medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Exploiting collider bias to apply two-sample summary data Mendelian randomization methods to one-sample individual level data

Ciarrah Barry[1,2], Junxi Liu[1,2], Rebecca Richmond[1,2]
Martin K Rutter[3,4], Deborah A Lawlor[1,2], Frank Dudbridge[5]
& Jack Bowden[6,1,*]

[1] *MRC Integrative Epidemiology Unit at the University of Bristol, U.K.*
[2] *Population Health Sciences, University of Bristol, U.K.*
[3] *Division of Diabetes, Endocrinology and Gastroenterology, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, UK*
[4] *Diabetes, Endocrinology and Metabolism Centre, Manchester University NHS Foundation Trust, Manchester*
[5] *Department of Health Sciences, University of Leicester, Leicester, U.K.*
[6] *Exeter Diabetes Group (ExCEED), College of Medicine and Health, University of Exeter, Exeter, U.K.*

*Address for correspondence:
Jack Bowden, University of Exeter College of Medicine and Health
RILD Building, RD&E Hospital,
Barrack Road, Exeter, Devon, EX2 5DW
j.bowden2@exeter.ac.uk

Abstract

Over the last decade the availability of SNP-trait associations from genome-wide association studies data has led to an array of methods for performing Mendelian randomization studies using only summary statistics. A common feature of these methods, besides their intuitive simplicity, is the ability to combine data from several sources, incorporate multiple variants and account for biases due to weak instruments and pleiotropy. With the advent of large and accessible fully-genotyped cohorts such as UK Biobank, there is now increasing interest in understanding how best to apply these well developed summary data methods to individual level data, and to explore the use of more sophisticated causal methods allowing for non-linearity and effect modification.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

In this paper we describe a general procedure for optimally applying any two sample summary data method using one sample data. Our procedure first performs a meta-analysis of summary data estimates that are intentionally contaminated by collider bias between the genetic instruments and unmeasured confounders, due to conditioning on the observed exposure. A weighted sum of these estimates is then used to correct the standard observational association between an exposure and outcome. Simulations are conducted to demonstrate the method's performance against naive applications of two sample summary data MR. We apply the approach to the UK Biobank cohort to investigate the causal role of sleep disturbance on HbA1c levels, an important determinant of diabetes.

Our approach is closely related to the work of Dudbridge et al. (*Nat. Comm.* **10**: 1561), who developed a technique to adjust for index event bias when uncovering genetic predictors of disease progression based on case-only data. Our paper serves to clarify that in any one sample MR analysis, it can be advantageous to estimate causal relationships by artificially inducing and then correcting for collider bias.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Background

Mendelian randomisation (MR) is a technique used to test for, and quantify, the causal relationship between a modifiable exposure and health outcome with observational data, by using genetic variants as instrumental variables [1, 2]. MR circumvents the need to measure and adjust for all variables which confound the exposure-outcome association, and is therefore seen as an attractive additional analysis to perform alongside more traditional epidemiological methods [3]. The following Instrumental Variable assumptions are usually invoked in order justify testing for a causal effect of an exposure X on a health outcome Y using a set of genes, G:

- IV1: G must be associated with X;
- IV2: G must be independent of unmeasured confounding between X and Y;
- IV3: G must be independent of Y conditional on X and all confounders of the X-Y relationship.

These assumptions are encoded in the causal diagram in Figure 1. Further linearity and homogeneity assumptions are needed in order to consistently estimate the magnitude of the causal effect. When performing an MR-analysis it is best practice to select SNPs for use as instruments using external data, in order to avoid bias due to the winner's curse [4]. Subsequently, if an individual genetic variant is no longer strongly associated with an exposure in the MR data, assumption IV1 will only be weakly satisfied, which leads to so-called weak instrument bias [5, 6]. This issue is mitigated as the sample size increases as long as the true association is still non-zero. When a genetic variant is in fact associated with the outcome through pathways other than the exposure, a phenomenon known as horizontal pleiotropy [7], this is a violation of assumptions IV2 and/or IV3. Horizontal pleiotropy is not necessarily mitigated by an increasing sample size and is also harder to detect. Its presence can therefore render very precise MR estimates hopelessly biased. Pleiotropy-robust MR methods have been a major focus of research in recent years for this reason [8, 9, 10, 11].

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

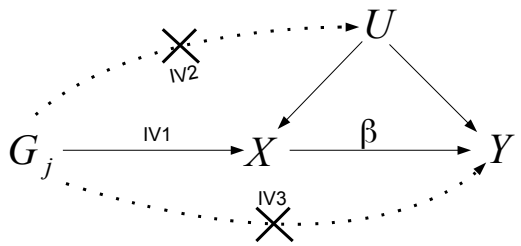

Figure 1: The IV assumptions for a genetic variant G are represented by solid lines in the directed acyclic graph (DAG). Dotted lines represent violations of IV assumptions as described in IV2 and IV3. The causal effect of a unit increase of the exposure,X, on the outcome,Y, is denoted by β. U represents unobserved confounders of X and Y

One-sample versus Two-sample MR: pros and cons

Obtaining access to a single cohort with measured genotype, exposure and outcome data that is large enough to furnish an MR analysis has been difficult, historically. It has instead been far easier to obtain summary data estimates of gene-exposure and gene-outcome associations from two independent studies, and to perform an analysis within the 'two-sample summary data MR' framework (see Figure 2). [12, 13]. This has made it an attractive option for the large scale pursuit of MR, through software platforms such as MR-Base [14]. The relative simplicity of these methods (which resemble a standard meta-analysis of study results) and their ability to furnish graphical summaries for the detection and adjustment of pleiotropy [15] has also acted to increase their popularity. Indeed, the array of pleiotropy robust two sample summary data methods far outstrips those available for one sample individual level data MR analysis [16].

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

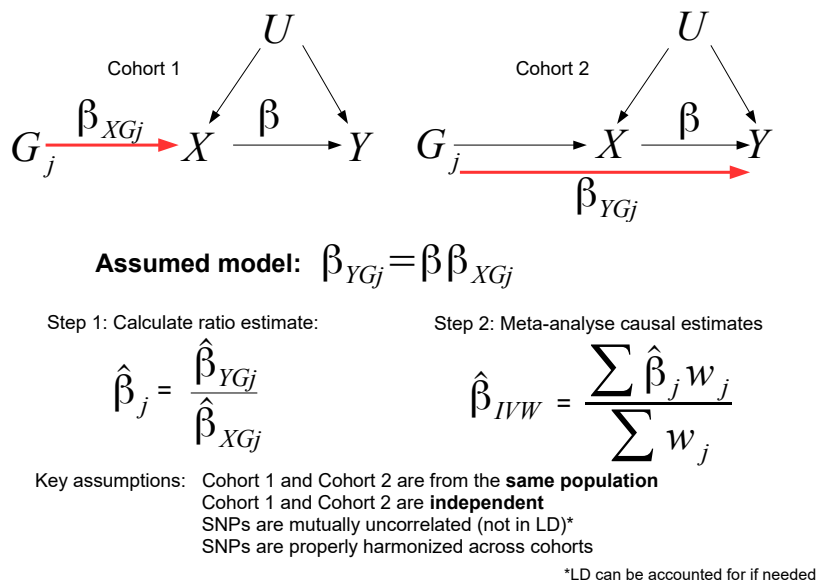

Figure 2: In two sample summary data MR, $(G - X)$ association estimates, $\hat{\beta}_{XGj}$, from one cohort are combined with $(G - Y)$ association estimates, $\hat{\beta}_{YGj}$ from a separate, non-overlapping cohort, to produce a set of SNP-specific causal estimates, $\hat{\beta}_j$. These are combined using inverse variance weighted meta-analysis (w_j being the weight) to obtain an overall estimate $\hat{\beta}_{IVW}$ for the true causal effect β.

A further advantage of two-sample over one-sample MR is that weak instruments bias causal estimates towards the null, which is conservative [17]. Furthermore, this bias can easily be quantified and corrected for, because uncertainty in the SNP-exposure association estimates obtained from one cohort is independent of the uncertainty in SNP-outcome association estimates from a non-overlapping cohort (Figure 2). This makes the bias problem akin to 'classical' measurement error [18] and enables standard approaches such as Simulation Extrapolation [5, 19] or modified weighting [6] to be used to adjust for its presence. In contrast, weak instruments bias MR estimates obtained from a one sample analysis towards the observational association because uncertainty in the SNP-exposure and SNP-outcome association estimates are correlated. This bias is harder to correct for and is potentially anti-conservative.

There are, however, many disadvantages of using two sample summary data compared to individual level data on from a single sample MR. The the two-sample approach assumes the two cohorts are perfectly homogeneous [13]. If, for example, the distribution of confounders is different between the samples, this can result in severe bias [20]. Alternatively, it may be that the independence assumption is violated due to an unknown number of shared subjects across the

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

two studies [21], which cannot be easily removed [22]. Even when the homogeneity assumption is satisfied, two sample methods can give misleading results if the two sets of associations are not properly harmonized [23].

It is often the case that summary data from a GWAS has been adjusted for factors that might bias MR results, and the unadjusted data are not available [24]. Alternatively, it may not be possible to source summary data on the exact population needed for a particular analysis, for example on either or men or women only when looking at sex-specific outcomes) [25]. Finally, a richer array of analyses are possible with individual level data. For example, the estimation of non-linear causal effects across the full range of the exposure and the exploration of effect modification via covariates.

It is of course possible to naively apply summary data MR methods to the one-sample context, estimating both the gene-exposure and gene-outcome associations in the same sample [7], an analysis made increasingly easy by the advent of large open-access cohort studies such as the UK Biobank [26]. This has the advantage of requiring no homogeneity assumptions but the disadvantage of potentially anti-conservative weak instrument bias due to correlated error. A preliminary investigation has found that this naive approach is particularly bad for pleiotropy robust approaches such as MR-Egger regression [27, 28]. So far, there is no consensus on how best to implement summary data approaches in the one sample setting.

In this paper we propose a general method that can reliably apply two-sample summary data MR methods to one-sample data, whilst maintaining the simplicity and appeal of the two-sample approach. Our method builds on the work of Dudbridge et. al. [29], who proposed a method to correct for 'index event' (or collider) bias in genetic studies of disease progression, when all subjects included in the analysis have been diagnosed with the disease. In this setting, the analysis is open to contamination from collider bias. Our work serves to clarify that the procedure can be extended to any MR analysis by artificially inducing and then correcting for collider bias. By construction it ensures that there is no correlation between sampling errors in the $G - X$ and $G - Y$ associations, thus recreating the conditions of the two sample MR. This allows two sample methods to be used in a one sample design. It therefore provides an attractive and efficient vehicle for applying two sample approaches to one sample data beyond the standard two-stage least squares procedure, thereby benefiting from plethora of weak instrument and pleiotropy robust approaches available.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Methods

To motivate ideas, we assume the following individual level data model for the exposure X and continuous outcome Y for subject i:

$$X_i|G_i, U_i = \sum_{j=1}^{k} \beta_{XG_j} G_{ij} + \beta_{UX} U_i + \varepsilon_{X_i} \quad (1)$$

$$Y_i|X_i, G_i, U_i = \beta X_i + \sum_{j=1}^{k} \alpha_j G_{ij} + \beta_{UY} U_i + \varepsilon_{Y_i} \quad (2)$$

$$= \sum_{j=1}^{k} (\alpha_j + \beta\beta_{XGj}) G_{ij} + (\beta\beta_{UX} + \beta_{UY}) U_i + \beta\varepsilon_{X_i} + \varepsilon_{Y_i}$$

$$= \sum_{j=1}^{k} \beta_{YGj} G_{ij} + \varepsilon^*_{Y_i} \quad (3)$$

Here, $G_i = (G_{i1}, \ldots, G_{ik})'$ represents a set of k variants that predict X_i, β represents the true causal effect of the exposure on the outcome we wish to estimate, and U represents unmeasured confounding predicting both X and Y. The variables $\varepsilon_{X_i}, \varepsilon_{Y_i}$ represent independent residual error terms. Since the unmeasured confounder U is common to both X and Y, the total residual errors around $X|G$, $Y|X, G$ and $Y|G$ in equations (1) - (3) are correlated.

The standard approach to estimating β with individual level data is Two Stage Least Squares (TSLS). This assumes that all instruments are valid (not pleiotropic), so that $\alpha_j = 0$ for all j. TSLS firstly regresses the exposure on all k genotypes under model (1) to derive an estimate for subject i's genetically predicted exposure: $\hat{X}_i = \sum_{j=1}^{k} \hat{\beta}_{XG_j} G_{ij}$. The outcome Y is then regressed on $\hat{X}_i$ assuming model (2) and its regression coefficient is taken as the causal estimate $\hat{\beta}$. When the set of k SNPs which predict X are mutually independent (i.e. not in linkage disequilibrium), the TSLS estimate is asymptotically equivalent to the IVW estimate [30] obtained by:

- Calculating the causal estimate $\hat{\beta}_j$ by dividing the SNP-outcome association $\hat{\beta}_{YGj}$ obtained from model (3) by the SNP-exposure estimate $\hat{\beta}_{XGj}$ from model (1) for each SNP and;
- Performing an inverse variance weighted meta-analysis of the k individual causal estimates, $\hat{\beta}_1, \ldots, \hat{\beta}_k$.

We will refer to this as the 'standard' IVW approach. It is commonly used in two sample summary data MR out of necessity, but is not typically used in the one sample setting because its finite sample performance is not currently well understood [27].

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Inducing collider bias into SNP-outcome associations

Consider a regression of the outcome Y on G and X together (but not U):

$$Y_i|X_i, G_i = \beta^* X_i + \sum_{j=1}^{k} \alpha_j^* G_{ij} + \epsilon_i^{'}, \tag{4}$$

yielding estimated coefficients $\hat{\beta}^*$ and $\hat{\alpha}_1^*, \ldots, \hat{\alpha}_k^*$. Since X represents an inverted fork in Figure (1), conditioning on it breaks the d-separation between G and U [31] thus inducing collider bias in the estimated G_j-Y association (or pleiotropic effect through a spurious association with U) so that $\hat{\alpha}_j^*$ is not a consistent estimate for α_j. Likewise, $\hat{\beta}^*$ is not a consistent estimate for β. It instead reflects the causal effect, plus a contribution via the back-door path from X to Y via U. Such 'collider-biased' analyses are usually avoided for this reason [32]. However, it is in a special sense advantageous to fit model (4) because it can be shown that, under model's (1) and (2), α_j^*, α_j, β^* and β are linked by the following linear relation:

$$\alpha_j^* = \alpha_j + (\beta - \beta^*)\beta_{XG_j}, \tag{5}$$

(see *Supplementary information* for a detailed derivation). This suggests the following algorithm for estimating the causal effect:

1. Regress Y on X and G to obtain the collider-biased parameter estimates $\hat{\beta}^*$ and $\hat{\alpha}_1^*, \ldots, \hat{\alpha}_k^*$.

2. Regress X on G to obtain estimates $\hat{\beta}_{XG1}, \ldots, \hat{\beta}_{XGk}$.

3. Fit the linear model:

$$\hat{\alpha}_j^* = \alpha_0 + (\beta - \beta^*)\hat{\beta}_{XG_j} + \epsilon_j \tag{6}$$

   under a user-specified loss function in order to obtain an estimate for the collider-correction term $(\beta - \beta^*)$.

4. Adjust the observational estimate to obtain an estimate for the causal effect β via:

$$\hat{\beta} = \hat{\beta}^* + \widehat{\beta - \beta^*} \tag{7}$$

The above procedure, which we call 'collider-Correction' is a simple modification of the Dudbridge approach [29]. In step 1 we simultaneously adjust for all genetic instruments. In step 4 we instead focus on estimation of the causal parameter β rather than, as Dudbridge et al do, the pleiotropic effects. The major advantage of enacting this algorithm is that, if the sample covariance $\widehat{Cov}(\alpha_j, \beta_{XG})$ is zero (the InSIDE assumption [9]), then $\widehat{Cov}(\alpha_j^*, \beta_{XGj})$ will also be zero. This means that weak instrument bias is akin to classical measurement error, which will induce a dilution in the estimated collider-correction slope $\widehat{\beta - \beta^*}$ from (6) towards zero. This dilution is easy to quantify and correct for

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

using standard methods.

The collider-correction algorithm is generalisable in theory to any MR analysis method, but we now describe several canonical implementations.

Implementing collider-Correction

IVW implementation

To calculate the IVW estimate using our approach we make the assumption that the mean pleiotropic effect is zero and the InSIDE assumption is satisfied [9]. This justifies setting the parameter α_0 to zero in equation (6) and estimating the slope $(\beta - \beta^*)$ using least squares. To quantify the impact of weak instrument bias in the estimated slope we calculate the mean F statistic:

$$\bar{F} = \sum_{j=1}^{k} \frac{\hat{\beta}^2_{XGj}}{\sigma^2_{XGj}} \tag{8}$$

where σ^2_{XGj} equals the squared standard error of $\hat{\beta}_{XGj}$. The expected dilution is then given by $(\bar{F} - 1)/\bar{F}$. Note that, whilst the collider-Correction slope is diluted towards zero in the presence of weak instrument bias, the causal estimate itself is still biased toward the observational association estimate $\hat{\beta}^*$, because the causal effect is the sum of $\hat{\beta}^*$ and $\widehat{\beta - \beta}^*$.

In order to formally test for the presence of pleiotropy (i.e. any α_j being non-zero) whilst accounting for weak instrument bias, we can calculate the analogue of the familiar Cochran's Q statistic as described by Bowden et al. [33]. Specifically, this equals:

$$Q(0) = \sum_{j=1}^{k} Q_j(0) = \sum_{j=1}^{k} w_j(0)(\{\widehat{\beta^* - \beta_j}\} - \{\widehat{\beta^* - \beta}\})^2 \tag{9}$$

where

$$\widehat{\beta^* - \beta_j} = \frac{\hat{\alpha}^*_j}{\hat{\beta}_{XGj}}, \quad w_j(0) = \frac{\hat{\beta}^2_{XGj}}{(\widehat{\beta^* - \beta})^2 \sigma^2_{Xj} + \mathbf{0} \times \sigma^2_{\alpha^*}}, \quad \sigma^2_\alpha = Var(\alpha_j), \tag{10}$$

and where the value of $\widehat{\beta^* - \beta}$ is chosen that minimises $Q(0)$. Under the null hypothesis that all α_j equal zero, $Q(0)$ is approximately χ^2_{k-1} distributed and each individual component, Q_j is approximately χ^2_1 distributed. If $Q(0)$ or any of its components are significantly large with respect to these null distributions then we can reject the null hypothesis of no pleiotropy. In order to estimate the collider-correction slope whilst accounting for weak instrument bias and pleiotropy we can find the value of $\widehat{\beta^* - \beta}$ that minimises the statistic $Q(1)$ using weights $w^*_j(1)$ which allow for a non-zero variance in the pleiotropic effect,

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

σ_α^2. This is a direct analogue of the procedure described in detail by Bowden et al. [6].

An alternative method for weak instrument bias adjustment that can be applied directly to the collider-corrected IVW estimate is Simulation Extrapolation (SIMEX) [34, 13]. Under SIMEX, a parametric bootstrap is used to generate SNP-exposure associations with an increasing amount of weak instrument bias, and hence increasing dilution in the estimate for $(\beta - \beta^*)$ from fitting model (6). A global model is then fitted to the entire set of simulated data in order to extrapolate back to the estimate for $(\beta - \beta^*)$ that would have been obtained if there were no uncertainty in the SNP-exposure associations. SIMEX is attractive because it can be applied to any regression model (and hence many MR methods), and reliable software is available in standard software packages, such as R and Stata.

MR-Egger implementation

In order to account for pleiotropy with a non-zero mean, we can instead allow the intercept α_0 to be freely estimated along with the slope in equation (6), again using least squares. This is equivalent to performing MR-Egger regression [9]. To assess the vulnerability of MR-Egger regression to weak instrument bias we use the I_{GX}^2 statistic [5]:

$$I_{GX}^2 = \frac{Q_{GX} - (k-1)}{Q_{GX}}, \text{where} \quad Q_{GX} = \sum_{j=1}^{k} \frac{(\hat{\beta}_{XGj} - \bar{\beta}_{XGj})^2}{\sigma_{XGj}^2} \tag{11}$$

The expected dilution in the collider correction due to weak instruments is equal to $(\beta - \beta^*)I_{GX}^2$. It can easily be adjusted for by applying SIMEX to the MR-Egger fit of equation (6).

A robust regression implementation

Both the IVW and MR-Egger approaches rely on the InSIDE assumption to consistently estimate the causal effect. This may be violated in practice, hence the rationale for the development of alternative, robust methods such as the Weighted Median [10]. In the two-sample summary data context it can consistently estimate the causal effect if the majority of the weight in the MR analysis stems from genetic variants that are not pleiotropic, even if InSIDE is violated. In order to calculate a causal estimate within our collider-Correction algorithm that is close in spirit to the Weighted Median, we propose to fit step 3 using Least-Absolute Deviation (LAD) regression instead of least squares, setting α_0 to zero. Currently, no summary statistic exists to quantify the weak instrument bias of such a procedure, but because the approach is implemented as a regression model it is straightforward to apply the SIMEX algorithm to it to guard against this.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Simulation studies

In order to confirm our theoretical results and assess the performance of the collider-correction algorithm, data sets of between 5000 and 50,000 individuals were generated under models (1) and (2). The genetic data comprised k=50 uncorrelated SNPs with an allele frequency set at 0.3 in Hardy-Weinburg equilibrium, which collectively explained approximately 1.5% of the variance in the exposure X. The parameter β, which reflects the population average causal effect of inducing a one-unit change in the exposure on the outcome, was set to 0.5. The shared confounder U was simulated from a zero-centred normal distribution, giving rise to an observational estimate for $\hat{\beta}^*$ of approximately 1.12. The true collider correction term is therefore $\beta - \beta^*$ = -0.62.

Standard IVW approach

We first simulated individual level data in a one-sample design under a balanced pleiotropy model. Pleiotropic effect parameters α_j were generated with a zero mean under InSIDE. Figure 3 (top-left) shows, for a range of sample sizes, the distribution of: (a) The standard IVW estimate (black line); (b) the SIMEX adjusted standard IVW estimate (blue line); (c) the collider corrected IVW estimate (red line); (d) the collider Corrected IVW estimate with SIMEX correction (green line); and (e) the TSLS estimate (orange line). We see that methods (a), (c) and (e) give essentially the same answer, and can therefore not be individually distinguished in the figure. The approximate equivalence of the TSLS and IVW approaches with uncorrelated SNPs is well known, but it is also reassuring that our two step approach is also equivalent. We also see that applying a direct SIMEX correction to method (a) (i.e. method (b)) dramatically increases the bias of the causal estimate beyond even that of the observational estimate for small sample sizes. This bias is slow to diminish as the sample size grows. Conversely, we see that applying a SIMEX correction indirectly to method (c) (i.e method (d)) results in the least biased estimate of all.

Figure 3 (top-right) gives further intuition on why the correction process works. The black line shows the estimated collider correction $\widehat{\beta - \beta^*}$ as a function of the given sample size. The blue line shows the true collider correction multiplied by the expected dilution factor $\frac{\bar{F}-1}{\bar{F}}$, which varies as a function of the sample size. The fact that the two lines are in good agreement indicates that the bias in $\widehat{\beta - \beta^*}$ is indeed a dilution, which can be perfectly predicted by $\bar{F}$. This underlines why SIMEX can be used to correct for it.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

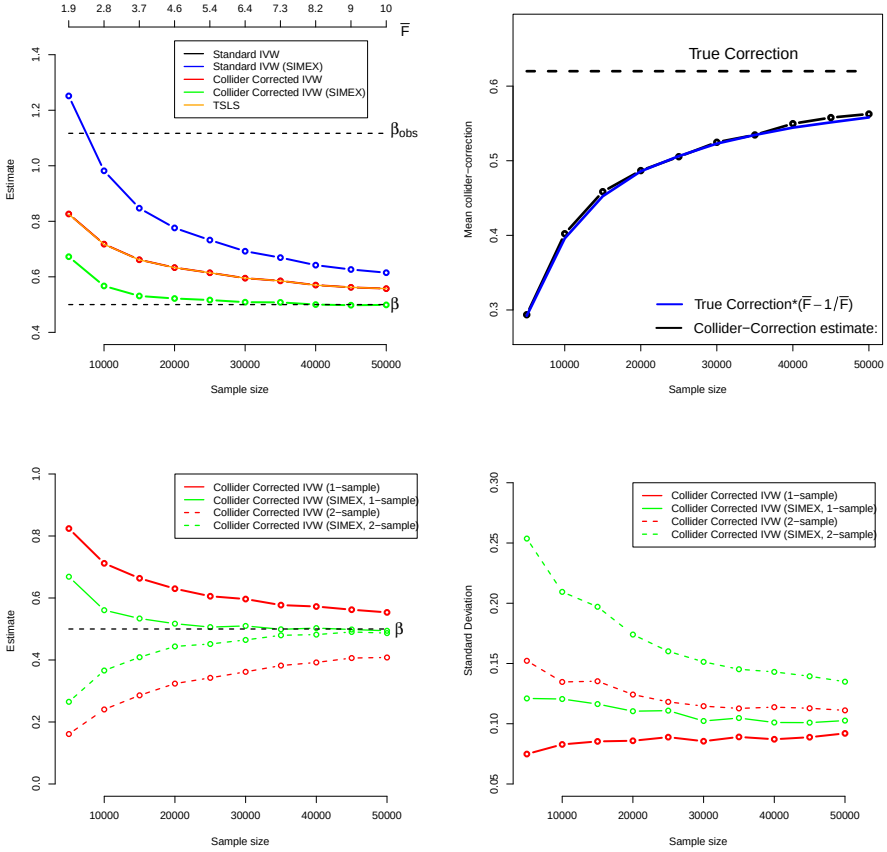

Figure 3: Top: Performance of IVW implementations (including the the collider-Correction algorithm) using one-sample data. Bottom: comparison of the one sample collider correction versus two-sample IVW approaches in terms of bias (bottom-left) and efficiency (bottom-right).

Figure 3 (bottom-left) shows the performance of the IVW estimate implemented using the (one sample) collider-Correction algorithm, versus that obtained from artificially splitting the data in two, calculating SNP-exposure associations in one half, SNP-outcome associations in the other half and combining in the usual manner. This is the most obvious way to ensure the independence of errors assumed by two sample approaches is satisfied. We show results for each method with and without SIMEX correction. We see that the absolute bias of the collider-Correction implementations is less than that of the two-sample implementation. However, the two estimation strategies differ more substantially in terms of precision, as shown in Figure 3 (bottom-right). Collider-correction is shown to be far more efficient.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

MR-Egger implementations

We next simulated individual level data under a directional pleiotropy model. That is, the pleiotropic effect parameters α_j were generated with a non-zero mean, but are still independent of the SNP-exposure association parameters β_{XGj} (i.e InSIDE satisfied). This enabled the assessment of the MR-Egger implementations of our collider correction algorithm. The results are shown in Figure 4 (top-left). Again, we see that standard and collider-corrected MR-Egger methods give the same results, but the two approaches differ greatly under SIMEX correction, with the SIMEX adjusted collider-corrected SIMEX estimate being least biased. In Figure 4 (top-right) we show how weak instrument bias in the collider-corrected slope for MREgger can be accurately quantified using the I^2_{GX} statistic, which again explains why SIMEX works.

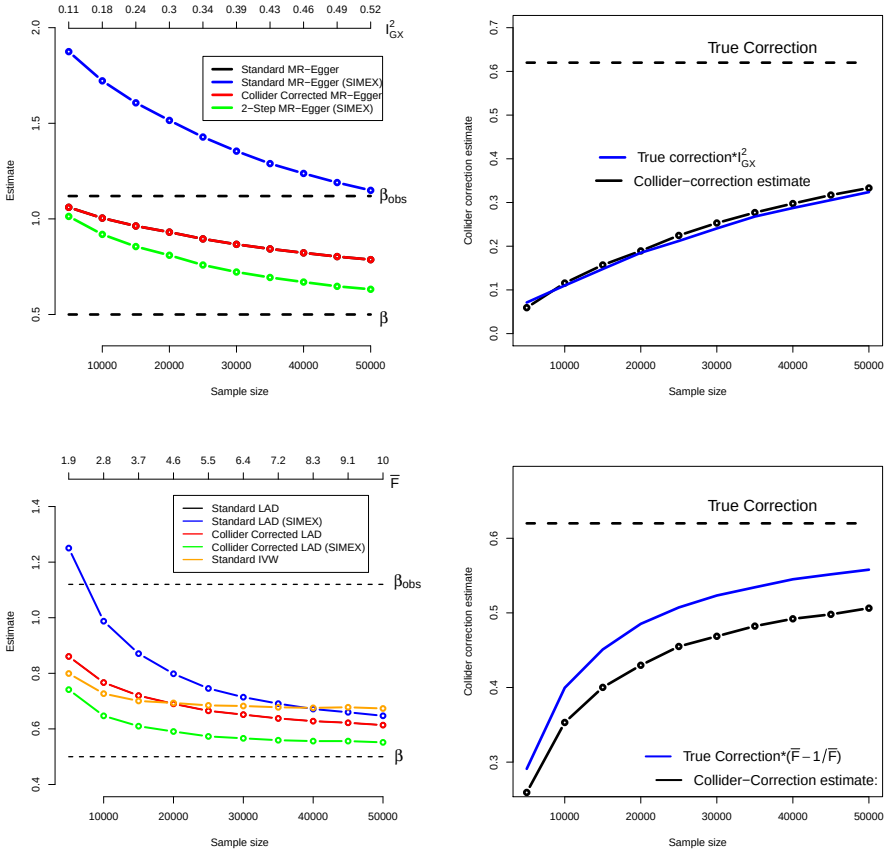

Figure 4: Top: Performance of the MR-Egger implementation of the collider-Correction algorithm under a directional pleiotropy scenario. Bottom: Performance of the LAD-regression implementation

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

LAD-regression implementation

We next simulated data under a model with directional pleiotropy induced by 15 large outliers, which constitute 30% of the k=50 SNPs. Figure 4 (bottom-left) show results for the collider-corrected LAD regression procedure compared to Standard LAD regression, with and without SIMEX correction. For comparison we also show the results for the Standard IVW implementation. Its bias does not approach zero as the sample size increases, which serves to motivate the utility of a robust regression procedure. Figure 4 (bottom-right) plots the mean dilution in the collider-Corrected LAD regression estimate, versus that predicted by the IVW dilution factor $\frac{\bar{F}-1}{\bar{F}}$. The fact that the observed dilution is below the expected IVW dilution illustrates that LAD regression is more vulnerable to weak instrument bias, because it is a less efficient but more robust technique. This emphasises the importance of being able to address its weak instrument bias.

Figure 6 in *Supplementary Information* shows the Monte-Carlo standard deviations of all estimators across the three scenarios considered. We see that the Standard and collider Corrected IVW, MR-Egger and LAD regression estimates have identical precision, which is higher than their SIMEX corrected analogues. This illustrates the bias-variance trade off made when enacting a weak-instrument bias adjustment.

Results: Assessing the causal role of Insomnia on HbA1c

Observationally, sub-optimal sleep (i.e., low sleep quantity and quality) has been found to be associated with hyperglycaemia [35, 36, 37] and increased diabetes risk [38]. Insomnia, defined as difficulty initiating or maintaining sleep, is one of the most important indices of sleep quality [39]. It has been associated with type 2 diabetes in observational studies [39] and in a previous Mendelian randomization study [40]. However, it is unclear whether associations with insomnia are mediated through HbA1c in the general population, whose glucose levels may not meet the threshold criteria for a formal diabetes diagnosis. As such, we focus on a potentially causal role of insomnia on HbA1c, a well-established clinical assessment of long-term glycaemic regulation that is central to the diagnosis of diabetes [41]. To address this question we use individual level data on approximately 320,000 individuals in UK Biobank to furnish a one sample Mendelian randomization study.

Two hundred and forty-eight independent genetic variants at 202 loci were associated with self-reported insomnia at or above the standard genome-wide significance threshold (p-value$<5\times10^{-8}$) in a recent GWAS of over 1.33 million UK Biobank and 23andMe individuals reported by Jansen [40] which collectively explained 2.6% of the total trait variance. SNP-exposure associations

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

were measured on the log-odds scale using logistic regression. Among this set of variants, 240 SNPs were in principle available for use as instruments in UK Biobank. In this cohort, participants were asked: "Do you have trouble falling asleep at night or do you wake up in the middle of the night?" with responses "Never/rarely", "Sometimes", "Usually", or "Prefer not to answer". Those who responded "Prefer not to answer" were set to missing. To reflect the Jansen analysis, the remaining entries were treated as a binary variable for insomnia symptoms, with "Never/rarely", "Sometimes", and "Usually" coded as 0, 0, and 1, respectively and a logistic regression performed. HbA1c measurements were obtained from a panel of biomarkers assayed from blood samples collected at baseline from UK Biobank participants. HbA1c (mmol/mol) was measured in red blood cells by HPLC analysis using Bio-Rad VARIANT II Turbo and log-transformed.

Instrument selection and winner's curse

The mean F statistic for the 240 genetic instruments in the original GWAS was 41, but in UK Biobank the same SNPs had an $\bar{F}$ of approximately 8.3 and an I^2_{GX} statistic of approximately 40%, meaning that the MR analysis was susceptible to bias due to both weak instrument and pleiotropy. This motivates the use of our collider-correction method for causal estimation. However, the original Jansen GWAS combined data from the UK Biobank (n=386,533) and 23andMe (n=944,477) using METAL [42]. As such, there was an approximate 23% overlap between data used for SNP discovery and for estimation in our MR model. This is known to induce a degree of winner's curse into the SNP-exposure associations [4]. For this reason we performed our subsequent analysis using (a) all 240 SNPs and (b) a subset of 112 SNPs that were only genome-wide significant using only the 23andMe portion of the Jansen data. This latter set is protected by design from winner's curse but, with an $\bar{F}$ of 6.8, were even weaker instruments.

Methods used

We applied the TSLS, IVW, MR-Egger and LAD regression approaches to the data. In addition, we applied two further methods that extend the basic IVW approach. The first is the Robust Adjusted Profile Score (MR-RAPS) approach of Zhao et al [43]. In its most basic form it delivers a causal estimate equivalent to that obtained by minimising Q-statistic (9) using w_j^* weights and a least squares loss function. However, it allows the user to incorporate alternative loss functions to give added robustness to outliers. We opted to use Tukey loss for this analysis. The second is a Bayesian implementation of MR-RAPS termed BayEsian Set IDentification and Estimation (MR-BESIDE) [44]. Rather than obtaining robustness to outliers through penalization, MR-BESIDE averages over models in which different collections of SNPs are included or discarded from the model. Although the technique can identify clusters of SNPs and estimate separate causal effects for each, we chose the one-parameter imple-

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

mentation of the algorithm with the DerSimonian and Laird (DL) pleiotropy variance estimation option. Both MR-RAPS and MR-BESIDE incorporate exact weighting an hence are naturally immune to weak instrument bias.

We employed three variants of the IVW, MR-Egger and LAD regression approaches: The 'Standard' 1-sample approach, collider-Correction and collider-Correction with weak instrument adjustment (with this latter being what we determine to be the gold standard implementation). For MR-Egger and LAD-regression we used SIMEX for weak instrument adjustment, whereas for IVW, MR-RAPS and MR-BESIDE, exact weighting was used.

Causal estimates

SNP exposure associations $\hat{\beta}_{XGj}$ were obtained from a logistic regression of insomnia on the set of SNPs as well age at recruitment, sex, assessment centre, 10 genetic principal components, and genotyping chip. Estimates for collider-biased SNP outcome associations $\hat{\alpha}_j^*$ were obtained from a multivariable regression of HbA1c on observed insomnia severity, all genetic variants and the same additional covariates. This second regression additionally yielded an estimate for the collider biased observational association between insomnia severity and HbA1c of $\hat{\beta}^* = 0.012$ (se $= 0.00057$).

Figure 5 (top) plots the collider-biased SNP-outcome associations versus the SNP-exposure associations. Overlaid on the plot are the weak-instrument-adjusted collider-Correction slopes estimated by the five summary data MR methods. We calculate the exact Q statistic [6], $Q(0)$, which measures global pleiotropy adjusted for weak instruments. It is 809 (df $= 239$) providing overwhelming evidence of heterogeneity due to pleiotropy. The 13 SNPs circled in black contribute a component to this global statistic with a bonferroni corrected p-value below $(5/240)\%$ and could therefore be classed as outliers. A list of these SNPs is provided in *Supplementary Information*.

For comparison SNPs coloured red have less than a $(5/240)\%$ probability of inclusion in the selected set of valid instruments within MR-BESIDE. These 41 SNPs are also given in *Supplementary information*. Adjusted causal effect estimates can be found in Table 1 (rows 1-5), with those in bold matching the slopes in Figure 5 (top).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

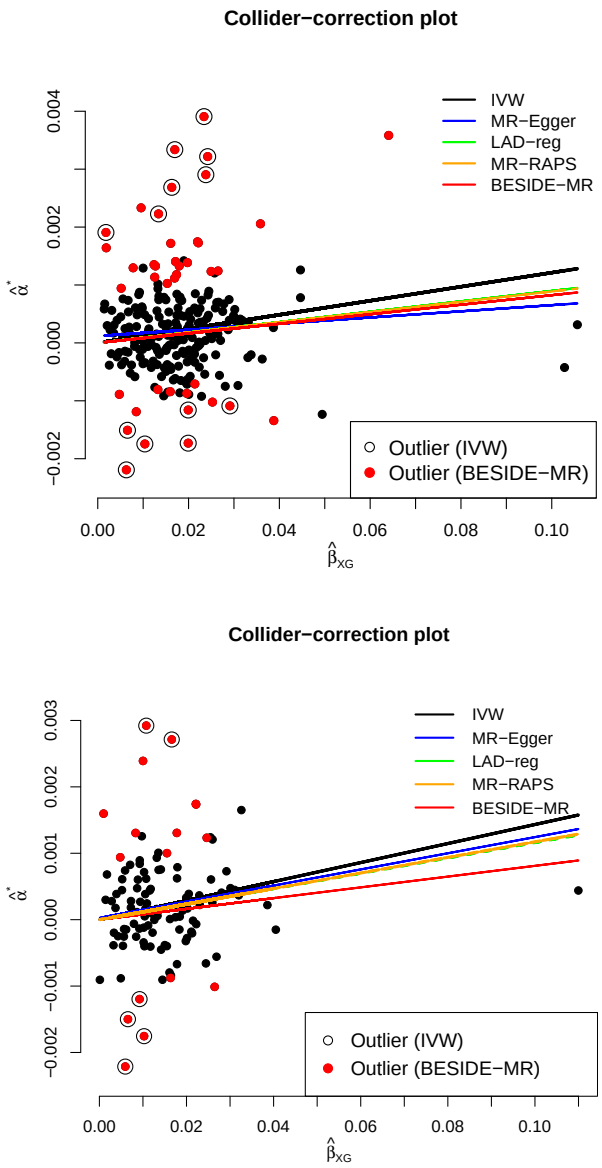

Figure 5: Collider-biased SNP outcome associations, $\hat{\alpha}_j^*$, versus SNP-exposure associations, $\hat{\beta}_{XGj}$ for: 240 SNPs that were genome-wide significant using 23andMe and UKB data (top); the 112 SNPs that were genome-wide significant using 23andMe data only (bottom)

Across all methods, we see a consistent picture: a unit increase in the log-odds of insomnia leads to an increase of between 0.17 and 0.24 units of log

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Method	Estimate	S.E	p-value
23andMe + UK Biobank significant SNPs			
# SNPs: 240, $\bar{F}$=8.36, $Q(0)(p-value)$ =809 ($<2\times10^{-16}$), I^2_{GX}=41.0%			
β^*	0.012	0.00057	$< 2\times10^{-16}$
TSLS	0.016	0.002	$< 1\times10^{-16}$
Standard IVW	0.013	0.008	5.04×10^{-7}
Col-Cor IVW (1)	0.022	0.00278	1.1×10^{-15}
Col-Cor IVW (2)	0.024	0.0029	6.7×10^{-16}
Standard MR-Egger	0.007	0.005	1.3×10^{-1}
Col-Cor MR-Egger (1)	0.015	0.006	1.33×10^{-2}
Col-Cor MR-Egger (2)	0.017	0.0086	4.9×10^{-2}
Standard LAD	0.011	0.004	2.09×10^{-3}
Col-Cor LAD (1)	0.020	0.0036	2.00×10^{-8}
Col-Cor LAD (2)	0.020	0.0027	1.3×10^{-14}
Col-Cor MR-RAPs	0.020	0.0026	3.1×10^{-15}
Col-Cor BESIDE-MR	0.020	0.0019	$<2\times10^{-16}$
23andMe significant SNPs only			
# SNPs: 112, $\bar{F}$=6.88, $Q(0)(p-value)$ =385 ($<2\times10^{-16}$), I^2_{GX}=52.1%			
β^*	0.012	0.00057	$< 2\times10^{-16}$
TSLS	0.017	0.003	2.39×10^{-10}
Standard IVW	0.014	0.004	5.23×10^{-4}
Col-Cor IVW (1)	0.024	0.0045	1.21×10^{-7}
Col-Cor IVW (2)	0.026	0.0048	6.5×10^{-8}
Standard MR-Egger	0.008	0.006	1.76×10^{-1}
Col-Cor MR-Egger (1)	0.020	0.008	1.79×10^{-2}
Col-Cor MR-Egger (2)	0.024	0.0110	3.3×10^{-2}
Standard LAD	0.012	0.006	3.30×10^{-2}
Col-Cor LAD (1)	0.021	0.0056	1.52×10^{-4}
Col-Cor LAD (2)	0.023	0.0042	4.2×10^{-8}
Col-Cor MR-RAPs	0.023	0.0043	3.6×10^{-8}
Col-Cor BESIDE-MR	0.020	0.0030	6.9×10^{-11}

Table 1: *Point estimates, standard errors and p-values for the: IVW, MR-Egger, MR-RAPS and BESIDE-MR method. Estimates reflect the average causal effect of a unit increase in the log-odds of insomnia on HbA1c levels across the population. 'Standard' = standard 1-sample analysis. 'Col-Cor (1)' = collider correction analysis. 'Col-Cor (2)' = collider correction + weak instrument correction. Top rows: All 240 SNPs from Jansen et al used. Bottom rows: only genome wide significant SNPs from 23andMe data (ignoring UK Biobank) used.*

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

mmol/mol of HbA1c. All estimates are further from the null than the collider-biased observational association, $\hat{\beta}^*$. However, the results highlight that, without weak-instrument adjustment, all summary data MR-methods are biased in the direction of $\hat{\beta}^*$.

Table 1 (rows 6:10) and Figure 5 (bottom) show the MR results when using only the 112 SNPs identified in Jansen from 23andMe data, which are immune to winner's curse. These SNPs have a weaker mean F statistic of 6.88 but a higher I^2_{GX} statistic of 52%. All causal estimates are seen to increase, because winner's curse leads to an over-estimation of the SNP-exposure association, which forms the denominator of the standard ratio estimate for β. Although fewer outliers are detected using individual heterogeneity statistics and MR-BESIDE, they represent a similar proportion of the total number of SNPs as in the previous analysis. Again, across all methods, we see consistent evidence that increased Insomnia leads to higher HbA1c levels.

Discussion

In this paper we clarify how the principle of collider correction offers a vehicle for applying any two-sample summary data MR method to one sample data, making it easy to account for both pleiotropy and weak instrument bias. Our method is closely related to the approach of Dudbridge et al [29] for genetic studies of disease progression, and primarily serves to emphasise that this procedure is in fact applicable to any MR analysis. We used our new method to provide important insights into the role of insomnia on glycated haemoglobin and, by extension, on incident diabetes.

We showcased the collider-correction approach using five univariate MR approaches. A current trend in MR analysis is to employ methods that attempt to estimate causal effects identified by different clusters of SNPs, see for example MRMix [45], MR-ConMix[46] and the two-parameter extension of MR-BESIDE[44]. All of these methods can be applied directly to one sample data using our approach, and an examination of their performance in this arena is an important topic of future research.

The insomnia data was affected by a small amount of winner's curse, which we removed by design in a sensitivity analysis by restricting our SNP set to those obtained from a purely independent data source. More sophisticated approaches to adjusting for winner's curse are possible by incorporating the original Discovery data. For example, Bowden and Dudbridge [4] show that winner's curse bias can be perfectly eliminated by combining SNP discovery and validation data from two non-overlapping GWAS studies. As further work, we plan to extend this approach and combine it with collider-correction.

Although our method ostensibly requires individual level data, it can be en-

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

acted purely with summary statistics on the SNP-exposure associations, collider biased SNP-outcome associations, and the biased observational association. For more complicated analyses that investigate effect modification and multi-variable non-linear or sex-specific effects, the same principle remains although a much larger set of summary measures would be needed. In future work we will aim to clarify these details in the full generalization of the collider correction approach.

Acknowledgement James Liu is funded by a Diabetes UK project grant (17/0005700). Deborah A Lawlor, James Liu and Rebecca Richmond all work in a Unit that receives support from the University of Bristol and UK Medical Research Council (MCUU00011/6). Deborah A Lawlor is a National Institute of Research Senior Investigator (NF-0616-10102). Jack Bowden is funded by an Establishing Excellence in England (E3) research award. None of the funders influenced the research presented here and the views expressed in this paper are those of the authors and not necessarily any funders acknowledged here.

Author contributions

Barry and Bowden developed the collider-correction algorithm and its rationale was further refined by Dudbridge. Liu, Richmond and Bowden applied the method to UK Biobank and 23andMe data to produce the results detailed in the paper. Bowden, Barry and Dudbridge produced an initial draft of the paper which was further refined by Lawlor, Rutter, Liu and Richmond. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

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* 32.1 (Feb. 2003), pp. 1–22. ISSN: 0300-5771. DOI: 10.1093/ije/dyg070. eprint: http://oup.prod.sis.lan/ije/article-pdf/32/1/1/11213726/320001_dyg070.pdf. URL: https://doi.org/10.1093/ije/dyg070.

[2] Nuala A Sheehan et al. "Mendelian Randomisation and Causal Inference in Observational Epidemiology". In: *PLOS Medicine* 5.8 (Aug. 2008), pp. 1–6. DOI: 10.1371/journal.pmed.0050177. URL: https://doi.org/10.1371/journal.pmed.0050177.

[3] George Davey Smith et al. "Clustered Environments and Randomized Genes: A Fundamental Distinction between Conventional and Genetic Epidemiology". In: *PLOS Medicine* 4.12 (Dec. 2007), pp. 1–8. DOI: 10.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

1371/journal.pmed.0040352. URL: https://doi.org/10.1371/journal.pmed.0040352.

[4] Jack Bowden and Frank Dudbridge. "Unbiased estimation of odds ratios: combining genomewide association scans with replication studies". In: *Genetic Epidemiology* 33.5 (2009), pp. 406–418. DOI: 10.1002/gepi.20394. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/gepi.20394. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/gepi.20394.

[5] Jack Bowden et al. "Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic". In: *International Journal of Epidemiology* 45.6 (Sept. 2016), pp. 1961–1974. ISSN: 0300-5771. DOI: 10.1093/ije/dyw220. eprint: http://oup.prod.sis.lan/ije/article-pdf/45/6/1961/24170827/dyw220.pdf. URL: https://doi.org/10.1093/ije/dyw220.

[6] Jack Bowden et al. "Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption". In: *International Journal of Epidemiology* 48.3 (Dec. 2018), pp. 728–742. ISSN: 0300-5771. DOI: 10.1093/ije/dyy258. eprint: http://oup.prod.sis.lan/ije/article-pdf/48/3/728/29005355/dyy258.pdf. URL: https://doi.org/10.1093/ije/dyy258.

[7] Gibran Hemani, Jack Bowden, and George Davey Smith. "Evaluating the potential role of pleiotropy in Mendelian randomization studies". In: *Human Molecular Genetics* 27.R2 (May 2018), R195–R208. ISSN: 0964-6906. DOI: 10.1093/hmg/ddy163. eprint: http://oup.prod.sis.lan/hmg/article-pdf/27/R2/R195/25229892/ddy163.pdf. URL: https://doi.org/10.1093/hmg/ddy163.

[8] Hyunseung Kang et al. "Instrumental Variables Estimation With Some Invalid Instruments and its Application to Mendelian Randomization". In: *Journal of the American Statistical Association* 111.513 (2016), pp. 132–144. DOI: 10.1080/01621459.2014.994705. eprint: https://doi.org/10.1080/01621459.2014.994705. URL: https://doi.org/10.1080/01621459.2014.994705.

[9] Jack Bowden, George Davey Smith, and Stephen Burgess. "Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression". In: *International Journal of Epidemiology* 44.2 (June 2015), pp. 512–525. ISSN: 0300-5771. DOI: 10.1093/ije/dyv080. eprint: http://oup.prod.sis.lan/ije/article-pdf/44/2/512/2266954/dyv080.pdf. URL: https://doi.org/10.1093/ije/dyv080.

[10] J. Bowden et al. "Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator". In: *Genetic Epidemiology* 40 (4 2016), pp. 304–14. DOI: 10.1002/gepi.21965.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[11] Fernando P. Hartwig, Jack Bowden, and George Davey Smith. "Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption". In: *International journal of epidemiology* 46.6 (2017), pp. 1985–1998. DOI: 10.1093/ije/dyx102.

[12] Stephen Burgess, Adam Butterworth, and Simon G. Thompson. "Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data". In: *Genetic Epidemiology* 37.7 (2013), pp. 658–665. DOI: 10.1002/gepi.21758. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/gepi.21758. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/gepi.21758.

[13] Jack Bowden et al. "A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization". In: *Statistics in Medicine* 36.11 (2017), pp. 1783–1802. DOI: 10.1002/sim.7221. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.7221. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.7221.

[14] G Hemani et al. "The MR-Base platform supports systematic causal inference across the human phenome." In: *E-life* 7 (2018), e34408. DOI: 10.7554/eLife.34408.

[15] Jack Bowden et al. "Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression". In: *International Journal of Epidemiology* 47.4 (June 2018), pp. 1264–1278. ISSN: 0300-5771. DOI: 10.1093/ije/dyy101. eprint: http://oup.prod.sis.lan/ije/article-pdf/47/4/1264/26568165/dyy101.pdf. URL: https://doi.org/10.1093/ije/dyy101.

[16] Deborah Lawlor et al. *"A Mendelian Randomization Dictionary Useful Definitions and Descriptions for Undertaking, Understanding and Interpreting Mendelian Randomization Studies."*. Feb. 2019. DOI: 10.31219/osf.io/6yzs7.

[17] Atsushi Inoue and Gary Solon. "TWO-SAMPLE INSTRUMENTAL VARIABLES ESTIMATORS". In: *The Review of Economics and Statistics* 92.3 (2010), pp. 557–561. ISSN: 00346535, 15309142. URL: http://www.jstor.org/stable/27867559.

[18] R. Hyslop and Guido W. Imbens. "Bias from Classical and Other Forms of Measurement Error". In: *Journal of Business Economic Statistics* 19.4 (2001), pp. 475–481. ISSN: 07350015. URL: http://www.jstor.org/stable/1392282.

[19] James W. Hardin, Henrik Schmiediche, and Raymond J. Carroll. "The Simulation Extrapolation Method for Fitting Generalized Linear Models with Additive Measurement Error". In: *The Stata Journal* 3.4 (2003), pp. 373–385. DOI: 10.1177/1536867X0400300407. eprint: https://doi.org/10.1177/1536867X0400300407. URL: https://doi.org/10.1177/1536867X0400300407.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[20] Qingyuan Zhao et al. "Two-Sample Instrumental Variable Analyses Using Heterogeneous Samples". In: *Statist. Sci.* 34.2 (May 2019), pp. 317–333. DOI: 10.1214/18-STS692. URL: https://doi.org/10.1214/18-STS692.

[21] The CARDIoGRAMplusC D. Consortium, Panos Deloukas, and Nilesh J Samani. "Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique". In: *International journal of epidemiology* 45.6 (2016), pp. 1717–1726. URL: 10.1093/ije/dyx028.

[22] Marissa LeBlanc et al. "A correction for sample overlap in genome-wide association studies in a polygenic pleiotropy-informed framework". In: *BMC Genomics* 19.1 (2018), p. 494. ISSN: 1471-2164. DOI: 10.1186/s12864-018-4859-7. URL: https://doi.org/10.1186/s12864-018-4859-7.

[23] Fernando P. Hartwig et al. "Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique". In: *International journal of epidemiology* 45.6 (2016), pp. 1717–1726. URL: 10.1093/ije/dyx028.

[24] Fernando Pires Hartwig et al. "Bias in two-sample Mendelian randomization by using covariable-adjusted summary associations". In: *bioRxiv* (2019). DOI: 10.1101/816363. eprint: https://www.biorxiv.org/content/early/2019/10/24/816363.full.pdf. URL: https://www.biorxiv.org/content/early/2019/10/24/816363.

[25] Debbie A. Lawlor. "Commentary: Two-sample Mendelian randomization: opportunities and challenges". In: *International journal of epidemiology* 45.3 (2016), pp. 908–15. URL: doi:10.1093/ije/dyw127.

[26] Cathie Sudlow et al. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age". In: *PLOS Medicine* 12.3 (Mar. 2015), pp. 1–10. DOI: 10.1371/journal.pmed.1001779. URL: https://doi.org/10.1371/journal.pmed.1001779.

[27] Cosetta Minelli et al. "The use of two-sample methods for Mendelian randomization analyses on single large datasets". In: *bioRxiv* (2020). DOI: 10.1101/2020.05.07.082206. eprint: https://www.biorxiv.org/content/early/2020/05/07/2020.05.07.082206.full.pdf. URL: https://www.biorxiv.org/content/early/2020/05/07/2020.05.07.082206.

[28] Fernando Pires Hartwig and Neil Martin Davies. "Why internal weights should be avoided (not only) in MR-Egger regression". In: *International Journal of Epidemiology* 45.5 (Sept. 2016), pp. 1676–1678. ISSN: 0300-5771. DOI: 10.1093/ije/dyw240. eprint: https://academic.oup.com/ije/article-pdf/45/5/1676/7383008/dyw240.pdf. URL: https://doi.org/10.1093/ije/dyw240.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[29] Frank Dudbridge et al. "Adjustment for index event bias in genome-wide association studies of subsequent events". In: *Nature Communications* 10.1 (2019), p. 1561. DOI: 10.1038/s41467-019-09381-w. URL: https://doi.org/10.1038/s41467-019-09381-w.

[30] Stephen Burgess and Jack Bowden. *Integrating summarized data from multiple genetic variants in Mendelian randomization: bias and coverage properties of inverse-variance weighted methods*. 2015. arXiv: 1512.04486 [stat.AP].

[31] Judea Pearl. "Causal inference in statistics: An overview". In: *Statist. Surv.* 3 (2009), pp. 96–146. DOI: 10.1214/09-SS057. URL: https://doi.org/10.1214/09-SS057.

[32] Marcus R Munafò et al. "Collider scope: when selection bias can substantially influence observed associations". In: *International Journal of Epidemiology* 47.1 (Sept. 2017), pp. 226–235. ISSN: 0300-5771. DOI: 10.1093/ije/dyx206. eprint: http://oup.prod.sis.lan/ije/article-pdf/47/1/226/25088130/dyx206.pdf. URL: https://doi.org/10.1093/ije/dyx206.

[33] Jack Bowden et al. "Improving the accuracy of two-sample summary data Mendelian randomization: moving beyond the NOME assumption". In: *bioRxiv* (2018). DOI: 10.1101/159442. eprint: https://www.biorxiv.org/content/early/2018/10/11/159442.full.pdf. URL: https://www.biorxiv.org/content/early/2018/10/11/159442.

[34] J. R. Cook and L. A. Stefanski. "Simulation-Extrapolation Estimation in Parametric Measurement Error Models". In: *Journal of the American Statistical Association* 89.428 (1994), pp. 1314–1328. DOI: 10.1080/01621459.1994.10476871. eprint: https://doi.org/10.1080/01621459.1994.10476871. URL: https://doi.org/10.1080/01621459.1994.10476871.

[35] K Spiegel, R Leproult, and E Van Cauter. "Impact of sleep debt on metabolic and endocrine function". In: *Lancet (London, England)* 354.9188 (Oct. 1999), pp. 1435–1439. ISSN: 0140-6736. DOI: 10.1016/s0140-6736(99)01376-8. URL: https://doi.org/10.1016/S0140-6736(99)01376-8.

[36] Katherine A. Stamatakis and Naresh M. Punjabi. "Effects of sleep fragmentation on glucose metabolism in normal subjects". English (US). In: *Chest* 137.1 (Jan. 2010), pp. 95–101. ISSN: 0012-3692. DOI: 10.1378/chest.09-0791.

[37] Arlet V. Nedeltcheva et al. "Exposure to Recurrent Sleep Restriction in the Setting of High Caloric Intake and Physical Inactivity Results in Increased Insulin Resistance and Reduced Glucose Tolerance". In: *The Journal of Clinical Endocrinology Metabolism* 94.9 (Sept. 2009), pp. 3242–3250. ISSN: 0021-972X. DOI: 10.1210/jc.2009-0483. eprint: https://academic.oup.com/jcem/article-pdf/94/9/3242/9064984/jcem3242.pdf. URL: https://doi.org/10.1210/jc.2009-0483.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[38] Zhilei Shan et al. "Sleep Duration and Risk of Type 2 Diabetes: A Meta-analysis of Prospective Studies". In: *Diabetes Care* 38.3 (2015), pp. 529–537. ISSN: 0149-5992. DOI: 10.2337/dc14-2073. eprint: https://care.diabetesjournals.org/content/38/3/529.full.pdf. URL: https://care.diabetesjournals.org/content/38/3/529.

[39] Michael J. Green et al. "Insomnia symptoms as a cause of type 2 diabetes Incidence: a 20?year cohort study". In: *BMC Psychiatry* 17.1 (2017), p. 94. ISSN: 1471-244X. DOI: 10.1186/s12888-017-1268-4. URL: https://doi.org/10.1186/s12888-017-1268-4.

[40] Philip R. Jansen et al. "Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways". In: *Nature Genetics* 51.3 (2019), pp. 394–403. ISSN: 1546-1718. DOI: 10.1038/s41588-018-0333-3. URL: https://doi.org/10.1038/s41588-018-0333-3.

[41] "Diagnosis and Classification of Diabetes Mellitus". In: *Diabetes Care* 27.1 (2004), s5–s10. DOI: 10.2337/diacare.27.2007.S5.

[42] Cristen J. Willer, Yun Li, and Gonçalo R. Abecasis. "METAL: fast and efficient meta-analysis of genomewide association scans". In: *Bioinformatics* 26.17 (July 2010), pp. 2190–2191. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btq340. eprint: https://academic.oup.com/bioinformatics/article-pdf/26/17/2190/501468/btq340.pdf. URL: https://doi.org/10.1093/bioinformatics/btq340.

[43] Qingyuan Zhao et al. *Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score: to appear in Annals of Statistics*. 2018. arXiv: 1801.09652 [stat.AP].

[44] Chin Yang Shapland, Qingyuan Zhao, and Jack Bowden. "Profile-likelihood Bayesian model averaging for two-sample summary data Mendelian randomization in the presence of horizontal pleiotropy". In: *bioRxiv* (2020). DOI: 10.1101/2020.02.11.943712. eprint: https://www.biorxiv.org/content/early/2020/02/11/2020.02.11.943712.full.pdf. URL: https://www.biorxiv.org/content/early/2020/02/11/2020.02.11.943712.

[45] Guanghao Qi and Nilanjan Chatterjee. "Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects". In: *Nature Communications* 10.1 (2019), p. 1941. ISSN: 2041-1723. DOI: 10.1038/s41467-019-09432-2. URL: https://doi.org/10.1038/s41467-019-09432-2.

[46] Stephen Burgess et al. "A robust and efficient method for Mendelian randomization with hundreds of genetic variants". In: *Nature Communications* 11.1 (2020), p. 376. ISSN: 2041-1723. DOI: 10.1038/s41467-019-14156-4. URL: https://doi.org/10.1038/s41467-019-14156-4.

Supplementary information

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

A Derivation of Equation (5)

The asymptotic least squares estimates of the effects of X and G on Y , without conditioning on U , are

$$\begin{bmatrix} \beta^* \\ \alpha_1^* \\ \vdots \\ \alpha_k^* \end{bmatrix} = \begin{bmatrix} var\,(X) & cov\,(X,G_1) & \cdots & cov\,(X,G_k) \\ cov\,(X,G_1) & var\,(G_1) & \cdots & cov\,(G_1,G_k) \\ \vdots & \vdots & \ddots & \vdots \\ cov\,(X,G_k) & cov\,(G_1,G_k) & \cdots & var\,(G_k) \end{bmatrix}^{-1} \begin{bmatrix} cov\,(X,Y) \\ cov\,(G_1,Y) \\ \vdots \\ cov\,(G_k,Y) \end{bmatrix}$$

Assuming no LD between SNPs, so $cov\,(G_i,G_j) = 0$ where $i \neq j$, the variance-covariance matrix has block form with a diagonal matrix in the lower right quadrant. Block-wise inversion gives

$$\begin{bmatrix} var\,(X) & cov\,(X,G_1) & \cdots & cov\,(X,G_k) \\ cov\,(X,G_1) & var\,(G_1) & \cdots & cov\,(G_1,G_k) \\ \vdots & \vdots & \ddots & \vdots \\ cov\,(X,G_k) & cov\,(G_1,G_k) & \cdots & var\,(G_k) \end{bmatrix}^{-1} =$$

$$\frac{1}{var\,(X) - \sum_j \frac{cov(X,G_j)^2}{var(G_j)}} \begin{bmatrix} 1 & \frac{-cov(X,G_1)}{var(G_1)} & \cdots & \frac{-cov(X,G_k)}{var(G_k)} \\ \frac{-cov(X,G_1)}{var(G_k)} & \frac{cov(X,G_1)^2}{var(G_1)^2} & \cdots & \frac{cov(X,G_1)cov(X,G_k)}{var(G_1)var(G_k)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-cov(X,G_k)}{var(G_k)} & \frac{cov(X,G_1)cov(X,G_k)}{var(G_1)var(G_k)} & \cdots & \frac{cov(X,G_k)^2}{var(G_k)^2} \end{bmatrix} +$$

$$\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{var(G_1)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{var(G_k)} \end{bmatrix}$$

Then

$$\beta^* = \frac{cov\,(X,Y) - \sum_j \frac{cov(X,G_j)cov(G_j,Y)}{var(G_j)}}{var\,(X) - \sum_j \frac{cov(X,G_j)^2}{var(G_j)}}$$

And

$$\alpha_i^* = \frac{\frac{-cov(X,G_j)}{var(G_j)} \left(cov\,(X,Y) - \sum_j \frac{cov(X,G_j)cov(G_j,Y)}{var(G_j)} \right)}{var\,(X) - \sum_j \frac{cov(X,G_j)^2}{var(G_j)}} + \frac{cov\,(G_i,Y)}{var\,(G_i)}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

$$= -\beta_{XG_j}\beta^* + \frac{cov\left(G_i, Y\right)}{var\left(G_i\right)}$$

From equation 2, $cov\left(G_i, Y\right) = \left(\alpha_i + \beta\beta_{XG_i}\right) var\left(G_i\right)$. Therefore

$$\alpha_j^* = \alpha_j + \beta_{XG_j}\left(\beta - \beta^*\right)$$

The causal effect β is therefore the observational effect β^* , plus the slope of the regression of α_j^* on β_{XG_j} .

B Standard deviation plots for Section 3 simulation study

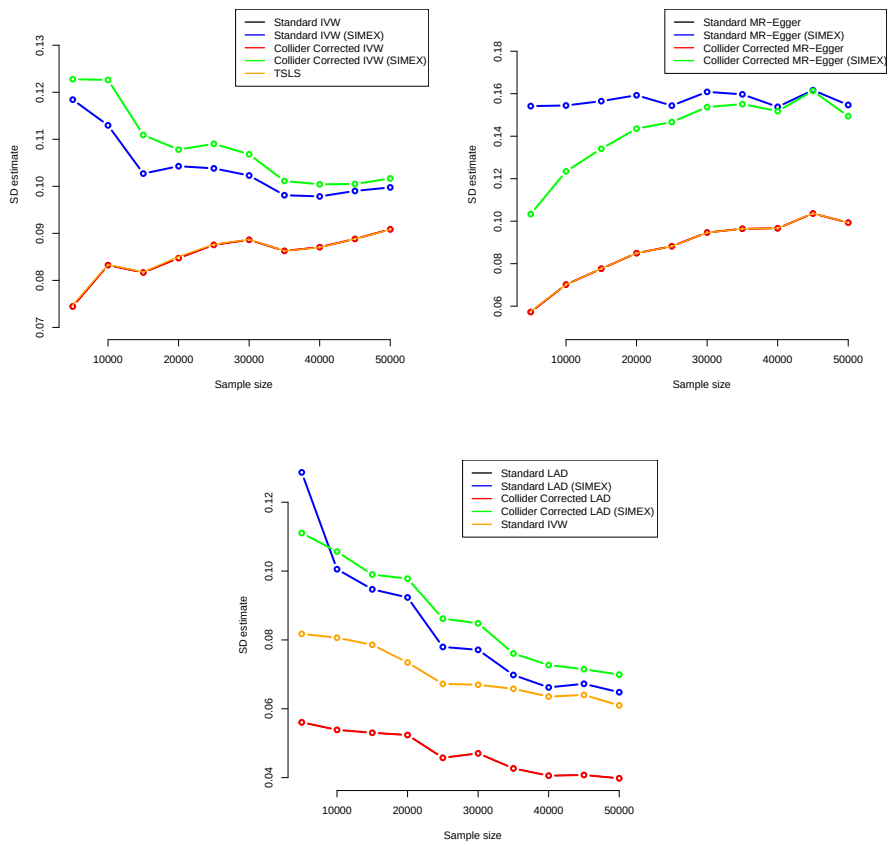

Figure 6: Monte-Carlo standard deviations for all IVW (top-left), MR-Egger (top-right) and LAD regression (bottom) estimators

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

C Outlier SNPs sets for the Insomnia analysis of Section 4.

SNP set detected as outliers using a Bonferroni corrected exact Q statistic (23andMe data only)

1	rs10758593
2	rs1264419
3	rs214934
4	rs2792990
5	rs4788203
6	rs647905

SNP set detected as outliers using a Bonferroni corrected BMA inclusion probability statistic (23andMe data only)

1	rs10758593
2	rs1264419
3	rs214934
4	rs2286729
5	rs2792990
6	rs28582096
7	rs2903385
8	rs4502882
9	rs4788203
10	rs55772859
11	rs60565673
12	rs6119267
13	rs647905
14	rs694786
15	rs742760
16	rs8180457
17	rs874168

SNP set detected as outliers using a Bonferroni corrected exact Q statistic (23andMe + UK Biobank data)

1	rs10758593
2	rs1264419
3	rs12917449
4	rs12924275
5	rs1861412
6	rs214934
7	rs2737240
8	rs2792990
9	rs3131638
10	rs34490907

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

11	rs429358
12	rs4788203
13	rs6888135

SNP set detected as outliers using a Bonferroni corrected BMA inclusion probability statistic (23andMe + UK Biobank data)

1	rs1064939
2	rs10758593
3	rs11650304
4	rs12187443
5	rs12251016
6	rs1264419
7	rs12917449
8	rs12924275
9	rs1519102
10	rs1861412
11	rs214934
12	rs2216427
13	rs2286729
14	rs238869
15	rs2737240
16	rs2792990
17	rs28582096
18	rs28611339
19	rs2903385
20	rs3131638
21	rs3184470
22	rs34214423
23	rs34490907
24	rs429358
25	rs4502882
26	rs4643373
27	rs4788203
28	rs55772859
29	rs60565673
30	rs6119267
31	rs647905
32	rs6606731
33	rs6888135
34	rs694786
35	rs72820274
36	rs742760
37	rs7486418
38	rs8180457
39	rs823247
40	rs874168

medRxiv preprint doi: https://doi.org/10.1101/2020.10.20.20216358; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

41	rs908668