An examination of multivariable Mendelian randomization in the single sample and two-sample summary data settings.

Eleanor Sanderson[*a,b], *George Davey Smith*[a,b] *Frank Windmeijer*[a,c], *Jack Bowden*[a,b]

a MRC Integrative Epidemiology Unit, University of Bristol, b Population Health Sciences, University of Bristol, c Department of Economics, University of Bristol

Abstract

Mendelian Randomisation (MR) is a powerful tool in epidemiology which can be used to estimate the causal effect of an exposure on an outcome in the presence of unobserved confounding, by utilising genetic variants that are instrumental variables (IVs) for the exposure. This can be extended to Multivariable MR (MVMR) to estimate the effect of two or more exposures on an outcome. We use simulations and theoretical arguments to clarify the interpretation of estimated effects in a MVMR analysis under a range of underlying scenarios, where a secondary exposure acts variously as a confounder, a pleiotropic pathway, a mediator and a collider. We then describe how instrument strength and validity can be assessed for an MVMR analysis in the single sample setting, and how such tests can be extrapolated to the popular two-sample summary data setting. We illustrate our methods using data from UK biobank to estimate the effect of education and cognitive ability on body mass index. We show that MVMR analysis consistently estimates the effect of an exposure, or exposures, of interest and provides a powerful tool for determining causal effects in a wide range of scenarios with either individual or summary level data.

Introduction

In many scenarios where we wish to estimate the causal effect of an exposure X on an outcome Y, a conventional regression analysis can be misleading, as the observational association between the two variables could easily be affected by unobserved confounding. If genetic variants – usually single nucleotide polymorphisms (SNPs) - are available which reliably predict the exposure variable but do not have an effect on the outcome through any other pathway, then they are valid instrumental variables (IVs) and can be used in a Mendelian randomization (MR) analysis [1,2] to test whether the exposure causally affects the outcome, as illustrated in Figure 1.

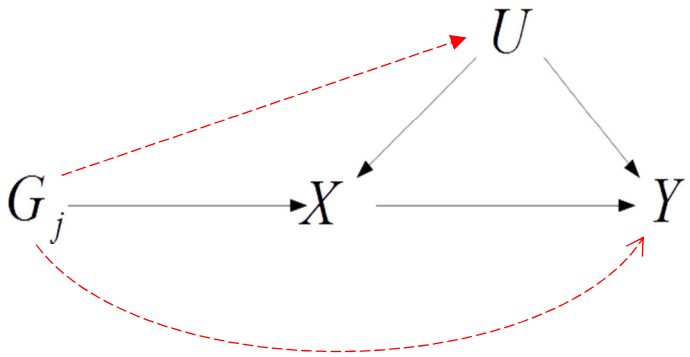

Figure 1: *Hypothesised relationship between genetic variant G_j, modifiable exposure X and outcome Y in the presence of unobserved confounder, denoted by U. Dashed lines represent violations of the instrumental variable assumptions.*

To state the IV assumptions more formally with reference to Figure 1: For a single SNP G_j to be a valid IV it must be: associated with X (IV1); independent of all confounders of X and Y, as represented by U (IV2); and (IV3) independent of Y given X and U. These assumptions are encoded by the solid arrows in Figure 1. If IV1-IV3 are satisfied for a single SNP G_j, or more generally a set of SNPs $G = (G_1, \dots, G_L)$, then traditional IV methods can be employed to reliably test for a causal effect of X on Y using G, X and Y alone, without any attempt to adjust for U at all. For example, suppose the variables G, X, U and Y are linked via the following models:

$$Y = \beta_0 + \beta X + U + v_y \quad (1)$$

$$X = \gamma_0 + \pi G + U + v_x \quad (2)$$

Here v_x and v_y represent independent error terms, π represents the parameter vector $\pi_1, \dots, \pi_L$, and β is the true causal effect of X on Y we wish to estimate. We will assume throughout this paper that $(G_1, \dots, G_L)$ are mutually uncorrelated (by design). A naïve regression of Y on X will not yield a consistent estimate for β because the explanatory variable in the regression, X, is correlated with U. However, regressing Y instead on $\hat{X}$ - the predicted value of X from a regression of X on G – *will* yield a consistent estimate for β, because $\hat{X}$ is independent of U. This procedure is referred to as two-stage least squares (TSLS).[3]

TSLS relies on individual level data, but the sharing of such data is often impractical. In recent years it has become much more common to attempt MR analyses using summary data estimates of SNP-exposure and SNP-outcome associations gleaned from two

independent but homogeneous study populations.[4] The SNPs in question are usually identified as genome-wide significant `hits' in distinct genomic regions via a genome wide association study (GWAS) for the exposure. This is referred to as `two-sample summary data MR'.

Let π_j and Γ_j represent the true association for SNP G_j in G with the exposure and the outcome respectively. From models (1) and (2) we can link the j'th SNP outcome association to the j'th SNP exposure association via the model

$$\Gamma_j = \beta\pi_j + v_{Yj} \qquad (3)$$

It follows that the ratio estimator $\hat{\beta}_j = \frac{\hat{\Gamma}_j}{\hat{\pi}_j}$, is a consistent estimate for β also. When the SNPs are uncorrelated, taking an inverse variance weighted (IVW) average of the ratio estimates will yield an overall estimate for β, $\hat{\beta}_{IVW}$, that closely approximates the TSLS estimate that would have been obtained if individual level data were available.[5]

Detecting `weak' instruments and `invalid' instruments in MR

If assumptions IV1 – IV3 are fulfilled for all SNPs in G, and linear models (1)-(2) hold, then either a TSLS or IVW analysis (with uncorrelated SNPs) will consistently estimate the causal effect. In order to satisfy IV1, the SNPs in G should strongly predict the exposure X. This can be quantified using the F-statistic from the first stage regression of X on G. Using instruments that are only weakly associated with the exposure (i.e. which have a small F-statistic) will result in weak instrument bias.

Secondly, SNPs should not exert a direct effect on Y, i.e. they should not affect Y other than through X. Horizontal pleiotropy could easily be responsible for such a violation in the MR setting.[5] This would represent a violation of IV2 and/or IV3. Any such violation is likely to lead to bias and potentially erroneous conclusions in both the TSLS and IVW estimates.[3] This can be evaluated using the Sargan test[6] using individual level data and Cochran's Q statistic[7,8,9] using summary data, which will be discussed in more detail in later sections.

Multivariable Mendelian randomization

MR can be extended to estimate the effect of multiple exposure variables on an outcome[10] and is particularly useful in cases where a standard MR analysis would fail due to violation of assumptions IV2-3. It is also useful in cases where two or more correlated exposures are of interest[11] and may help to understand if both exposures exert a causal effect on the outcome, or if one in fact mediates the effect of the other on the outcome[12,13]. `Multivariable MR' (MVMR) requires a set of SNPs, G, which are associated with the exposure variables but do not affect the outcome other than through these variables. In the same way as

standard (single variable) MR, these SNPs can be used to predict each of the exposure variables in the model and these predicted values can be used to estimate the effect of the exposures on the outcome in a multivariable regression analysis. The setup for MVMR is illustrated for an MVMR analysis involving two exposure variables (X_1 and X_2) in Figure 2. The arrows linking X_1 with X_2, and X_2 with Y have been left bi-directional (and coloured red), to acknowledge the fact that many underlying causal relationships are possible. That is, they could point in either direction or be completely absent. Indeed, many of these options will be subsequently explored.

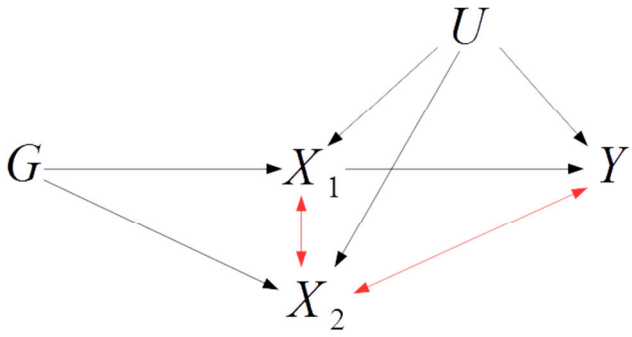

Figure 2: *Hypothesised relationship between genetic variant(s) G, modifiable exposures, X_1, X_2 and outcome Y in the presence of unobserved confounder U. Bi-directional arrows represent possible violations of the IV assumptions induced by X_2 that are explored in this paper.*

Although it is the simplest possible MVMR setting, two exposures suffice to illustrate all the scenarios and ideas described in this paper. From Figure 2, we can write the following general model linking Y, X_1, X_2 and U:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U + v_Y \qquad (4)$$

For example, suppose that X_1 and X_2 are in fact independent given G (so there is **no** arrow in Figure 2 between X_1 and X_2) and X_2 affects Y independently of X_1 (so that there is a **direct arrow** from X_2 to Y). If true, then models (5) and (6) for X_1 and X_2 would, jointly with (4), describe the data:

$$X_1 = \pi_1 G + U + v_{x_1} \qquad (5)$$

$$X_2 = \pi_2 G + U + v_{x_2} \qquad (6)$$

The purpose of an MVMR analysis is to determine the direct causal effect of both X_1 and X_2 on the outcome Y, when conditioned on one another. Without loss of generality we will focus primarily on the effect of X_1 (and the parameter β_1) with the direct effect of X_2 on Y denoted by β_2 being of secondary importance.

With individual level data, regressing each exposure on the full set of SNPs would yield genetically predicted estimates for X_1 and X_2. The outcome Y would then be regressed on these predicted estimates for X_1 and X_2 jointly to obtain a consistent estimate of $\beta_1$and β_2. This can be conducted by simply using the *ivreg2* command in Stata or *ivpack* in R.

In the two sample summary data setting, MVMR can still be implemented using summary data estimates of the association between SNP j (out of L) and: the outcome, $\hat{\Gamma}_j$; exposure X_1, $\hat{\pi}_{1j}$; and exposure X_2, $\hat{\pi}_{2j}$, by fitting the following model:

$$\hat{\Gamma}_{\mathrm{j}} = \beta_1\hat{\pi}_{1,\mathrm{j}} + \beta_2\hat{\pi}_{2,\mathrm{j}} + v_{\mathrm{Yj}} \qquad (7)$$

This is a straightforward generalization of the IVW estimation framework, as first described by Burgess and colleagues.[10]

Important considerations

To conduct an MVMR analysis it is necessary to have at least as many genetic instruments as there are exposures to be instrumented in the model, this is true regardless of whether single sample or two sample summary data are used.[3] It is possible to include genetic instruments that are associated with more than one exposure variable, providing all of those exposure variables are included in the estimation. Instruments must not, however, exert a direct effect on the outcome, except through the included exposures. There is no benefit to excluding instruments that are only associated with one exposure, as this will lead to a loss of precision in the estimates obtained. This also avoids any potential bias that could arise due to selecting instruments based on their strength.[14]

What quantity do MR and MVMR estimate and when does this differ?

MR and MVMR target different causal effects of the exposure on the outcome. In general, MR estimates the *total* effect of the exposure on the outcome, whereas MVMR estimates the *direct* effect of each exposure on the outcome.

For example, if Figure 3 describes the truth, the total effect of exposure X_1 on the outcome is the effect of X_1 on the outcome Y directly plus the effect of X_1 on Y via X_2, and is equal to $\beta_1 + \alpha\beta_2$. The direct effect of the exposure X_1on the outcome Y is the effect X_1 has on Y not via any other exposure variables included, and so is equal to β_1. Whether or not these effects differ in general depends on the underlying relationship between the exposures and between each exposure and the outcome. If there is no effect of X_1 on X_2 or of X_2 on Y, i.e. either α or β_2 is equal to zero, these effects will be the same.

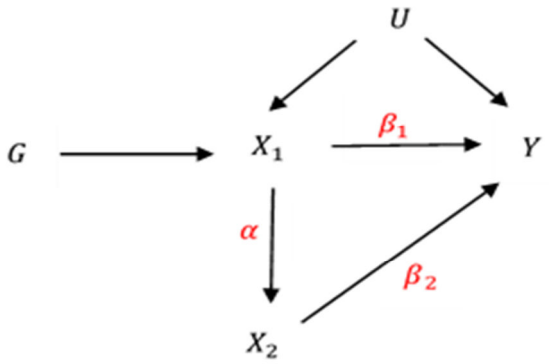

Figure 3. Illustration of the direct effect and total effect of X_1 on the outcome Y.

To highlight the potential differences between MR and MVMR, and the potential benefits of MVMR, we now consider the application of MVMR to four different scenarios which are commonly encountered, or at least suspected, in epidemiological studies.

In the first scenario X_2 is a **confounder** of the relationship between X_1 and Y. That is, there is a direct causal path from X_2 to X_1 and from X_2 to Y. Along with model (4), model (6) above and (8) below underlie the individual level data:

$$X_1 = \pi_1 G + \gamma_1 X_2 + U + v_{x_1} \qquad (8)$$

In the second scenario X_2 is a **collider** of the relationship between X_1 and Y. That is, there is a direct causal path from X_1 to X_2 and from Y to X_2. When an exposure and outcome both influence another variable controlling for that variable in in conventional analysis will introduce collier bias into the observed association between the exposure and the outcome.[15] Along with model (4) (with β_2 set to 0), model (5) above and (9) below are used to generate the individual level data:

$$X_2 = \pi_2 G + \gamma_1 X_1 + \gamma_y Y + U + v_{x_2} \qquad (9)$$

In the third scenario X_2 is an independent **pleiotropic** pathway from G to Y. This corresponds to the scenario first described in the previous section. Along with model (4), model (5) and (6) above are used to generate the individual level data.

In the fourth scenario X_2 is a **mediator** of the relationship between X_1 and Y. Along with model (4), model (5) above and (10) below are used to generate the individual level data:

$$X_2 = \pi_2 G + \gamma_1 X_1 + U + v_{x_2} \qquad (10)$$

Each of these scenarios are shown in Figure. 5. Datasets of 10,000 individuals are simulated under all four scenarios discussed using $L = 30$ genetic variants. The variants are assumed

to be uncorrelated but, for added realism and complexity, are further subdivided into three categories:

- 10 SNPs that only predict X_1: G_1 (with a non-zero π_1 element but zero π_2 element);
- 10 SNPs that only predict X_2: G_2 (with a non-zero π_2 element but zero π_1 element);
- 10 SNPs that predict X_1 and X_2: G_{12} (with non-zero π_1 and π_2 elements).

G therefore represents the complete vector(G_1, G_2, G_{12}). For each scenario the causal parameter of interest, β_1, is set to 1.

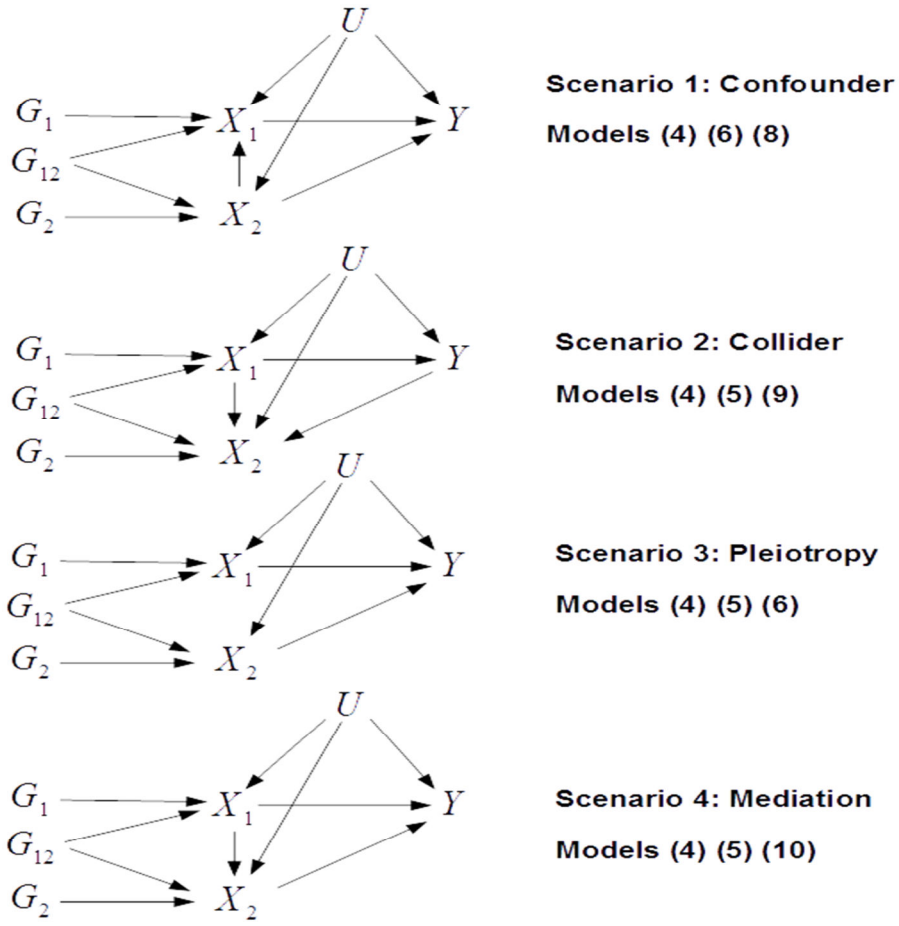

Figure 4: *Causal diagrams for scenarios 1-4.*

For each scenario, we estimate the causal effect of X_1 and X_2 on Y, (β_1 and β_2) using a range of estimation methods. With **single sample individual level data,** we implemented:

- OLS, both for X_1 and X_2 individually (i.e. univariate regressions) and together (i.e. a multivariable regression);
- MR for X_1 and X_2 individually, each time using all the available SNPs as instruments,
- Multivariable MR including both X_1 and X_2 in the same analysis.
- MR using for X_1 and X_2 individually only the SNPs truly associated with each exposure included in the estimation.

With **two sample summary level data,** we implemented:

- MR for X_1 and X_2 individually using all of the instruments available;
- MVMR including both X_1 and X_2;
- MR X_1 and X_2 individually using only the SNPs truly associated with each exposure included in the estimation.

All estimation methods are described in Table A.1 in the appendix. In all of the scenarios considered the exposure variables are strongly predicted by the instruments and the instruments have no additional pleiotropic effects on the outcome, other than through the exposures included in the model.

Results

Focusing our attention on exposure 1, the results from these simulations show that MVMR always gives an unbiased estimate of the *direct* effect of X_1 on the outcome. In the hypothetical case where only the SNPs truly associated with X_1 (G_1) are used as instruments in a single variable MR the estimated effect of X_1 on Y is the *total* effect of a change in X_1 on the outcome. Whether the direct or total effect is of more interest to practitioners will depend on the particular situation being considered. In many of the scenarios explored the direct effect equals the total causal effect, however when X_2 is a mediator of the relationship between X_1 and the outcome, the direct and total effects of X_1 may be substantially different. The results from the simulations are given in Table A.2 in the appendix and a summary table of what is estimated by each method in each scenario is given in Table 1.

Table 1 – Summary of estimated effects for β_1

	Scenario/which estimand is targeted?			
Method	I	2	3	4
Individual level data				
OLS	x	x	x	x
Univariate MR	x	Direct/total	x	x
MVMR	Direct/total	Direct/total	Direct/total	*Direct*
Univariate MR – subset of SNPS	Direct/total	Direct/total	Direct/total	*Total*
Two-sample summary data analysis				
Univariate MR	x	Direct/total	x	x
MVMR	Direct/total	Direct/total	Direct/total	*Direct*
Univariate MR – subset of SNPS	Direct/total	Direct/total	Direct/total	*Total*

When each method of estimation estimates the direct and total effects for β_1 in each of the scenarios considered. An 'x' represents a biased method of estimation

When conducting the univariate MR estimation with a subset of SNPs we have, for illustration, assumed `oracle' knowledge on which SNPs truly predict which exposure. This will, of course, not be possible in practice. Table 1 shows that when all SNPs in G are used for a univariate MR analysis, it will deliver a biased estimate of the total causal effect in scenarios 1, 3 and 4. MVMR will then provide a consistent estimator of the direct effect of

the exposure on the outcome. These simulation results also highlight that MVMR does not introduce collider bias to the results when X_2 is a collider of the relationship between X_1 and Y. This is because the predicted value of X_2 , $\hat{X}_2$, is instead used in the analysis. This is an important benefit of MVMR.

Testing the assumptions of MVMR.

In the simulations above we assumed, for clarity, that the instruments were both strong and valid for the purposes of an MVMR analysis. However, violation of these assumptions can give misleading results in practice, so it is necessary to test these assumptions. We now describe how instrument strength and validity can be scrutinised for an MVMR analysis in the individual and two sample summary data settings.

The Individual level data MVMR setting.

Instrument strength

In any MR analysis the set of genetic instruments G must be strong in order to avoid weak instrument bias (assumption IV1). However, the assessment of instrument strength is more complicated in the multivariate setting. It is necessary for G to strongly predict both X_1 and X_2 (as quantified by strong F-statistics), but not sufficient. In addition, G must also *jointly* predict both X_1 and X_2. That is, once the secondary exposure X_2 has been predicted using G, G must *still* be able to predict the primary exposure X_1. Figure 5 illustrates three scenarios (A – C) where this may not be the case even when both exposures appear to be strongly predicted individually by G and a fourth scenario (D) where both exposures are strongly predicted.

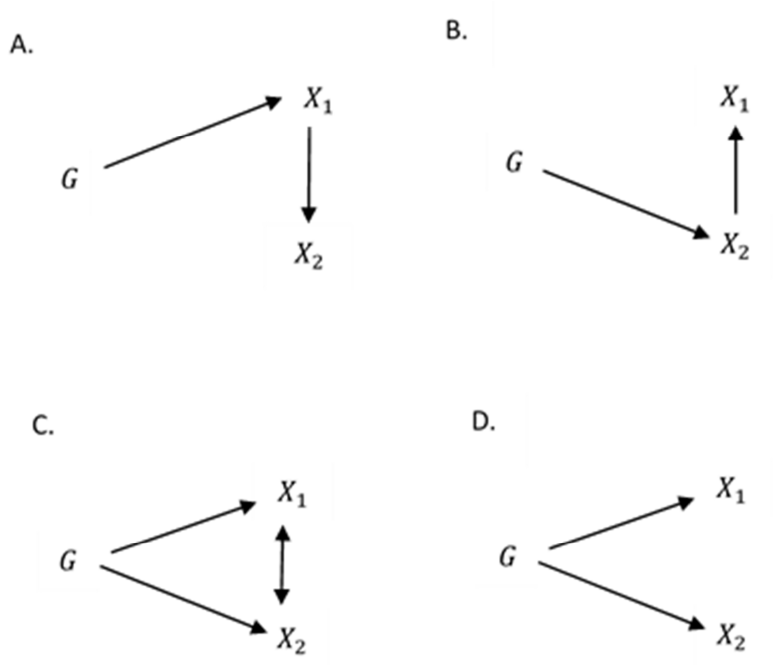

Figure 5 *– Potential setups of instruments and exposures. In A – B the exposures are individually strongly predicted but are not jointly predicted. In C the exposures are individually strongly predicted but weakly predicted in a joint sense. . In D; the exposures are individually and jointly strongly predicted. Specifically: A: G predicts X_1 which is a predictor of X_2. B: G predicts X_2 which is a predictor of X_1. C: G predicts X_1 and X_2 which are highly correlated. D: G predicts X_1 and X_2 which are uncorrelated (given G).*

Joint strength can be assessed using the Sanderson-Windmeijer conditional F statistic[16], F_c, that is available as part of ivreg2 in Stata. F_c is calculated in the following manner:

- X_2 is regressed on the full set of genetic instruments (and any control variables included in the estimation) and the predicted value of X_2, $\hat{X}_2$ is calculated;
- X_1 is then regressed on $\hat{X}_2$ (and any control variables) to yield the TSLS estimate $\hat{\delta}$and the residual error terms $X_1 - \hat{\delta}X_2$ are saved;
- The errors are then regressed on the full set of instruments (and any control variables). The conditional F statistic is obtained as the F statistic for the effect of the instruments in this regression;
- The conditional F statistic must be adjusted for a degrees of freedom correction, and can be compared to the conventional weak instrument critical values.[17]

If the conditional F statistic for all of the exposure variables is sufficiently large then the instruments can be considered strong for the purposes of MVMR.

Instrument validity

If no pleiotropy exists amongst the genetic variants then each one should identify the same causal parameter. This can be evaluated using the Sargan test.[6] Specifically, it tests whether the instruments can explain any of the variation in the outcomes that has not been explained by the value of the exposure variables. It is calculated by the following steps;

- Regress the outcome Y on the exposures using TSLS to yield causal estimates $\hat{\beta}_1$ and $\hat{\beta}_2$
- Calculate the residual error term $Y - (\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2)$ and then regress the residuals on the full set of instruments. The Sargan test is then the sample size times the R2 of this regression.
- Evaluating with the Sargan statistic with respect to a χ^2 distribution with degrees of freedom equal to the number of instruments minus the number of predicted exposure variables (i.e. the null hypothesis that all of the instruments are valid).[3]

This test is again available as part of the ivreg2 command in Stata, and the ivpack package in R. In order to conduct this test the model must be over-identified, i.e. there must be more instruments than exposure variables (so that the degrees of freedom of the χ^2 test is positive).[18] This `global' test does not give any indication as to which of the genetic instruments are invalid if the test rejects the null.

The two-sample summary data setting

Assessment of instrument validity and strength has, to the best of our knowledge, yet to be described in the two sample summary data setting that is relevant to the majority of contemporary Mendelian randomization studies, and consequently it is not implemented in any standard software. We therefore describe the necessary procedures in fine detail so that it can be confidently implemented by others.

Assessing instrument strength: heterogeneity is `good'

Suppose that all of the genetic instruments predict both exposure variables, so that models (4), (5) and (6) hold, but there are at least two elements of $\pi_1$and π_2 in (5) and (6) which differ. If true, then the model will be *at least exactly identified*. That is, there will be at least as many independent genetic instruments (i.e. 2) as there are exposure variables to be instrumented. This implies that model (11):

$$X_1 = \delta X_2 + u_1 \qquad (11)$$

$$X_2 = \pi_2 G + u_2,$$

must be over-identified (or equivalently miss-specified), because X_2 cannot then be simply a scalar multiple, δ, of X_1 . Therefore, we can test for under-identification in our estimation model by testing for over-identification in model (11) using the Sargan test as described above. The equivalence of this test with the Sanderson-Windmeijer approach has been shown formally elsewhere[20]. The null of this Sargan test is that of underidentification.

Extending this to two-sample analysis; $\hat{\pi}_{1,j} = \delta\hat{\pi}_{2,j} + v$ is analogous to equation (11) estimated by IV using individual level data with $\hat{X}_2$ predicted using G, therefore it should be possible to test for under-identification in two-sample MVMR estimation by testing for overidentification in the model $\hat{\pi}_{1,j} = \delta\hat{\pi}_{2,j} + v$. We recommend that this test is conducted using a modified version of Cochran's Q statistic, as shown in equation (12) below:

$$Q_{x_1} = \sum_{j=1}^{L}\left(\frac{1}{\sigma^2_{x1j}}\right)\left(\hat{\pi}_{1j} - \hat{\delta}\hat{\pi}_{2j}\right)^2. \qquad (12)$$

The variance term for Q_{x_1}, $\sigma^2_{x1j} = \sigma^2_{1j} + \hat{\delta}^2\sigma^2_{2j} - 2\hat{\delta}\sigma_{12j}$, where σ^2_{1j} is the variance of $\hat{\pi}_{1,j}$, σ^2_{2j} is the variance of $\hat{\pi}_{2,j}$, σ_{12j} is the covariance of $\hat{\pi}_{1,j}$ and $\hat{\pi}_{2,j}$, and $\hat{\delta}$ is an efficient estimator for δ Estimation of the σ^2_{x1j}terms in practice depends on the type of model used to obtain $\hat{\pi}_{1,j}$ and $\hat{\pi}_{2,j}$. When each exposure is regressed on the entire set of SNPs simultaneously (i.e. via multivariate regressions without an intercept):

$$\sigma_{1j}^2 = \frac{(G^T G)_{jj}^{-1}}{n}\sum_{i=1}^{n}\hat{v}_{1i}^2 \ , \qquad \sigma_{2j}^2 = \frac{(G^T G)_{jj}^{-1}}{n}\sum_{i=1}^{n}\hat{v}_{2i}^2, \qquad \text{and } \sigma_{12j} = \frac{(G^T G)_{jj}^{-1}}{n}\sum_{i=1}^{n}\hat{v}_{1i}\,\hat{v}_{2i}$$

Where n is the number of subjects, and $(\hat{v}_{1i}, \hat{v}_{2i})$ are the estimated residuals from these regressions. If $\hat{\pi}_{1j}$ and $\hat{\pi}_{2j}$ are obtained separately (i.e. via univariate regressions without an intercept), then the error terms are obtained from the equivalent expressions:

$$\sigma_{1j}^2 = \frac{\left(G_j^T G_j\right)^{-1}}{n}\sum_{i=1}^{n}\hat{v}_{1ij}^2 \ , \qquad \sigma_{2j}^2 = \frac{\left(G_j^T G_j\right)^{-1}}{n}\sum_{i=1}^{n}\hat{v}_{2ij}^2, \qquad \text{and } \sigma_{12j} = \frac{\left(G_j^T G_j\right)^{-1}}{n}\sum_{i=1}^{n}\hat{v}_{1ij}\,\hat{v}_{2ij}$$

Respectively, $\hat{v}_{1ij}$ and $\hat{v}_{2ij}$ are the estimated residuals from the j'th regression.

Under the null hypothesis the instruments do not contain enough information to predict both exposure variables, Q_{x_1} will be asymptotically χ^2_{L-1} distributed when δ is estimated using an asymptotically efficient estimator, where L is the number of instruments. Rejection of the null hypothesis (i.e. detection of `heterogeneity') indicates that the model we wish to estimate *is* identified for X_1 .

All the above can be repeated for X_2 by swapping the roles of $\hat{\pi}_1$and $\hat{\pi}_2$ and calculating an equivalent Q statistic for X_2, Q_{x_2} say. If Q_{x_1} and Q_{x_2} are larger than the chosen critical value then the null hypothesis of under-identification can be rejected and the test suggests that the instruments can predict variation in both exposures. Table 2 shows the distribution of Q_{x_1} and $Q_{x_2}$for four different scenarios with two exposure variables and $L = 30$ SNPs. X_1 and X_2 are both functions of a set of SNPs and independent confounding variables. In the first simulation the model has been set up as given in Scenario 3 in Figure 4 and in Figure 5D with each of the exposure variables predicted by a set of SNPs and a common confounding variable. This model is identified as both exposure variables can be predicted by the set of instruments. In the second and third simulations the model has been set up in the same way but with no effect of the SNPs on either X_1 or X_2 respectively. That is, the model is under identified with one of the exposure variables not being predicted by the instruments in each case. In the final simulation the model has been set up with the effect of the SNPs on the exposures as given in Figure. 5A and a common confounder. This setup leads to neither exposure being predicted by the SNPs when they are both included in an MVMR estimation as the SNPs in the model cannot predict both of the exposure variables jointly. The results from these simulations show that this test has the required distribution under the null hypothesis.

Table 2 - The distribution of the modified Q statistic as a test for under-identification

	Q_{x_1}			Q_{x_2}		
	Mean	Std. dev	Rej. Rate (%)	Mean	Std. dev	Rej. Rate (%)
x_1 strongly identified x_2 strongly identified	588 11.5	132 59.4	100	58790. 6	13573 .5	100
x_1 unidentified x_2 strongly identified	30.0	7.9	6.3	58154. 2	93864 .4	100
x_1 strongly identified x_2 unidentified	597 78.8	956 31.8	100	30.2	7.3	5.7
x_1 strongly identified x_2 strongly identified Jointly unidentified $x_1 = \delta x_2,\ \delta = 1$	29.7	7.7	4.8	29.7	7.7	4.8

N = 5,000. Repetitions = 1000, 30 SNPs as instruments. Rejection rates give the proportion of times each Q statistic is larger than the 95th percentile of a Chi-squared distribution on 29 degrees of freedom (42.56).

Testing instrument validity: heterogeneity is `bad'

Cochran's Q statistic for the regression of interest has been proposed as a method for identifying invalid instruments (e.g. due to horizontal pleiotropy) in two-sample summary data MR analysis, with a single exposure.[9] Specifically, if all instruments are valid IVs, and the modelling assumptions necessary for two-sample MR are satisfied, then each genetic instrument should give the same estimate of the effect of the exposure on the outcome. Excessive heterogeneity in the causal effect estimates obtained by each SNP individually now becomes an indicator of invalid instruments. We propose testing for invalidity in two sample summary data MVMR using an adjusted version of the Cochran Q statistic given by:

$$Q_A = \sum_{j=1}^{L} \left(\frac{1}{\sigma_{Aj}^2}\right) \left(\hat{\Gamma}_j - \left(\hat{\beta}_1 \hat{\pi}_{1j} + \hat{\beta}_2 \hat{\pi}_{2j}\right)\right)^2. \quad (13)$$

Where $\sigma_{Aj}^2 = \sigma_{yj}^2 + \hat{\beta}_1^2\, \sigma_{1j}^2 + \hat{\beta}_2^2\, \sigma_{2j}^2 + 2\hat{\beta}_1\hat{\beta}_2\sigma_{12j}$. To clarify, σ_{yj}^2 is the variance of $\hat{\Gamma}_j$, and $\hat{\beta}_1$ and $\hat{\beta}_2$ are efficient estimates of β_1 and β_2 (for example as obtained from fitting model (7)). Under the null hypothesis that the genetic instruments do not have pleiotropic effects on the outcome, Q_A is asymptotically χ^2 distributed with $(L-2)$ degrees of freedom. The standard implementation of Cochran's Q would merely have a weighting of σ_{yj}^2, and is not therefore asymptotically χ^2 distributed. It is a straightforward generalisation of the adjusted Q statistic recently proposed by Bowden et al in the univariate MR setting.[8] Excessive heterogeneity in Q_A therefore brings assumptions IV2 and IV3 into doubt.

Figure 6 shows the distribution of Q_A compared to the standard Q statistic and a χ^2 distribution with 98 degrees of freedom for a model with 2 exposure variables and 100 genetic instruments. For simplicity the estimated effects of the SNPs on the exposures each

have a common variance of 0.02 and have a common covariance of 0. Q_A is seen to have the correct distribution under the null hypothesis of no pleiotropy in the model.

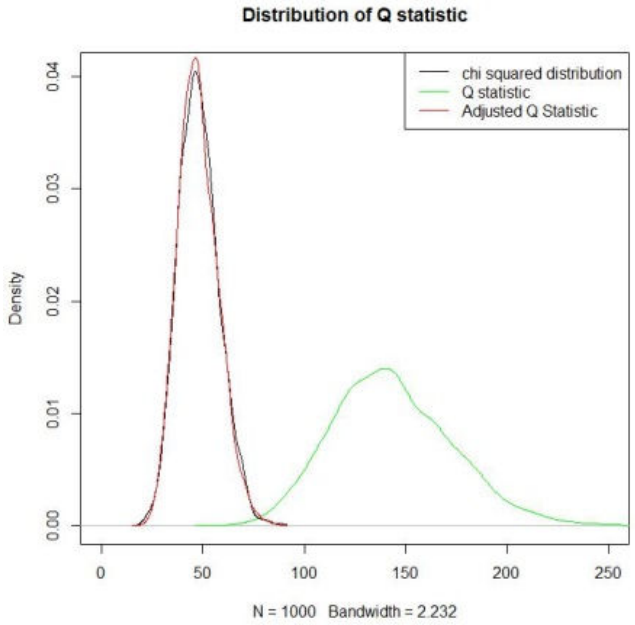

Figure 6: *The distribution of the adjusted and standard Q statistics under the null hypothesis of no heterogeneity. 5000 repetitions, 100 SNPs.* Here $\beta_1 = \beta_2 = 1$. $\sigma_{1j}^2 = \sigma_{2j}^2 = 0.02$, $\sigma_{12j} = 0$ *for all j.*

Approximating $Q_{x1,}Q_{x2}$ and Q_A with incomplete information

The covariance vector σ_{12j} that is necessary for correct specification of $Q_{x1,}Q_{x2}$ and Q_A can only be calculated from the individual participant data. If this information is not available, one solution would be to ensure that σ_{12j} is zero, by estimating the genetic associations with each exposure *and* the outcome in separate samples. This would correspond to a `three-sample' summary data MR-analysis when two exposures constitute the MVMR analysis.

Another pragmatic solution would be to assume that each σ_{12j} term is zero. This will give a good approximation for Q_{x1} and Q_{x2} whenever $\delta\sigma_{12j}$ is small and for Q_A whenever $\hat{\beta}_1\hat{\beta}_2\sigma_{12j}$ is small.

Application to education, cognitive ability and Body Mass Index

In this section we apply all of the methods discussed above to investigate whether there is evidence for a causal effect of education and cognitive ability on body mass index (BMI) using data from UK biobank. Education and cognitive ability have both been previously associated with BMI, with higher levels of education and cognitive ability being associated with lower levels of BMI However, there is also a high level of correlation observed between

completed education and measured cognitive ability, therefore it is not clear whether, once this correlation has been controlled for, both education and cognitive ability have a causal effect on BMI.

Data

UK biobank recruited 502,641 individuals aged 37-73 years between 2006 and 2010 from across the UK. Individuals where invited to a clinic where they answered a questionnaire and interview about a range of health topics and provided anthropomorphic measurements and gave samples of blood, urine and saliva. This study has been described in full previously.[20]

Individuals in UK biobank were asked to report the highest educational qualification they had obtained. For each individual we assigned an age at which they left education based their reported qualification. A breakdown of educational qualifications and associated ages across the cohort is given in Table A.3.

Cognitive ability was measured among a subset of the UK biobank participants as the number of correct answers recorded in a series of 13 questions designed to measure cognitive ability that where completed as part of the initial clinic. The cognitive ability variable was then standardised to have mean zero and variance 1. BMI was calculated based on the height and weight of the individuals in the sample. Throughout the analysis we analysed this variable on the natural log scale because of its skewed distribution.

Analysis

We first conducted MR analyses for the effect of education and cognitive ability on BMI separately using single variable MR. A single composite instrument for education was created using the polygenic score of 74 SNPs from a recent GWAS of educational attainment.[21] A single composite instrument for cognitive ability was created using the polygenic score of 18 SNPs from a recent GWAS of cognition.[22] As this GWAS was conducted using the interim release of UK Biobank we restricted our analysis to individuals not included in the interim release.

We then conducted a multivariable MR analysis of the effect of education and cognitive ability on BMI. This analysis included both the composite instruments for education and cognitive ability used in the single variable MR analyses.

The results from this analyses, along with a multivariable OLS regression of BMI on education and cognitive ability, are given in Table 4. The OLS results show that each extra year of education is associated with a decrease in BMI MR and MVMR results suggests a causal effect in the same direction, but with a larger magnitude. The results for cognitive ability are more mixed with no association seen in the OLS results, a negative total effect of

cognitive ability on BMI in the MR analysis and potentially a positive direct effect of cognitive ability on BMI observed in the MVMR analysis. Our empirical and theoretical investigation helps to clarify why the the high level of correlation between education and cognitive ability would lead to the conclusion that there is a negative effect of cognitive ability on BMI in MR analysis. The MVMR results show that, if anything, the direct effect of increasing cogntive ability is to increase BMI. These results highlight the potential benefits of MVMR. However, before giving much credence to this result it is necessary to assess the strength of our SNPs to jointly predict education and cognitive ability.

Table 4 – The effect of education and cognitive ability on BMI

		OLS	**Single variable MR**		**Multivariable MR**
Age completed Education	*Effect*	-0.008	-0.028		-0.044
	Std.Error	*(0.0003)*	*0.005*		*0.013*
	95% C.I.	*[-0.0085, -0.0074]*	*[-.0391 -.0179]*		*[-.0704 -.0187]*
	F-statistic		*188.2*		*195.0*
	Partial F-statistic				*35.7*
Standardised cognitive ability Score	*Effect*	0.0001		-0.023	0.048
	Std.Error	*(0.0007)*		*0.008*	*0.025*
	95% C.I.	*[-0.0013, 0.0014]*		*[-.0380 -.0082]*	*[-0.001 0.098]*
	F-statistic			*542.2*	*309.7*
	Partial F-statistic				*37.0*

Dependent variable is log(BMI).
Estimates of the effect of education and cognitive ability on BMI from OLS, single variable MR and multivariable MR analysis of individual level data.
All regressions also include a full set of control variables: age, gender, income and 10 genetic principal components
Instruments are constructed from GWAS scores for education and cognitive ability. The regressions are weighted so that individuals who left school at 15 are given an 80% upweighting. All non-European and related individuals have been excluded from the analysis. Total sample size included in all regressions: 74,309.

Testing the instrument strength in the single sample setting

As a measure of the strength of the instruments we calculate the standard F-statistic for both education and cognitive ability and the Sanderson – Windmeijer partial F-statstic[16] for the multivariable MR analysis. As all F-statistics are much larger than the rule-of-thumb cut off of 10 we are reassured that the instruments are not individually weak. However, the partial F-statistic for both education and cognitive ability is significantly lower, showing that the power of the instruments to predict both variables simultanously is greatly reduced.

The Sargan test for invalid instruments can only be calculated for estimation models with more instruments than exposure variables. In this estimation we have two exposure and two instruments and so it is not possible to calculate the Sargan statistic.

Two-Sample Multivariable MR

To illustrate two-sample MVMR we randomly divided the sample used for the individual analysis into three equal-sized groups. For each SNP used in the polygenic score, we then calculated its effect on log(BMI), education and cognitive ability using different parts of the sample., The results were then used to conduct a two-sample MVMR analysis. The results are given in Table A.4. They show that increased education has a direct effect which decreases BMI and cognitive ability has no direct effect on BMI. The results are in line with those obtained from the individual level analysis.

Testing instrument strength in the two-sample setting

To test for weak instruments in this analysis we have calculated the weak-instrument Q statistics for education and cognitive ability. The Q_{edu} statistic for education is 1724.4. The Q_{cog} statistic for cognitive ability is 1488.8. The critical value for a χ^2 distribution with 88 degrees of freedom at the 5% level is 110.9. Therefore we reject the null hypothesis that the SNPs do not explain any of the variation in the exposures education and cognitive ability in this two sample analysis and can conclude that these SNPs can predict both education and cognitive ability in the data.

Testing for pleiotropy in the two-sample setting.

To illustrate the two tests for pleiotropy discussed earlier we report the Q_A statistic for MVMR. The value of Q_A for this regression is 129.5. The critical value for a χ^2 distribution with 87 degrees of freedom is 109.77. Therefore, the null hypothesis that there is no heterogeneity is rejected for this value of Q_A.

Multivariable MR Egger regression

An alternative procedure that has been recently proposed to adjust for pleiotropy beyond that explainable by genetically predictable exposures (e.g. X_1 and X_2) is a Multivariable MR Egger regression[23] This is a natural extension of the original MR Egger approach[24] and is calculated by fitting the two sample MVMR model with a constant included;

$$\hat{\Gamma}_j = \beta_0 + \beta_1 \hat{\pi}_{1j} + \beta_2 \hat{\pi}_{2j} + U_{Yj} + v_{Yj}$$

If the constant is different from zero this suggests that additional pleiotropy is meaningfully biasing the analysis. However a generalisation of the InSIDE assumption is required in order for it to deliver unbiased causal estimates. These are described in detail elsewhere.[24]

The two sample results were used to fit multivariable MR Egger regression, the results of which are given in Table A.5. Its constant intercept parameter is estimated to be small, and consequently the estimated effects of the exposures do not differ from those in the two-sample MVMR estimation. This supports the suggestion that the SNPs do not exert a direct effect on BMI apart from through education or cognitive ability. As MR-Egger is dependent on the orientation of the SNP exposure associations, we repeated this analysis with the associations orientated so that the SNP education associations where all positive and then with the SNP cognitive ability associations all positive. These changes had no substantive effect on the results obtained.

The difference between the Q-statistic and Multivariable MR Egger estimation suggest an inconsistency between these two tests however this may have arisen due to a high level of variation in the effect of the SNPs on each exposure leading to a higher Q statistic. This is supported by Figure. 7a and 7b which gives individual MR plots for each exposure, and shows that there is a large amount of variation of the SNPs on each of the exposures. Repeating this analysis with the outlying SNP excluded makes no substantive difference to the results obtained.

The MVMR Egger analysis was repeated using the effect of each SNP on education, cognitive ability and BMI taken from GWAS estimates.[21,22,25] The magnitude of the estimated effects differ in this analysis as the outcome variable is BMI rather than the natural log of BMI, however these results also show no pleiotropic effect of the SNPs on the outcome and a negative effect of higher education on BMI. Results from this analysis are given in Table A5.

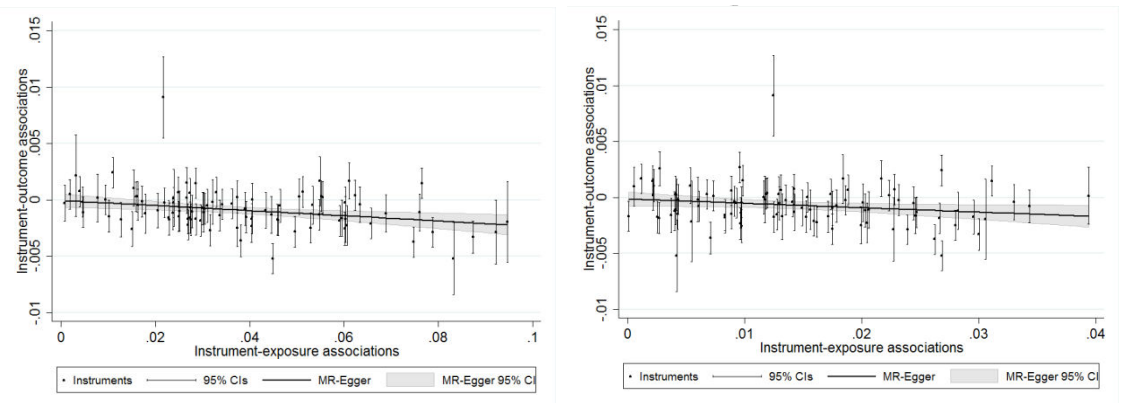

Figure 7: *Left: MR Egger plot for the association between educational attainment and BMI. Right: MR Egger plot for the association between cognitive ability and BMI. All SNPs that affect either education or cognitive ability are included*

Discussion

In this paper we have attempted to explain the principled application and interpretation of instrumental variable analysis to the epidemiological setting with multiple exposures. We first focused on the individual data setting, for which it is possible to borrow well-established methods (and related software) from the econometrics literature. We then clarified how these methods can be faithfully transcribed to the two-sample summary data setting, with particular attention paid to assessing the validity and relevance of the genetic instruments. In particular, we propose

- Modified Q statistics, (in our case Q_{x1} , and Q_{x2}) that detect `good' heterogeneity if a set of SNPs can jointly and reliably predict all intermediate exposures of interest;
- A modified Q statistic, Q_A that detects `bad' heterogeneity if a set of SNPs contains invalid instruments.

We finally illustrated the application of MVMR using individual and summary level data to estimate the effect of education and cognitive ability on BMI. The results from this analysis show that increasing education leads to lower BMI and the size of this effect increases when cognitive ability is controlled for. Comparing the single exposure MR analysis results (with all SNPs that affect educational attainment excluded) to the MVMR results for cognitive ability shows a large change in the size and direction of the effect. This result suggests that education is a mediator of the relationship between cognitive ability and BMI and any direct effect of cognitive ability is minimal.

Our applied results highlight that even when the instruments appear to be very strong for each of the exposures individually, this does not guarantee that they will be equally as strong for the exposures when estimated jointly in a MVMR model. For example, the F-statistics decrease from 195 and 310 to 36 and 37 for educational attainment and cognitive ability respectively.

References

1 Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. *International Journal of Epidemiology* **32**, 1-22, doi:10.1093/ije/dyg070 (2003).

2 Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in medicine* **27**, 1133-1163 (2008).

3 Wooldridge, J. M. *Econometric analysis of cross section and panel data*. (MIT press, 2010).

4 Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic epidemiology* **37**, 658-665 (2013).
5 Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human molecular genetics* **23**, R89-R98 (2014).
6 Sargan, J. D. The Estimation of Economic Relationships using Instrumental Variables. *Econometrica* **26**, 393-415, doi:10.2307/1907619 (1958).
7 Cochran, W. G. The combination of estimates from different experiments. *Biometrics* **10**, 101-129 (1954).
8 Bowden, J. *et al.* Improving the accuracy of two-sample summary data Mendelian randomization: moving beyond the NOME assumption. *bioRxiv*, 159442 (2017).
9 Greco, M., Del, F., Minelli, C., Sheehan, N. A. & Thompson, J. R. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Statistics in medicine* **34**, 2926-2940 (2015).
10 Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *American journal of epidemiology* **181**, 251-260 (2015).
11 Burgess, S., Freitag, D. F., Khan, H., Gorman, D. N. & Thompson, S. G. Using Multivariable Mendelian Randomization to Disentangle the Causal Effects of Lipid Fractions. *PLOS ONE* **9**, e108891, doi:10.1371/journal.pone.0108891 (2014).
12 Burgess, S. *et al.* Dissecting Causal Pathways Using Mendelian Randomization with Summarized Genetic Data: Application to Age at Menarche and Risk of Breast Cancer. *Genetics*, doi:10.1534/genetics.117.300191 (2017).
13 Relton, C. L. & Davey Smith, G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *International journal of epidemiology* **41**, 161-176 (2012).
14 Burgess, S. & Thompson, S. G. Avoiding bias from weak instruments in Mendelian randomization studies. *International Journal of Epidemiology* **40**, 755-764, doi:10.1093/ije/dyr036 (2011).
15 Munafo, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Smith, G. D. Collider Scope: How selection bias can induce spurious associations. *bioRxiv*, 079707 (2016).
16 Sanderson, E. & Windmeijer, F. A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics* **190**, 212-221 (2016).
17 Stock, J. & Yogo, M. in *Identification and Inference for Econometric Models* (ed Donald W.K. Andrews) 80-108 (Cambridge University Press, 2005).
18 Murray, M. P. Avoiding invalid instruments and coping with weak instruments. *The journal of economic perspectives* **20**, 111-132 (2006).
19 Windmeijer, F., Farbmacher, H. & Davies, N. On the Use of the Lasso for Instrumental Variables Estimation with Some Invalid Instruments. (2017).
20 Collins, R. What makes UK Biobank special? *The Lancet* **379**, 1173-1174 (2012).
21 Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539 (2016).
22 Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics* (2017).
23 Rees, J., Wood, A. & Burgess, S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Statistics in Medicine* (2017).
24 Bowden, J., Smith, G. D., Burgess, S. & Unit. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* **44**, 512–525 (2015).
25 Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-948, doi:10.1038/ng.686 (2010).

Table A.1 – Methods of estimation used in the simulations.

OLS estimation			
Single variable regression	$y = \alpha + \beta_1 x_1$	$y = \alpha + \beta_2 x_2$	
Multivariable regression	$y = \alpha + \beta_1 x_1 + \beta_2 x_2$		
Single sample methods of estimation			
1. Single variable MR	$y = \alpha + \beta_1 \hat{x}_1 + u$ $x_1 = \gamma_1 + \pi_1 G + v_1$	$y = \alpha + \beta_2 \hat{x}_2 + u$ $x_2 = \gamma_1 + \pi_2 G + v_2$	Individual single variable Mendelian Randomisation for each exposure variable, each predicted using all of the genetic variants available.
2. Multivariable MR	$y = \alpha + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2 + u$ $x_1 = \gamma_1 + \pi_1 G + v_1$ $x_2 = \gamma_2 + \pi_2 G + v_2$		Multivariable Mendelian Randomisation with both the exposures included and predicted using all of the genetic variants available.
3. Single variable MR Using a subset of the SNPs available.	$y = \alpha + \beta_1 \hat{x}_1 + u$ $x_1 = \gamma_1 + \pi_1 G_1 + v_1$	$y = \alpha + \beta_1 \hat{x}_1 + u$ $x_1 = \gamma_1 + \pi_1 G_2 + v_1$	Individual single variable Mendelian Randomisation for each exposure variable, each predicted using only the SNPs known to be associated with only that exposure.
Two-sample methods of estimation			
4. Single variable MR	$\hat{\Gamma} = \beta_1 \hat{\pi}_1 + u$ $y = \Gamma_j G_j + \epsilon_j$ $x_1 = \pi_{1,j} G_j + v_{1,j}$	$\hat{\Gamma} = \beta_2 \hat{\pi}_2 + u$ $y = \Gamma_j G_j + \epsilon_j$ $x_2 = \pi_{2,j} G_j + v_{2,j}$	Individual single variable Mendelian Randomisation for each exposure variable, each predicted using all of the genetic variants available.
5. Multivariable MR	$\hat{\Gamma} = \beta_1 \hat{\pi}_1 + \beta_2 \hat{\pi}_2 + u$ $y = \Gamma_j G_j + \epsilon_j$ $x_1 = \pi_{1,j} G_j + v_{1,j}$ $x_2 = \pi_{2,j} G_j + v_{2,j}$		Multivariable Mendelian Randomisation with both exposure variables included and predicted using all of the genetic variants available.
6. Single variable MR Using a subset of the SNPs available.	$\hat{\Gamma} = \beta_1 \hat{\pi}_1 + u$ $y = \Gamma_j G_{1,j} + \epsilon_j$ $x_1 = \pi_{1,j} G_{1,j} + v_{1,j}$	$\hat{\Gamma} = \beta_2 \hat{\pi}_2 + u$ $y = \Gamma_j G_{2,j} + \epsilon_j$ $x_2 = \pi_{2,j} G_{2,j} + v_{2,j}$	Individual single variable Mendelian Randomisation for each exposure variable, each

			predicted using only the SNPs known to be associated with only that exposure.

Table A.2 –Simulation results

Method of Estimation	$\widehat{\beta}_1$	**std. error** $\widehat{\beta}_1$	$\widehat{\beta}_2$	**std. error** $\widehat{\beta}_2$
Confounding Setup				
OLS estimation				
Single variable	1.88	0.046	2.31	0.067
Multivariable	1.25	0.032	1.13	0.044
Single Sample MR estimation				
1. Univariate MR	1.53	0.132	1.84	0.132
2. MVMR	1.00	0.017	1.00	0.022
3. Univariate MR – subset of SNPS	1.01	0.044	1.51	0.045
Two Sample MR estimation				
4. Univariate MR	1.52	0.141	1.82	0.144
5. MVMR	0.99	0.072	0.99	0.091
6. Univariate MR – subset of SNPS	0.99	0.088	1.49	0.092
Collider setup				
OLS estimation				
Single variable	1.36	0.031	0.71	0.019
Multivariable	0.65	0.054	0.43	0.032
Single Sample MR estimation				
1. Univariate MR	1.00	0.016	0.51	0.084
2. MVMR	0.99	0.028	0.01	0.018
3. Univariate MR – subset of SNPS	1.00	0.025	0.02	0.045
Two Sample MR estimation				
4. Univariate MR	0.99	0.038	0.50	0.087
5. MVMR	1.00	0. 066	0.00	0.040
6. Univariate MR – subset of SNPS	0.99	0.055	0.00	0.045
Pleiotropic setup				
OLS estimation				
Single variable	1.81	0.063	1.81	0.067
Multivariable	1.25	0.032	1.25	0.033
Single Sample MR estimation				
1. Univariate MR	1.34	0.129	1.34	0.133
2. MVMR	1.00	0.018	1.00	0.018
3. Univariate MR – subset of SNPS	1.00	0.045	1.01	0.045
Two Sample MR estimation				
4. Univariate MR	1.32	0.137	1.33	0.138
5. MVMR	0.99	0.060	0.99	0.060
6. Univariate MR – subset of SNPS	0.99	0.072	0.99	0.073
Mediation setup				
OLS estimation				
Single variable	2.31	0.063	1.88	0.048
Multivariable	1.12	0.043	1.25	0.033
Single Sample MR estimation				
1. Univariate MR	1.84	0.129	1.54	0.133
2. MVMR	1.00	0.021	1.00	0.018
3. Univariate MR – subset of SNPS	1.51	0.045	1.01	0.045
Two Sample MR estimation				
4. Univariate MR	1.82	0.142	1.52	0.142
5. MVMR	0.99	0.091	0.99	0.073
6. Univariate MR – subset of SNPS	1.49	0.090	0.99	0.088

n= 20,000. 1000 repetitions.

.

Table A.3 – Educational qualifications

Highest Educational qualification	Age completed education	% of final sample
None	15	12.29
CSE's/O levels/GCSEs	16	27.61
NVQ/HND/HNC	18	6.41
A levels	18	12.23
Other professional qualification (e.g. Nursing/Teaching etc)	20	4.98
College or University degree	21	36.47

The highest reported educational qualification and associated age for completing education for the individuals from UK biobank included in this analysis.

Table A.4 – Two-sample Multivariable MR estimation

	Effect	Std. Error	95% Confidence Interval	P-value
Log BMI				
Age completed Education	-0.022	0.006	[-0.034, -0.010]	<0.001
Standardised Cognitive ability score	0.007	0.016	[-0.026, 0.039]	0.684

Estimates of the effect of Education and cognitive ability on Log BMI from a two-sample analysis
3 SNPs in the education GWAS which are in LD with SNPs from the cognitive ability GWAS have been excluded.
Each sample includes on third of the observations in the total sample
The effects of each SNP on log BMI, education, cognitive ability have each been calculated from one sub–sample only.

Table A.5 - Multivariable MR Egger estimation

	Effect	Std. Error	95% Confidence Interval	P-value
Biobank data - Log BMI				
Age completed Education	-0.022	0.006	[-0.033 -0.010]	<0.001
Standardised Cognitive ability score	0.007	0.016	[-0.025 0.039]	0.667
Constant	-0.0003	0.0002	[-0.0008 0.0001]	0.166
Summary GWAS data – BMI				
Age completed Education	-0.345	0.098	[-0.542 -0.146]	0.001
Standardised Cognitive ability score	0.046	0.073	[-0.102 0.194]	0.532
Constant	-0.001	0.001	[-0.004 0.001]	0.254

Multivariable MR Egger estimates for the effect of education and cognitive ability on BMI.
The first section shows the estimated effects calculated using UK biobank. the estimation has been constructed in the same way as given in Table A.4.
The second section shows the results calculated using summary GWAS data for education, cognitive ability and BMI [37,38] [39],