Improving the accuracy of two-sample summary data Mendelian randomization: moving beyond the NOME assumption

Jack Bowden[1,2*], Fabiola Del Greco M[3], Cosetta Minelli[4]
Debbie A Lawlor[1,2], Nuala A Sheehan[5] John Thompson[5]
George Davey Smith[1,2]

[1] *MRC Integrative Epidemiology Unit at the University of Bristol, U.K*
[2] *School of Social and Community Medicine, University of Bristol, U.K*
[3] *Center for Biomedicine, EURAC research, Bolzano, Italy*
[4] *Population Health and Occupational Disease, NHLI, Imperial College, London, U.K*
[5] *Department of Health Sciences, University of Leicester, Leicester, U.K*

*Address for correspondence:
Jack Bowden
MRC Integrative Epidemiology Unit
Oakfield House, Bristol, BS8 2BN, U.K
jack.bowden@bristol.ac.uk.

Abstract

Two-sample summary data Mendelian randomization (MR) incorporating multiple genetic variants in a meta-analysis framework is a popular technique for assessing causality in epidemiology. If all genetic variants satisfy the instrumental variable (IV) assumptions, then their individual causal ratio estimates should be homogeneous. Observed heterogeneity, therefore, supports the notion that a portion of variants violate the IV assumptions due to pleiotropy. Model fitting and heterogeneity assessment in MR requires an approximation for the variance of each ratio estimate. We show that the most popular approximation can lead to an inflation in the chances of detecting heterogeneity when in fact it is not present. Conversely, an ostensibly more accurate approximation can dramatically increase the chances of failing to detect heterogeneity, when it is truly present. Here we derive a 'modified 2nd order' approximation to the variance that makes use of the derived causal estimate to mitigate both of these adverse effects. Using Monte Carlo simulations, we show that the modified 2nd order approximation outperforms both its 1st and 2nd order counterparts in the presence of weak instruments or a large causal effect. We illustrate the utility of the new method using data from a recent two-sample summary data MR analysis to assess the causal role of systolic blood pressure on coronary heart disease risk. Modified 2nd order weighting should be used as standard within two-sample summary data MR studies for model fitting, the quantification of heterogeneity and the detection of outliers. `R` code is provided to apply these weights in practice.

Key words: Mendelian randomization, two-sample MR, Cochran's Q statistic, MR-Egger regression, Rücker's $Q^{'}$ statistic.

Introduction

Mendelian randomization (MR) [1] is an instrumental variable approach that uses genetic data, typically in the form of single nucleotide polymorphisms (SNPs), to assess whether a modifiable exposure exerts a causal effect on a health outcome in the presence of unmeasured confounding. Traditionally, researchers have assumed that SNPs used for MR studies are valid instrumental variables (IVs) for the purposes of inferring the causal effect of an exposure, X, on an outcome, Y. Specifically, the SNP is: associated with X (IV assumption 1 (IV1)); not associated with any confounders of X and Y (IV2); and can only be associated with Y through X (IV3). The IV assumptions are represented by the solid lines in Figure 1 for a SNP G_j, with unobserved confounding represented by U. Dotted lines represent dependencies between G and U, and G and Y that are prohibited by the IV assumptions.

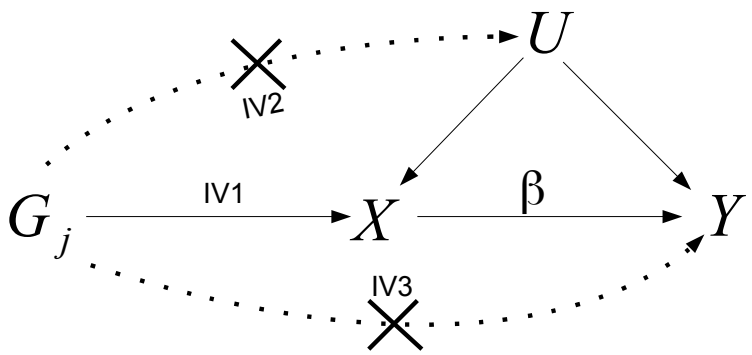

Figure 1: *Causal diagram representing the IV assumptions (and violations thereof) for a SNP G_j, an exposure X and an outcome Y. The causal effect of X on Y, denoted by β, is the quantity we wish to estimate.*

Suppose initially that a SNP, G_j, is a valid IV. Further assume that the association of G_j with both X and Y is linear with no effect modification, and X affects Y linearly with no effect modification. The underlying SNP-outcome association Γ_j - the increase in Y for a unit increase in G_j - can then be expressed as a scalar multiple of the underlying SNP-exposure association estimate, γ_j - the increase in X for a unit increase in G_j. That is: $\Gamma_j = \beta\gamma_j$, where β denotes the causal effect of a unit increase in X on the outcome Y.

Figure 1 encodes the assumptions that are traditionally required for a single sample of individuals for whom G, X and Y are measured. A particular MR study design gaining in popularity instead combines publically available summary data on SNP-exposure and SNP-outcome associations from two separate studies for large numbers of uncorrelated variants G_1,...,G_L within the framework of a meta-analysis. These studies should ideally contain no overlapping individuals (to ensure independence) but should also originate from the same source population. This is referred to as two-sample summary data MR [2]. Providing the aforementioned modelling assumptions

are met and each SNP is a valid IV, when SNP-exposure and SNP-outcome associations are estimated from their respective samples the ratio $\hat{\beta}_j = \hat{\Gamma}_j/\hat{\gamma}_j$ for any single variant should also provide a consistent estimate for β. Combining the set of L ratio estimates obtained across all variants into an overall inverse variance weighted (IVW) estimate using the standard meta-analytic formula:

$$\hat{\beta}_{IVW} = \frac{\sum_{j=1}^{L} w_j \hat{\beta}_j}{\sum_{j=1}^{L} w_j} \quad \text{where} \quad w_j = var(\hat{\beta}_j)^{-1}, \tag{1}$$

then provides an efficient and consistent estimate for β. For a more detailed description of the assumptions required by two-sample summary data MR, see Bowden et al [3].

Heterogeneity assessment

If the aforementioned modelling and IV assumptions hold, then Cochran's Q statistic:

$$Q = \sum_{j=1}^{L} Q_j = \sum_{j=1}^{L} w_j(\hat{\beta}_j - \hat{\beta}_{IVW})^2. \tag{2}$$

should follow, asymptotically, a χ^2 distribution on L-1 degrees of freedom. Excessive heterogeneity could therefore indicate the modelling assumptions have been violated, or that some of the genetic variants violate the IV assumptions. For example, even if all SNPs are valid IVs, causal effect heterogeneity will exist whenever the SNP-outcome association is an odds ratio, due to issues of non-collapsibility [4]. Specifically, the causal effect identified by a SNP will depend on its strength of association with the outcome (which is itself a function of the SNP-exposure association and the SNP-outcome standard error), rather than being equal to a constant value β. It could instead be the case that a SNP actually increases the exposure for one group of individuals and decreases it for another, which would be a violation of the monotonicity assumption [5]. Another potential source of heterogeneity that has received a lot of attention in the literature in recent years is that some or all of the SNPs could in fact exert a direct effect on the outcome not through the exposure [6] by violating IV2, IV3 or both, which is termed 'horizontal pleiotropy' [7, 8]. 'Vertical pleiotropy' - in which the effect of a SNP on the exposure of interest is actually mediated through other, earlier exposures, does not pose a problem. For brevity we will refer to problematic horizontal pleiotropy simply as pleiotropy from now on.

The presence of heterogeneity does not necessarily invalidate an MR study. For example if the underlying cause of the heterogeneity is pleiotropy but, across all variants, the amount of pleiotropy is independent of instrument strength (the InSIDE assumption [9]), then a standard random effects meta-analysis will still yield reliable inferences [3, 9]. If the pleiotropy is instead 'directional' (it has a non-zero mean) then a random effects meta-analysis will be biased, but MR-Egger regression [9] can still yield reliable inferences. MR-Egger regression has been used extensively as a

sensitivity analysis tool since its proposal for this reason (see [10, 11, 12, 13, 14] for some recent examples).

Choice of weights in two-sample summary data MR

Two popular choices for the inverse variance weights used to calculate both the IVW estimate in (1) and Cochran's Q in (2) are:

$$\text{1st order weights:} \quad w_j = \frac{\hat{\gamma}_j^2}{\sigma_{Yj}^2} \tag{3}$$

$$\text{2nd order weights:} \quad w_j = \left(\frac{\sigma_{Yj}^2}{\hat{\gamma}_j^2} + \frac{\hat{\Gamma}_j^2 \sigma_{Xj}^2}{\hat{\gamma}_j^4} \right)^{-1} \tag{4}$$

where σ_{Yj}^2 represents the variance of $\hat{\Gamma}_j^2$ and σ_{Xj}^2 represents the variance of $\hat{\gamma}_j^2$. 2nd order weights, which are derived via a Taylor series expansion, attempt to acknowledge uncertainty in both the numerator and denominator of the ratio estimate (note that in the two-sample setting, it is not necessary to include terms in the Taylor series expansion involving the covariance of $\hat{\gamma}_j$ and $\hat{\Gamma}_j$, because they are obtained from separate samples). 1st order weights ignore uncertainty in the denominator of the ratio estimate, which is equivalent to making the NO Measurement Error (NOME) assumption ($\hat{\gamma}_j = \gamma_j$) as defined in [15] within the context of a two-sample MR analysis. The same simplifying approximation has also been used extensively for general IV analyses in economics [16, 17]. In practice when SNP-exposure association estimates are very precise, for example when they are derived from a study with a large sample size, the two weighting schemes can be very similar. Unfortunately, this is often not the case.

Further remarks on Cochran's Q

Provided that the two samples used in the analysis are homogeneous, and the SNPs used as IVs are mutually independent, the IVW estimate obtained using 1st order weights is asymptotically equivalent to the two-stage least squares (TSLS) estimate for the causal effect obtained using individual level data on G, X and Y from either sample, if such data were available (see for example Section 2.2 in [18]). Simulations conducted by Del Greco et al [6] and methodological work by Windmeijer [19] have also clarified the fact that Cochran's Q statistic for detecting heterogeneity amongst summary data estimates is asymptotically equivalent to the Sargan's test for over-identification with individual level data. A further fact noted by Windmeijer is that both Cochran's Q and the Sargan test statistic are minimised at the IVW and TSLS estimates respectively. We can see this for Cochran's Q by noting that $\hat{\beta}_{IVW}$ in (1) satisfies:

$$\frac{\partial Q}{\partial \beta}(\beta = \hat{\beta}_{IVW}) = 0.$$

This expression presents Cochran's Q as an estimating equation for the causal parameter β. Given that 2nd order weights provide an ostensibly more accurate reflection of the variance of the ratio estimate, it would seem obvious that they should be used as standard within an MR study to calculate the IVW estimate and Cochran's Q. However, Thompson et al [20] showed that 2nd order weights produce causal estimates which are generally more biased than using 1st order weights. The reason for this apparent paradox is that 2nd order weights can be highly correlated with the ratio estimates themselves. Strict independence is required between the w_j and $\hat{\beta}_j$ terms in (1) in order for the IVW estimate to be consistent.

In this paper we derive a simple approximation to the variance of the ratio estimate that circumvents the deficiencies of 1st and 2nd order weighting. We show by Monte Carlo simulation that it can dramatically improve the reliability of Cochran's Q statistic for detecting heterogeneity amongst a set of ratio estimates. We then go on to demonstrate how our modified weights can also improve the assessment of heterogeneity about the MR-Egger estimate in the presence of directional pleiotropy, when using a generalized heterogeneity measure known as Rücker's $Q^{'}$ [3, 21]. In the Results section we apply our improved heterogeneity statistics to a two-sample summary data MR study to determine the causal effect of systolic blood pressure on coronary heart disease originally published by Ference et al [22], with recent application of IVW and MR-Egger approaches in Lawlor et al[13].

Materials and Methods

We start by motivating the derivation of the IVW estimate using 1st and 2nd order weights. We assume the basic underlying model generating the observed SNP-outcome association estimates:

$$\text{True model:} \quad \hat{\Gamma}_j = \beta\gamma_j + \sigma_{Yj}\epsilon_j, \quad \text{var}(\epsilon_j) = 1 \tag{5}$$

Note that model (5) assumes no heterogeneity (e.g. due to pleiotropy), and is a function of the true underlying SNP-exposure association γ_j. In practice, when fitting this model we must work with the SNP-exposure association estimate $\hat{\gamma}_j$ (with variance σ^2_{Xj}) instead. Substituting $\hat{\gamma}_j$ into (5) in place of γ_j therefore yields:

$$\text{Fitted model:} \quad \hat{\Gamma}_j = \beta\hat{\gamma}_j + \sqrt{\beta^2\sigma^2_{Xj} + \sigma^2_{Yj}}\epsilon_j \tag{6}$$

We can derive an expression for the ratio estimate $\hat{\beta}_j$ and its variance that is consistent with 2nd order weighting, by replacing β with $\hat{\Gamma}_j/\hat{\gamma}_j$ in equation (6), and by dividing through by $\hat{\gamma}_j$ to give:

$$\hat{\beta}_j = \beta + \sqrt{\frac{\hat{\Gamma}^2_j}{\hat{\gamma}^4_j}\sigma^2_{Xj} + \frac{\sigma^2_{Yj}}{\hat{\gamma}^2_j}}\epsilon_j. \tag{7}$$

Setting σ^2_{Xj} in formula (7) to zero yields an expression for the ratio estimate $\hat{\beta}_j$ and its variance that is consistent with 1st order weighting.

Modified 2nd order weights

Replacing β with $\hat{\Gamma}_j/\hat{\gamma}_j$ in equation (6), as suggested by 2nd order weighting, means that the variance of each ratio estimate will be a function of the ratio estimate itself. It is easy to see that this will induce a negative bias in the IVW estimate because whenever $\hat{\beta}_j$ is randomly large, its contribution to (1) will be down-weighted (likewise its contribution to (1) will be up-weighted when $\hat{\beta}_j$ is randomly small). This negative bias will also effect Cochran's Q statistic. This problem can be avoided when using 1st order weights by setting σ^2_{Xj} to zero, but the obvious downside is that the variance of each $\hat{\beta}_j$ is then under-estimated. We therefore suggest the following scheme to address both shortcomings, by plugging in an overall estimate for β in model (6) instead. The procedure for calculating the weights is as follows:

1. Use 1st order weights and formula (1) to derive the IVW estimate, $\hat{\beta}_{IVW}$;
2. Calculate the modified 2nd order weights via the formula:

$$w_j(\hat{\beta}_{IVW}) = \left(\frac{\sigma^2_{Yj} + \hat{\beta}^2_{IVW}\sigma^2_{Xj}}{\hat{\gamma}^2_j} \right)^{-1} \tag{8}$$

   where $\hat{\beta}_{IVW}$ is obtained from step 1.

The modified 2nd order weights can then be used to re-calculate $\hat{\beta}_{IVW}$ to (a) give an overall measure of causal effect, and (b) to evaluate Cochran's Q statistic to look for the presence of heterogeneity. If performed recursively, the above procedure would find the value of β satisfying:

$$\frac{\partial Q_m(\beta)}{\partial \beta} = 0, \quad \text{for} \quad Q_m(\beta) = \sum_{j=1}^{L} w_j(\beta)(\hat{\beta}_j - \beta)^2, \tag{9}$$

and where $w_j(\beta)$ is taken from formula (8). A brief investigation revealed that our simple procedure is essentially idempotent after one iteration, and therefore additional steps are unnecessary.

IVW and Cochran's Q analysis when no pleiotropy present

Naturally we would like Cochran's Q to follow a χ^2_{L-1} distribution as closely as possible when no heterogeneity is present, to guard against the erroneous detection of pleiotropy. We now assess the performance of all three weighting schemes to achieve this aim via simulation. Two-sample summary data MR studies comprising L=25 SNP-exposure and SNP outcome association estimates $(\hat{\Gamma}_j, \hat{\gamma}_j)$ were generated from the following normal models:

$$\hat{\gamma}_j \sim N(\gamma_j, \sigma^2_{Xj}), \quad \hat{\Gamma}_j \sim N(\beta\gamma_j, \sigma^2_{Yj}) \tag{10}$$

Mean	1st order w_j			2nd order w_j			Modified 2nd order w_j		
F	Q	T1E(Q)	$\hat{\beta}_{IVW}$ (S.E.)	Q	T1E(Q)	$\hat{\beta}_{IVW}$ (S.E.)	Q	T1E(Q)	$\hat{\beta}_{IVW}$ (S.E.)
				No heterogeneity, β=0					
100	23.9	0.05	0.00 (0.011)	22.8	0.028	0.00 (0.011)	23.9	0.05	0.00 (0.011)
61	23.9	0.046	0.00 (0.011)	21.8	0.016	0.00 (0.011)	23.9	0.046	0.00 (0.011)
40	24.0	0.051	0.00 (0.011)	20.4	0.006	0.00 (0.011)	24.0	0.050	0.00 (0.011)
25	24.0	0.049	0.00 (0.011)	17.6	0.001	0.00 (0.01)	23.9	0.048	0.00 (0.011)
10	24.0	0.049	0.00 (0.009)	12.3	0.000	0.00 (0.008)	23.7	0.045	0.00 (0.009)
				No heterogeneity, β=0.05					
100	24.1	0.048	0.049 (0.011)	22.8	0.023	0.048 (0.011)	23.9	0.044	0.049 (0.011)
61	24.5	0.057	0.049 (0.012)	21.9	0.016	0.047 (0.011)	24.1	0.049	0.049 (0.012)
40	24.8	0.063	0.048 (0.012)	20.3	0.007	0.045 (0.011)	23.9	0.046	0.048 (0.012)
25	25.9	0.092	0.046 (0.012)	17.9	0.002	0.041 (0.011)	24.1	0.054	0.046 (0.012)
10	31.5	0.278	0.033 (0.012)	13.5	0.00	0.027 (0.011)	26.1	0.114	0.034 (0.012)
				No heterogeneity, β=0.1					
100	24.8	0.065	0.099 (0.012)	22.9	0.028	0.097 (0.012)	24.0	0.048	0.099 (0.012)
61	25.7	0.086	0.098 (0.012)	21.9	0.019	0.095 (0.012)	24.0	0.052	0.098 (0.012)
40	27.2	0.13	0.096 (0.013)	20.4	0.009	0.091 (0.012)	24.0	0.050	0.096 (0.013)
25	31.7	0.277	0.092 (0.014)	18.3	0.003	0.083 (0.014)	24.5	0.062	0.092 (0.014)
10	53.8	0.788	0.065 (0.018)	15.8	0.004	0.055 (0.015)	30.1	0.228	0.070 (0.017)

Table 1: *Mean Q statistic and IVW estimate $\hat{\beta}_{IVW}$ calculated 1st order, 2nd order and modified 2nd order weights. Results calculated over 10,000 simulated data sets. Type I error rate (T1E(Q)) refers to the proportion of times Q is greater than the upper 95th percentile of a χ^2_{24} distribution.*

given parameter values for $(\gamma_j, \sigma^2_{Xj}, \sigma^2_{Yj})$ and β. Data generated under Model (10) furnishes a set of ratio estimates between which **no** heterogeneity should exist asymptotically, as the variance of $\hat{\gamma}_j$, σ^2_{Xj}, reduces to zero. To highlight this the magnitude of σ^2_{Xj} was varied in order to mimic a range of MR studies, from very small (with mean F-statistic 10), to very large (with mean F-statistic 100). Note in this scenario the F-statistic for SNP j can be approximated by the squared t-statistic $\hat{\gamma}^2_j/\sigma^2_{Xj}$.

Table 1 (columns 2-9) show the mean Q statistic, IVW estimate and the probability of the Q statistic detecting heterogeneity at the 5% significance level (the type I error rate), when using 1st and 2nd order weights. Five different mean F-statistic values were considered for β=0 (no causal effect), β=0.05 and β=0.1, giving 15 scenarios in total.

From equation (4) we can see that 1st order weights are valid when σ^2_{Xj}=0, which would imply that its F-statistic were infinite. Whenever σ^2_{Xj} is non-negligible, 1st order weights under-estimate the true variability amongst the ratio estimates. One might therefore suspect that their associated Q statistics would be too large on average (i.e. positively biased beyond their expected value of 24). This would inflate the type I error rate for detecting pleiotropy beyond nominal levels. From Table 1 we can

see that this is indeed the case when β is non-zero, when $\beta = 0$ the correct behaviour is observed. One might also suspect that 2nd order weights would naturally eliminate this undesirable property, furnishing a Q statistic following the required distribution. From Table 1 we can see that this is, unfortunately, not the case: 2nd order weighting appears to over-correct the Q statistic so that it is negatively biased, thereby removing *any* ability to detect heterogeneity at all.

Table 1 columns 10-13 shows the performance of modified 2nd order weights with respect to estimation of β and Q. When β is non-zero, modified 2nd order weights are much more effective at preserving the type I error rate of the Q statistic at its nominal level, unless the mean F-statistic is very low (indicating weak instruments). Figure 2 (left and right) shows the distribution of Q statistics under each weighting scheme for β=0.1 when the mean F-statistic is 100 and 10, to illustrate this point.

The IVW estimate is known to suffer from regression dilution bias towards zero by an amount approximately proportional to the inverse of the mean F-statistic. For example, if the mean F-statistic is 100 and the true causal effect is 0.1, the expected dilution would be 1% (giving an expected estimate of 0.099). In these simulations, 2nd order weighting yields estimates with the largest dilution, and modified 2nd order weights yield estimates with the smallest dilution. This dilution can be mitigated to a large degree by applying a bias correction. For example Bowden et al [15] proposed the use of SIMulation EXtrapolation (SIMEX) [23] to account for regression dilution bias in MR studies, and applied it to the IVW estimate in [3]. For brevity, we do not additionally assess SIMEX correction here.

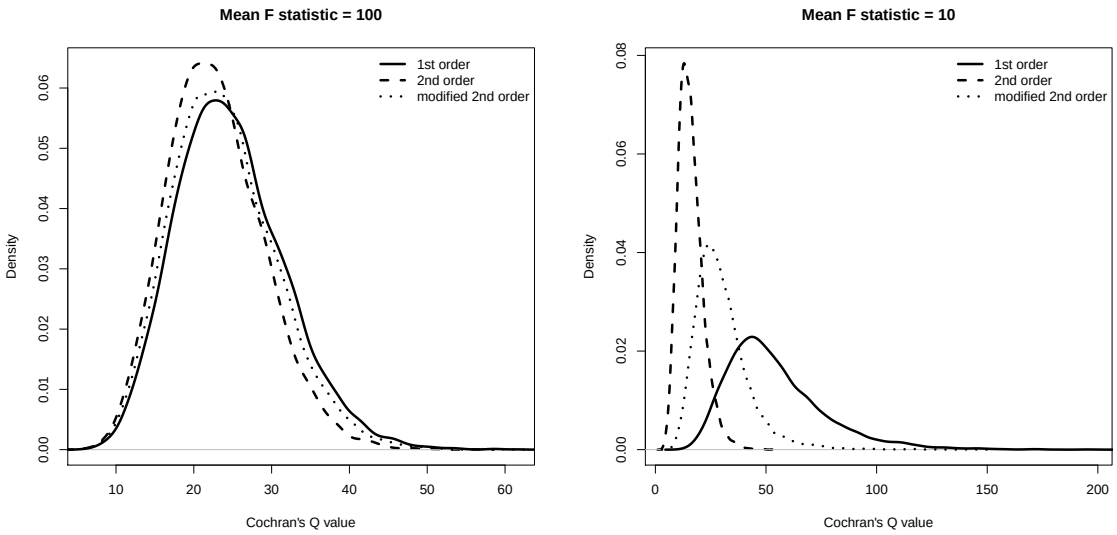

Figure 2: *Distribution of Q statistics using 1st order, 2nd order and modified second weights in simulation scenario 1 (left) and scenario 5 (right) of Table (1) respectively.*

IVW and Cochran Q analysis in the presence of heterogeneity

Let α_j represent the pleiotropic effect of SNP j on the outcome not via the exposure and let μ_α and σ_α^2 denote the sample mean and variance, respectively, of these L pleiotropic effects. Suppose that the pleiotropic effects collectively satisfy the InSIDE assumption, and that the mean pleiotropic effect $\mu_\alpha = 0$. This is referred to as 'balanced' pleiotropy. Two common frameworks to account for heterogeneity are the additive random effects model [24] and the multiplicative random effects model [25, 26]:

$$\text{Additive pleiotropy model: } \hat{\Gamma}_j = \mu_\alpha + \beta\gamma_j + \sqrt{\sigma_\alpha^2 + \sigma_{Yj}^2}\epsilon_j \tag{11}$$

$$\text{Multiplicative pleiotropy model: } \hat{\Gamma}_j = \mu_\alpha + \beta\gamma_j + \sqrt{1 + \sigma_\alpha^2}\sigma_{Yj}\epsilon_j, \tag{12}$$

When balanced pleiotropy is assumed, μ_α is constrained to zero in both (11) and (12). The multiplicative model is far less intuitive than the additive model, but it has the following attractive property. if heterogeneity is detected, point estimates for β remain unchanged (i.e. they are the same as those obtained from a fixed effect model), but their standard errors are scaled up accordingly. Additive random effects models yield different point estimates and standard errors compared to fixed effect models, and are known to be more prone to bias in the presence of model misspecification. For a more lengthly discussion see [3].

In Table 1 the type I error rate of Cochran's Q statistic to detect heterogeneity using 2nd order weights was below its nominal level. This is detrimental if it translates into a low statistical power to detect heterogeneity when it *is* truly present. Figure 3 (left) shows the power of Cochran's Q to detect heterogeneity as a function of all three weighting schemes when data are simulated under additive random effects model (11) with balanced pleiotropy for increasing values of σ_α. We see that the power of Cochran's Q to detect heterogeneity approaches 100% for all weighting schemes as σ_α increases. However, the power of 2nd order weighting is considerably lower than either the 1st order or modified 2nd order weights with more moderate heterogeneity. Results for data simulated under the multiplicative pleiotropy model were highly similar (data available from the authors on request).

Comparison with the work of Thompson et al.

In related work, Thompson et al [20] also noted the poor performance of 2nd order weights when estimating the causal effect via the IVW approach in two-sample summary data MR. They also proposed two methods for improving the performance of 1st order weights. Firstly, they derived a bias reduced estimate for each ratio estimate, by multiplying it by a function of its F statistic. Secondly, they noted that a more precise estimate for the association between SNP j and the exposure could be derived by combining the original estimate $\hat{\gamma}_j$ with a second estimate $\hat{\Gamma}_j/\hat{\beta}_{IVW}$ via an inverse variance weighted average. We conducted an investigation into the effectiveness of these two procedures, a full description of which is given in the Appendix. In

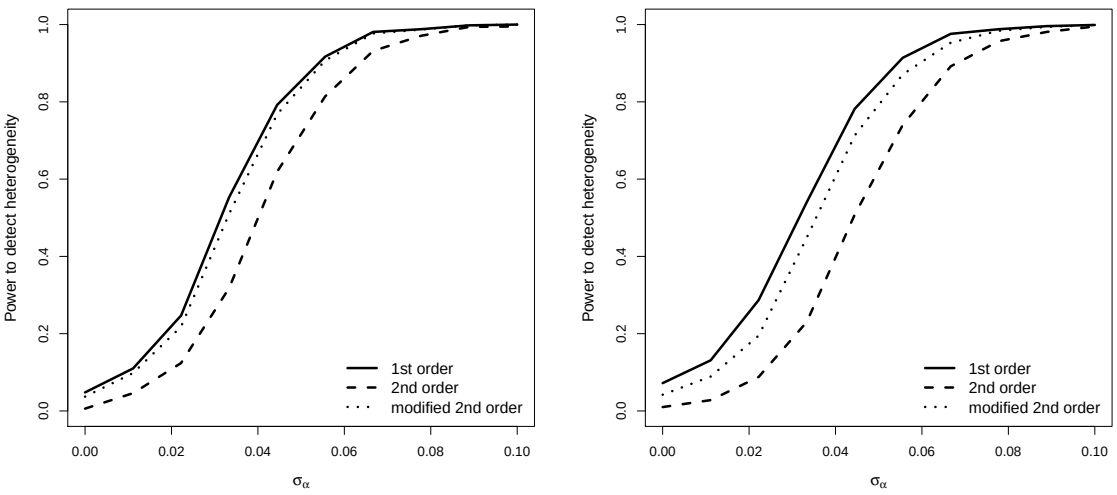

Figure 3: *Left: Power of Cochran's Q statistic to detect heterogeneity as a function of σ_α using 1st order, 2nd order and modified 2nd weights under a balanced pleiotropy model. Right: Power of Rücker's Q statistic to detect heterogeneity as a function of σ_α^2 using 1st order, 2nd order and modified 2nd order weights under a directional pleiotropy model.*

summary, Thompson et al's improved precision formula is generally effective for IVW analyses: It produces Q statistics and associated p-values that are conservative, but the level of conservatism is far less severe than 2nd order weighting. However, the addition of Thompson et al's bias correction formula leads to a dramatic instability in the distribution of Cochran's Q statistic - its mean value is highly unstable unless the mean F statistic is very large. Due to this apparent limitation we do not consider this procedure further, although it is an interesting avenue for future research.

MR-Egger regression and Rücker $Q^{'}$ analysis

When Cochran's Q statistic detects significant amounts of heterogeneity, it is prudent to test whether it is meaningfully biasing the analysis. This would indeed be the case if the heterogeneity were caused by directional pleiotropy with a non-zero mean. Equations (11) and (12) can still serve as underlying models for the data by allowing the parameter space of μ_α to be unconstrained, and making use of MR-Egger regression [9, 3]. This approach simply regresses the SNP-outcome association estimates on the SNP-exposure association estimates, where the SNPs are coded to make each $\hat{\gamma}_j$ positive. Traditionally, it is initially fitted as a fixed model using 1st order weighting [9], which for the purposes of this paper, is equivalent to:

$$\hat{\beta}_j = \frac{\beta_{0E}}{\hat{\gamma}_j} + \beta_{1E} + w_j^{-\frac{1}{2}}\epsilon_j. \tag{13}$$

where w_j are taken from (3). Model (13) assumes that there is no heterogeneity about the MR-Egger fit after the addition of an intercept, so that σ_α^2 is implicitly set to zero. The observed heterogeneity around the MR-Egger fit can then be quantified using Rücker's $Q^{'}$ statistic: [21, 3]:

$$Q^{'} = \sum_{j=1}^{L} Q_j^{'} = \sum_{j=1}^{L} w_j(\hat{\beta}_j - \frac{\hat{\beta}_{0E}}{\hat{\gamma}_j} - \hat{\beta}_{1E})^2, \tag{14}$$

where, again, w_j are taken from (3). If $Q^{'}$ is larger than L-2, its expected value under the assumption of no heterogeneity, standard errors of the MR-Egger intercept and slope estimates are inflated by factor of $\sqrt{1+\hat{\sigma}_\alpha^2}$, where $\hat{\sigma}_\alpha^2$ is estimated directly from $Q^{'}$. This is consistent with assuming a multiplicative random effects model. See Bowden et al [3] for further details as well as a description of how MR-Egger regression can be fitted within an additive random effects framework.

In truth, it highly implausible to think that non-zero mean pleiotropy could exist, but have a zero variance, and therefore that the null χ^2_{L-2} distribution of $Q^{'}$ will ever hold. However, this does not mean that $Q^{'}$ is useless: we can still use it to ascertain the amount that variant j contributes to the overall heterogeneity via $Q_j^{'}$, and assess the relative goodness-of-fit of MR-Egger over the IVW model (e.g. via $Q_R = Q^{'}/Q$ as defined in [3]).

2nd order and modified 2nd order weights for MR-Egger

To fit MR-Egger regression and evaluate Rücker's $Q^{'}$ statistic using 2nd order weights, we simply replace w_j in equation's (13) and (14) with weights defined in equation (4). Modified 2nd order weights for MR-Egger regression can also be substituted into these formulae, after being derived in the following manner:

1. Use 1st order weights and formula (13) to derive the MR-Egger slope and intercept estimates, $(\hat{\beta}_{0E}, \hat{\beta}_{1E})$;

2. Calculate modified 2nd order weights via the formula:

$$w_j(\hat{\beta}_{1E}) = \left(\frac{\sigma_{Yj}^2 + \hat{\beta}_{1E}^2 \sigma_{Xj}^2}{\hat{\gamma}_j^2} \right)^{-1} \tag{15}$$

   where $\hat{\beta}_{1E}$ is obtained from step 1.

The modified weights can then be used to re-calculate $(\hat{\beta}_{0E}, \hat{\beta}_{1E})$ and Rücker's $Q^{'}$ statistic. As for the IVW case previously, recursive application of the above rule would find the value of $\boldsymbol{\beta}_E = (\beta_{0E}, \beta_{1E})$ satisfying:

$$\frac{\partial Q_m^{'}}{\partial \boldsymbol{\beta_E}} = (0,0), \quad \text{for} \quad Q_m^{'} = \sum_{j=1}^{L} w_j(\beta_{1E})(\hat{\beta}_j - \beta_{0E}/\hat{\gamma}_j - \beta_{1E})^2, \tag{16}$$

and where $w_j(\beta_{1E})$ is taken from formula (15). As before, only one iteration of this procedure is sufficient. In the Appendix we provide `R` code to fit MR-Egger regression using 1st order, 2nd order and modified 2nd order weights.

Performance of modified weights

In order to test the reliability of each weighting scheme for MR-Egger, we simulate two-sample summary data in the following manner:

$$\hat{\gamma}_j \sim N(\gamma_j, \sigma^2_{Xj}), \quad \alpha_j \sim N(\mu_\alpha, \sigma^2_\alpha), \quad \hat{\Gamma}_j \sim N(\beta\gamma_j + \alpha_j, \sigma^2_{Yj}) \tag{17}$$

and evaluate the performance of $\hat{\beta}_{1E}$ and Q' just as for $\hat{\beta}_{IVW}$ and Q in Table 1, with one notable change. Whereas the data in Table 1 were generated given parameter values for $(\gamma_j, \sigma^2_{Xj}, \sigma^2_{Yj})$ that induced a range of mean F-statistic values in the data, we now choose parameters that induce a range of I^2_{GX} values instead. It is this statistic, first defined in [15], which both encapsulates instrument strength for MR-Egger and indicates the likely dilution towards 0 when 1st order weights are used. Put simply, I^2_{GX} quantifies the proportion of the observed variation amongst the SNP-exposure association estimates (the $\hat{\gamma}_j$s) that is due to true differences in their underlying parameter values (the γ_js). An I^2_{GX} value of 90% indicates an expected dilution of 10% in the MR-Egger causal effect estimate. SIMEX can be used to adjust MR-Egger regression estimates for regression dilution, regardless of the weighting used.

Data were simulated to give I^2_{GX} values ranging from 99% to 82%. For the same data, mean F-statistic values ranged from roughly 600 (when $I^2_{GX} = 99\%$) to 60 (when $I^2_{GX} = 82\%$). This highlights the fact that MR-Egger regression is far less efficient than an IVW analysis (as discussed at length in [3, 9, 27] and elsewhere) and requires instruments that are both strong as well as having variable magnitudes of effect on exposure in order to function effectively.

No heterogeneity around the MR-Egger fit

Table 2 shows the results when the mean pleiotropic effect, μ_α equals 0.1 and the pleiotropy variance σ^2_α is set to zero. In this case, MR-Egger regression should perfectly adjust for the directional pleiotropy and no residual pleiotropy should remain. Rücker's Q' should therefore not detect heterogeneity beyond nominal levels. We again see that first order weights can inflate the chances of erroneously detecting heterogeneity, with the type I error increasing as the magnitude of β increases, especially when I^2_{GX} is low. Likewise, 2nd order weights are always too conservative, failing to detect heterogeneity at the nominal level of the test, especially when β is small. Modified 2nd order weights appear to offer a good compromise, with good behaviour across all settings. In terms of point estimation, all methods are unbiased under the causal null, and I^2_{GX} gives a good prediction as to the magnitude of dilution when the causal effect is non-zero. Unlike in the previous simulation, the dilution of the MR-Egger estimate is at its least when 2nd order weights are used. For illustration

Figure 4 shows the distribution of $Q^{'}$ statistics using the three weighting schemes when $I^2_{GX} = 97\%$ (left) and $I^2_{GX} = 82\%$ (right). As I^2_{GX} increases towards 100%, their distributions converge on the truth, but can be very different for smaller I^2_{GX} values.

Mean	1st order w_j			2nd order w_j			Modified 2nd order w_j		
I^2_{GX}	$Q^{'}$	T1E($Q^{'}$)	$\hat{\beta}_{1E}$ (S.E.)	$Q^{'}$	T1E($Q^{'}$)	$\hat{\beta}_{1E}$ (S.E.)	$Q^{'}$	T1E($Q^{'}$)	$\hat{\beta}_{1E}$ (S.E.)
				σ^2_α=0, μ_α = 0.1, β=0					
99	23.0	0.048	0.00 (0.012)	22.1	0.034	0.00 (0.012)	22.9	0.048	0.00 (0.012)
98	22.9	0.048	0.00 (0.012)	21.4	0.026	0.00 (0.012)	22.9	0.048	0.00 (0.012)
97	23.1	0.051	0.00 (0.012)	20.6	0.020	0.00 (0.012)	23.0	0.051	0.00 (0.012)
95	23.0	0.051	0.00 (0.012)	18.8	0.008	0.00 (0.013)	22.9	0.049	0.00 (0.012)
82	23.0	0.051	0.00 (0.010)	14.3	0.001	0.00 (0.013)	22.7	0.046	0.00 (0.010)
				σ^2_α=0, μ_α = 0.1, β=0.05					
99	23.2	0.053	0.05 (0.012)	21.7	0.031	0.051 (0.012)	23.0	0.050	0.05 (0.012)
98	23.4	0.058	0.049 (0.012)	20.8	0.022	0.051 (0.013)	23.0	0.051	0.049 (0.012)
97	23.7	0.063	0.048 (0.012)	19.5	0.011	0.051 (0.013)	22.9	0.049	0.048 (0.012)
95	24.7	0.090	0.046 (0.013)	17.6	0.005	0.050 (0.014)	23.1	0.054	0.046 (0.012)
82	29.7	0.254	0.032 (0.013)	13.3	0.000	0.041 (0.016)	24.8	0.100	0.033 (0.013)
				σ^2_α=0, μ_α = 0.1, β=0.1					
99	23.7	0.056	0.099 (0.013)	21.3	0.021	0.100 (0.013)	22.9	0.042	0.099 (0.013)
98	24.5	0.081	0.098 (0.013)	20.4	0.017	0.100 (0.013)	23.1	0.047	0.098 (0.013)
97	25.9	0.119	0.096 (0.013)	19.0	0.009	0.100 (0.014)	23.0	0.050	0.096 (0.013)
95	29.9	0.260	0.091 (0.015)	17.2	0.005	0.098 (0.016)	23.5	0.062	0.092 (0.014)
83	49.9	0.765	0.065 (0.019)	14.2	0.004	0.080 (0.022)	28.7	0.221	0.069 (0.018)

Table 2: *Mean $Q^{'}$ statistic and MR-Egger estimate $\hat{\beta}_{1E}$ calculated 1st order, 2nd order and modified 2nd order weights. Results calculated over 10,000 simulated data sets. Type I error rate (T1E($Q^{'}$)) refers to the proportion of times $Q^{'}$ is greater than the upper 95th percentile of a χ^2_{23} distribution.*

Heterogeneity around the MR-Egger fit

Figure 3 (right) shows the power of Rücker's Q to detect heterogeneity as a function of all three weighting schemes when data are simulated with mean I^2_{GX}=98% and β=0.1, but the pleiotropy variance σ^2_α is allowed to be non-zero. As before, second order weighting is seen to result in the lowest power, but the power loss is less dramatic than in the IVW case.

Results

Figure 5 (left) shows a scatter plot of summary data estimates for the associations of 26 genetic variants with systolic blood pressure (SBP, the exposure) and coronary heart disease (CHD, the outcome). SNP-exposure association estimates were obtained from the International Consortium for Blood Pressure consortium (ICBP) [28]. SNP-CHD association odds ratios were collected from Coronary ARtery Disease Genome-Wide

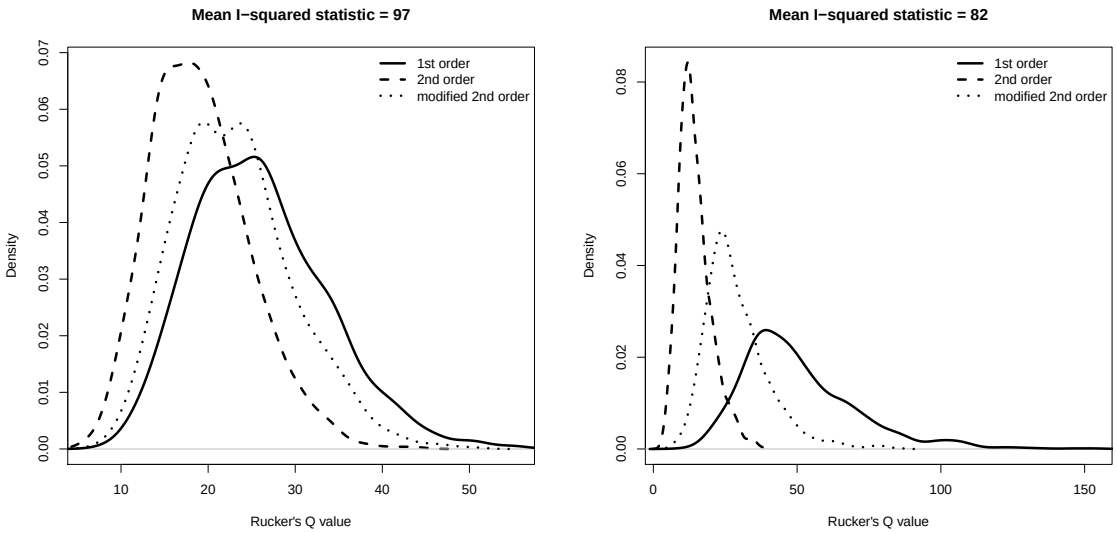

Figure 4: *Distribution of $Q^{'}$ statistics using 1st order, 2nd order and modified 2nd order weights in simulation scenario 3 (left) and scenario 5 (right) of Table (2) respectively.*

Replication And Meta-Analysis (CARDIoGRAM) consortium [29]. These data have previously been used in a two-sample summary data MR analysis by Ference et al [22] and Lawlor et al [13], but we extend their original analysis here in by applying our modified 2nd order weights and conducting a more in depth inspection of each variants contribution to the overall heterogeneity. The mean F statistic for these data is 61, and the I^2_{GX} statistic is 84%.

The ratio estimate for any individual variant is the slope joining its data point to the origin. Using 1st order weights the IVW estimate for these data, which represents the causal effect of a 1mmHg increase in SBP on the log-odds ratio of CHD is 0.053. This is shown as the slope of a solid black line passing through the origin (note: the origin is not visible in either plot because of a truncated x-axis). The MR-Egger regression causal estimate obtained using 1st order weights is very close to zero (-0.002) due to the detection of positive directional pleiotropy. If the causal effect were larger, it would be tempting to apply a SIMEX correction to the MR-Egger estimate given I^2_{GX} predicts an expected dilution of 16% in its estimate. However, this would have little impact in absolute terms because the estimate itself is so close to zero.

Table 3 shows the results of applying the IVW and MR-Egger regression approaches to the data using all three weighting schemes. Point estimates and standard errors are in good agreement across the different weights. All three schemes detect significant heterogeneity about the IVW and MR-Egger fits (as quantified by Q and $Q^{'}$ respectively). As expected, the observed heterogeneity is largest when using 1st order weights, smallest when using 2nd order weights, and in between the two when using modified 2nd order weights.

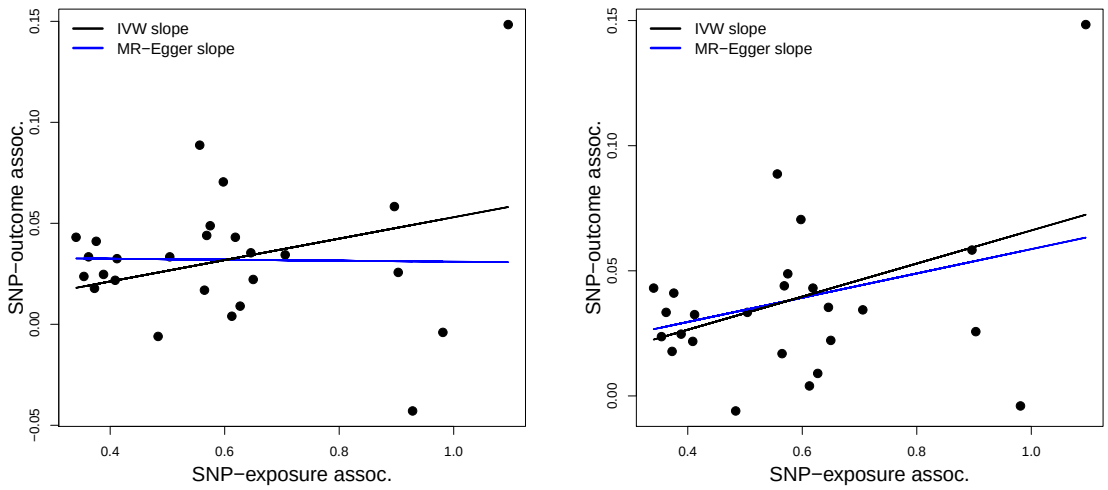

Figure 5: *Scatter plots of SNP-outcome associations* $\hat{\Gamma}_j$ *versus SNP-exposure associations* $\hat{\gamma}_j$*. IVW slope shown as a black line, MR-Egger slope shown as a blue line. 1st order weights used. Left: Full data. Right: with SNP rs17249754 removed.*

When heterogeneity is detected by the IVW model, and is only partially mitigated by applying MR-Egger regression, it is important to investigate whether the heterogeneity is contributed to by all SNPs, or if a small number of SNPs are responsible. Under their respective null hypotheses, Q and Q' should follow χ^2_{L-1} and χ^2_{L-2} distributions. Following on from this, each component of Q and Q' (Q_j and Q'_j) can be loosely approximated by χ^2_1 distribution. Both statements of course require the specification of 'correct' weights. Figure 6 (top-left and top-right) shows Q_j and Q'_j under each weighting scheme. Horizontal lines have been drawn to indicate the location of the 5th, 1st and 0.19th percentile of a χ^2_1 in order to help assess the magnitude of the contributions. The 0.19th percentile is derived as a 0.05 threshold adjusted for multiple testing using the Bonferroni correction. From Figure 6, we see that the eighth SNP in our list (rs17249754) is responsible for the vast majority of the excess heterogeneity in both Q and Q'. Q_8 ranges from approximately 24.5 to 28 and Q'_8 ranges from approximately 16 to 19, depending on weighting. Variant rs17249754 sits in the ATPase plasma membrane Ca2+ transporting 1 (*ATP2B1*) gene, which is involved in intracellular calcium homeostasis, and is strongly associated with higher SBP. However, in the CARDIoGRAM consortium it is associated with reduced CHD. It could be that rs17249754 truly increases SBP in the ICBP population but decreases it in CARDIoGRAM, which would be a violation of the monotonicity assumption. Alternatively, rs17249754 could be exerting a pleiotropic effect on CHD not through SBP in a consistent manner for both the ICBP and CARDIoGRAM study populations, which is then reflected in the CARDIoGRAM estimate. As previously discussed, incorporating odds ratios into an MR analysis can lead to heterogeneity amongst causal estimates due to non-collapsibility. However, this could only ever do so by shrinking estimates towards zero, not changing their sign. We can therefore rule out this explanation here.

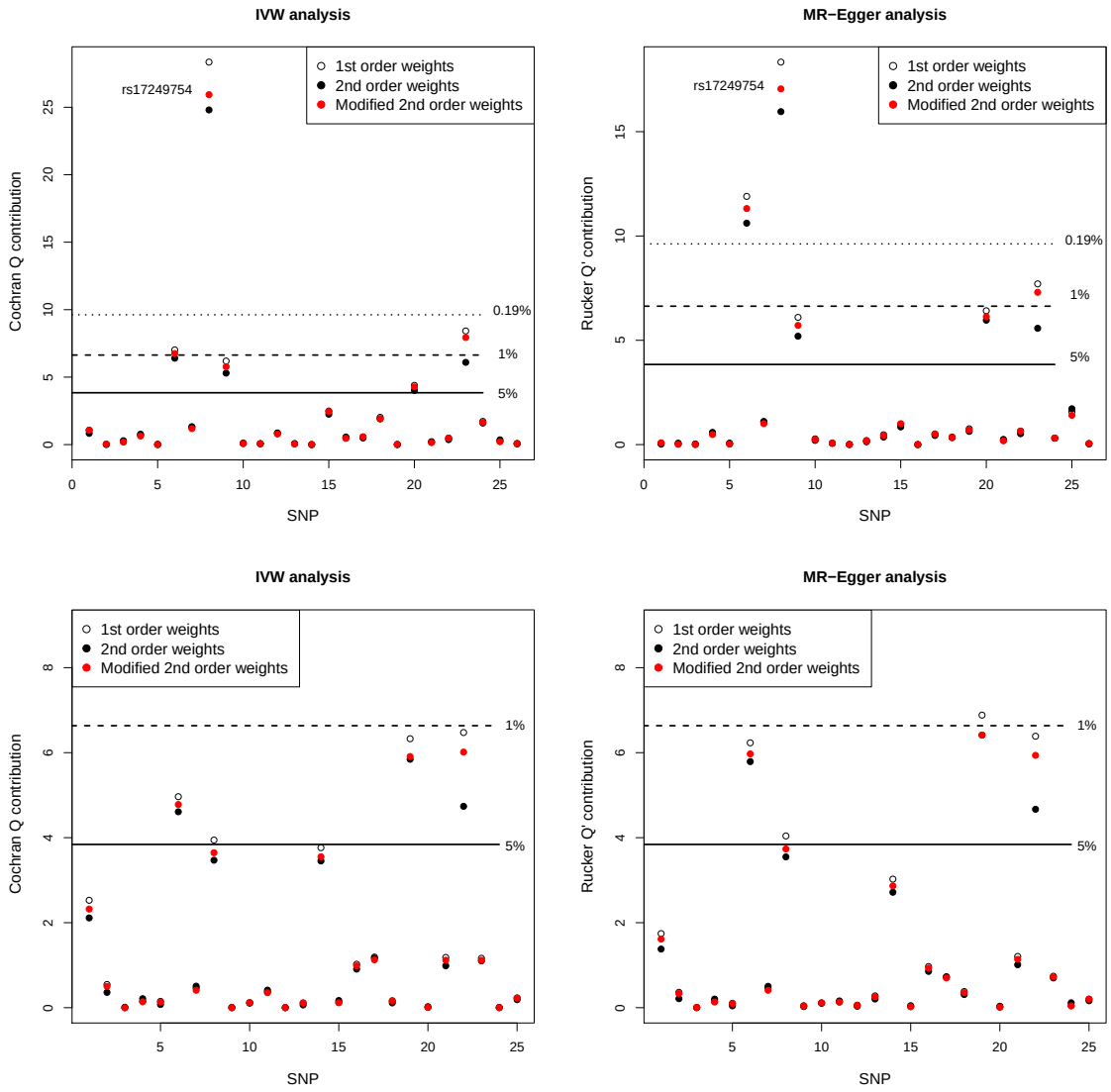

Figure 6: *Contribution to Cochran's Q statistic and Rücker's Q' statistic for the complete data (top-left and top-right) and with SNP rs17249754 removed (bottom-left and bottom-right).*

Method (weights)	Estimate	S.E.	P-value	Het. Stat (p)
Complete data				
Causal estimate				
IVW (1st)	$\hat{\beta}_{IVW}$: 0.053	0.010	3.01×10^{-5}	$Q = 67.1$ (1.03×10^{-5})
MR-Egger (1st)	$\hat{\beta}_{1E}$: -0.002	0.031	0.94	$Q' = 58.6$ (1.00×10^{-4})
IVW (2nd)	$\hat{\beta}_{IVW}$: 0.052	0.010	4.54×10^{-5}	$Q = 58.8$ (1.54×10^{-4})
MR-Egger (2nd)	$\hat{\beta}_{1E}$: -0.004	0.030	0.88	$Q' = 51.1$ (1.03×10^{-3})
IVW (Mod 2nd)	$\hat{\beta}_{IVW}$: 0.054	0.010	2.40×10^{-5}	$Q = 62.8$ (4.31×10^{-5})
MR-Egger (Mod 2nd)	$\hat{\beta}_{1E}$: -0.000	0.031	1.00	$Q' = 55.2$ (2.92×10^{-4})
MR-Egger intercept				
MR-Egger (1st)	$\hat{\beta}_{0E}$: 0.033	0.018	0.075	-
MR-Egger (2nd)	$\hat{\beta}_{0E}$: 0.033	0.017	0.069	-
MR-Egger (Mod 2nd)	$\hat{\beta}_{0E}$: 0.032	0.018	0.083	-
Weighted median (mod 2nd)				
Weighted Median	$\hat{\beta}_{WM}$: 0.064	0.010	1.56×10^{-6}	-
SNP rs17249754 removed				
Causal estimate				
IVW (1st)	$\hat{\beta}_{IVW}$: 0.066	0.008	2.63×10^{-8}	$Q = 35.0$ (0.068)
MR-Egger (1st)	$\hat{\beta}_{1E}$: 0.0490	0.028	0.09	$Q' = 34.3$ (0.061)
IVW (2nd)	$\hat{\beta}_{IVW}$: 0.063	0.008	4.06×10^{-8}	$Q = 30.6$ (0.164)
MR-Egger (2nd)	$\hat{\beta}_{1E}$: 0.0447	0.027	0.11	$Q' = 30.0$ (0.151)
IVW (Mod 2nd)	$\hat{\beta}_{IVW}$: 0.066	0.008	2.90×10^{-8}	$Q = 32.8$ (0.108)
MR-Egger (Mod 2nd)	$\hat{\beta}_{1E}$: 0.049	0.028	0.09	$Q' = 32.2$ (0.095)
MR-Egger intercept				
MR-Egger (1st)	$\hat{\beta}_{0E}$: 0.010	0.015	0.51	-
MR-Egger (2nd)	$\hat{\beta}_{0E}$: 0.011	0.015	0.48	-
MR-Egger (Mod 2nd)	$\hat{\beta}_{0E}$: 0.010	0.015	0.52	-
SIMEX adjusted MR-Egger				
MR-Egger (Mod 2nd)	$\hat{\beta}_{0E}$: 0.006	0.018	0.76	-
	$\hat{\beta}_{1E}$: 0.057	0.032	0.09	-
Weighted median (mod 2nd)				
Weighted Median	$\hat{\beta}_{WM}$: 0.065	0.010	1.18×10^{-6}	-

Table 3: *IVW, MR-Egger and Weighted Median analyses of the causal effect of SBP on CHD risk using 1st order, 2nd order and modified 2nd order weights for the complete data (top) and with SNP rs17249754 removed (bottom).* $\hat{\beta}_{IVW}$ *is the IVW estimate.* $\hat{\beta}_{0E}$ *and* $\hat{\beta}_{1E}$ *are the MR-Egger intercept and slope parameter estimates respectively.* $\hat{\beta}_{WM}$ *is the Weighted Median estimate.* Q *equals Cochran's* Q*,* Q' *equals Rücker's* Q'*. SIMEX refers to estimates obtained by the method of simulation extrapolation.*

Since rs17249754 is driving the analysis to a large degree we opt to remove it in a further sensitivity analysis. Figure 5 (right) and Table 3 shows the results. After removal of rs17249754, MR-Egger regression does not detect the presence of substantial directional pleiotropy. Consequently, its slope is in broad agreement with that of the IVW estimate. The agreement is even closer after applying SIMEX correction (results for the modified 2nd order weights only are shown in the last two rows of Table 3). Only borderline evidence of residual heterogeneity in the data remains, with the strongest evidence suggested by 1st order weighting. Figure 6 (bottom-left and bottom-right) shows the updated contributions of each SNP to Q and Q'. If only 1st order weighting were available, it might be tempting to exclude further variants from the analysis. This is appropriately tempered by using the modified 2nd order weights instead.

First order weights are also used as standard by other MR methods that are robust to pleiotropy, for example the Weighted Median estimate [31], $\hat{\beta}_{WM}$, that can consistently estimate the causal effect when up to (but not including) half of the information in the analysis stems from genetic variants that are invalid IVs. Here, for the first time, we calculate this estimate using modified 2nd order weights. These are defined by first calculating $\hat{\beta}_{WM}$ using 1st order weights, and then plugging them into formula (8) in place of $\hat{\beta}_{IVW}$. Table 3 shows the results. Its estimate for the causal effect both with and without rs17249754 is 0.063 with 0.065 respectively. This analysis nicely illustrates a major strength of the weighted median is its robustness to outliers, and why it naturally compliments both the IVW and MR-Egger approaches.

Discussion

In this paper we have demonstrated the limitations of 1st and 2nd order weighting when used for IVW and MR-Egger regression analysis in two-sample summary data Mendelian randomization, and suggested a simple modification to help address the problem. Our simulations show that the new approximation will be most useful when the causal effect is large, or the instruments are relatively 'weak', as measured by the F statistic for IVW and I^2_{GX} for MR-Egger.

Modified 2nd order weights should also prove a more reliable tool for the detection and removal of outliers in a given data set, as apposed to 1st order weights (which may detect too many outliers) and 2nd order weights (that may detect too few). In this paper we used heterogeneity statistics for outlier detection, but many other test statistics calculated from a fitted model such as Cook's distance and studentized residuals can and have been used for this purpose in MR (see for example [30]). No doubt our modified 2nd order weights would improve their performance too.

We demonstrated the use of modified 2nd order weights within the Weighted Median estimator. A closely related pleiotropy-robust regression strategy termed the Mode Based Estimate (MBE) [32] has recently been proposed that can consistenly estimate

the true causal effect when the most common (modal) pleiotropic effect amongst a set of SNPs is zero. 2nd order weights have been shown to be most effective for this estimator so far, but our modified weights may improve its performance further still, which is another topic for further investigation.

Modified 2nd order weights avoid having to make the NOME assumption when fitting MR-Egger regression. This does not mean its estimates are immune to regression dilution bias, although our simulations suggest the bias is smaller compared to 1st order weights. Applying a SIMEX correction will attenuate the dilution further; `R` code is provided in the appendix to implement this. We also provide code to calculate a more general form of the I^2_{GX} statistic, that is specific to the particular weighting scheme used.

Acknowledgements

Jack Bowden is supported by an MRC Methodology Research Fellowship (grant MR/N501906/1). Jack Bowden, Debbie A Lawlor and George Davey Smith all work in a Unit that receives support from the University of Bristol and UK Medical Research Council (MCUU12013/1 and MCUU12013/5). The authors declare no conflict of interest.

References

[1] Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* 2003; **32**:1–22.

[2] Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology* 2013; **37**:658–665.

[3] Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization *Statistics in Medicine* **36**: 1783–1802

[4] Vansteelandt S, Bowden J, Babanezhad M, Goetghebeur E. On instrumental variables estimation of causal odds ratios. *Statistical Science* 2011; **26**:403–422.

[5] Swanson SA, Hernn MA. The challenging interpretation of instrumental variable estimates under monotonicity *International Journal of Epidemiology* 2017 . DOI: https://doi.org/10.1093/ije/dyx038

[6] Del Greco M F, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Statistics in Medicine* 2015; **34**:2926–2940.

[7] Davey Smith G, Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human molecular genetics* 2014, **23**: 89–98.

[8] Stearns FW. One hundred years of pleiotropy: a retrospective. *Genetics* 2010; **186**:767–773.

[9] Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* 2015; **44**:512–525.

[10] Borges MC, Lawlor DA, de Oliveira C, White J, Horta BL, Barros AJD. The Role of Adiponectin in Coronary Heart Disease Risk: A Mendelian Randomization Study. *Circulation Research* 2016; **119**:491–499

[11] White J, Sofat R, Hemani G, Shah T, Engmann J, Dale C, Shah S et al. Plasma urate concentration and risk of coronary heart disease: a Mendelian randomisation analysis. *The Lancet Diabetes & Endocrinology* 2016; **4**:327–336.

[12] Noyce AJ, Kia DA, Hemani G, Nicolas A, Price TR, De Pablo-Fernandez E, Haycock PC, Lewis PA, Foltynie T, Davey Smith G, IPDGC collaborators, Schrag A, Lees AJ, Hardy J, Singleton A, Nalls MA, Pearce N, Lawlor DA, Wood NW. Estimating the causal influence of BMI on risk of Parkinson's disease: a Mendelian randomization study. *PLoS Medicine* 2017; **14**: e1002314

[13] Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *International Journal of Epidemiology* 2016 **45**:1866–86.

[14] Haycock PC, Burgess S, Nounu A, Zheng J, Okoli GN, Bowden J, Wade KH, Timpson NJ, Evans DM, Willeit P, Aviv A. Association between telomere length and risk of cancer and non-neoplastic diseases: a Mendelian randomization study. *JAMA oncology* 2017 **3**: 636-651.

[15] Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for Mendelian randomization analyses using MR-Egger regression: the role of the I^2 statistic. *International Journal of Epidemiology* 2016. **45**: 1961–74.

[16] Bloom HS. Accounting for no-shows in experimental evaluation designs. *Evaluation review* 1984, **8**: 225–246

[17] Yang F, Zubizaretta J, Small DS, Lorch S, Rosenbaum P. Dissonant conclusions when testing the validity of an Instrumental Variable. *The American Statistician* 2014, **68**: 253–263

[18] Burgess S, Bowden J. Integrating summarized data from multiple genetic variants in Mendelian randomization: bias and coverage properties of inverse-variance weighted methods Technical report, University of Cambridge 2015. `https://arxiv.org/pdf/1512.04486.pdf`

[19] Windmeijer, F. Two-Stage Least Squares as Minimum Distance Working paper, department of economics, University of Bristol, 2017. Available at: `http://www.efm.bris.ac.uk/economics/working_papers/pdffiles/dp17683.pdf`

[20] Thompson J, Minelli C, Del Greco M F. Mendelian randomization using public data from genetic consortia. *International Journal of Biostatistics* 2016 **12**:2

[21] Rücker G, Schwarzer G, Carpenter J, Binder H, Schumacher M. Treatment-effect estimates adjusted for small study effects via a limit meta-analysis. *Biostatistics* 2011; **12**:122–142.

[22] Ference BA, Julius S, Mahajan N, Levy PD, Williams KASr, Flack JM. Clinical effect of naturally random allocation to lower systolic blood pressure beginning before the development of hypertension. *Hypertension* 2014; **63**: 1182-88.

[23] Cook JR, Stefanski LA. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* 1994; **89**:1314–1328.

[24] DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.

[25] Thompson SG, Sharp S. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**:2693–2708.

[26] Baker R, Jackson D. Meta-analysis inside and outside particle physics: two traditions that should converge? *Research Synthesis Methods* 2013; **4**:109–124.

[27] Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *European Journal of Epidemiology* 2017, **19**: 1–13.

[28] International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011, **478**:103-9.

[29] CARDIoGRAMplusC4D. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genetics* 2003; **45**:25–33.

[30] Corbin LJ, Richmond RC, Wade KH, Burgess S, Bowden J, Davey Smith G, Timpson NJ. Body mass index as a modifiable risk factor for type 2 diabetes: Refining and understanding causal estimates using Mendelian randomisation. *Diabetes* 2016 db160418.

[31] Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic epidemiology* 2016 **40**: 304–14

[32] Hartwig FP, Davey Smith G, Bowden J. Robust inference in two-sample Mendelian randomisation via the zero modal pleiotropy assumption. *International Journal of Epidemiology* 2017, In press.

R code

R code to fit MR-Egger regression using the three weighting schemes is given below

```
# Pre-processing steps to ensure all
# gene--exposure estimates are positive

BetaYG  = BetaYG*sign(BetaXG)
BetaXG  = abs(BetaXG)

# 1st order weights

W1        = 1/seBetaYG^2
MREfitR1  = summary(lm(BetaYG ~BetaXG,weights=W1))

# 2nd order weights

W2        = 1/(seBetaYG^2 + (BIV^2)*seBetaXG^2)
MREfitR2  = summary(lm(BetaYG ~BetaXG,weights=W2))

# Modified 2nd order weights

BhatE1   = MREfitR1$coef[2,1]
W3       = 1/(seBetaYG^2 + (Bhat1^2)*seBetaXG^2)
MREfitR3 = summary(lm(BetaYG ~BetaXG,weights=W3))
```

Here `BetaYG`,`seBetaYG`, `BetaXG` and `seBetaYG` refer to the vectors of: SNP outcome associations, their standard errors, the SNP exposure associations and their standard errors respectively. In order to calculate the corresponding $Q^{'}$ statistic for modified 2nd order weights, the following code can be used:

```
phi_E3  = MREfitR3$sigma^2
QE3     = DF*phi_E3 # Q' statistic
QEp3    = 1-pchisq(QE3,DF) # pvalue
# individual Q' contribution vector
QE3ind  = W3*(BetaYG - MREfitR3$coef[1,1] - MREfitR3$coef[2,1]*BetaXG)^2
```

`R` code to peform an IVW analysis using modified 2nd order weights is given below

```
# IVW analysis

BIV       = BetaYG/BetaXG
W1        = 1/(seBetaYG^2/BetaXG^2)
BIVw1     = BIV*sqrt(W1)
sW1       = sqrt(W1)
```

```
IVWfitR1  = summary(lm(BIVw1 ~ -1+sW1))
Bhat1     = IVWfitR1$coef[1]
DF        = length(BIV)-1

# modified 2nd order weights

W3       = 1/(seBetaYG^2/BetaXG^2 + (Bhat1^2)*seBetaXG^2/BetaXG^2)
BIVw3    = BIV*sqrt(W3)
sW3      = sqrt(W3)
IVWfitR3 = summary(lm(BIVw3 ~ -1+sW3))
Bhat3    = IVWfitR3$coef[1]

phi_IVW3  = IVWfitR3$sigma^2
QIVW3     = DF*phi_IVW3 # Q statistic
Qp3       = 1-pchisq(QIVW3,DF) # p-value
# individual Q contribution vector
Q3ind     = W3*(BIV - Bhat3)^2
```

In order to apply SIMEX correction to MR-Egger regression, the following code can be used

```
library(simex)
Fit      = lm(BetaYG~BetaXG,weights=W,x=TRUE,y=TRUE)
mod.sim1 =  simex(Fit,B=1000,measurement.error = seBetaXG,
            SIMEXvariable="BetaXG",fitting.method ="quad",asymptotic="FALSE")
summary(mod.sim1)
```

The value `W` can be substituted with any weighting scheme defined above.

We now provide `R` code to calculate the I^2_{GX} statistic under any scheme

```
Isq = function(y,s){
k          = length(y)
w          = 1/s^2; sum.w  = sum(w)
mu.hat     = sum(y*w)/sum.w
Q          = sum(w*(y-mu.hat)^2)
Isq        = (Q - (k-1))/Q
Isq        = max(0,Isq)
return(Isq)
}

# General I^2 measure

Isq(BXG*sqrt(W),seBetaXG*sqrt(W))
```

```
# e.g. W=W1 or W=W3
```

Comparison with the work of Thompson *et. al*

In related work, Thompson et al [20] also noted the poor performance of 2nd order weights and, when estimating the causal effect in two-sample summary data MR. They also proposed two methods for improving the performance of 1st order weights, which we now summarise. Firstly, they showed that the ratio estimate for SNP j, $\hat{\beta}_j$ is positively biased by a factor approximately equal to $1 + 1/F_j$, where F_j is the F-statistic $\hat{\gamma}_j^2/\sigma_{Xj}^2$. They used this result to derive a bias reduced estimate for the ratio estimate

$$\hat{\beta}_{j*} = \hat{\beta}_j \frac{F_j}{1 + F_j} \tag{18}$$

Secondly, they noted that a more precise estimate for the association between SNP j and the exposure could be derived by combining the original estimate $\hat{\gamma}_j$ with a second estimate $\hat{\Gamma}_j/\hat{\beta}$ via the inverse-variance weighted average:

$$\hat{\gamma}_{*j} = \frac{\hat{\gamma}_j/\sigma_{Xj}^2 + \hat{\beta}^2\hat{\Gamma}_j/\sigma_{Yj}^2}{1/\sigma_{Xj}^2 + \hat{\beta}^2/\sigma_{Yj}^2}, \tag{19}$$

We assessed the performance of using the improved precision formula (19) in previous simulation set up, firstly on its own. This was achieved by plugging in the IVW estimate for $\hat{\beta}$ into (19), which was calculated using 1st order weights. The resulting estimate $\hat{\gamma}_{j*}$ was then used to calculate (a) updated ratio estimates $\hat{\Gamma}_j/\hat{\gamma}_{j*}$ and (b) updated 1st order weights $\hat{\gamma}_{j*}^2/\sigma_{Yj}^2$ before re-estimating the causal effect and Cochran's Q. Next we assessed the additional benefit of incorporating bias correction via formula (18) into the analysis. Note that this bias adjustment also made use of the updated estimate $\hat{\gamma}_{j*}$ in both the definition of F_j and $\hat{\beta}_j$. The Results for both procedures are shown in Table (4).

Thompson et al's improved precision formula is generally effective: It produces Q statistics and associated p-values that are slightly conservative - the level of conservatism increasing with either an increasing causal effect or a decreasing mean instrument strength. Point estimates for the causal effect are also less affected by regression dilution bias also. As expected, incorporating bias correction formula (18) into the analysis produces causal estimates with minimal bias. However, it also has leads to a dramatic instability in the distribution of Cochran's Q statistic - its mean value explodes unless the mean F statistic is very large.

We investigated two further methods: implementing bias correction on its own, as

well as combining Thompson's improved precision formula with our modified 2nd order weights. Both approaches also led to severe instability in Cochran's Q statistic, thus ruling them out of further consideration. Further research is required to ascertain whether this deficiency is intrinsic, or if can be circumvented with a suitable modification. Further work would also be required to extend their approach to the MR-Egger regression context.

Mean	1st order w_j			1st order w_j		
	Improved precision			Improved precision + bias correction		
F	Q	T1E(Q)	$\hat{\beta}_{IVW}$ (S.E.)	Q	T1E(Q)	$\hat{\beta}_{IVW}$ (S.E.)
			No heterogeneity, β=0.1			
100	23.2	0.037	0.100 (0.012)	24.1	0.052	0.101 (0.012)
61	22.7	0.032	0.100 (0.012)	4.5×10^{6}	0.082	0.102 (0.013)
40	21.7	0.022	0.099 (0.013)	$7.\times10^{12}$	0.222	0.103 (0.086)
25	20.3	0.014	0.097 (0.014)	1.5×10^{13}	0.588	0.110 (0.26)
10	19.4	0.031	0.077 (0.016)	1.4×10^{22}	0.984	-0.444 (53.4)
			No heterogeneity β=0.05			
100	23.7	0.041	0.050 (0.012)	25.3	0.057	0.05 (0.012)
61	23.6	0.043	0.050 (0.012)	2.3×10^{5}	0.107	0.051 (0.012)
40	23.2	0.036	0.050 (0.012)	3.7×10^{11}	0.287	0.052 (0.045)
25	22.6	0.033	0.049 (0.013)	1.2×10^{14}	0.697	0.05 (0.392)
10	21.9	0.041	0.041 (0.013)	1.1×10^{20}	0.993	0.057 (5.823)
			No heterogeneity β=0			
100	23.9	0.050	0.00 (0.012)	27.3	0.067	0.00 (0.012)
62	23.9	0.046	0.00 (0.012)	1.6×10^{6}	0.111	0.00 (0.013)
40	24.0	0.050	0.00 (0.012)	7.7×10^{16}	0.321	0.02 (1.918)
25	23.8	0.046	0.00 (0.012)	3.4×10^{12}	0.763	0.00 (0.291)
10	23.3	0.041	0.00 (0.012)	4.0×10^{14}	0.996	0.01 (0.873)

Table 4: *Mean Q statistic and IVW estimate $\hat{\beta}_{IVW}$ calculated using Thompson et al's improved precision formula (19) on its own, and in conjunction with bias correction formula (18). Results calculated over 10,000 simulated data sets. Type I error rate (T1E(Q)) refers to the proportion of times Q is greater than the upper 95th percentile of a χ^2_{24} distribution.*