0.29)
ret → ra	0.12 ↓	**0.0034** ↓	0.082 ↑	**0.018** ↓	0.084 ↑	**0.00023** ↓	0.29(0.49)	0.08	−0.03(0.29)
hct → ra	0.98 ↓	0.88 ↓	0.94 ↑	**0.022** ↑	**0.025** ↑	0.61 ↓	0.17(0.64)	0.03	−0.04(0.31)
mch → asthma	0.1 ↓	**0.01** ↓	0.24 ↓	0.067 ↓	0.14 ↓	**0.011** ↓	0.29(0.37)	0.07	−0.03(0.35)
ret → allg	1 ↓	0.15 ↓	0.8 ↓	0.68 ↓	0.38 ↑	0.38 ↓	−0.01(0.82)	0.03	−0.02(0.42)
mch → ra	0.12 ↓	$\mathbf{1.1 \cdot 10^{-6}}$ ↓	$\mathbf{6.9 \cdot 10^{-6}}$ ↓	$\mathbf{2 \cdot 10^{-4}}$ ↓	$\mathbf{3.1 \cdot 10^{-8}}$ ↓	$\mathbf{2.8 \cdot 10^{-6}}$ ↓	0.43(0.54)	0.12	0.02(0.48)
hct → allg	0.8 ↑	0.93 ↑	0.45 ↑	0.17 ↑	0.13 ↑	0.64 ↑	−0.01(0.95)	0.03	−0.02(0.5)
irf → ibd	0.32 ↓	0.16 ↓	0.53 ↓	$\mathbf{1.2 \cdot 10^{-5}}$ ↓	$\mathbf{2.6 \cdot 10^{-5}}$ ↓	**0.0015** ↓	0.03(0.78)	0.03	−0.03(0.52)
hct → sle	0.71 ↓	0.86 ↓	0.67 ↑	**0.036** ↑	0.082 ↑	0.84 ↑	0.34(0.52)	0.04	0.03(0.52)
rdw → ibd	1 ↑	0.5 ↑	0.73 ↑	0.58 ↑	0.83 ↓	0.38 ↑	−0.19(0.67)	0.03	0.03(0.55)
irf → asthma	0.49 ↓	0.16 ↓	0.7 ↓	0.12 ↓	0.34 ↓	0.2 ↓	0.17(0.69)	0.05	0.02(0.64)
ret → ibd	1 ↑	0.86 ↑	0.37 ↓	0.05 ↑	0.14 ↓	0.41 ↑	−0.04(0.96)	0.03	0.02(0.68)
irf → allg	0.36 ↓	**0.023** ↓	0.17 ↓	**0.0095** ↓	**0.016** ↓	**0.038** ↓	0.17(0.93)	0.06	0.01(0.68)
rdw → allg	1 ↑	0.86 ↓	0.19 ↓	0.62 ↓	0.77 ↓	0.85 ↓	0.12(0.7)	0.03	0.01(0.81)
mch → allg	0.84 ↓	0.58 ↓	0.88 ↑	0.3 ↓	0.85 ↓	0.38 ↓	0.24(0.7)	0.03	0(0.84)
irf → ra	1 ↓	0.79 ↓	0.91 ↑	0.42 ↓	0.24 ↓	0.67 ↓	0.01(0.88)	0.03	0(0.95)
Platelet Traits									
pdw → ibd	1 ↓	0.42 ↑	0.97 ↑	**0.008** ↑	**0.015** ↑	0.055 ↑	−0.33(0.76)	0.03	0.09(0.06)
pdw → sle	0.4 ↑	0.45 ↑	0.18 ↓	0.54 ↑	0.53 ↓	0.42 ↑	0.03(0.99)	0.05	0.12(0.065)
pdw → allg	0.19 ↓	**0.00011** ↓	0.28 ↓	**0.00095** ↓	**0.027** ↓	**0.00016** ↓	0.11(0.66)	0.21	−0.05(0.072)
mpv → sle	0.54 ↑	0.77 ↑	0.8 ↑	0.76 ↑	0.63 ↑	0.58 ↑	0.02(0.49)	0.04	0.06(0.14)
plt → ra	1 ↑	0.91 ↑	0.37 ↓	**0.00031** ↓	$\mathbf{6.1 \cdot 10^{-5}}$ ↓	0.7 ↓	−0.18(0.44)	0.03	0.03(0.23)
plt → asthma	1 ↓	0.61 ↑	0.82 ↓	0.59 ↑	0.6 ↑	0.55 ↑	0(0.77)	0.03	0.05(0.24)
mpv → asthma	1 ↓	0.63 ↓	0.53 ↓	0.99 ↓	0.79 ↓	0.68 ↓	−0.04(0.67)	0.03	−0.05(0.25)
plt → sle	1 ↓	0.35 ↑	0.31 ↑	**0.023** ↑	0.12 ↑	0.85 ↑	0.05(0.69)	0.03	−0.04(0.33)
mpv → allg	0.18 ↓	0.071 ↓	0.34 ↓	**0.029** ↓	**0.0079** ↓	0.088 ↓	−0.04(0.88)	0.07	−0.02(0.42)
pdw → ra	1 ↓	0.6 ↓	0.17 ↓	**0.0048** ↓	**0.0018** ↓	0.5 ↓	0.08(0.81)	0.03	−0.02(0.55)
pdw → asthma	0.29 ↓	**0.015** ↓	0.057 ↓	0.07 ↓	0.12 ↓	**0.0085** ↓	0.09(0.97)	0.05	−0.03(0.59)
plt → ibd	0.8 ↑	0.75 ↑	0.77 ↑	0.93 ↓	0.67 ↑	0.49 ↑	−0.15(0.8)	0.03	0.03(0.62)
mpv → ibd	0.63 ↑	0.39 ↑	0.66 ↓	0.4 ↑	0.49 ↑	0.16 ↑	−0.08(0.79)	0.03	0.03(0.63)
mpv → ra	0.99 ↑	0.46 ↓	0.86 ↓	**0.05** ↓	**0.0039** ↓	0.82 ↓	0.1(0.81)	0.03	0.01(0.76)
plt → allg	1 ↓	0.22 ↓	0.075 ↓	0.56 ↓	0.61 ↓	0.38 ↓	−0.06(0.92)	0.03	0(0.89)

Table S4: Summary of results for blood cell traits and immune-mediated diseases grouped by blood cell category. 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. The LCV column gives the estimated GCP from LCV with p-value testing that GCP$= 0$ in parenthesis. Values are bold if estimated GCP> 0.6. The "Cause q" column gives the posterior median of q in the CAUSE sharing model and the GC column gives the genetic correlation estimated by LD score regression with p-value testing that genetic correlation is zero in parenthesis. In each section, pairs are ordered by increasing genetic correlation p-value.

Supplementary Note for Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics.

Jean Morrison *[1], Nicholas Knoblauch[1], Joseph Marcus[1], Matthew Stephens †[1,2], and Xin He ‡[1]

[1]Department of Human Genetics, University of Chicago
[2]Department of Statistics, University of Chicago

January 28, 2020

SN1 Empirical Parameter Estimation

Analysis with CAUSE is comprised of two main steps (see Figure S1). The first step is to estimate nuisance parameters ρ, which accounts for overlap in the two GWAS samples and the parameters defining the empirical joint prior distribution for $\beta_{M,j}$ and θ_j. We do this in two sub-steps. First, we select a panel of candidate covariance matrices $\Sigma_0, \ldots, \Sigma_K$. Second, we fix $\gamma = \eta = 0$ and compute the maximum a posteriori (MAP) values of ρ and $\pi_0, \ldots, \pi_K$.

The set of candidate covariance matrices should be large enough that a flexible joint distribution can be fit for $\beta_{M,j}$ and θ_j, but not so large that evaluating the likelihood in (7) becomes burdensome. To choose this set, we first apply the Adaptive Shrinkage (ASH) method proposed by Stephens [1] to estimate the distributions of $\boldsymbol{\beta}_M$ and $\boldsymbol{\beta}_Y$ separately. Briefly, given a set of summary statistics for a single study, ASH estimates a sparse unimodal distribution for the marginal effects. This distribution is flexible and parameterized as a mixture of univariate normal distributions centered at 0. ASH uses the model

$$\beta_{\cdot,j} | \varpi_0, \ldots, \varpi_L, \varsigma_0, \ldots, \varsigma_L \sim \sum_{l=0}^{L} \varpi_l N(0, \varsigma_l^2),$$

where $\varsigma_0, \ldots, \varsigma_L$ are a fixed grid of variances with $\varsigma_0 = 0$. ASH estimates the mixing proportions $\varpi_0, \ldots, \varpi_L$ using a prior on $\boldsymbol{\varpi}$ that encourages more weight to be given to ϖ_0, the proportion of effects equal to 0. Despite starting with a large number of candidate variances, ASH solutions tend to place most of the weight on only a few values. The resulting solution is sparse (most of the estimated effects are 0) and parsimonious (there are few components in the model with non-zero mixing proportion).

Let $\varsigma_{M,0}, \ldots, \varsigma_{M,l_M}$ and $\varsigma_{Y,0}, \ldots, \varsigma_{Y,l_Y}$ be the set of variances with non-zero weight in the ASH estimates for traits M and Y respectively. Because ASH encourages sparsity, in all cases $\varsigma_{M,0} = \varsigma_{L,0} = 0$. We construct the panel of candidate 2×2 covariance matrices by taking all pairs of these variances as diagonal elements and setting the off diagonal elements to be 0. Thus if $l_M = 4$ and $l_Y = 3$, our method produces a set of $(4+1)(3+1) = 20$ candidate covariance matrices.

In the second step, we fix $\gamma = \eta = 0$ and calculate the MAP values of ρ and $\pi_0, \ldots, \pi_K$. We use a Dirichlet$(10, 1, \ldots, 1)$ prior on $\pi_0, \ldots, \pi_K$ with π_0 corresponding to the covariance matrix of all zeros. This prior is the same prior used by default in ASH and encourages a sparse solution, however the weights may

*jeanm@uchicago.edu
†mstephens@uchicago.edu
‡xinhe@uchicago.edu

be adjusted by the user in the CAUSE software. We use a prior on $z = \tanh^{-1}(\rho)$ of $z \sim N(0, 0.25)$, which is a weak prior encouraging ρ to be close to zero. To calculate the MAP estimate, we use coordinate descent, alternating between optimization of ρ with $\boldsymbol{\pi}$ fixed and optimization of $\boldsymbol{\pi}$ with ρ fixed. As observed by Stephens [1] and others, maximization in $\boldsymbol{\pi}$ is a convex optimization problem that can be completed quickly. In practice, we find that convergence is usually reached within five iterations.

SN2 Prior Distributions for γ and η

In most cases, little prior information is available about the size of the causal or confounding effect. Differences in variable scaling and covariate adjustment across GWAS may make it difficult to predict the magnitude of these effects. Fortunately, we find that CAUSE results are robust to a wide range in prior distributions for these parameters. We require that the same prior be used for γ and η. If this is not the case false positives can arise when the true shared factor effect is better represented by the prior on γ than the prior on η. We use normal prior distributions with mean 0 and variance $\sigma^2_{\gamma\eta}$ for γ and η.

To assess the robustness of CAUSE to changes in $\sigma^2_{\gamma\eta}$, we analyze data simulated from three scenarios using a range of values for $\sigma^2_{\gamma\eta}$. The three scenarios are 1) a setting with a causal effect ($\gamma = \sqrt{0.05}$, $\eta = 0$, $q = 0$), 2) a setting with no causal effect but some correlated pleiotropy ($\gamma = 0$, $\eta = \sqrt{0.05}$, $q = 0.3$), and 3) a setting with neither a casual effect or correlated pleiotropy ($\gamma = 0$, $\eta = 0$, $q = 0$). We analyze simulated data using three values of $\sigma^2_{\gamma\eta}$. These are chosen so that $\sqrt{0.05}$, is at the 80th, 65th, and 51st quantile of the $N(0, \sigma^2_{\gamma\eta})$ distribution, giving small, medium, and large values of $\sigma_{\gamma\eta}$ respectively. All simulations are conducted with sample sizes $N_M = N_Y = 40,000$, the high power setting used in other simulations.

We compare p-values and posterior median estimates for γ, η, and q under the causal and sharing models across different values of $\sigma^2_{\gamma\eta}$. The Pearson correlation in p-values was greater than 0.95 for all settings and between all pairs of values of $\sigma^2_{\gamma\eta}$. The correlation in posterior medians was higher than 0.8 for all parameters and all settings except for estimates of q in setting 1 under the causal model. These had a somewhat lower correlation (minimum correlation 0.51). However, in all cases the posterior median of q in the causal model was very low – less than 0.05. These results demonstrate that very similar inference can be obtained using a wide range of prior distributions for γ and η.

By default, $\sigma^2_{\gamma\eta}$ is chosen using the data. We use a set of variants with trait M p-value $< 10^{-3}$ and compute $\hat{\gamma}_{max} = \max|\frac{\hat{\beta}_2}{\hat{\beta}_1}|$. This is the largest magnitude of causal estimate that could be achieved using only one variant. We then choose $\sigma^2_{\gamma\eta}$ so that the prior probability that γ or η has magnitude larger than $\hat{\gamma}_{max}$ is 0.05.

SN3 Approximating posteriors of γ, η, and q

We use an adaptive variation of a simple grid posterior approximation[2] to approximate the joint posterior distribution of γ, η, and q. To compute this approximation, we begin with initial bounds on γ and η of $(-1, 1)$. These will be adaptively expanded as needed. The bounds on q are fixed at $(0, 1)$.

The approximation proceeds as follows:

1. The domain of (γ, η, q) is divided into a coarse set of cubes. The approximate posterior probability of each cube is computed by approximating the likelihood within the cube as constant and equal to the likelihood at the midpoint of the cube.

2. After the first rough approximation, the bounds of γ and η are expanded so that less that 0.001 of the posterior mass falls in the cubes closest to the boundary. These bounds are then fixed.

3. The grid is then iteratively refined until no cube contains more than 1% of the posterior density. At each iteration, all cubes containing more than this are subdivided into nine smaller cubes and the posterior is re-estimated.

SN4 Effects of LD

In this section we explore the question of whether the CAUSE model and likelihood are valid for variants in LD. For simplicity, in this section only, we assume that the GWAS for traits M and Y have no overlapping samples and have the same LD structure. CAUSE relies on two assumptions. The first is that the joint likelihood of all pairs of summary statistics can be factorized into the product of the likelihood for each variant. Variants in LD are not independent, however, by pruning variants so that the set is nearly independent, we can approximate this condition. The second assumption is that

$$\mathrm{Cov}(\hat{\beta}_{Y,j}, \hat{\beta}_{M,j}|Z_j, \gamma, \eta, \beta_{M,j}) = (\gamma + Z_j\eta)\mathrm{Var}(\beta_{M,j}) = (\gamma + Z_j\eta)\left(\mathrm{Var}(\hat{\beta}_{M,j}) - s^2_{M,j}\right). \tag{12}$$

Without LD, this is a consequence of Equation (4) and the assumption that $\hat{\beta}_{M,j}$ and $\hat{\beta}_{Y,j}$ are unbiased estimates of $\beta_{M,j}$ and $\beta_{Y,j}$. In the presence of LD, the expectation of $\hat{\beta}_{\cdot,j}$ ($\cdot$ may be M or Y) is not $\beta_{\cdot,j}$, but a combination of $\beta_{\cdot,j}$ and a contribution from each variant in LD with variant j. Using results from [3],

$$E[\hat{\beta}_{\cdot,j}] = \sum_k \frac{r_{jk}s_{\cdot,j}}{s_{\cdot,k}}\beta_{\cdot,k} \equiv \beta^*_{\cdot,j}, \tag{13}$$

where r_{jk} is the correlation between variant j and variant k. We refer to $\beta^*_{\cdot,j}$ as the *LD-transformed effects*. Note that if allele frequencies are the same in the two GWAS populations then $s_{M,j} = cs_{Y,j}$ where c is a constant depending on sample size. Thus

$$\frac{r_{jk}s_{M,j}}{s_{M,k}} = \frac{r_{jk}s_{Y,j}}{s_{Y,k}} \equiv h_{j,k}. \tag{14}$$

We now derive $\mathrm{Cov}(\hat{\beta}_{Y,j}, \hat{\beta}_{M,j}|\mathbf{Z}, \gamma, \eta, \boldsymbol{\beta}_M)$ in the presence of LD, where $\mathbf{Z}$ is a vector with jth element equal to Z_j and $\boldsymbol{\beta}_M$ is a vector with jth element equal to $\beta_{M,j}$. We assume that direct effects are independent so $\mathrm{Cov}(\beta_{M,j}, \beta_{M,k}) = 0$ if $j \neq k$ and $\mathrm{Cov}(\beta_{M,j}, \theta_k) = 0$ for all j and k. We also assume that $\hat{\beta}_{M,j}$ and $\hat{\beta}_{Y,j}$ are independent conditional on $\beta^*_{M,j}$ and $\beta^*_{Y,j}$. With these assumptions,

$$\begin{aligned}\mathrm{Cov}\left(\hat{\beta}_{Y,j}, \hat{\beta}_{M,j}|\mathbf{Z}, \gamma, \eta, \boldsymbol{\beta}_M\right) =& \mathrm{Cov}\left(\sum_k h_{j,k}\beta_{Y,k}, \sum_k h_{j,k}\beta_{M,k}\middle|\mathbf{Z}, \gamma, \eta\right)\\ =& \sum_k h^2_{j,k}\mathrm{Cov}\left(\beta_{Y,k}, \beta_{M,k}|Z_k, \gamma, \eta\right)\\ =& \sum_k h^2_{j,k}\left(\gamma + Z_k\eta\right)\mathrm{Var}\left(\beta_{M,k}\right),\end{aligned} \tag{15}$$

and

$$\mathrm{Var}(\hat{\beta}_{M,j}|\mathbf{Z}) = \mathrm{Var}(\beta^*_{M,j}) + s^2_{M,j} = \sum_k h^2_{j,k}\mathrm{Var}(\beta_{M,k}) + s^2_{M,j}. \tag{16}$$

Suppose that the variant correlation structure can be decomposed into independent LD blocks and that there is at most one M effect variant per block. If this variant has index k' then, for any variant in the block

$$\begin{aligned}\mathrm{Cov}\left(\hat{\beta}_{Y,j}, \hat{\beta}_{M,j}\middle|\mathbf{Z}, \gamma, \eta, \boldsymbol{\beta}_M\right) =& h^2_{j,k'}\left(\gamma + Z_{k'}\eta\right)\mathrm{Var}(\beta_{M,k'})\\ =& \left(\gamma + Z_{k'}\eta\right)\left(\mathrm{Var}(\hat{\beta}_{M,j}) - s^2_{M,j}\right).\end{aligned} \tag{17}$$

This means that, if $Z_{k'} = 1$, then variant k' induces correlation between effect estimates for other variants in the block, even when $\gamma = 0$. However, if we use only one variant per LD block to estimate parameters then there is no distortion in the proportion of correlated variants. If $\gamma = 0$, then the proportion of variants with correlated effect estimates will be equal q, the proportion of true effect variants acting through U. In this case, CAUSE will not have an increased false positive rate but may have lower power if the variants selected to use in estimation are far from the true causal variants. To maximize power, we prune for LD prioritizing variants with low trait M p-values.

More generally, a block may contain multiple causal variants for M, with some acting on U (correlated pleiotropic variants) and others not. Equation 15 implies that a correlated pleiotropic variant in LD with variant j will induce non-zero correlation between $\hat{\beta}_{Y,j}$ and $\hat{\beta}_{M,j}$ even if $\gamma = 0$ and $Z_j = 0$. The presence of non-shared variants in a block reduces the correlation, but will not eliminate it. If M is highly polygenic and q is large, the proportion of blocks containing at least one shared variant may be much larger than the true value of q. In this case, LD can create an impression that a larger propotion of variants have correlated effects, which can lead to inflated estimates of q and higher false positive rates using CAUSE. We assume that these settings are rare. However, our simulations are conducted under moderately dense conditions using 1000 effect variants and only 1,170 LD blocks, demonstrating that CAUSE still performs well even under somewhat unfavorable conditions.

SN5 Connections with LCV

O'Connor and Price [4] propose an approach to identifying pairs of traits with causal relationships that uses a latent causal variable (LCV) model. This model is similar to the CAUSE model (Figure 3) with $\gamma = 0$. Rather than explicitly modeling both correlated pleiotropy and a causal effect, the LCV model includes only a shared factor (U in the CAUSE model) and estimates the "genetic causality proportion" (GCP). The GCP reflects the relative proportions of heritability of each trait that are explained by a shared factor. A causal effect (with no additional correlated pleiotropy from other sources) is equivalent to a model in which all variants act through a shared factor. In this case, the GCP is equal to 1 or -1 depending on the direction of the effect. LCV estimates the GCP and computes a test statistic testing whether GCP$= 0$. Models with non-zero GCP are not necessarily causal and this p-value is not intendd as a test of causality. O'Connor and Price [4] use an estimated GCP larger than 0.6 to suggest a possible causal relationship.

Because the LCV model is similar to the CAUSE model with $\gamma = 0$, we can derive an expression for GCP in terms of CAUSE parameters under this condition. The LCV model uses two parameters q_M^{lcv} and q_Y^{lcv}, the square root of the proportions of trait M and Y heritability explained by the shared factor. Here we use the same trait M and Y notation used in our discussion of CAUSE. In terms of CAUSE parameters these are

$$q_M^{lcv} = \sqrt{q} \tag{18}$$

$$q_Y^{lcv} = \frac{\sqrt{q}\eta h_M}{h_Y} \tag{19}$$

The GCP is then defined as

$$GCP = \frac{\log|q_Y^{lcv}| - \log|q_M^{lcv}|}{\log|q_M^{lcv}| + \log|q_Y^{lcv}|} = \frac{\log(\eta h_M/h_Y)}{\log(q\eta h_M/h_Y)}. \tag{20}$$

From this formula, we see that GCP is non-zero if η and q are both non-zero. For example, if the heritability of the two traits is equal, $q = 0.3$ and $\eta = \sqrt{0.05}$ then GCP$= 0.55$.

SN6 Additional Simulation Results

SN6.1 Effects of the prior on q

The choice of prior on q can affect the power and robustness of CAUSE. The default choice of a Beta$(1, 10)$ prior is used in the main Results. Here we compare simulation performance using Beta$(1, 2)$, Beta$(1, 10)$, and Beta$(1, 100)$ distributions. The Beta$(1, 2)$ distribution is quite lenient, placing 25% of the prior mass on values above 0.5. The Beta$(1, 100)$ distribution is sharply peaked close to zero, placing 95% of the prior mass below 0.03. In order for the CAUSE model to be identifiable, the prior on q must be asymmetric, placing more weight on values less than 0.5. Figure SN1 shows results for simulations described in the main Results (analogous to main text Figure 4). As we would expect, CAUSE with the Beta$(1, 100)$ prior has high false positive rate when the true proportion of correlated pleiotropy is large. CAUSE with the Beta$(1, 2)$ prior has fewer false positives for large values of q but also reduced power. We found the Beta$(1, 10)$ prior to be an attractive compromise because its power is only slightly worse than power obtained using the Beta$(1, 100)$

prior but the false positive rate is substantially improved. CAUSE shows the same performance in ROC curves for the three prior choices (Figure SN1c). One consequence of these results is that using a more permissive prior for q may be a better choice for higher powered studies. Results of CAUSE for pairs of GWAS traits using different priors can be seen online at the CAUSE website (see URLs).

SN6.2 Parameter Estimation

In the main Results, we focus on CAUSE's performance in terms of power to detect causal effects and false discovery rate. CAUSE also produces point estimates for parameters γ, η, and q as medians of marginal posterior distributions from either the sharing or causal model. Figure SN2 compares point estimation of the causal effect, γ, in simulations with a causal effect and no correlated pleiotropy using CAUSE and other methods. All methods display some shrinkage. For methods besides CAUSE, shrinkage is attributed to imperfect ascertainment of instruments. In every simulated data set, there is some chance of including variants that are genome-wide significant by chance and are not tagging a causal trait M variant. These variants are uncorrelated with Y and lead to underestimation of the causal effect. As the power of the trait M GWAS increases, shrinkage diminishes because there are more strong M effect variants meeting the inclusion criteria.

Some of the shrinkage in the CAUSE estimator is also due to uncertainty about which variants are M effect variants. CAUSE uses all variants, and the strength of each variant's affect on the estimate is proportionate to the strength of its association with M. Variants that do not tag an M effect variant but have strong associations with M by chance push the causal estimate downward. The CAUSE estimate is also shrunk by the prior on γ which is a wide normal distribution centered at zero. The prior on q does not affect the point estimate of γ in these experiments but does influence the width of the posterior distribution for γ, with narrower posteriors corresponding to stronger priors on q (Figure SN2(b)).

CAUSE obtains a smaller mean squared error (MSE) than Egger regression and the modal estimator for nearly every parameter value. MSE for Egger regression is not shown in Figure SN2c because it is much larger than the MSE for any other method and including it in the plot masks differences between other methods. CAUSE's MSE is smaller or comparable to other methods for small causal effects and grows for larger effects, reflecting its shrinkage.

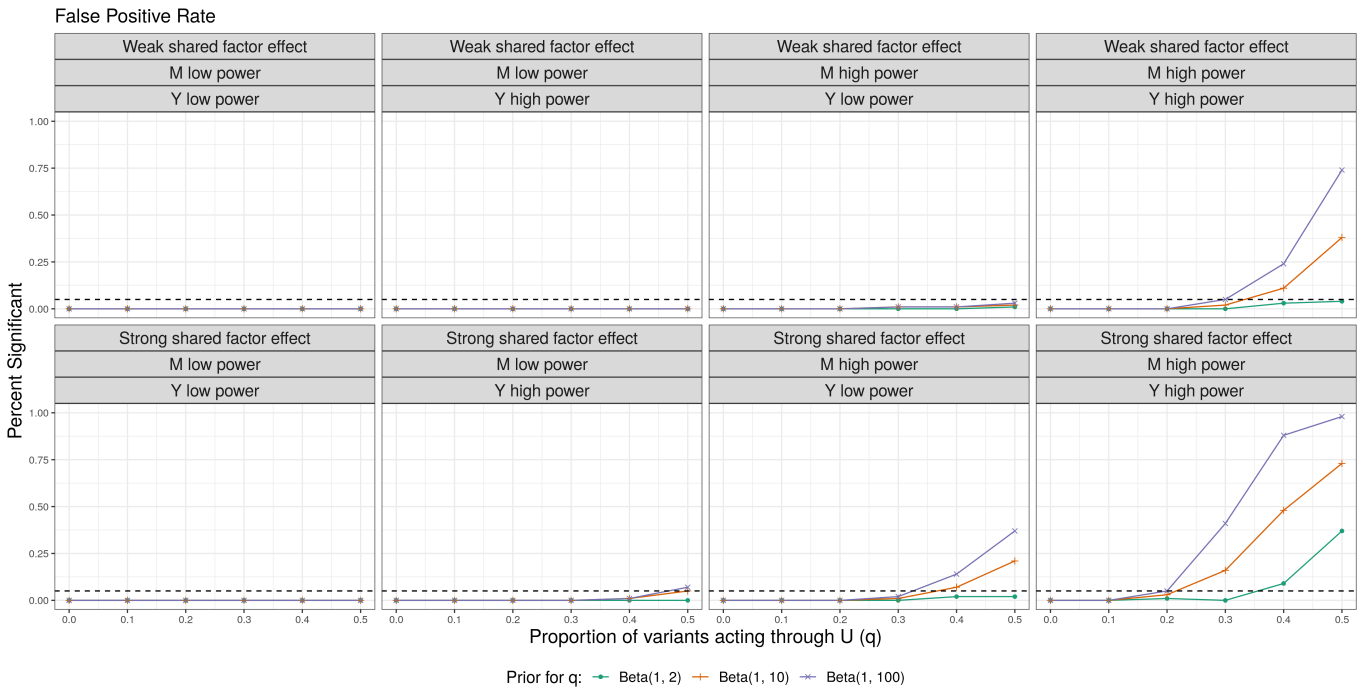
We find that accurately estimating q and η is more challenging than estimating of γ. Posterior distributions for q and η are strongly dependent and are affected by the prior distribution of q. In some cases, posterior distributions may be multimodal, making the posterior median a poor summary. Figure SN3 shows average posterior medians for q and η in simulations with correlated pleiotropy and no causal effect. As expected, posterior medians of q are shrunk more closely to zero when the prior places more weight on small values. This leads to wide credible intervals for η because there is little information about this parameter if q is small. Using the Beta$(1, 2)$ prior for q we find that credible intervals for η are slightly narrower and that the posterior median is less shrunk relative to the true value but that credible intervals for q are quite wide. We are only able to obtain accurate estimates of q and η in the highest power scenarios.

SN6.3 Causal effects with shared factors

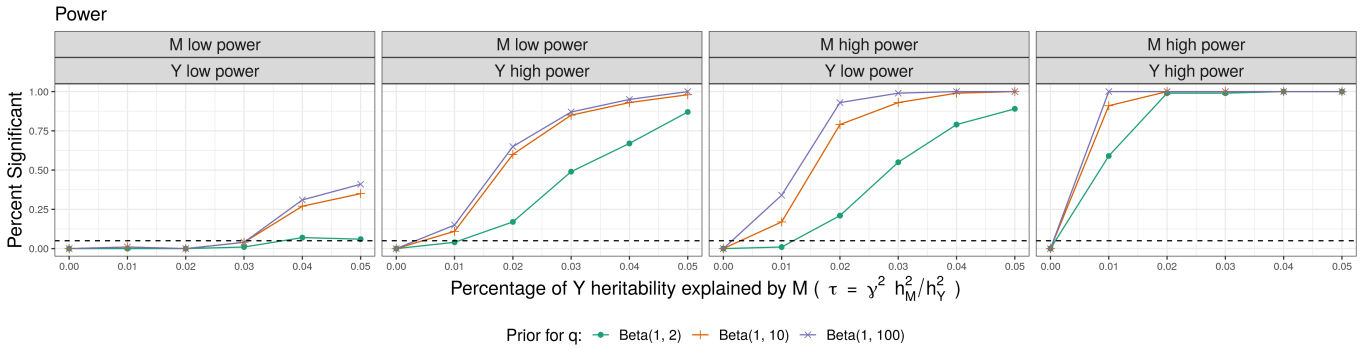
We explore simulation scenarios with both a causal effect and correlated pleiotropy. We focus primarily on settings with antagonistic correlated pleiotropy (correlation in the opposite direction from the causal effect), since we expect these scenarios to be the most difficult.

We first examine how a shared factor affects the power of each method. We simulate data with a causal of 0.2 and a proportion of variants, $q = 0.1$ or 0.3 acting through a shared factor with effect η ranging from -0.4 to 0.1. The power of CAUSE and other methods is shown in Figure SN4. A detection is only counted if the sign of the point estimate is positive. We did not find that erroneously detecting negative causal effects was a problem for any method (Egger regression detected one across all simulations and no other methods detected any). We find that CAUSE is able to maintain power when the shared factor effect has a smaller magnitude than the causal effect, but that when the shared factor effect is equal to the causal effect or larger and in the opposite direction, the power of CAUSE decreases. For larger antagonistic effects, correlation from the shared factor cancels out the correlation from the causal effect and the pattern of effect estimates becomes similar to the pattern created by a shared factor only.

(a) False Positive Rate



(b) Power



(c) False Positive Rate vs Power

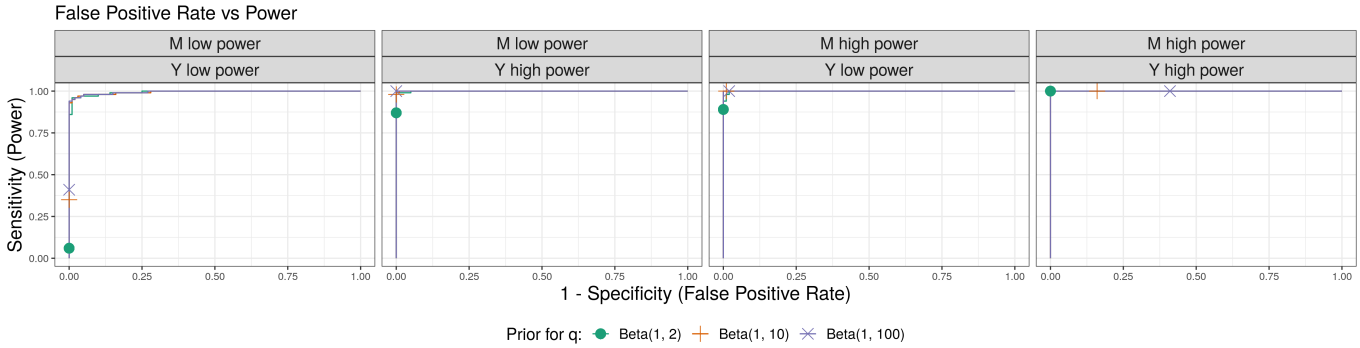


Figure SN1: Performance of CAUSE using different priors for q simulated data. (a) False positive rate averaged over 100 simulated data sets in settings with no causal effect and a proportion of correlated pleiotropic variants (q) ranging from 0 to 50%. (b) Power averaged over 100 simulated data sets 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 there is no shared factor to the false positive rate when there is no causal effect, but a proportion $q = 0.3$ of variants act through a shared factor with effect $\eta = \sqrt{0.05}$ on Y. 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 $p \leq 0.05$.

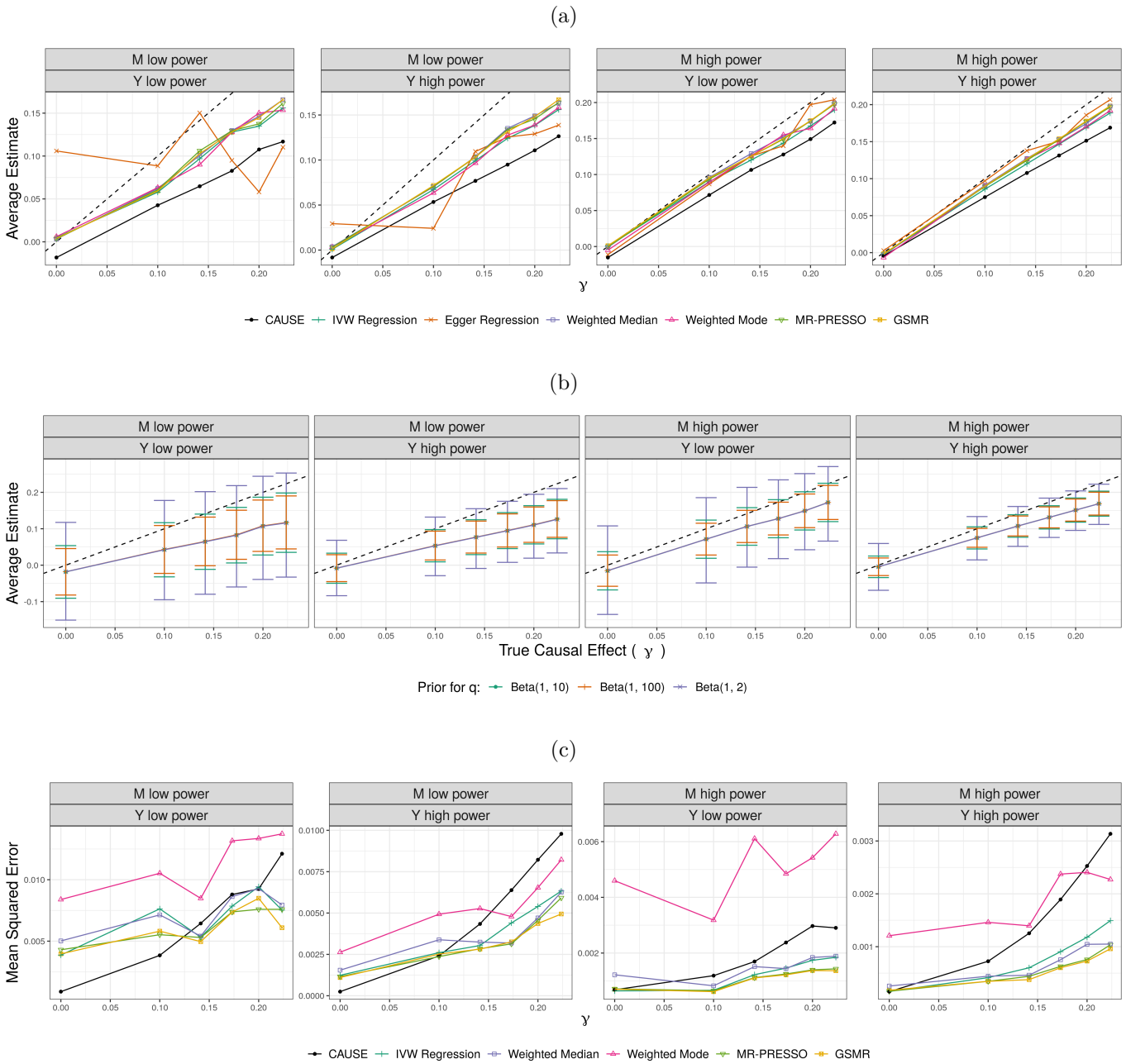
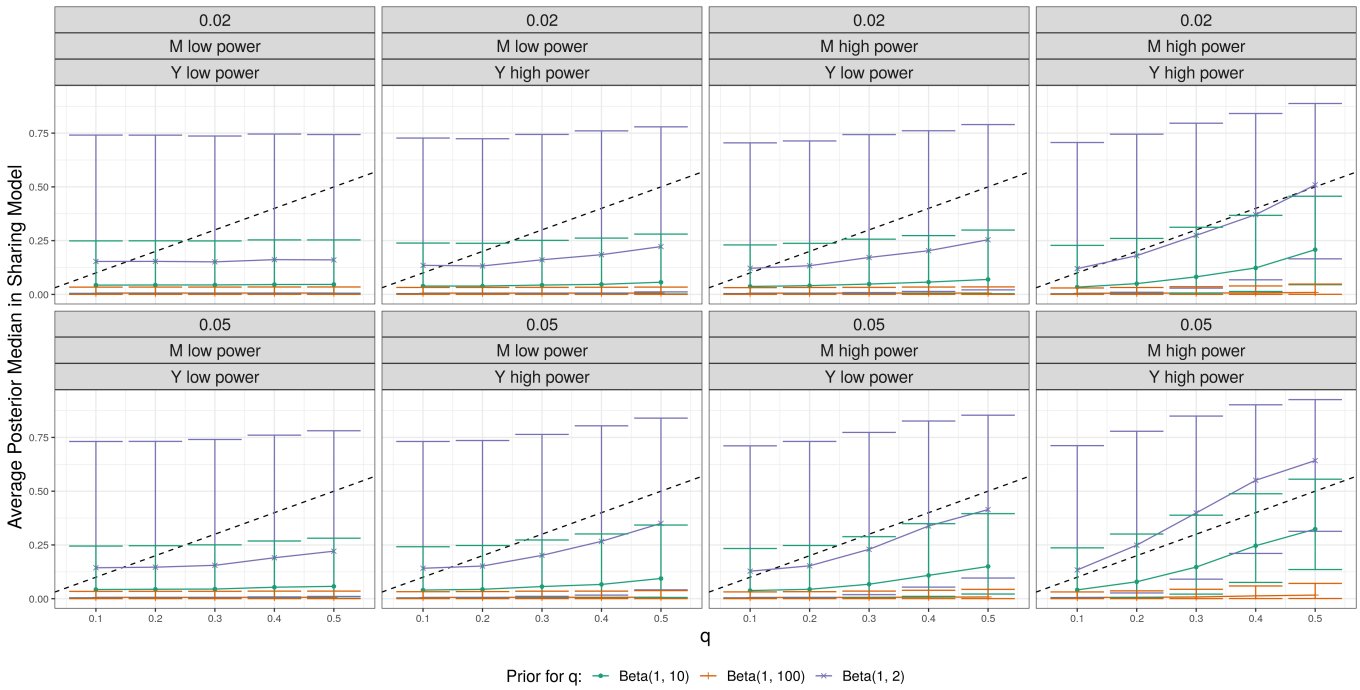


Figure SN2: Point estimation of γ for simulated data with a causal effect and no correlated pleiotropy. (a) Comparing CAUSE to other MR methods, points indicate the average point estimate over 100 simulations. CAUSE is run with the default Beta$(1, 10)$ prior for q. The dotted black line shows the true parameter value. (b) Comparing different priors for q. Error bars show average upper and lower extents of 95% credible intervals. CAUSE obtains the same point estimate using different priors but the posterior distribution is more peaked using a stronger prior on q. (c) Comparing mean squared error across methods. Egger regression is omitted because it obtains an MSE much larger than the other methods. CAUSE results using the default prior are shown. Note that the vertical axes in (c) are not the same across panels.

(a) Estimation of q



(b) Estimation of η

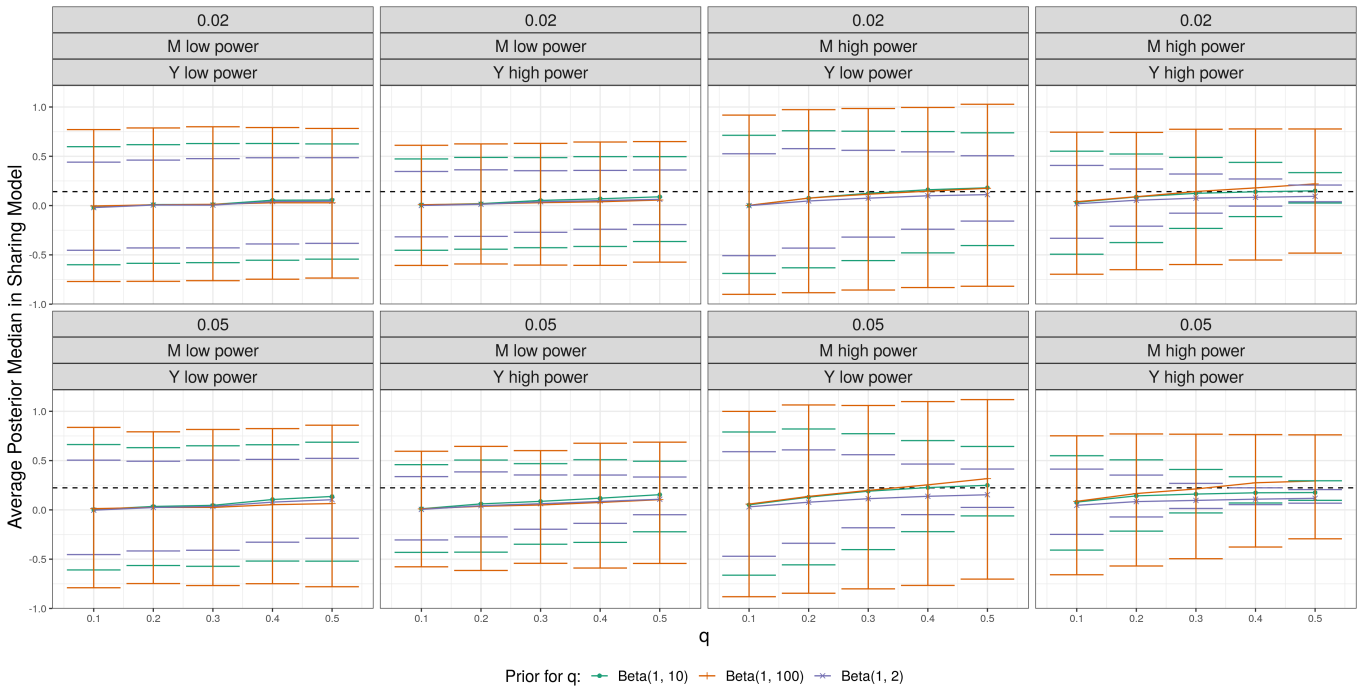


Figure SN3: Estimation of q and η from the sharing model in simulations with no causal effect. Points show the average posterior median over 100 simulations. Upper and lower ends of vertical bars indicate the average upper and lower boundary of 95% credible intervals. Dotted lines show the true value of the parameter with estimates shown, either q (a) or η (b).

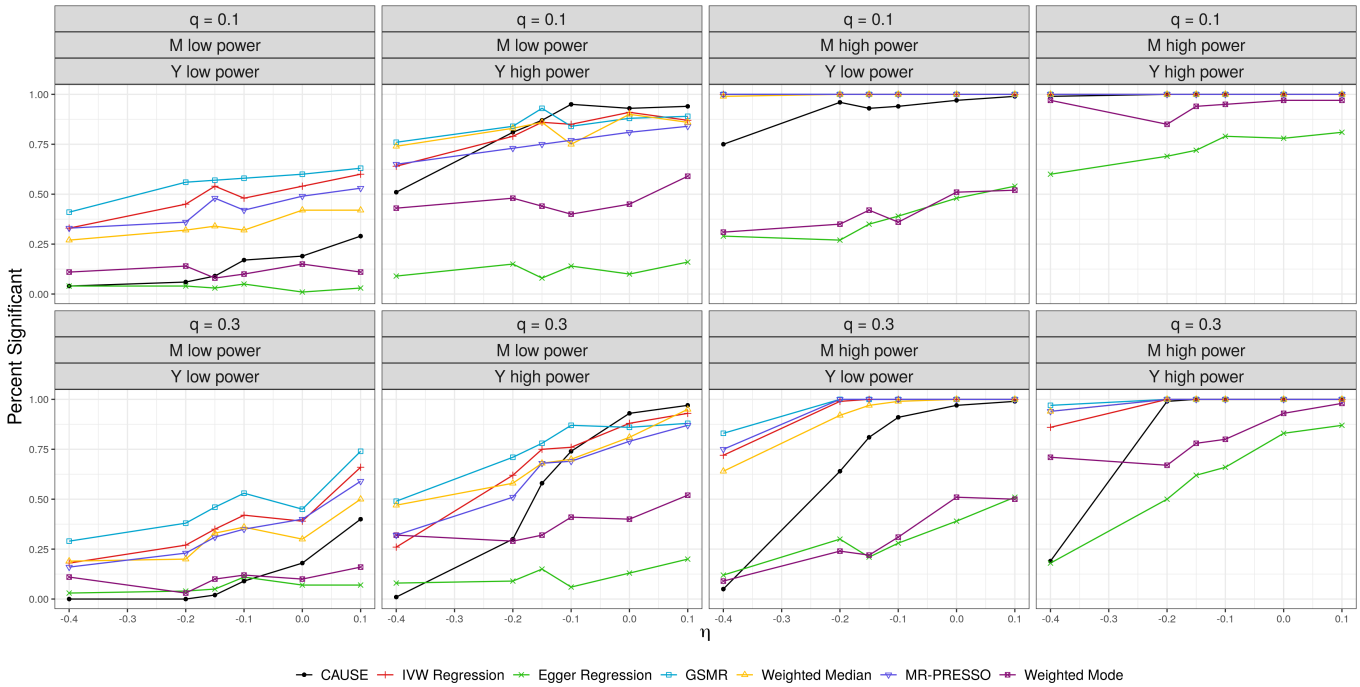


Figure SN4: Power comparison in simulations with both a causal effect and correlated pleiotropy. In all simulations, the causal effect is 0.2. Either 10% (top row) or 30% (bottom row) of variants display correlated pleiotropy with effect size given on the horizontal axis. Points indicate the proportion on simulations detected to have a positive causal effect by each method.

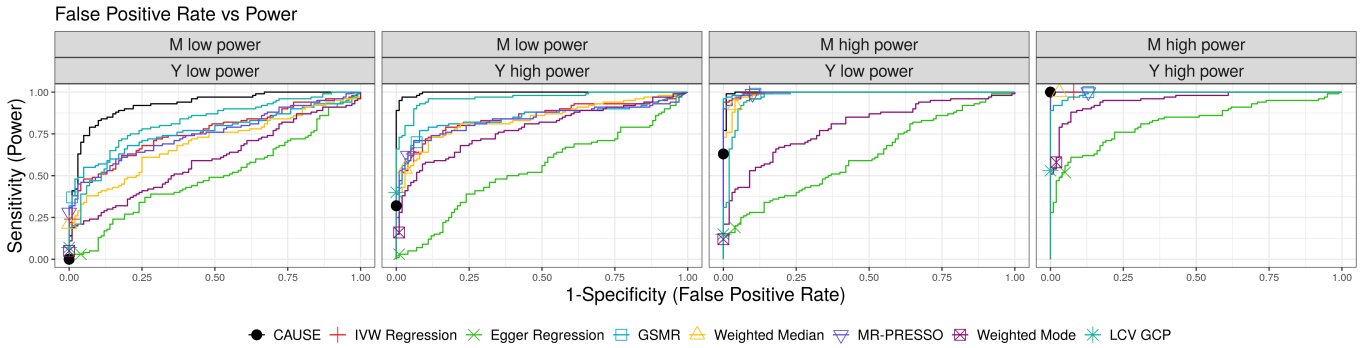
Next we compare the ability of CAUSE and other methods to discriminate a scenario with a causal effect and a shared factor from one with only a shared factor. The LCV GCP estimate is included in these comparisons. In each comparison, the non-null scenario is identical to the null scenario but with the addition of a causal effect. We consider shared factors accounting for 10% and 30% of trait M variants and causal effects of $\gamma = -1 * \eta$ or $\gamma = -2 * \eta$. The shared factor effect η is the same in all comparisons. Curves showing the trade-off between false and true positives are shown in Figure SN5. We find that, although CAUSE has lower power when there is an antagonistic shared factor, it is able to discriminate causal and non-causal scenarios at least as well as other methods.

SN7 Existing Information About Pairs of GWAS Traits

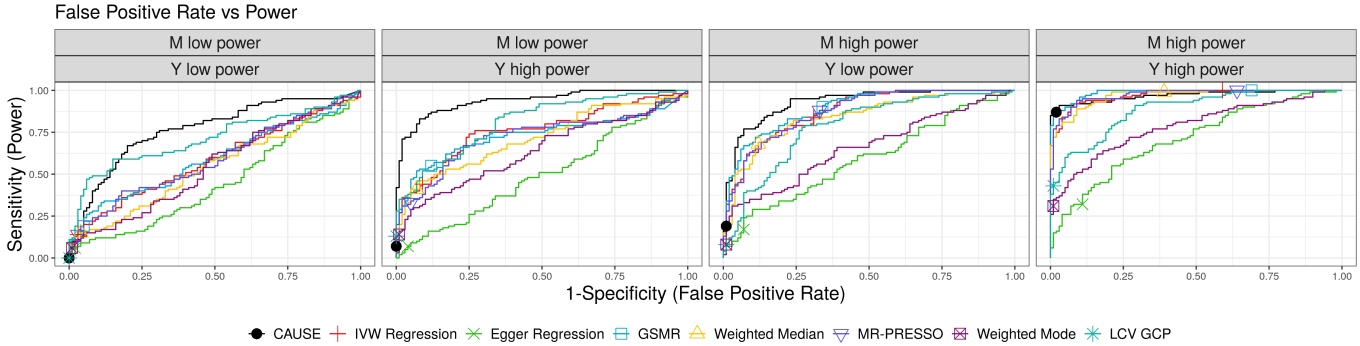
We provide brief summaries of literature about the relationships of the 12 risk factors and four diseases examined in the main Results. We have classified the relationship of each pair as considered causal (C), supported by literature (S), unknown (U), implausible and unsupported by literature (I), or considered non-causal (N). We classify a trait pair as considered causal if there are clinical trial or similar results supporting a causal effect and the effect is generally accepted in the literature. We classify pairs as supported by literature if there are strong molecular or physiological hypotheses supporting the effect and observational correlations. We classify pairs as implausible if there is no suggestion of a causal effect in the literature there is no genetic correlation. Effects of HDL on CAD and stroke are classified as considered to be non-causal, discussed below. All remaining trait pairs are classified as unknown or ambiguous. These include pairs that are correlated observationally but clinical trial or molecular evidence does not support a causal effect and pairs that are subject to ongoing debate. Below we summarize the literature used to make each classification, grouped by risk factor. At the end of the summary for each risk factor we list in parenthesis our classification of its relationship with CAD, stroke, T2D, and asthma in that order.

- Alcohol: Alcohol consumption is difficult to study due to its correlation with other dietary and lifestyle factors. Observational studies have suggested a protective effect of moderate alcohol consumption on cardiovascular disease including CAD and stroke [5]. This is the opposite direction of effect suggested

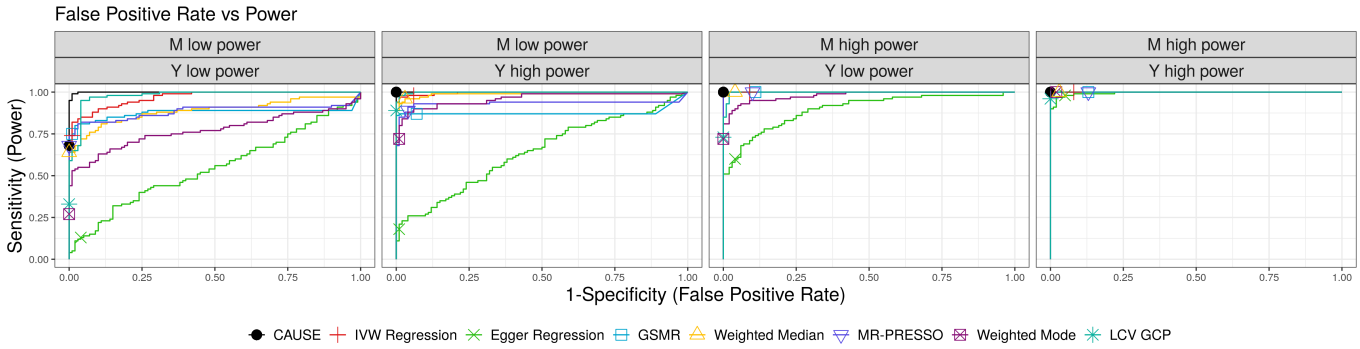
(a) (Null) $q = 0.1, \eta = -\sqrt{0.02}, \gamma = 0$ vs (Non-Null). $q = 0.1, \eta = -\sqrt{0.02}, \gamma = \sqrt{0.02}$



(b) (Null) $q = 0.3, \eta = -\sqrt{0.02}, \gamma = 0$ vs (Non-Null). $q = 0.3, \eta = -\sqrt{0.02}, \gamma = \sqrt{0.02}$



(c) (Null) $q = 0.1, \eta = -\sqrt{0.02}, \gamma = 0$ vs (Non-Null). $q = 0.1, \eta = -\sqrt{0.02}, \gamma = 2 * \sqrt{0.02}$



(d) (Null) $q = 0.3, \eta = -\sqrt{0.02}, \gamma = 0$ vs (Non-Null). $q = 0.3, \eta = -\sqrt{0.02}, \gamma = 2 * \sqrt{0.02}$

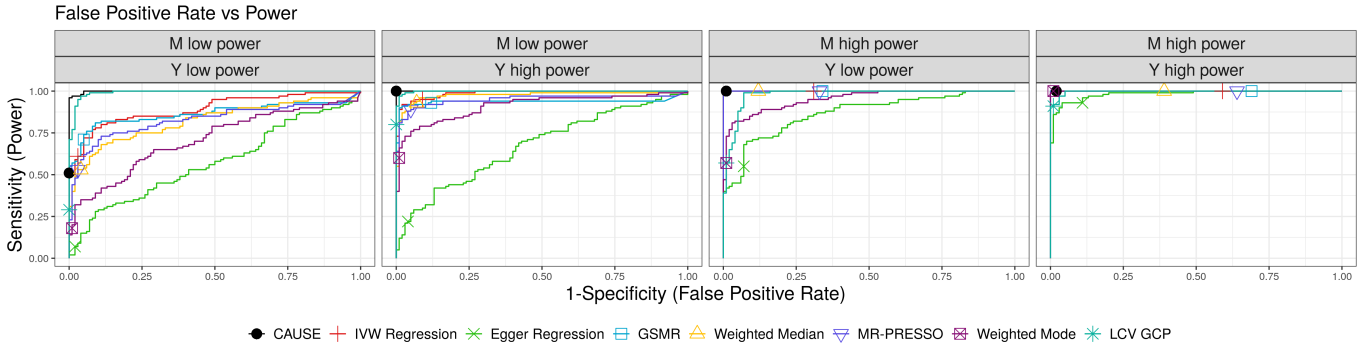


Figure SN5: Comparison of false positive-power trade-off. There are 100 simulations each in the causal and non-causal scenarios. Curves are created by varying the significance threshold. Points indicate the power and false positive rate achieved at a threshold of $p \leq 0.05$ or $G\hat{C}P0.06$ for LCV.

by modal MR of alcohol on CAD, the only method to find a causal effect of alcohol on any trait at $p < 0.05$. In a clinical trial, Davies et al. [6] find moderate alcohol consumption associated with increased insulin sensitivity, and decreased triglycerides, but no effect on fasting glucose. To our knowledge, there is no evidence linking alcohol consumption and asthma. (U, U, U, I)

- Smoking: Smoking is a well known risk factor for cardiovascular disease (CVD), including CAD and stroke, and has been studied extensively through observational and molecular studies over the last several decades [7]. This report also concludes that smoking is a cause of T2D, and that there may be a causal relationship between smoking and asthma but this remains inconclusive. No MR methods are able to detect effects of smoking on either diabetes risk or asthma. Associations between smoking behavior and asthma risk may be confounded by effects of parental smoking behavior. (C, C, C, S)

- Blood pressure (SBP, DBP): Two recent clinical trials have studied effects of intensive blood pressure lowering in groups with high risk of CVD. The ACCORD study [8] included individuals with T2D which is associated with elevated risk of CVD. The SPRINT trial [9] included individuals with baseline systolic blood pressure above 130 mm Hg and excluding those with prior stroke. Both studies had primary composite outcomes including both myocardial infarction and stroke. ACCORD did not find a significant effect of intensive blood pressure lowering on the composite outcome but did find evidence of an effect on stroke, a secondary outcome. The SPRINT trial was stopped early due to a significantly lower rate of cardiovascular events in the intensive treatment group suggesting a causal effect. The difference between treatment groups in SPRINT was driven primarily by a decrease in heart failure and a post-hoc analysis of stroke is not significant [10]. This may be attributable to the lower stroke risk of the SPRINT population, which may reduce power to detect an effect. Although the two studies have somewhat different outcomes, we classify the relationships of blood pressure with CAD and stroke both as considered causal due to strong observational evidence and support in clinical trials. To our knowledge, no randomized trials have examined relationships between blood pressure and T2D risk or asthma risk. Because blood pressure and asthma are also not genetically correlated, we classify that relationship as implausibly causal. Blood pressure and T2D are genetically correlated so this relationship is classified as unknown. (C, C, U, I)

- Fasting Glucose: Blood glucose control and insulin sensitivity are disrupted in T2D. Impaired fasting glucose is one of the diagnostic criteria for diabetes so it would be reasonable to expect these traits to share most of their genetic variants. CAUSE finds a significant result in both directions which is consistent with this expectation. It has been suggested that elevated levels of blood glucose can contribute to T2D risk by encouraging more insulin production which eventually overtaxes the pancreas. The causal ordering between elevated is unclear plausibly reciprocal (i.e. there may be causal relationships in both directions). In a meta analysis of cohort studies, Sarwar et al. [11] find a strong correlation between T2D and CVD. However, they find that the relationship between CVD and fasting glucose is non-linear and not significant in lower ranges. It is possible that this association is mediated by diabetic disruption of lipid profiles rather than elevated blood glucose. Fasting glucose is genetically correlated with stroke and CAD but existing literature does not provide clear evidence of a causal effect so these effects are classified as unknown. Fasting glucose and asthma are not genetically correlated and have not been linked in epidemiological literature. (U, U, S, I)

- Birth Weight: Horikoshi et al. [12] study genetic effects on birth weight and genetic correlation with other traits including CAD, T2D, and asthma. They find a strong negative genetic correlation between birth weight and both CAD and T2D and no significant genetic correlation with asthma. In our study we additionally find no significant genetic correlation with stroke. Explanations for the observed genetic correlations are unknown and may result from a combination of both maternal and fetal genetics. Tyrrell et al. [13] observe a negative correlation between paternal diabetes and birth weight and a positive correlation between maternal diabetes and birth weight. This is consistent with a model in which fetal inheritance of diabetes risk alleles lowers birth weight by impeding glucose metabolism but exposure to maternal high glucose during gestation increases birth weight. Horikoshi et al. [12] suggest that the relationship between birth weight and CAD may be mediated by effects on blood

pressure. If this were the case, we might also expect an effect on stroke risk, though no methods except Egger regression find evidence of this relationship in our data. (U, U, U, I)

- Height: Inverse correlations between adult height and both heart disease and stroke have been found in observational studies, with stronger and more consistent results for heart disease [14, 15]. In a Mendelian randomization study, Nüesch et al. [16] find evidence of a causal effect of height on CAD but not stroke. The authors suggest that the effect of height on CAD may be mediated by effects on lung function and lipid profiles. This study uses overlapping data and similar approaches to those used here so cannot be viewed as independent evidence. However, based on strong observational correlation and presence of a physiological explanation for a possible effect, we classify the effect of height of CAD as supported. We classify the effect of height on stroke as unknown. Height is not expected to be causally related to T2D or asthma and is not genetically correlated with either of these traits. (S, U, I, I)

- BMI and Body Fat Percentage: We use the same classifications for body fat percentage and BMI because they are closely related and, there is not always research examining the two traits separately, and we have not found qualitatively different evidence for the two traits. Overweight and obesity is often considered a risk factor for CVD. Recently, Khan et al. [17] combined data from 10 prospective longitudinal cohort studies in the Cardiovascular Disease Lifetime Risk Pooling Project to estimate lifetime risk of CVD in BMI categories. They find increased risk of CVD in the highest categories with risk increasing for groups with higher BMI. This trend was strongest for cardiac outcomes and was weaker for stroke. Observational associations are suggestive but don't establish causality or the mechanism of association. However, because the correlations are found across studies and remain after controlling for lifestyle factors such as smoking, we categorize the relationships between BMI/body fat and CAD/stroke as supported. Obesity is typically considered a risk factor for T2D, primarily supported by consistent observational associations with higher risk in higher BMI categories [18]. Additional research has supported body fat as a stronger predictor of T2D than BMI [19]. Associations have been found between BMI and asthma in children, including in a Mendelian randomization study [20, 21]. However, the mechanism of this association is unknown and complicated by other environmental factors such as maternal BMI during gestation. It is also possible for asthma to increase risk of high BMI through decreased physical activity [22]. In a longitudinal study of a Swedish birth cohort Ekström et al. [22] observe higher BMI throughout childhood for females with persistent asthma but not for males or for children with transient asthma or asthma which onset after 4 years. Based on these findings, we categorize the relationship between BMI and body fat and asthma as unknown. (S, S, S, U)

- HDL: Low levels of HDL cholesterol are correlated with increased risk of heart disease and stroke in observational studies. However, HDL is also correlated with other known CAD risk factors including LDL cholesterol. Voight et al. [23] identify a large effect monogenic variant affecting HDL independent of LDL and triglyceride levels and demonstrate no association between this variant and CAD. Additionally, Burgess and Bowden [24] use multivariable MR to adjust for LDL cholesterol and triglyceride levels and find no effect of HDL on CAD conditional on these variables. In addition to this evidence, there are several Mendelian variants with large effects on HDL cholesterol levels that do not have clear effects on cardiovascular disease, and clinical trials of HDL raising drugs have failed to consistently show an effect. In combination, this evidence has created significant doubts that HDL cholesterol is protective for atherosclerosis as had been previously hypothesized [25]. Dyslipidemia (high LDL cholesterol, high trigylycerides, low HDL cholesterol) is often associated with T2D. This observation has been attributed to secondary effects of insulin resistance [26] but there have also been suggestions that imbalanced lipids could play a role in disease development and that particularly, HDL may be protective for pancreatic β-cells [27]. A Mendelian randomization study [28] using overlapping data to the data used here finds nominal associations for some tests including association between genetically higher LDL and lower risk of diabetes and weaker associations for higher HDL and lower disease risk. We note that the association between genetically predicted LDL and T2D has opposite sign from the observational association. Fall et al. [28] conclude that they have not identified consistent evidence for a causal role of circulating lipids and cite several possible confounding factors that may affect MR

results. Based on this we classify the relationship of HDL with T2D as unknown. HDL cholesterol and asthma have not been linked and are not genetically correlated. (N, N, U, I)

- LDL: LDL cholesterol has been studied extensively through randomised trials of LDL lowering drugs and through MR studies of strong Mendelian LDL effect variants. Cholesterol Treatment Trialsists Collaborators [29] meta-analyze 14 statin trials and find a significant reduction in incidence of coronary heart disease and events related to vascular disease including myocardial infarction and stroke. In addition to this evidence, studies of individuals with monogenic forms of high LDL cholesterol find that these individuals are at high risk for atherosclerosis, or a build up of plaque in the arteries that can lead to heart disease and stroke and find higher incidence of CAD for individuals with monogenic high LDL cholesterol than for those without those mutations [30, 31]. Based on this evidence, we classify relationships of LDL with CAD and stroke as considered causal. Based on research summarized in the HDL section, we classify the relationship of LDL and T2D as unknown. We further note that genetic and observational studies suggest different directions of association between LDL and T2D with high LDL often observed in individuals with T2D but genetically lower LDL associated with higher risk using some methods. Statin therapy which lowers LDL has been seen to increase risk of T2D through an unknown mechanism, further complicating the picture [32]. LDL cholesterol and asthma have not been linked and are not genetically correlated. (C, C, U, I)

- Triglycerides: Like LDL cholesterol, high triglyceride levels are positively correlated with risk of heart disease and stroke. Whether this relationship is causal and if triglycerides should be a therapeutic target has been debated and there is conflicting evidence. In a summary from the American Heart Association, Miller et al. [33] note that associations with cardiovascular disease disappear after controlling for other factors including HDL and non-HDL cholesterol levels. They conclude from a range of epidemiological and molecular evidence that triglyceride levels are not directly atherogenic but are useful biomarkers for cardiovascular disease due to high correlation with levels of risk increasing particles. However, a recent meta analysis of clinical trials of lipid lowering therapies that reduce triglycerides more than they reduce LDL cholesterol does find a significant effect of triglycerides on risk of cardiovascular events [34]. This association was strongly influenced by a single outlying study and becomes non-significant when that study is removed. Two Mendelian randomization studies adjusting for LDL and HDL found a significant effect of triglycerides on CAD [35, 24]. Based on this research we classify the relationship of triglycerides and CAD as supported though we note that this issue is still debated and classify triglycerides and stroke as unknown. Based on research summarized under HDL we classify the effect of triglycerides on T2D as unknown. Triglycerides and asthma have not been linked and are not genetically correlated so we classify that relationship as implausible. (S, U, U, I)

SN8 Discussion of selected negative results of CAUSE

In Table 1, there are several pairs of traits that are often considered to be causal and have some literature support, but CAUSE does not find a significant result. In order to understand these results more thoroughly, we examine the data and point-wise contribution to the CAUSE test statistic for these trait pairs.

SN8.1 Body fat percentage and CAD and T2D

Effect size plots for BF $\rightarrow$ CAD and BF $\rightarrow$ T2D are shown in Figure SN6, illustrating why CAUSE obtains a negative result. For both CAD and T2D there is significant heterogeneity in effect size correlation and slightly sub-threshold variants are not consistent with a causal effect. Causal effects of body fat percentage are found more strongly by the weighted median, weighted mode, and in the case of T2D, MR-PRESSO, methods that down-weight variants with heterogeneous or "outlying" effect size correlations.

SN8.2 Triglycerides and CAD

The effect of triglycerides on heart disease risk has been debated with conflicting evidence and opinions in the literature (see SN Section SN7). It is complicated by the fact that triglycerides are both genetically

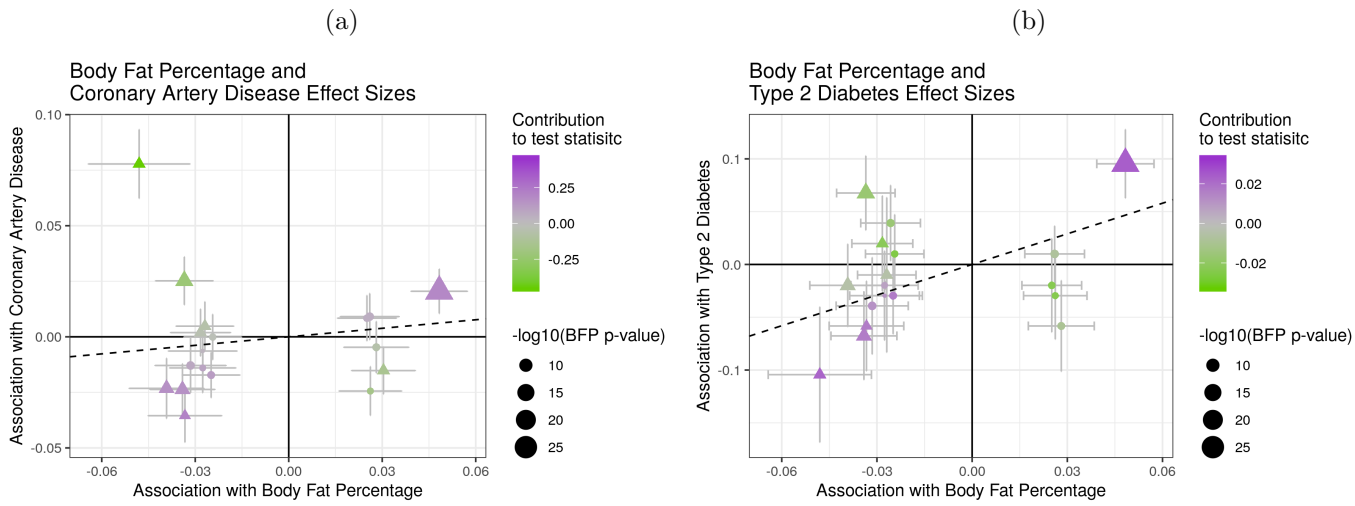


Figure SN6: Effect size estimates and variant level contribution to CAUSE test statistics for four trait pairs. Effect estimates for trait M (horizontal axis) are plotted against estimates for trait Y (vertical axis). Error bars have length 1.96 times the standard error of the estimate. Triangles indicate 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) Body fat percentage (M) and CAD (Y). (b) Body fat percentage (M) and T2D (Y).

and observationally correlated with several CAD risk factors including LDL cholesterol. Two Mendelian randomization studies, Do et al. [35] and Burgess and Bowden [24], have tested for a causal effect of triglycerides on heart disease controlling for LDL and HDL cholesterol. Both studies find in favor of a causal effect. In the data, we see that some variants are strongly associated with triglyceride levels but not with CAD or in the opposite direction from the main trend (Figure SN7a). Some of these variants are also strongly associated with LDL cholesterol or systolic blood pressure (Figures SN7b,c), and there appears to be less heterogeneity even when variants associated with these factors are removed, though this also eliminates all variants most strongly associated with triglycerides (Figure SN7d). In order to assess whether complex pleiotropy between triglycerides, LDL, and SBP could explain CAUSE's negative result, we ran CAUSE twice using two different subsets of variants. First using only variants with LDL p-value > 0.05 we obtained a p-value from CAUSE of 0.3 suggesting that LDL does not fully explain the heterogeneity of effect correlation. Second, using only variants with LDL p-value > 0.05 and SBP p-value > 0.05, we obtained a p-value from CAUSE of 0.044. These analyses are exploratory and more work needs to be done to fully understand this relationship.

SN9 Results using a low powered blood pressure GWAS

We use a lower-powered GWAS for systolic and diastolic blood pressure [36] to shed light on the performance of each method. This study includes only 69,395 individuals, about 10% of the sample size of the GWAS performed by Evangelou et al. used in the main Results. We note that the Ehret et al. study is a subset of the sample used by Evangelou et al. Results using the lower powered GWAS are shown in Table SN1. Interestingly, IVW regression and Egger regression are unable to detect the effect of blood pressure on CAD using the smaller study. An explanation for this can be seen in the effect estimates for SBP and CAD plotted in Figure SN8. Only seven variants reach genome wide significance. Six have correlated effects on both traits but one highly significant variant contradicts this pattern with a negative association with SBP and a strong positive association with CAD. The weighted median, weighted mode, and MR-PRESSO all down-weight outliers through different mechanisms and are therefore able to detect the effect. CAUSE uses variants that do not reach genome-wide significance which contribute to evidence of a causal effect and models multiple sources of pleiotropy allowing it to accommodate the outlying variant and detect the effect.

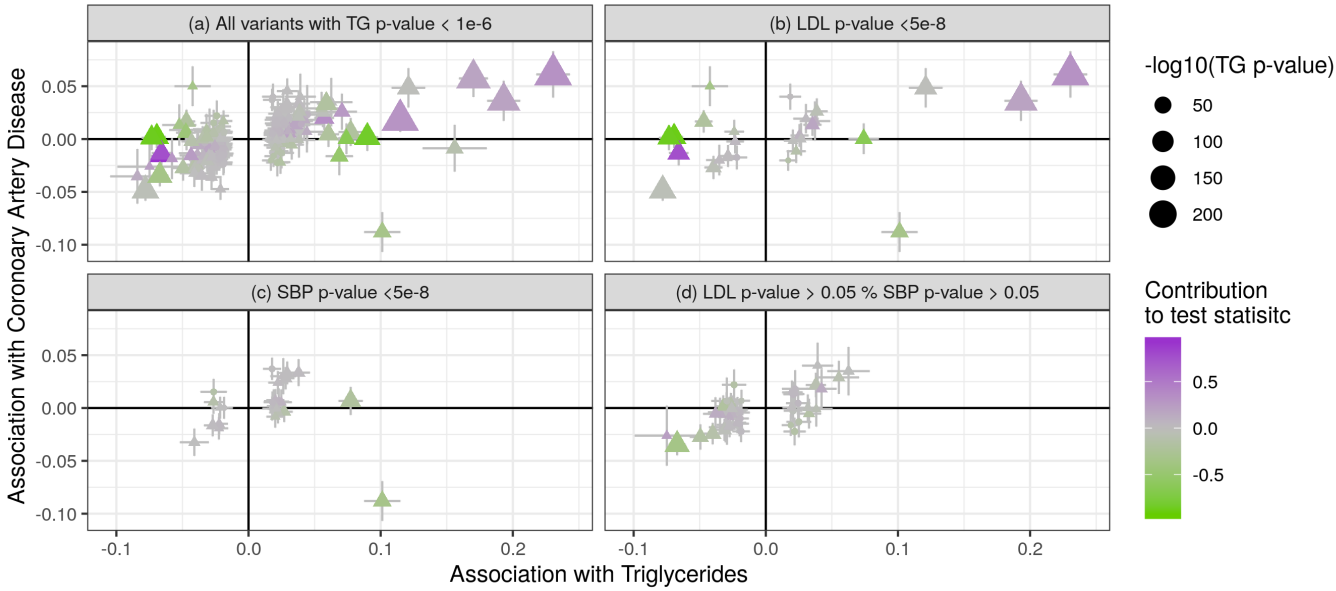


Figure SN7: Effect size estimates and variant level contribution to CAUSE test statistics for triglycerides and coronary artery disease. Symbols and error bars as in previous plots. Dotted lines show the IVW estimate obtained using only genome-wide significant variants. (a) All variants with triglycerides p-value $< 1 \cdot 10^{-}6$. Dotted line shows MR estimate using genome-wide significant variants. (b) Only variants with LDL p-value $< 5 \cdot 10^{-8}$. (c) Only variants with SBP p-value $< 5 \cdot 10^{-8}$. (d) Only variants with LDL p-value > 0.05 and SBP p-value > 0.05.

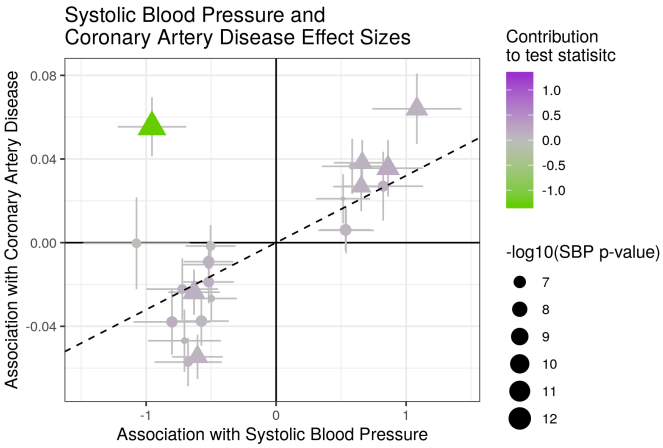


Figure SN8: Effect size estimates and variant level contribution to CAUSE test statistics for systolic blood pressure (SBP) and coronary artery disease (CAD). Effect estimates for SBP are taken from the GWAS of Ehret et al. which has lower power than the study used in the main Results. Symbols and error bars as in previous plots. The dotted line shows the IVW estimate obtained using only genome-wide significant variants.

Traits	CAUSE	IVW	Egger	Weighted Median	Weighted Mode	MR-PRESSO	LCV	CAUSE sharing q	Genetic Correlation
SBP → CAD	**0.0021** ↑	0.061 ↑	0.43 ↓	$\mathbf{1.1 \cdot 10^{-9}}$ ↑	$\mathbf{1.5 \cdot 10^{-11}}$ ↑	**0.00091** ↑	0.49(0.043)	0.54	$0.34(7.3 \cdot 10^{-15})$
DBP → CAD	**0.0017** ↑	0.064 ↑	0.87 ↑	$\mathbf{4.7 \cdot 10^{-8}}$ ↑	$\mathbf{1.8 \cdot 10^{-7}}$ ↑	**0.013** ↑	0.58(0.053)	0.38	$0.29(1.4 \cdot 10^{-10})$
SBP → Stroke	**0.003** ↑	**0.0053** ↑	0.26 ↓	**0.00044** ↑	**0.0037** ↑	**0.0075** ↑	0.13(0.34)	0.54	$0.37(1.8 \cdot 10^{-7})$
DBP → Stroke	**0.0014** ↑	$\mathbf{2.3 \cdot 10^{-6}}$ ↑	0.74 ↑	$\mathbf{8.7 \cdot 10^{-5}}$ ↑	**0.0048** ↑	$\mathbf{6.9 \cdot 10^{-5}}$ ↑	0.12(0.43)	0.38	$0.34(2.2 \cdot 10^{-5})$
SBP → T2D	0.22 ↑	**0.017** ↑	0.29 ↑	0.19 ↑	0.32 ↑	**0.022** ↑	−0.28(0.35)	0.1	0.18(0.0033)
DBP → T2D	0.096 ↑	0.081 ↑	**0.039** ↑	0.16 ↑	0.26 ↑	0.13 ↑	−0.19(0.42)	0.12	0.12(0.079)
DBP → Asthma	1 ↓	0.52 ↓	0.33 ↑	0.44 ↓	0.59 ↓	0.53 ↓	−0.06(0.77)	0.04	0.06(0.37)
SBP → Asthma	1 ↓	0.84 ↓	0.64 ↑	0.72 ↓	0.66 ↓	0.55 ↓	−0.14(0.79)	0.04	0.01(0.82)

Table SN1: Summary of results using the lower powered blood pressure GWAS of Ehret et al. 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. The LCV column gives the estimated GCP from LCV with p-value testing that GCP$= 0$ in parenthesis. Values are bold if estimated GCP> 0.6. The "Cause q" column gives the posterior median of q in the CAUSE sharing model and the GC column gives the genetic correlation estimated by LD score regression with p-value testing that genetic correlation is zero in parenthesis.

References

[1] Matthew Stephens. "False discovery rates: a new deal". In: *Biostatistics* 18.2 (2017), pp. 275–295.

[2] Richard McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan.* Boca Raton, FL: CRC Press, Taylor and Francis Group, 2016.

[3] Xiang Zhu and Matthew Stephens. "Bayesian large-scale multiple regression with summary statistics from genome-wide association studies". In: *Annals of Applied Statistics* 11.3 (2017), pp. 1561–1592.

[4] Luke J O'Connor and Alkes L. Price. "Distinguishing genetic correlation from causation across 52 diseases and complex traits". In: *Nature Genetics* 50 (2018), pp. 1728–1734.

[5] Paul E. Ronksley, Susan E. Brien, Barbara J. Turner, Kenneth J. Mukamal, and William A. Ghali. "Association of alcohol consumption with selected cardiovascular disease outcomes: A systematic review and meta-analysis". In: *Bmj* 342.7795 (2011), p. 479.

[6] Michael J. Davies, David J. Baer, Joseph T. Judd, Ellen D. Brown, William S. Campbell, et al. "Effects of moderate alcohol intake on fasting insulin and glucose concentrations and insulin sensitivity in postmenopausal women: A randomized controlled trial". In: *Journal of the American Medical Association* 287.19 (2002), pp. 2559–2562.

[7] U.S. Department of Health and Human Services. *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General.* Tech. rep. Atlanta: U.S. Department of Health et al., 2014.

[8] William C. Cushman, Gregory W. Evans, Robert P. Byington, David C. Goff, Richard H. Grimm, et al. "Effects of Intensive Blood-Pressure Control in Type 2 Diabetes Mellitus". In: *New England Journal of Medicine* 362.17 (2010), pp. 1575–1585.

[9] Jackson T. Wright, Jeff D. Williamson, Paul K. Whelton, Joni K. Snyder, Kaycee M. Sink, et al. "A Randomized Trial of Intensive Versus Standard Blood-Pressure Control". In: *New England Journal of Medicine* 373.22 (2015), pp. 2103–2116.

[10] Daniel T. Lackland, Robert M. Carey, Adriana B. Conforto, Clive Rosendorff, Paul K. Whelton, et al. "Implications of recent clinical trials and hypertension guidelines on stroke and future cerebrovascular research". In: *Stroke* 49.3 (2018), pp. 772–779.

[11] N. Sarwar, P. Gao, S. R. Kondapally Seshasai, R. Gobin, S. Kaptoge, et al. "Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: A collaborative meta-analysis of 102 prospective studies". In: *The Lancet* 375.9733 (2010), pp. 2215–2222.

[12] Momoko Horikoshi, Robin N. Beaumont, Felix R. Day, Nicole M. Warrington, Marjolein N. Kooijman, et al. "Genome-wide associations for birth weight and correlations with adult disease". In: *Nature* 538.7624 (2016), pp. 248–252.

[13] Jessica S. Tyrrell, Hanieh Yaghootkar, Rachel M. Freathy, Andrew T. Hattersley, and Timothy M. Frayling. "Parental diabetes and birthweight in 236 030 individuals in the UK Biobank study". In: *International Journal of Epidemiology* 42.6 (2013), pp. 1714–1723.

[14] D. A. Leon, G. D. Smith, M. Shipley, and D. Strachan. "Adult height and mortality in London: Early life, socioeconomic confounding, or shrinkage?" In: *Journal of Epidemiology and Community Health* 49.1 (1995), pp. 5–9.

[15] S. Goya Wannamethee, A. Gerald Shaper, P. H. Whincup, and M. Walker. "Adult height, stroke, and coronary heart disease". In: *American Journal of Epidemiology* 148.11 (1998), pp. 1069–1076.

[16] Eveline Nüesch, Caroline Dale, Tom M. Palmer, Jon White, Brendan J. Keating, et al. "Adult height, coronary heart disease and stroke: A multi-locus Mendelian randomization meta-analysis". In: *International Journal of Epidemiology* 45.6 (2016), pp. 1927–1937.

[17] Sadiya S. Khan, Hongyan Ning, John T. Wilkins, Norrina Allen, Mercedes Carnethon, et al. "Association of body mass index with lifetime risk of cardiovascular disease and compression of morbidity". In: *JAMA Cardiology* 3.4 (2018), pp. 280–287.

[18] Micael L Ganz, Neil WIntfeld, Qian Li, Veronica Alas, Jakob Langer, et al. "The association of body mass index with the risk of type 2 diabetes: A case-control study nested in an electronic health records system in the United States". In: *Diabetology and Metabolic Syndrome* 6.1 (2014), pp. 1–8.

[19] Ara Jo and Arch G. Mainous. "Informational value of percent body fat with body mass index for the risk of abnormal blood glucose: A nationally representative cross-sectional study". In: *BMJ Open* 8.4 (2018), pp. 1–7.

[20] K. G. Tantisira, A. A. Litonjua, S. T. Weiss, and A. L. Fuhlbrigge. "Association of body mass with pulmonary function in the Childhood Asthma Management Program (CAMP)". In: *Thorax* 58.12 (2003), pp. 1036–1041.

[21] Raquel Granell, A. John Henderson, David M. Evans, George Davey Smith, Andrew R. Ness, et al. "Effects of BMI, Fat Mass, and Lean Mass on Asthma in Childhood: A Mendelian Randomization Study". In: *PLoS Medicine* 11.7 (2014), pp. 1–14.

[22] Sandra Ekström, Jessica Magnusson, Inger Kull, Niklas Andersson, Matteo Bottai, et al. "Body Mass Index Development and Asthma Throughout Childhood". In: *American Journal of Epidemiology* 186.2 (2017), pp. 255–263.

[23] Benjamin F. Voight, Gina M. Peloso, Marju Orho-Melander, Ruth Frikke-Schmidt, Maja Barbalic, et al. "Plasma HDL cholesterol and risk of myocardial infarction: A mendelian randomisation study". In: *The Lancet* 380.9841 (2012), pp. 572–580.

[24] Stephen Burgess and Jack Bowden. "Integrating summarized data from multiple genetic variants in Mendelian randomization: bias and coverage properties of inverse-variance weighted methods". In: *arXiv* (2015). arXiv: `1512.04486`.

[25] Daniel J Rader and G Kees Hovingh. "HDL and cardiovascular disease". In: *The Lancet* 384.9943 (2014), pp. 618–625.

[26] Arshag D Moordian. "Dyslipidemia in type 2 diabetes mellitus". In: *Nature Clinical Practice* 5.3 (2009), pp. 150–159.

[27] Jannick Pétremand, Julien Puyal, Jean Yves Chatton, Jessica Duprez, Florent Allagnat, et al. "HDLs protect pancreatic β-cells against ER stress by restoring protein folding and trafficking". In: *Diabetes* 61.5 (2012), pp. 1100–1111.

[28] Tove Fall, Weijia Xie, Wenny Poon, Hanieh Yaghootkar, Reedik Magi, et al. "Using genetic variants to assess the relationship between circulating lipids and type 2 diabetes". In: *Diabetes* 64.7 (2015), pp. 2676–2684.

[29] Cholesterol Treatment Trialsists Collaborators. "Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90 056 participants in 14 randomised trials of statins". In: *The Lancet* (2005).

[30] Jonathan C. Cohen, Eric Boerwinkle, Thomas H. Mosley Jr, and Helen H. Hobbs. "Sequence Variations in PCSK0, Low LDL and Protection against Coronary Heart Disease". In: *Heart Disease* (2011), pp. 1264–1272.

[31] Daniel J. Rader, Jonathan Cohen, and Helen H. Hobbs. "Monogenic hypercholesterolemia: New insights in pathogenesis and treatment". In: *Journal of Clinical Investigation* 111.12 (2003), pp. 1795–1803.

[32] Daniel I. Swerdlow, David Preiss, Karoline B. Kuchenbaecker, Michael V. Holmes, Jorgen E.L. Engmann, et al. "HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: Evidence from genetic analysis and randomised trials". In: *The Lancet* 385.9965 (2015), pp. 351–361.

[33] Michael Miller, Neil J. Stone, Christie Ballantyne, Vera Bittner, Michael H. Criqui, et al. "Triglycerides and cardiovascular disease: A scientific statement from the American Heart Association". In: *Circulation* 123.20 (2011), pp. 2292–2333.

[34] Nicholas A. Marston, Robert P. Giugliano, Kyung Ah Im, Michael G. Silverman, Michelle L. O'Donoghue, et al. "Association Between Triglyceride Lowering and Reduction of Cardiovascular Risk Across Multiple Lipid-Lowering Therapeutic Classes: A Systematic Review and Meta-Regression Analysis of Randomized Controlled Trials". In: *Circulation* 140.16 (2019), pp. 1308–1317.

[35] Ron Do, Cristen J. Willer, Ellen M. Schmidt, Sebanti Sengupta, Chi Gao, et al. "Common variants associated with plasma triglycerides and risk for coronary artery disease". In: *Nature Genetics* 45.11 (2013), pp. 1345–1353.

[36] Georg B. Ehret, Patricia B. Munroe, Kenneth M. Rice, Murielle Bochud, Andrew D. Johnson, et al. "Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk". In: *Nature* 478.7367 (2011), pp. 103–109. arXiv: `NIHMS150003`.

[37] Evangelos Evangelou, Helen R. Warren, David Mosen-Ansorena, Borbala Mifsud, Raha Pazoki, et al. "Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits". In: *Nature Genetics* 50.12 (2018), pp. 1755–1755.