Published in final edited form as:
Nat Genet. 2018 May ; 50(5): 693–698. doi:10.1038/s41588-018-0099-7.

Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases

Marie Verbanck[1,2,3,*], **Chia-Yen Chen**[4,5,6,*], **Benjamin Neale**[4,5,6,§], and **Ron Do**[1,2,3,§]

[1]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, 1468 Madison Avenue, New York, NY, USA

[2]The Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, 1425 Madison Ave, New York, NY, USA

[3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, New York, NY, USA

[4]Analytic and Translational Genetics Unit, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA, USA

[5]Program in Medical and Population Genetics, Broad Institute, 75 Ames Street, Cambridge, MA, USA

[6]Stanley Center for Psychiatric Research, Broad Institute, 75 Ames Street, Cambridge, MA, USA

Abstract

Horizontal pleiotropy occurs when the variant has an effect on disease outside of its effect on the exposure in Mendelian randomization (MR). Violation of the ‘no horizontal pleiotropy’ assumption can cause severe bias in MR. We developed the Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-PRESSO) test to identify horizontal pleiotropic outliers in multi-instrument summary-level MR testing. We showed using simulations that MR-PRESSO is best suited when horizontal pleiotropy occurs in <50% of instruments. Next, we applied MR-PRESSO, along with several other MR tests to complex traits and diseases, and found that horizontal pleiotropy: (i) was detectable in over 48% of significant causal relationships in MR; (ii) introduced distortions in the causal estimates in MR that ranged on average from −131% to 201%; (iii)

Ron Do, Ph.D. The Charles Bronfman Institute for Personalized Medicine, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Pl., Box 1003, New York, NY 10029, ron.do@mssm.edu Tel: 212-241-6206. Benjamin Neale, Ph.D. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Richard B. Simches Research Center, 185 Cambridge Street, CPZN-6818, Boston, MA 02114, bneale@broadinstitute.org Tel: 617-643-5148.

[*]co-first authors: contributed equally

[§]co-senior authors: contributed equally

Author Contributions

M.V. and C-Y.C. contributed to study conception, data analysis, interpretation of the results and drafting of the manuscript. R.D. and B.N. contributed to study conception, interpretation of the results and critical revision of the manuscript.

Code and Data Availability Statement

We implemented MR-PRESSO using the R Project for Statistical Computing (version 3.3.1, Vienna, Austria). The MR-PRESSO software and full set of results are available (See URLs).

induced false positive causal relationships in up to 10% of relationships; and (iv) can be corrected in some but not all instances.

Epidemiological studies have established correlations between numerous exposures and complex diseases[1]. Drawing causal inferences from these studies can be challenging due to reverse causation, confounding and/or other biases[2].

Mendelian randomization (MR) is a commonly-used human genetics approach that can be used to infer causality of an exposure for a complex disease outcome[3,4]. MR presents a number of advantages over observational epidemiology, including the ability to control for non-heritable environmental confounders in such analyses and the use of genetic instruments to evaluate the impact of an exposure without necessitating the collection of that exposure in the outcome group. MR uses genetic variants as instrumental variables (IVs) that are robustly associated with the exposure of interest and tests whether the effects of the variants on the exposure result in proportional effects on the outcome.

In response to the advent of the genome-wide association (GWA) study and subsequent identification of thousands of trait-associated loci, multiple MR methods that leverage GWA summary statistics have been developed. These multi-instrument MR methods aggregate estimates from multiple IVs, testing for a causal relationship between a given exposure and outcome in a linear regression framework where the variants' effects on the outcome are regressed on the same variants' effects on the exposure[5,6].

A fundamental assumption of MR is the 'no horizontal pleiotropy' assumption (also called exclusion restriction criterion) which requires that the IV used for MR analysis acts on the target outcome exclusively through the exposure of interest[2]. Horizontal pleiotropy occurs when the variant has an effect on other traits outside of the pathway of the exposure of interest and has an impact on the target outcome, or when the variant has a direct effect on the target outcome[7]. As a violation of the 'no horizontal pleiotropy' assumption, horizontal pleiotropy can distort MR tests, leading to inaccurate causal estimates, loss of statistical power, and potential false positive causal relationships.

Emerging evidence has supported a pervasive role of pleiotropy amongst loci identified from GWA studies. Studies have shown that many traits are genetically correlated with each other[8]. Furthermore, studies have shown that hundreds of individual variants identified from GWA studies are associated with multiple traits[9–14].

As a result, there has recently been discussion regarding the potentially serious consequences horizontal pleiotropy may have on the validity of previous and current MR studies. Some have raised skepticism about the MR approach due to the pervasiveness of pleiotropy amongst trait-associated variants[15], while others have defended MR by noting that horizontal pleiotropy has long been known to impose limits on MR[16]. Regardless, the fact remains that the extent to which horizontal pleiotropy affects causal relationships inferred by MR is currently unknown. Importantly, a systematic evaluation of horizontal pleiotropy has not been performed.

Here, we conduct a systematic evaluation of the role of horizontal pleiotropy in MR. We developed the Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-PRESSO) approach to detect and correct for horizontal pleiotropic outliers in multi-instrument summary-level MR testing. In extensive simulations, we then evaluate the performance of MR-PRESSO and compare it to other complementary methods including methods that measure and correct for an average horizontal pleiotropic effect across all variants, and outlier-robust methods. Finally, we apply these methods to 4,250 MR tests of complex traits and diseases derived from 82 summary level GWA datasets.

Results

We developed MR-PRESSO to evaluate horizontal pleiotropy in multi-instrument summary-level MR. In brief, MR-PRESSO has three components (**Online Methods** and Figure 1), including: a) detection of horizontal pleiotropy (MR-PRESSO global test); b) correction for horizontal pleiotropy via outlier removal (MR-PRESSO outlier test); and c) testing of significant differences in the causal estimates before and after correction for outliers (MR-PRESSO distortion test). MR-PRESSO relies on a regression framework where the variants' effects on the outcome are regressed on the same variants' effects on exposure, with the slope of the regression line providing an estimate of the causal effect of the exposure on the outcome. The MR-PRESSO global test evaluates overall horizontal pleiotropy amongst all IVs in a single MR test by comparing the observed distance of all the variants to the regression line (residual sum of squares) to the expected distance under the null hypothesis of no horizontal pleiotropy. The MR-PRESSO outlier test evaluates the presence of specific horizontal pleiotropic outlier variants by using the observed and expected distributions of the tested variant. Finally, the MR-PRESSO distortion test evaluates the significance of the distortion between the causal estimate before and after removal of the horizontal pleiotropic outlier variants (detected from the outlier test of MR-PRESSO).

Comparison of statistical properties of MR methods

We assessed the statistical performances of several methods designed to either detect (MR-PRESSO global test, the Q test[17,18], Q (modified) test[19], Q' test [17,18] and Q' (modified) test[19]) or correct for horizontal pleiotropy (correction of the global average horizontal pleiotropic effect using MR-Egger regression[20,21] or Multi-variable MR[22,23]; outlier detection and removal approaches using MR-PRESSO outlier test, Cook's distance[24,25], Studentized residuals[24], Q (modified) outlier test or Q' (modified) outlier test) in multi-instrument summary-level MR. For these comparisons, we performed 10,000 simulations under different scenarios using the model described in Supplementary Figure 1b. We varied several parameters in the simulations including the main causal effect (β_{causal} = 0, 0.1, 0.2, 0.5 with $\beta_{pleiotropic}$ = 0.1), the percentage of horizontal pleiotropic outlier variants (0, 2, 4, 10, 50, 90%), the type of horizontal pleiotropy (positive or balanced) and whether or not we assume the Instrument Strength Independent of Direct Effect (InSIDE) condition[20] (**Online Methods**; **Supplementary Note**).

False positive rate/power to detect horizontal pleiotropy

Using simulations, we first assessed the false positive rate (type 1 error) and power (1 – type 2 error) of the above MR methods that detect horizontal pleiotropy, including the MR-PRESSO global test, the Q test[17,18], Q (modified) test[19], Q' test[17,18] and Q' (modified) test[19] (**Online Methods**; **Supplementary Note**).

Under the null hypothesis of no horizontal pleiotropy (Supplementary Table 1), the MR-PRESSO global test, the Q (modified) test and Q' (modified) test had controlled false positive rates (~5%) whereas the rates of the original Q and Q' tests were inflated (between 5 and 25%). Since the Q and Q' tests were found to be inflated, we proceeded with the MR-PRESSO global test, the Q (modified) test and the Q' (modified) test in power analyses. The power to detect horizontal pleiotropy was similar for the MR-PRESSO global test and the Q (modified) test (Table 1). We observed acceptable power to detect horizontal pleiotropy across all three tests under simulations when the percentage of horizontal pleiotropic variants was ≥ 10% (which corresponds to 5 horizontal pleiotropic variants out of 50). Similar results were observed across all magnitudes of causal estimates, including instances in which there was no causal relationship ($\beta_{causal} = 0$) (Table 1). In addition, we conducted several sensitivity analyses (violation of the InSIDE condition, percentage of horizontal pleiotropic variants ≥ 50% and perfectly overlapping samples to estimate the effect sizes on the outcome and exposure) to further evaluate the robustness of these methods. All three tests had high power to detect horizontal pleiotropy under a wide range of these parameters; however, simulations showed a reduction in power in those with perfectly overlapping samples (**Supplementary Results** and Supplementary Tables 2–4).

Evaluation of bias in the causal estimates

We investigated how the causal effect estimate (bias) and corresponding standard deviation (precision) of the MR-PRESSO outlier test were affected by horizontal pleiotropy. We then compared MR-PRESSO to other established MR methods that can correct for horizontal pleiotropy, including those that correct for a global average horizontal pleiotropic effect across all variants (e.g. inclusion of the intercept in MR-Egger regression[20,21], adjustment for multiple exposures in Multi-variable MR (MMR)[22,23]), horizontal pleiotropic outlier detection methods (Cook's distance[24,25], Studentized residuals[24], Q (modified) outlier test, Q' (modified) outlier test) and outlier-robust methods (weighted median[26], mode-based estimate[27]) (**Online Methods**).

The MR-PRESSO outlier test and Cook's distance had similar power to identify the correct horizontal pleiotropic outliers whereas Studentized residuals had low power (Supplementary Table 5; **Supplementary Note)**. However, Cook's distance had lower specificity to identify the correct horizontal pleiotropic outliers compared to the MR-PRESSO outlier test (inflated false positive rate (family-wise error rate), > 97% when there was no horizontal pleiotropy). The false positive rate (family-wise error rate) and power of the Q (modified) outlier test and the Q' (modified) outlier test were very similar to the MR-PRESSO outlier test (Supplementary Table 5).

MR-Egger regression generally had lower precision than the MR-PRESSO outlier test and MMR. The I_{GX}^2 index, which informs on when MR-Egger regression should be employed (i.e. I_{GX}^2>90%), was lower than 90% on average in most settings (Supplementary Table 6; **Supplementary Results**).

Cook's distance and Studentized residuals had similar bias and precision in the causal estimate compared to the MR-PRESSO outlier test (Supplementary Table 7). The weighted median had less bias but also less precision in the causal estimate compared to the MR-PRESSO outlier test, particularly when the percentage of horizontal pleiotropic variants was < 50%. The mode-based estimate generally had very low precision compared to the other methods.

All four methods (mode-based estimate, weighted median, Cook's distance and Studentized residuals) had similar limitations as the MR-PRESSO outlier test (e.g. inflated causal estimates and low precision) when there was a very high percentage of horizontal pleiotropic variants (≥ 50%) (Supplementary Table 7; **Supplementary Note**). Our simulations showed that when the InSIDE assumption was invalid, all methods exhibited bias in the causal estimate (Supplementary Tables 8 and 9; **Supplementary Note**).

The standard IVW approach showed an expected bias in the causal estimate due to horizontal pleiotropy. When the InSIDE assumption was valid and the percentage of horizontal pleiotropic variants was small (≤ 10%), the causal estimate of the MR-PRESSO outlier adjustment was less biased and had better precision (smaller standard deviation) than IVW, MR-Egger and MMR. However, when the percentage of horizontal pleiotropic variants was high (≥ 50%), the opposite was found. These trends were expected since outlier detection methods such as the MR-PRESSO outlier test require, as a baseline assumption, that at least 50% of the genetic variants to be valid instruments (no horizontal pleiotropy), have balanced pleiotropy and for the InSIDE assumption to be valid. Conversely, methods that correct for a global average horizontal pleiotropic effect amongst all variants (either by including the regression intercept in MR-Egger or covariate adjustment in MMR) are best suited when there is a large percentage of horizontal pleiotropic variants (> 50%).

Horizontal pleiotropy detection in MR for complex traits

We applied the MR-PRESSO global test, along with two methods (Q (modified) test, Q' (modified) test) that can detect horizontal pleiotropy in MR to all possible pairs of 82 complex traits and diseases retrieved from publicly available GWA datasets (Table 2; **Online Methods**). In total, we conducted 4,250 tests for each of the three MR approaches. We accounted for multiple testing of the 4,250 tests using the Bonferroni correction. We note that this correction is overly stringent since many of the traits and diseases are correlated. Using a Bonferroni-corrected threshold of $P < 1.17 \times 10^{-5}$, the MR-PRESSO global test was statistically significant in 21.69% (n = 922) of the 4,250 tests. When restricting to statistically significant causal estimates in the IVW meta-analysis, we detected significance of the MR-PRESSO global test at a higher rate of 48.69% (n = 93 of 191 tests). Both the Q (modified) test and the Q' (modified) test provided similar estimates. As a sensitivity analysis, we restricted to a subset of traits and diseases that were less correlated (e.g.

Pearson r < 0.30). We observed that 24.17% of MR tests were statistically significant for the MR-PRESSO global test amongst significant causal relationships (Supplementary Table 10).

Horizontal pleiotropy correction in MR for complex traits

We evaluated five methods to correct for horizontal pleiotropy in MR (Table 3). This included: 1) removing outliers detected by the MR-PRESSO outlier test, Q (modified) outlier test and Q’ (modified) outlier test (at Bonferroni-corrected threshold); and 2) adjusting for significant covariates individually or all together from the main MR test (e.g. significant causal effect in IVW meta-analysis at Bonferroni-corrected threshold). As shown in the simulations, the MR-PRESSO global test can be used to determine if there is any remaining horizontal pleiotropy after applying correction strategies to minimize initial horizontal pleiotropy in the MR test (**Supplementary Note**; Supplementary Table 11). In Table 3, we observed that the outlier removal approach using the MR-PRESSO outlier test was effective in eliminating statistical significance in the MR-PRESSO global test in 46% of the 922 tests. The Q (modified) outlier test and Q’ (modified) outlier test provided similar estimates. Furthermore, the covariate adjustment approach – defined by accounting for traits that were shown to have a significant causal effect on the same outcome – eliminated significance in the MR-PRESSO global test in 22% (n = 20) of the 93 tests when adjusting for a single covariate. When adjusting for all significant covariates in the same model, the covariate adjustment approach eliminated significance in 34% (n = 22) of the 42 tests. Taken together, these two correction strategies (MR-PRESSO outlier removal and MMR) were successful in 47% (n = 438) of the 922 tests total. Furthermore, we note that the covariate adjustment approach is limited in that it requires a priori knowledge of the trait responsible for the horizontal pleiotropic effect.

Outliers effect on the distortion of MR causal estimates

We evaluated the extent to which outliers cause distortion in the causal estimates resulting from MR. Using the MR-PRESSO distortion test, we compared the causal estimates from the IVW meta-analysis before and after removal of outlier variants detected by the MR-PRESSO outlier test (**Online Methods**). Using a Bonferroni-corrected threshold, we observed a significant distortion (of −93% and 35%) in 2.5% (n = 2) of significant causal estimates (n = 81 total). Since the Bonferroni correction is overly stringent, we considered the commonly-used nominal threshold of $P < 0.05$ that the majority of MR studies to date have used for statistical significance. A significant distortion was observed in almost 10% (n = 22) of the causal relationships (n = 229 total) with a distortion between −131% and 201% on average (Figure 2).

Below, we provide one example to highlight the role of pleiotropy in MR (Supplementary Figure 2). We observed that the causal effect of body mass index (BMI) on C-reactive protein was estimated to be 0.39 ($P = 7.02 \times 10^{-8}$) by IVW. The MR-PRESSO global test showed statistical significance ($P < 10^{-6}$) and the MR-PRESSO outlier test identified one significant outlier variant ($P < 10^{-6}$, rs2075650 in the *APOE* locus). Examining this further, we observed that this variant was highly pleiotropic, having associations with several traits and diseases including Alzheimer’s Disease, body mass index, C-reactive protein, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, plasma triglycerides, waist circumference, hip circumference and waist-hip ratio at $P = 6 \times 10^{-4}$ (Bonferroni-corrected $P = 0.05 / 82$; Supplementary Table 12). Furthermore, this variant was associated with several other traits and diseases in the public NHGRI-EBI GWAS catalog[28] ($P < 5 \times 10^{-8}$; Supplementary Table 13). We believe some of these genetic associations with other traits and diseases will be due to horizontal pleiotropy while others will be in the same causal pathway (e.g. vertical pleiotropy). After removing this outlier variant, we observed a lower estimation of the causal estimate for BMI on C-reactive protein ($\beta_{causal} = 0.35$, $P = 3.45 \times 10^{-16}$) with this single variant alone causing a 12% distortion in the causal estimate (Supplementary Figure 2).

Outliers effect on false positive causal relations in MR

We evaluated the extent to which outliers (as detected by the MR-PRESSO outlier test) can induce false positive causal relationships. A false positive causal relationship was defined as an exposure-outcome pair in which the causal estimate was no longer statistically significant in the outlier-corrected IVW model but was previously significant in the naïve IVW model. According to this definition, we identified false positive relationships in 10% of putatively causal relationships (n = 24 out of 229 total tests) using the common nominal $P < 0.05$ threshold and 1.2% (n = 1 out of 81 total tests) using the stringent Bonferroni-corrected threshold ($P < 1.17 \times 10^{-5}$). We note that the outlier removal approach decreases the power of the outlier-corrected IVW model test because a smaller number of variants is included after removal of horizontal pleiotropic outlier variants. Therefore, we expect that a small number of false positive causal relationships may be due to reduced statistical power.

Causal relationships inferred from MR testing

We identified 191 significant causal relationships out of a total number of 4,250 MR tests using the Bonferroni-corrected threshold of $P < 0.05 / 4{,}250 < 1 \times 10^{-5}$. We note that many of the traits examined are closely related to each other (e.g. BMI/waist circumference/hip circumference, and low-density lipoprotein cholesterol (LDL-C)/total cholesterol) and therefore the tests are not completely independent of each other. After correcting for horizontal pleiotropy via outlier removal using the MR-PRESSO outlier test, we validated known causal relationships including LDL-C on coronary artery disease (CAD) ($\beta_{causal} = 0.52$, $P = 5.15 \times 10^{-12}$)[29], systolic blood pressure on CAD ($\beta_{causal} = 0.05$, $P = 1.78 \times 10^{-6}$)[30–32], BMI on type 2 diabetes ($\beta_{causal} = 0.76$, $P = 2.19 \times 10^{-9}$)[33] and BMI on C-reactive protein ($\beta_{causal} = 0.35$, $P = 3.45 \times 10^{-16}$), amongst the strongest findings. Furthermore, we observed an effect of BMI on uric acid ($\beta_{causal} = 0.31$, $P = 3.29 \times 10^{-15}$)[34] and plasma triglycerides ($\beta_{causal} = 0.20$, $P = 6.9 \times 10^{-15}$)[12], although these are significant for the MR-PRESSO global test even after correction via outlier removal.

Discussion

In summary, we have evaluated horizontal pleiotropy in the context of MR testing across pairwise comparisons of a large number of complex traits and diseases. We have: i) developed the MR-PRESSO method to detect and correct for horizontal pleiotropic outliers in MR and compared it to several established methods; (ii) applied several of these MR methods to complex traits and diseases, and showed that horizontal pleiotropy occurs in over

48% of causal relationships between complex traits and diseases inferred by MR; (iii) observed average distortion between −131% and 201% in the causal estimates of MR due to horizontal pleiotropy; and (iv) showed that horizontal pleiotropy can be minimized and corrected in some cases through outlier detection and/or secondary phenotype adjustment when the mediating trait is known.

By applying the MR-PRESSO global test to detect horizontal pleiotropy in a wide array of complex traits and diseases, we observed horizontal pleiotropy in approximately over 48% of inferred causal relationships. This is consistent with emerging evidence that many disease-associated variants identified from GWA studies have effects on multiple traits[10]. Since these variants are used as IVs in multi-instrument MR, it is likely that a non-negligible number of these variants do not meet the ‘no horizontal pleiotropy assumption’ in MR. Furthermore, Hemani et al.[35] have evaluated horizontal pleiotropy in a study concomitant to the present study. The study proposed a mixture-of-experts machine learning framework to select the most appropriate MR method amongst a variety of standard and horizontal pleiotropy-robust MR methods associated with an IV selection procedure. The framework selected a method that involved horizontal pleiotropy (pleiotropy-robust method or IV filtering) in 90% of the MR tests. These results indicate that horizontal pleiotropy is commonplace and highlight the need to evaluate horizontal pleiotropy for variants acting as instrumental variables as a necessary and standard test when performing MR.

Horizontal pleiotropy in MR has direct implications for genetics-guided drug discovery and validation. Accurate estimates of causal effects between biomarkers and diseases can inform dose-response curves for drug efficacy and safety[36]. In the present study, we show that horizontal pleiotropy can induce distortion in the causal estimates in MR and that this distortion is pervasive amongst many causal relationships. Secondly, there is increasing interest in using surrogate endpoints for drugs in clinical trials. Identifying true causal relationships using MR can pinpoint biomarkers that are causal and hence identify those surrogate endpoints that are most relevant to disease[37].

The current study has several strengths. Outlier detection methods are useful because they work within the framework of IVW. Furthermore, the MR-PRESSO global test (as well as the Q (modified) test and Q’ (modified) test) is adequately powered to detect horizontal pleiotropy amongst even a small subset of loci. Finally, outlier detection methods can be used in several different MR tests including IVW, MMR and even within the framework of MR-Egger regression. The MR-PRESSO method also has limitations. There were instances for which correction strategies (outlier removal or covariate adjustment) could not completely remove horizontal pleiotropy as detected by the MR-PRESSO global test. Possible reasons include violation of the InSIDE condition (as shown in the simulations) which is untestable, a percentage of horizontal pleiotropic outlier variants that is not suitable for the particular correction strategy (outlier detection method with > 50% of horizontal pleiotropic variants), no applicable covariate adjustment, or other sources of heterogeneity in the effect sizes other than horizontal pleiotropy such as gene-gene and gene-environment interactions. Furthermore, several GWA consortia use the same cohorts and study samples; therefore, some GWA summary statistics may have overlapping samples. In simulations, we evaluated the effect of perfectly overlapping samples on the power of our global test across a range of scenarios. A reduction in power for detecting horizontal pleiotropy was observed in the model with perfectly overlapping samples. Finally, because MR-PRESSO requires simulations, the processing time to apply the method can be slower compared to other methods.

In summary, we have shown through a series of analyses, that horizontal pleiotropy is pervasive in MR testing between complex traits and diseases, highlighting the need to employ approaches that minimize horizontal pleiotropy. Rigorous analysis and careful interpretation of causal inference testing in MR is warranted as a result of these observations.

Online Methods

General assumptions of Mendelian randomization

MR relies on genetic variants that are robustly associated with the tested exposure. A variant can be used as an instrumental variable (IV) through its effect size on the exposure and its proportional effect on the outcome. The causal estimate is an inferred estimate of the predicted response of the outcome caused by a modification of the exposure. Supplementary Figure 3 illustrates a standard MR framework (**Supplementary Note**). The single instrument approach can be extended to multiple instruments. Multi-instrument MR can be performed using an inverse variance-weighted, fixed effects meta-analysis[38] (IVW meta-analysis). IVW meta-analysis consists of fitting a weighted linear regression with a fixed intercept of 0 between the set of effect sizes on the outcome (either continuous or dichotomous response) and the effect sizes on the exposure (either continuous or dichotomous predictor), with the inverse of the variance of the effect sizes on the outcome as weights. When the intercept is fixed to 0, the slope of the regression then provides an estimate of the causal effect of the exposure on the outcome[5].

The validity of MR analysis relies on three assumptions (Supplementary Figure 3): 1) the variant (i.e. IV) is associated with the exposure; 2) the variant is independent of all confounders of the exposure-outcome relationship; and 3) the variant is independent of the outcome conditional on the exposure and all confounders of the exposure-outcome association (i.e. exclusion restriction criterion). Violation of the third assumption, the exclusion restriction criterion, is a direct consequence of horizontal pleiotropy (Supplementary Figure 1).

Existing methods to detect and correct for horizontal pleiotropy in MR

Several existing methods have been developed to detect horizontal pleiotropy in MR. Methods to detect horizontal pleiotropy include the Q test and the Q' test, as well as the recently proposed modified versions of these tests[19], which are traditionally used to identify over dispersion and have been applied in the context of MR[17].

Several existing MR methods have also been designed to correct for horizontal pleiotropy. Methods that we evaluated in the current study are described below.

Outlier methods: Traditional methods to detect outliers have been applied in the context of MR to detect and remove invalid IVs[24]. We applied Cook's distance where we defined outliers using a threshold of greater than four divided by the number of variants[25]. We also applied Studentized residuals where we performed Student tests on the Studentized residuals and defined outliers using a Bonferroni-corrected *P-value* threshold (0.05 divided by the number of variants). Finally, the Q (modified) outlier test and the Q' (modified) outlier test use a χ^2distribution with one degree of freedom to test each variant separately[19] in association with a Bonferroni-corrected *P-value* threshold (0.05 divided by the number of variants).

Correction of the average horizontal pleiotropic effect methods: The intercept of MR-Egger regression[20] can correct for the average horizontal pleiotropic effect across all IVs. The I_{GX}^2 index, which encapsulates the collective strength of a set of variants, can be used to inform on the use of MR-Egger regression[21]. MMR (Multi-variable MR)[22,23] takes into account known causal exposures to the outcome (E_2 in Supplementary Figure 1b). If the IV is subject to horizontal pleiotropy and has a significant effect on not only the tested exposure E_1 but also on a second exposure E_2, then MMR allows to adjust for this secondary horizontal pleiotropic effect. In practice, MMR is implemented by fitting a weighted linear regression by regressing $\hat{\Gamma}_{1,j}$ on both $\hat{\gamma}_{1,j}$ and $\hat{\gamma}_{2,j}$ (with the inverse of the variance of the effect sizes on the outcome as weights). The $\hat{\gamma}_{2,j}$ are the genetic effects of the variants on the known secondary causal exposure E_2. The model can be extended to include multiple exposures. In this study, a fixed intercept of 0 was used in all MMR models.

Outlier-robust methods: Additional methods are naturally robust to horizontal pleiotropic outlier variants in MR. The weighted median[26] estimator provides a consistent estimate of the causal effect when up to 50% of genetic variants are invalid instruments. The mode-based estimate (MBE) method[27] uses the property that valid instruments should provide the largest number of similar individual-instrument causal estimates even if the majority of instruments is invalid.

MR-PRESSO: MR Pleiotropy RESidual Sum and Outlier

We developed the MR Pleiotropy RESidual Sum and Outlier (MR-PRESSO) test to detect and correct for horizontal pleiotropic outliers. MR-PRESSO is comprised of three components (Figure 1):

a) detection of horizontal pleiotropy (and violation of the exclusion restriction assumption, Supplementary Figure 1) in MR (*global test*);

b) correction by removal of offending IVs that are due to horizontal pleiotropy (*outlier test*);

c) testing of significant differences in the causal estimates before and after outlier removal (*distortion test*).

MR-PRESSO extends the framework of IVW meta-analysis and is based on the following rationale. IVW meta-analysis consists of fitting a weighted linear regression between the set of effect sizes on the outcome and the effect sizes on the exposure. The slope of the regression provides an estimate of the causal effect of the exposure on the outcome. Under the null hypothesis where horizontal pleiotropy does not exist, all variants are expected to be close to the regression line (i.e. to have small residuals in the regression). However, when a variant is subject to horizontal pleiotropy, the effect size on the outcome can be larger or smaller than the effect size mediated by the exposure in question and therefore the variant can deviate from the true slope of the regression line. MR-PRESSO is designed to detect whether a subset of variants is significantly deviating from the regression line. The MR-PRESSO outlier test requires that at least 50% of the variants are valid instruments and relies on the InSIDE (Instrument Strength Independent of Direct Effect) condition[20], which states that the effect sizes of the variants on the exposure should not depend on the horizontal pleiotropic effects on the outcome.

MR-PRESSO global test

The MR-PRESSO global test evaluates for the presence of horizontal pleiotropy and is comprised of four steps (Figure 1a):

1) for each variant *j*, we removed the variant in question and refit a IVW regression. This allows us to calculate the slope of the regression line on the remaining variants, denoted $\hat{\beta}_{-j}$ which represents the causal estimate without variant *j*;

2) the estimated causal effect (slope) without variant *j* is used to predict the expected effect size on the outcome as the product of$\hat{\beta}_{-j}$ and the effect size of the same variant on the exposure $\hat{\gamma}_j$. Then, we calculated the observed residual sum of squares (RSS) as the difference between the observed effect size of the variant on the outcome ($\hat{\Gamma}_j$) and the predicted effect size of the same variant on the outcome $RSS_{obs}(j) = \left(\hat{\Gamma}_j - \hat{\beta}_{-j}\hat{\gamma}_j\right)^2$. The global observed RSS is then obtained by summing over the $J\,RSS_{obs}(j)$:

$$RSS_{obs} = \sum_j RSS_{obs}(j) = \sum_j \left(\hat{\Gamma}_j - \hat{\beta}_{-j}\hat{\gamma}_j\right)^2;$$

3) the observed RSS is compared to a simulated expected distribution of RSSs. The expected distribution is simulated under the null hypothesis (0% of variants are outliers). First, we simulated a distribution of effect sizes on the exposure $\hat{\gamma}_j^{random}$ from a Gaussian distribution $\mathcal{N}\left(\hat{\gamma}_j, \mathbb{V}\left(\hat{\gamma}_j\right)\right)$. Second, we simulated a distribution of effect sizes on the outcome $\hat{\Gamma}_{j\mathrm{random}}$ around the predicted effect size on the outcome ($\hat{\beta}_{-j}\hat{\gamma}_j$) by drawing in a Gaussian distribution $\mathcal{N}\left(\hat{\beta}_{-j}\hat{\gamma}_j, \mathbb{V}\left(\hat{\Gamma}_j\right)\right)$. The expected RSS is then obtained as $\sum_j RSS_{exp}(j) = \left(\hat{\Gamma}_j^{random} - \hat{\beta}_{-j}\hat{\gamma}_j^{random}\right)^2$. The procedure was repeated multiple times (K) to obtain a null distribution of the K expected RSSs, $RSS^k_{exp} = \sum_j RSS^k_{exp}(j) = \sum_j \left(\hat{\Gamma}^{random}_{jk} - \hat{\beta}_{-j}\hat{\gamma}^{random}_{jk}\right)^2$;

4) an empirical *P-value* is computed as the number of expected RSSs greater than the observed RSS divided by the total number of times the procedure is repeated:$P = \frac{\sum_k 1_{> RSS_{obs}}\left(RSS^k_{exp}\right)}{K}$.

K corresponds to the number of random drawings used to generate the expected distribution. The magnitude of K depends on the desired precision of the *P-value* which is $1/K$. We recommend to use at least K = 1,000, which will produce a precision of at least 10^{-3}.

MR-PRESSO outlier test

The MR-PRESSO outlier test allows for the detection of specific horizontal pleiotropic outlier variants. For a given variant *j*, we compared the observed j^{th} RSS $RSS_{obs}(j)$ (obtained in step 2 of the global test) with the distribution of K expected j^{th} RSSs $RSS^k_{exp}(j)$ (obtained in step 3 of the global test). Finally, an empirical *P-value* is computed as $P_j = \frac{\sum_k 1_{> RSS_{obs}(j)}\left(RSS^k_{exp}(j)\right)}{K}$ (Figure 1b) which is then multiplied by the number of variants J to account for multiple testing using the Bonferroni correction.

Given that J outlier tests are performed, we recommend to use a Bonferroni correction of the J *P-values*.

MR-PRESSO distortion test

The MR-PRESSO distortion test quantifies the distortion in the causal estimate due to significant horizontal pleiotropic outlier variants (Figure 1c). Distortion (D) is defined as the percentage of the causal estimate that is due to significant horizontal pleiotropic outlier variants. It is calculated as$D = 100 \times \frac{\hat{\beta}_{causal,o} - \hat{\beta}_{causal}}{\left|\hat{\beta}_{causal}\right|}$, with $\hat{\beta}_{causal,o}$the original causal effect estimated using all variants and $\hat{\beta}_{\text{causal}}$ the corrected causal estimate obtained after removing outliers identified by MR-PRESSO. Normalizing by the absolute value of the corrected causal estimate provides a direction for the magnitude of D. To test for statistical significance of D, we calculated an empirical *P-value* by generating a null distribution (the null hypothesis corresponds to the expected distortion due to a random set of variants). We defined n_O as the number of variants detected as outliers by the MR-PRESSO outlier test and n_E as the total number of variants robustly associated with the exposure. The null distribution is generated by substituting n_O variants detected as outliers by the MR-PRESSO outlier test with $n_E - 2n_O$ non-outliers, which are drawn with replacement from the entire set of non-outlier variants. This results in the total number of variants being fixed at $n_E - n_O$. We repeated this procedure K times to generate the null distribution. An empirical *P-value* is then calculated as the number of times that the observed distortion is greater than the expected distortion under the null hypothesis divided by *K*.

Simulation framework

We performed simulations to evaluate the statistical properties (false positive rate and power) to detect and correct for horizontal pleiotropy. We simulated the standard MR framework shown in Supplementary Figure 1b with an outcome as well as two exposures E_1 and E_2. A total of 50 variants was simulated per case. Horizontal pleiotropy was induced by a certain percentage of the 50 variants G_j having a significant effect on both E_1 (through $\gamma_{1,j}$) and E_2 (through $\gamma_{2,j}$). We varied the following parameters:

- the main causal effect (of exposure 1 on the outcome), $\beta_{causal>} = 0$ (no main causal effect) or $\beta_{causal} = 0.1, 0.2, 0.5$ whereas the causal effect of exposure 2 was always set to $\beta_{pleiotropic}= 0.1$;
- percentage of horizontal pleiotropy (0, 2, 4, 10, 50, 90);
- type of horizontal pleiotropy: positive (all $\gamma_{2,j}$ positive) or balanced horizontal pleiotropy (approximately half of the $\gamma_{2,j}$ positive and half negative);
- verification of the InSIDE[20] (Instrument Strength Independent of Direct Effect) assumption (**Supplementary Note**).

For each scenario, 10,000 simulations were performed.

Collection of genome-wide association (GWA) summary statistics

We retrieved publicly available genome-wide association (GWA) summary statistics data for 82 complex traits and diseases (**Supplementary Note**). We performed the following steps to ensure that all datasets were uniform and standardized. For each, we retrieved the appropriate variant annotation (build, rsid, chromosome, position, reference and alternate alleles) and summary statistics (effect size, standard errors, *P-values* and sample size of the study). All variant coordinates (chr, pos) were lifted over to hg19 using the UCSC Genome Browser LiftOver Tool. We imputed Z-scores of variants using ImpG[39] using 1000 Genomes Phase 3 European panel[40] (n = 503) as a reference panel. Effect sizes, standard errors and *P-values* were then calculated using the variance of the trait estimated from genotyped variants and allele frequencies calculated on the same subset of individuals from the 1000 Genomes reference panel. Sets of GWA-significant variants were manually retrieved from the corresponding GWA manuscripts. In total, we retrieved GWA summary statistics for 82 traits and diseases (Supplementary Table 14).

Detection and correction of horizontal pleiotropy using MR-PRESSO or covariate adjustment in MMR

We applied MR-PRESSO to all possible exposure-outcome pairs of 82 traits and diseases, and then compared the results of this test to those obtained from other MR methods. In total, we performed 4,250 MR tests. Only 53 distinct traits had a sufficient set of genome-wide significant variants that could be used as instrumental variables. This led to 53 × (82–1) =

4,293 possible exposure-outcome pairs. The remaining pairs were removed because of missing values in the summary data, which led to a total of 4,250 MR tests. We compared these results to other approaches including the Q (modified) test[19] and Q' (modified) test[19] as well as MR-Egger regression[20,21]. Next, we evaluated five strategies to correct for significant horizontal pleiotropy detected from our MR-PRESSO global test. The first approach included covariates in our MMR model, either one by one or all at the same time. We considered only covariates with a statistically significant causal effect (causal estimate of the IVW meta-analysis using a Bonferroni-corrected cut-off). Furthermore, to account for co-linearity, we also included only covariates with a correlation coefficient < 0.3. We note that not all pairs were eligible for MMR analysis due to either a lack of relevant covariates to adjust on in the one by one model or too many covariates to adjust on in the full model. The second approach corrected for horizontal pleiotropy by removing offending variants that are statistically significant outliers according to the MR-PRESSO outlier test. The MR-PRESSO global test was performed on the adjusted MR models to determine if there was any remaining horizontal pleiotropy. Finally, in a similar fashion, the Q (modified) outlier test and Q' (modified) outlier test were used to correct for horizontal pleiotropy by removing significant outliers and the Q (modified) test and Q' (modified) test were respectively applied to test for any remaining horizontal pleiotropy. 1,000,000 simulations to calculate the empirical *P-values* were performed for the MR-PRESSO global and outlier tests to obtain a precision of $P = 10^{-6}$ which allows for a Bonferroni-corrected cut-off of $P = 0.05/4{,}250 = 1.17\times10^{-5}$. 10,000 simulations were computed to calculate the empirical *P-values* for the MR-PRESSO distortion test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the various genome-wide association consortia for generously sharing the genome-wide association summary statistics. We thank Amanda Dobbyn and Daniel Jordan for critical comments on the manuscript. R.D is supported by R35GM124836 from the National Institute Of General Medical Sciences of the National Institutes of Health, R01HL139865 from the National Heart, Lung, Blood Institute of the National Institutes of Health, research grants from AstraZeneca and Goldfinch Bio and previously an American Heart Association Cardiovascular Genome-Phenome Discovery grant (15CVGPSD27130014). B.N. is supported by 1R01MH094469, 1R01MH107649–01 and 5U01HG009088–02 from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing Financial Interests Statement

B.N. is a member of Deep Genomics Scientific Advisory Board, has received travel expenses from Illumina, and also serves as a consultant for Avanir and Trigeminal solutions. R.D. has received research support from AstraZeneca and Goldfinch Bio.

References

1. Vasan RS Biomarkers of cardiovascular disease: molecular basis and practical considerations. Circulation 113, 2335–2362 (2006).16702488
2. Ebrahim S & Davey Smith G Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? Hum. Genet. 123, 15–33 (2008).18038153

3. Smith GD & Ebrahim S 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 32, 1–22 (2003).12689998

4. Evans DM & Smith GD Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality. Annual Review of Genomics and Human Genetics 16, 327–350 (2015).

5. Burgess S , Bowden J , Fall T , Ingelsson E & Thompson SG Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants: Epidemiology 28, 30–42 (2017).27749700

6. Burgess S et al. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol 30, 543–552 (2015).25773750

7. Solovieff N , Cotsapas C , Lee PH , Purcell SM & Smoller JW Pleiotropy in complex traits: challenges and strategies. Nat. Rev. Genet. 14, 483–495 (2013).23752797

8. Bulik-Sullivan B et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 47, 1236–1241 (2015).26414676

9. Sivakumaran S et al. Abundant pleiotropy in human complex diseases and traits. Am. J. Hum. Genet. 89, 607–618 (2011).22077970

10. Gratten J & Visscher PM Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. Genome Med 8, 78 (2016).27435222

11. Parkes M , Cortes A , van Heel DA & Brown MA Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat. Rev. Genet. 14, 661–673 (2013). 23917628

12. Pickrell JK et al. Detection and interpretation of shared genetic influences on 42 human traits. Nat. Genet. 48, 709–717 (2016).27182965

13. Grassmann F et al. Genetic pleiotropy between age-related macular degeneration and 16 complex diseases and traits. Genome Med 9, 29 (2017).28347358

14. Webb TR et al. Systematic Evaluation of Pleiotropy Identifies 6 Further Loci Associated With Coronary Artery Disease. J. Am. Coll. Cardiol. 69, 823–836 (2017).28209224

15. Pickrell J Fulfilling the promise of Mendelian randomization. bioRxiv 018150 (2015). doi: 10.1101/018150

16. Smith GD Mendelian randomization: a premature burial? bioRxiv 021386 (2015). doi: 10.1101/021386

17. Greco M FD , Minelli C , Sheehan NA & Thompson JR Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Statist. Med. 34, 2926–2940 (2015).

18. Bowden J et al. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Statist. Med. n/a-n/a (2017). doi:10.1002/sim.7221

19. Bowden J et al. Improving the accuracy of two-sample summary data Mendelian randomization: moving beyond the NOME assumption. bioRxiv 159442 (2017). doi:10.1101/159442

20. Bowden J , Smith GD & Burgess S Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int. J. Epidemiol. 44, 512–525 (2015). 26050253

21. Bowden J et al. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. Int J Epidemiol 45, 1961–1974 (2016).27616674

22. Do R et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. Nat Genet 45, 1345–1352 (2013).24097064

23. Burgess S & Thompson SG Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects. Am J Epidemiol 181, 251–260 (2015).25632051

24. Corbin LJ et al. Body mass index as a modifiable risk factor for type 2 diabetes: Refining and understanding causal estimates using Mendelian randomisation. Diabetes 65, 3002–3007 (2016). 27402723

25. Fox J & Long JS Modern methods of data analysis. (Sage Publications, 1990).

26. Bowden J , Davey Smith G , Haycock PC & Burgess S Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet. Epidemiol. 40, 304–314 (2016).27061298

27. Hartwig FP , Davey Smith G & Bowden J Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol doi:10.1093/ije/dyx102

28. MacArthur J et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 45, D896–D901 (2017).27899670

29. Cohen JC , Boerwinkle E , Mosley TH & Hobbs HH Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N. Engl. J. Med. 354, 1264–1272 (2006).16554528

30. Warren HR et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat. Genet. 49, 403–415 (2017).28135244

31. Ehret GB et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. Nat. Genet. 48, 1171–1184 (2016).27618452

32. Liu C et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. Nat. Genet. 48, 1162–1170 (2016).27618448

33. Holmes MV et al. Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis. Am. J. Hum. Genet. 94, 198–208 (2014).24462370

34. Lyngdoh T et al. Serum uric acid and adiposity: deciphering causality using a bidirectional Mendelian randomization approach. PLoS ONE 7, e39321 (2012).

35. Hemani G et al. Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. bioRxiv 173682 (2017). doi:10.1101/173682

36. Plenge RM , Scolnick EM & Altshuler D Validating therapeutic targets through human genetics. Nat Rev Drug Discov 12, 581–594 (2013).23868113

37. Mokry LE , Ahmad O , Forgetta V , Thanassoulis G & Richards JB Mendelian randomisation applied to drug development in cardiovascular disease: a review. J. Med. Genet. 52, 71–79 (2015). 25515070

Methods-only References

38. Burgess S , Butterworth A & Thompson SG Mendelian randomization analysis with multiple genetic variants using summarized data. Genet. Epidemiol. 37, 658–665 (2013).24114802

39. Pasaniuc B et al. Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. Bioinformatics 30, 2906–2914 (2014).24990607

40. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. Nature 526, 68–74 (2015).26432245

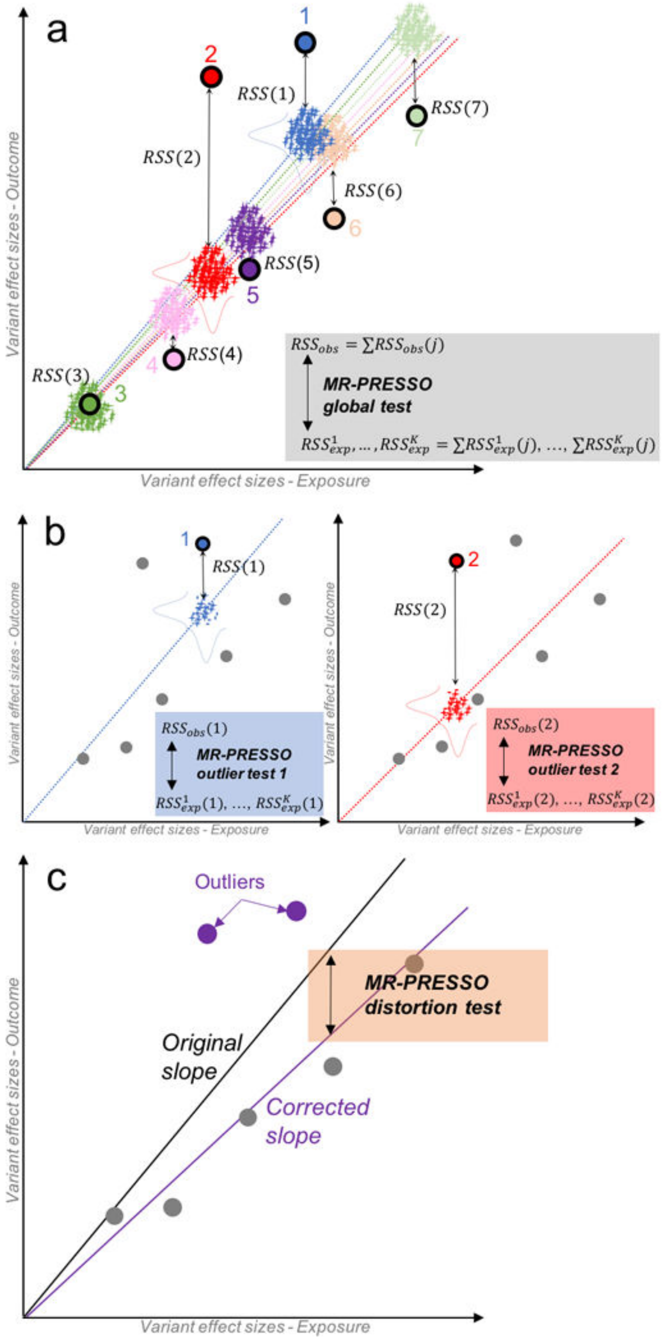
a
1
2
RSS(1)
RSS(7)
RSS(2)
7
RSS(6)
6
RSS(5)
5
RSS(4)
4
RSS(3)
3
$RSS_{obs} = \sum RSS_{obs}(j)$
MR-PRESSO global test
$RSS_{exp}^{1}, \ldots, RSS_{exp}^{K} = \sum RSS_{exp}^{1}(j), \ldots, \sum RSS_{exp}^{K}(j)$
Variant effect sizes - Outcome
Variant effect sizes - Exposure
b
1
RSS(1)
$RSS_{obs}(1)$
MR-PRESSO outlier test 1
$RSS_{exp}^{1}(1), \ldots, RSS_{exp}^{K}(1)$
Variant effect sizes - Outcome
Variant effect sizes - Exposure
2
RSS(2)
$RSS_{obs}(2)$
MR-PRESSO outlier test 2
$RSS_{exp}^{1}(2), \ldots, RSS_{exp}^{K}(2)$
Variant effect sizes - Outcome
Variant effect sizes - Exposure
c
Outliers
MR-PRESSO distortion test
Original slope
Corrected slope
Variant effect sizes - Outcome
Variant effect sizes - Exposure

Figure 1: Description of the Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-PRESSO) method.

MR-PRESSO is comprised of three components. Panel a represents the global test. For each variant *j*, a slope, representing the causal estimate, is computed without the variant using standard inverse variance weighted meta-analysis (seven colored dotted lines; each color of the regression line corresponds to the line obtained by excluding the variant of the same color). The observed residual sum of squares RSS_{obs}(j) is computed as the squared difference between the observed effect size of variant *j* on the outcome and the effect size predicted using the slope computed without *j*. In addition, *K* pairs of random effect sizes for the exposure (x-axis) and the outcome (y-axis) (represented as crosses) are drawn from two Gaussian distributions (horizontal and vertical bell curves respectively for the exposure and outcome) from the predicted effect sizes and standard errors using the slope computed without *j*. A distribution of *K* expected $\mathrm{RSS}_{\mathrm{exp}}^{\mathrm{k}}(\mathrm{j})$ is then calculated. By summing up the *J* $\mathrm{RSS}_{\mathrm{obs}}(\mathrm{j})$, we calculate a global statistic that is compared to the *K* expected sum of $\mathrm{RSS}_{\mathrm{exp}}^{\mathrm{k}}(\mathrm{j})$. Panel b represents the outlier test. A test is performed for each variant *j* by comparing the observed $\mathrm{RSS}_{\mathrm{obs}}(\mathrm{j})$ to the *K* expected $\mathrm{RSS}_{\mathrm{exp}}^{\mathrm{k}}(\mathrm{j})$. Here, only variants (1 and 2) are shown for simplicity. Panel c represents the distortion test. The panel shows how removing significant outliers detected by the MR-PRESSO outlier test (variant 1 and 2) leads to an unbiased slope (causal estimate).

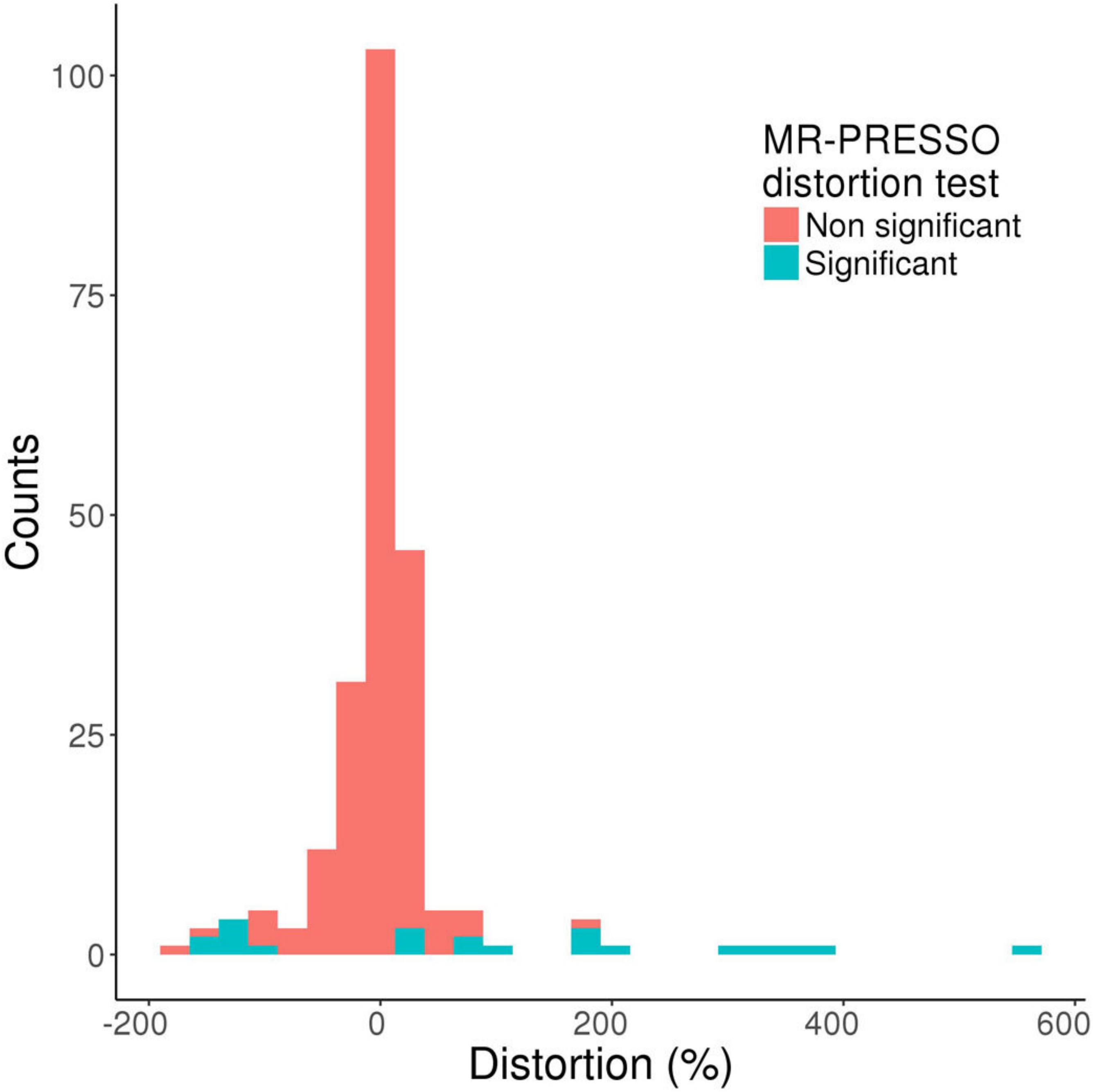

Figure 2: Distribution of the distortion of causal estimates before and after correction for horizontal pleiotropy using the MR-PRESSO distortion test.

The distortion coefficient was calculated for all exposure-outcome pairs with a significant causal estimate (P < 0.05) using an inverse variance weighted meta-analysis (n = 229). The distortion coefficients were then tested using MR-PRESSO (Mendelian Randomization Pleiotropy RESidual Sum and Outlier) distortion test which provides an empirical P-value. The distortion coefficients are colored according to whether the distortion is statistically significant (blue) or not (red) at a Bonferroni-corrected threshold of P < 0.05 / 229 in the MR-PRESSO distortion test. The distortion estimate represents the change in the causal estimate as a result of horizontal pleiotropic outlier variants (Online Methods). A positive distortion represents a decrease in the outlier-corrected causal estimate.

Table 1:

Power to detect horizontal pleiotropy in Mendelian randomization for different methods.

Causal effect	Percent	Pleiotropy	MR-PRESSO global test	Q (modified) test	Q’ (modified) test
0	2	balanced	25.34	25.40	22.04
0	2	positive	25.01	25.01	22.00
0	4	balanced	51.79	51.96	47.80
0	4	positive	50.88	51.32	47.34
0	10	balanced	95.53	95.58	94.47
0	10	positive	94.27	94.26	92.85
0.1	2	balanced	24.10	24.41	21.58
0.1	2	positive	23.60	23.89	20.96
0.1	4	balanced	51.17	51.49	47.51
0.1	4	positive	50.67	50.59	46.69
0.1	10	balanced	95.56	95.56	94.31
0.1	10	positive	93.73	93.77	92.08
0.2	2	balanced	22.42	22.70	19.74
0.2	2	positive	22.95	22.92	19.97
0.2	4	balanced	48.37	48.29	44.45
0.2	4	positive	46.89	46.82	43.18
0.2	10	balanced	94.08	94.11	92.78
0.2	10	positive	91.85	91.91	90.24
0.5	2	balanced	16.72	16.70	14.90
0.5	2	positive	16.76	16.55	15.15
0.5	4	balanced	33.71	33.79	31.00
0.5	4	positive	32.52	32.66	29.93
0.5	10	balanced	81.15	81.22	79.07
0.5	10	positive	76.99	77.09	74.75

A total of 50 variants was simulated in each case. The InSIDE (Instrument Strength Independent of Direct Effect) condition is satisfied in all reported scenarios. MR-PRESSO: Mendelian Randomization Pleiotropy RESidual Sum and Outlier.

Table 2:

Application of methods to detect horizontal pleiotropy in Mendelian randomization analysis from 82 summary-level genome-wide association traits and diseases.

	4,250 pairs		**191 ‘putatively causal’ pairs**	
	Percentage	**Counts**	**Percentage**	**Counts**
MR-PRESSO global test	21.69%	922	48.69%	93
Q (modified) test	20.59%	875	45.03%	86
Q’ (modified) test	18.85%	801	42.93%	82

MR-PRESSO: Mendelian Randomization Pleiotropy RESidual Sum and Outlier. ‘Putatively causal’ pairs are exposure-outcome pairs with a statistically significant causal estimate in the inverse variance weighted meta-analysis (at the Bonferroni-corrected cut-off).

Table 3:

Correction for horizontal pleiotropy in Mendelian randomization using two different approaches: MR-PRESSO outlier test and covariate adjustment for 922 exposure-outcome pairs with significant horizontal pleiotropy.

	MR-PRESSO outlier test	Q (modified) outlier test	Q’ (modified) outlier test	Single-covariate adjustment (MMR)	All-covariate adjustment (MMR)
Total corrected	422	354	356	20	22
Total remaining horizontal pleiotropy	500	511	445	73	42
Fraction corrected horizontal pleiotropy	46%	41%	44%	22%	34%

MR-PRESSO: Mendelian Randomization Pleiotropy RESidual Sum and Outlier; MMR: Multi-variable Mendelian Randomization.