IEA

METHODS

Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome

Stephen Burgess*

Department of Public Health and Primary Care, University of Cambridge

*Corresponding author. Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, 2 Worts Causeway, Cambridge, CB1 8RN, UK. E-mail: sb452@medschl.cam.ac.uk

Abstract

Background: Sample size calculations are an important tool for planning epidemiological studies. Large sample sizes are often required in Mendelian randomization investigations.
Methods and results: Resources are provided for investigators to perform sample size and power calculations for Mendelian randomization with a binary outcome. We initially provide formulae for the continuous outcome case, and then analogous formulae for the binary outcome case. The formulae are valid for a single instrumental variable, which may be a single genetic variant or an allele score comprising multiple variants. Graphs are provided to give the required sample size for 80% power for given values of the causal effect of the risk factor on the outcome and of the squared correlation between the risk factor and instrumental variable. R code and an online calculator tool are made available for calculating the sample size needed for a chosen power level given these parameters, as well as the power given the chosen sample size and these parameters.
Conclusions: The sample size required for a given power of Mendelian randomization investigation depends greatly on the proportion of variance in the risk factor explained by the instrumental variable. The inclusion of multiple variants into an allele score to explain more of the variance in the risk factor will improve power, however care must be taken not to introduce bias by the inclusion of invalid variants.

Key words: Mendelian randomization, sample size, power, binary outcome, allele score

---

Key Messages

- Resources are provided for investigators to perform sample size and power calculations for Mendelian randomization with a binary outcome.
- The sample size required for a given power level is greater with a binary outcome than a continuous outcome, and is highly dependent on the proportion of the variance in the risk factor explained by the instrumental variable.

Introduction

Sample size calculations are an important part of experimental design. They inform an investigator of the expected power of a given analysis to reject the null hypothesis. If the power of an analysis is low, then not only is the probability of rejecting the null hypothesis low, but when the null hypothesis is rejected, the posterior probability that the rejection of the null hypothesis is not simply a chance finding is low.[1]

Mendelian randomization is the use of genetic variants as instrumental variables for assessing the causal effect of a risk factor on an outcome from observational data.[2] Genetic variants are chosen which are specifically associated with a risk factor of interest, and not associated with variables which may be confounders of the association between the risk factor and outcome.[3] Such a variant divides the population into groups which are similar to treatment arms in a randomized controlled trial.[4] Under the instrumental variable assumptions,[5,6] a statistical association between the genetic variant and the outcome implies that the risk factor has a causal effect on the outcome.[7] However, as genetic variants typically explain a small proportion of the variance in risk factors, the power to detect a significant association between the variant and outcome in an applied Mendelian randomization context can be low.[8] Sample size analysis is particularly important to inform whether a null finding is representative of a true null causal relationship, or simply a lack of power to detect an effect size of clinical interest.

Sample size calculations have been previously presented for Mendelian randomization experiments with continuous outcomes. Calculations based on asymptotic statistical theory have been presented with a single instrumental variable (IV), whether that IV is a single genetic variant or an allele score.[9] An allele score (also called a genetic risk score) is a single variable summarizing multiple genetic variants as a weighted or unweighted sum of risk factor-increasing alleles.[10] A simulation study for estimating power has also been presented with both single and multiple IVs.[11] These approaches have shown good agreement. However, in many cases, the outcome in a Mendelian randomization experiment is binary (dichotomous), such as disease. In this paper, we present power calculations for Mendelian randomization studies with a binary outcome. We assume the context of a case-control study where the causal parameter of interest is an odds ratio, although the calculations are also valid for other study designs.

Methods and Results

We give results for the asymptotic variance of IV estimators with a single IV, and for the resulting sample size needed in a Mendelian randomization study to obtain a given power level. We initially present formulae with a continuous outcome (this reviews material previously covered by Freeman *et al.*[9]) and then analogous formulae with a binary outcome. We concentrate on estimates from the ratio (or Wald) method, as this method makes few parametric assumptions, relying only on a linear relationship between the conditional expectation of the outcome (or in the binary case, the logistic function of the probability of the outcome) and the risk factor.[12] If the imprecision in the estimate of the genetic association with the risk factor is negligible, then estimates of power and sample size from the ratio method also correspond to those from assessment of the causal relationship of the risk factor on the outcome by testing the association between the genetic variant and outcome.

Other estimation approaches are possible with a binary outcome[13] but these either give equivalent estimates to the ratio method with a single IV (the two-stage predictor substitution method[14]) or are not recommended for general use in applied practice. These include the two-stage residual inclusion method, due to inconsistency for a parameter with a natural interpretation,[15] and the generalized method of moments (GMM) and structural mean models (SMM) methods, due to potential lack of identifiability of the causal parameter (S Burgess *et al.*, unpublished data).

Power with a continuous outcome

With a single IV and a continuous outcome, the IV estimates from the ratio, two-stage least squares (2SLS) and limited information maximum likelihood (LIML) methods coincide.[16] The estimator can be expressed as the ratio between the coefficient from the regression of the outcome (Y) on the genetic variant (G), divided by the coefficient from the regression of the risk factor (X) on the variant:

$$\hat{\beta}_{IV} = \frac{\hat{\beta}_{GY}}{\hat{\beta}_{GX}}. \tag{1}$$

The asymptotic variance of this IV estimator is given by the formula:

$$\mathrm{var}(\hat{\beta}_{IV}) = \frac{\mathrm{var}(R_Y)}{N\mathrm{var}(X)\rho_{GX}^2} \tag{2}$$

where $R_Y = Y - \beta_1$. X is the residual of the outcome on subtraction of the causal effect of the risk factor, and ρ_{GX}^2 is the square of the correlation between the risk factor X and the IV G.[17] The coefficient of determination (R^2) in the regression of the risk factor on the IV is an estimate of ρ_{GX}^2. The IV in these calculations could either be a single genetic variant or an allele score.[10]

The asymptotic variance of the conventional regression (ordinary least squares, OLS) estimator of the association

between the risk factor X and the outcome Y is given by the formula:

$$\text{var}(\hat{\beta}_{\text{OLS}}) = \frac{\text{var}(R_Y)}{N\,\text{var}(X)}. \quad (3)$$

The sample size necessary for an IV analysis to demonstrate a non-zero association for a given magnitude of causal effect is therefore approximately equal to that for a conventional epidemiological analysis to demonstrate the same magnitude of association divided by the ρ^2_{GX} value for the IV.[18] If the significance level is a and the power desired to test the null hypothesis is $1-\beta$, then the sample size required to test a causal effect of size β_1 using IV analysis is:[9]

$$\text{Sample size} = \frac{(z_{(1-\frac{\alpha}{2})} + z_\beta)^2 \text{var}(R_Y)}{\text{var}(X)\beta_1^2 \rho^2_{\text{GX}}} \quad (4)$$

where z is a quantile function, so that z_α is the 100α percentile point on the standard normal distribution. If the significance level is 0.05 and the power is 0.8, then the sample size to test for a change of β_1 standard deviations in Y per standard deviation increase in X is:

$$\text{Sample size} = \frac{7.848}{\beta_1^2 \rho^2_{\text{GX}}}. \quad (5)$$

For a given sample size N, the power to detect a causal effect (in the same direction the true effect) can be calculated as:

$$\Phi(\beta_1 \rho_{\text{GX}} \sqrt{N} - z_{(1-\frac{\alpha}{2})}) \quad (6)$$

where Φ is the cumulative distribution function of the standard normal distribution. This is the inverse function of the quantile function ($\Phi(z_\alpha) = \alpha$).

We use these formulae to construct power curves for Mendelian randomization using a significance level of 0.05. In Figure 1 (left), we fix the squared correlation ρ^2_{GX} at 0.02, meaning the variant explains on average 2% of the variance of the risk factor, and vary the size of the effect $\beta_1 = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3$ and the sample size $N = 1000$ to $10\,000$. In Figure 1 (right), we fix the size of the effect at $\beta_1 = 0.2$ and vary the squared correlation $\rho^2_{\text{GX}} = 0.005, 0.01, 0.015, 0.02, 0.025, 0.03$ and the sample size as before. In each of the figures, the power to detect a positive causal relationship is displayed; this tends to 0.025 as the sample size tends to zero. We see that the power increases as the causal effect increases, and as the IV explains more of the variance in the risk factor (the ρ^2_{GX} parameter or the expected value of the R^2 statistic increases).

Similar formulae to these have been made available in an online tool for calculating either power for a given sample size or sample size needed for a given power, taking the causal effect (β_1) and squared correlation (ρ^2_{GX}) parameters, as well as the variance of the risk factor and outcome, and the observational (OLS) coefficient of the risk factor from regression on the outcome.[19]

Power with a binary outcome

With a single IV and a binary outcome, the same IV estimator[1] as in the continuous outcome case can be evaluated, except that a logistic model is typically used in the regression of the outcome on the genetic variant.[12] The asymptotic variance of this estimator can be approximated
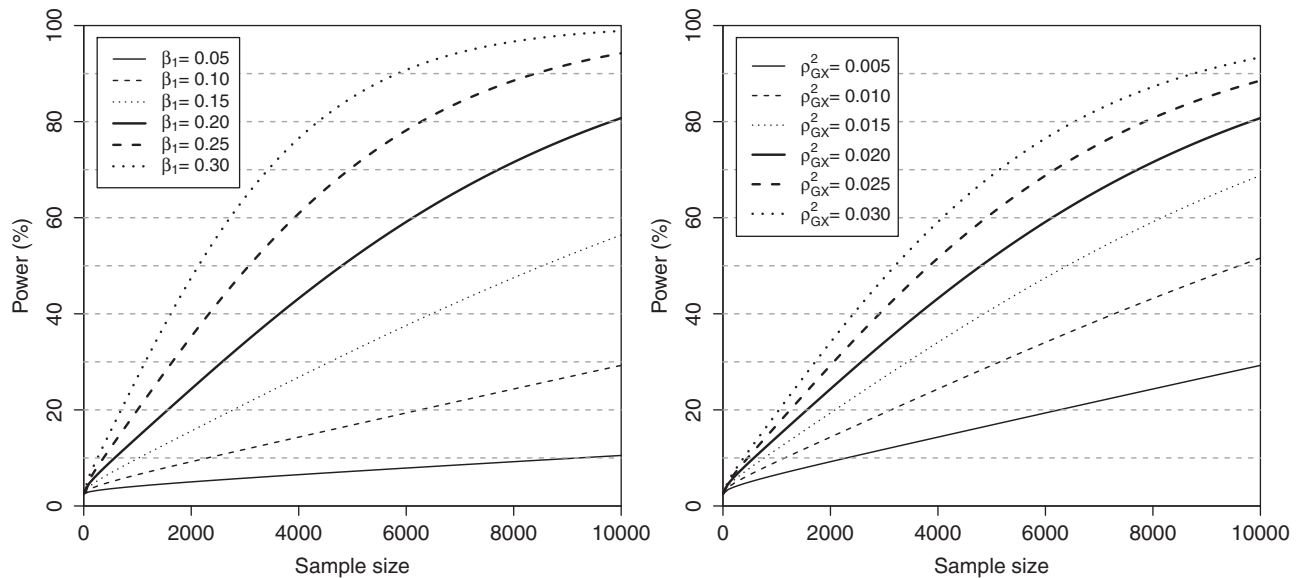

Figure 1. Power curves varying the sample size with continuous outcome and a single instrumental variable. Left panel: for a fixed value of the IV strength ($\rho^2_{\text{GX}} = 0.02$) and different values of the size of the causal effect ($\beta_1 = 0.05, 0.1,..., 0.3$). Right panel: for a fixed value of the causal effect ($\beta_1 = 0.2$) and varying the size of the IV strength ($\rho^2_{\text{GX}} = 0.005, 0.01,..., 0.03$)

using the delta method for the ratio of two estimates.[20] The leading term in the expansion is:

$$\text{var}(\hat{\beta}_{\text{IV}}) = \frac{\text{var}(\hat{\beta}_{\text{GY}})}{\hat{\beta}_{\text{GX}}}. \quad (7)$$

Although further terms from the delta method could be included, these are usually much smaller in magnitude. In the simulation example later in the paper, if the association between the risk factor and IV is estimated using data on the entire sample of control participants, the second and third terms in the expansion are two orders of magnitude smaller than the leading term (Figure 1). The asymptotic variance of the coefficient $\hat{\beta}_{\text{GY}}$ from logistic regression is:

$$\text{var}(\hat{\beta}_{\text{GY}}) = \frac{1}{\mathbb{E}\left[\sum_i g_i^2 \mathbb{P}(Y=1|G=g_i)\mathbb{P}(Y=0|G=g_i)\right]} \quad (8)$$

where i indexes individuals. This expression is obtained by differentiation of the log-likelihood. If the probability of an event does not depend greatly on the value of the genetic IV, then $\mathbb{P}(Y=1 \mid G=g_i) \approx \mathbb{P}(Y=1)$ which is the ratio of cases to participants in the sample. This approximation will be reasonable if the genetic variant does not explain a large proportion of the variance in the risk factor, and/or the effect of the risk factor on the outcome is not extreme. We assume (without loss of generality) that the mean of G is 0 and the variance is 1, so that $\mathbb{E}(\Sigma_i\, g_i^2) = N$, where N is the sample size. The square of the coefficient $\hat{\beta}_{\text{GX}}$ is approximately equal to $\text{var}(X)\,\rho_{\text{GX}}^2$. This gives:

$$\text{var}(\hat{\beta}_{\text{IV}}) = \frac{1}{N\,\text{var}(X)\rho_{\text{GX}}^2\,\mathbb{P}(Y=1)\,\mathbb{P}(Y=0)}. \quad (9)$$

The sample size required to detect an effect of size β_1 per standard deviation increase in X for 80% power with a significance level of 0.05 is therefore

$$\text{Sample size} = \frac{7.848}{\beta_1^2 \rho_{\text{GX}}^2\, \mathbb{P}(Y=1)\,\mathbb{P}(Y=0)} \quad (10)$$

where the effect β_1 is a log odds ratio. If there are to be an equal number of cases and controls, $\mathbb{P}(Y=1) = \mathbb{P}(Y=0) = 0.5$, and:

$$\text{Sample size} = \frac{31.392}{\beta_1^2 \rho_{\text{GX}}^2}. \quad (11)$$

The corresponding power to detect a causal effect of size β_1 with a significance level of 0.05 is:

$$\Phi(\beta_1 \rho_{\text{GX}} \sqrt{(N\mathbb{P}(Y=1)\mathbb{P}(Y=0))} - 1.96). \quad (12)$$

Similar power curves to Figure 1 in the binary outcome setting are given in the Web Appendix (available as Supplementary data at *IJE* online).

We use these approximations to calculate the number of cases needed to obtain 80% power in a Mendelian randomization analysis with a binary outcome for different values of β_1 and ρ_{GX}^2, assuming a 1:1 ratio of cases to controls. The results are displayed in Figure 2. We note that when the genetic variants explain a small proportion of the variance in the risk factor, large sample sizes are required to detect even moderately large causal effects with reasonable power.

An R[21] script for performing sample size and power calculations is provided in the Appendix (available as Supplementary data at *IJE* online). This code enables the calculation of the sample size required for a chosen power level given the values of β_1 and ρ_{GX}^2, as well as the power given the values of β_1, ρ_{GX}^2 and the chosen sample size. A calculator using this code is available online.[22]

Validation simulation

In order to validate the estimates of sample size and power, we simulate data on a genetic variant, a continuous risk factor and an outcome. The data-generating model for individuals indexed by i is:

$$\begin{aligned} g_i &\sim N(0,1) \\ x_i &\sim N(g_i\rho_{\text{GX}}, 1 - \rho_{\text{GX}}^2) \\ y_i &\sim \text{Binomial}(1,\, \text{expit}(\beta_0 + \beta_1 x_i)) \end{aligned} \quad (13)$$

where $\text{expit}(x) = (\exp(x)\,/\,1 + \exp(x))$ is the inverse of the logit function and β_1 is the log odds ratio per unit (which here equals 1 standard deviation) increase in the risk factor. The genetic variant is modelled by a standard normal distribution; it can be regarded as a standardized weighted allele score. The parametric relationship between X, G and ρ_{GX} ensures that the proportion of variance in the risk factor explained by the instrumental variable in a large sample is ρ_{GX}^2. We also simulate data with a dichotomous risk factor; details are given in the Web Appendix (available as Supplementary data at *IJE* online).

We set $\beta_0 = -3$ so that the outcome has a prevalence of about 5% in the population from which the case-control sample is taken. We take three values of $\beta_1 = 0.1, 0.2, 0.3$, three values of $\rho_{\text{GX}}^2 = 0.01, 0.02, 0.03$, three sample sizes (10 000, 20 000, and 30 000 cases), and two values of the ratio of cases to controls (1:1 and 1:2). For each set of parameter values, we calculate the estimate of the power from equation (12) using a significance level of 0.05, and compare this with the number of times the 95% confidence interval for the ratio estimate excludes the null based on 10 000 simulated datasets.

The 95% confidence interval for the ratio method used in calculating the power of the simulation method is
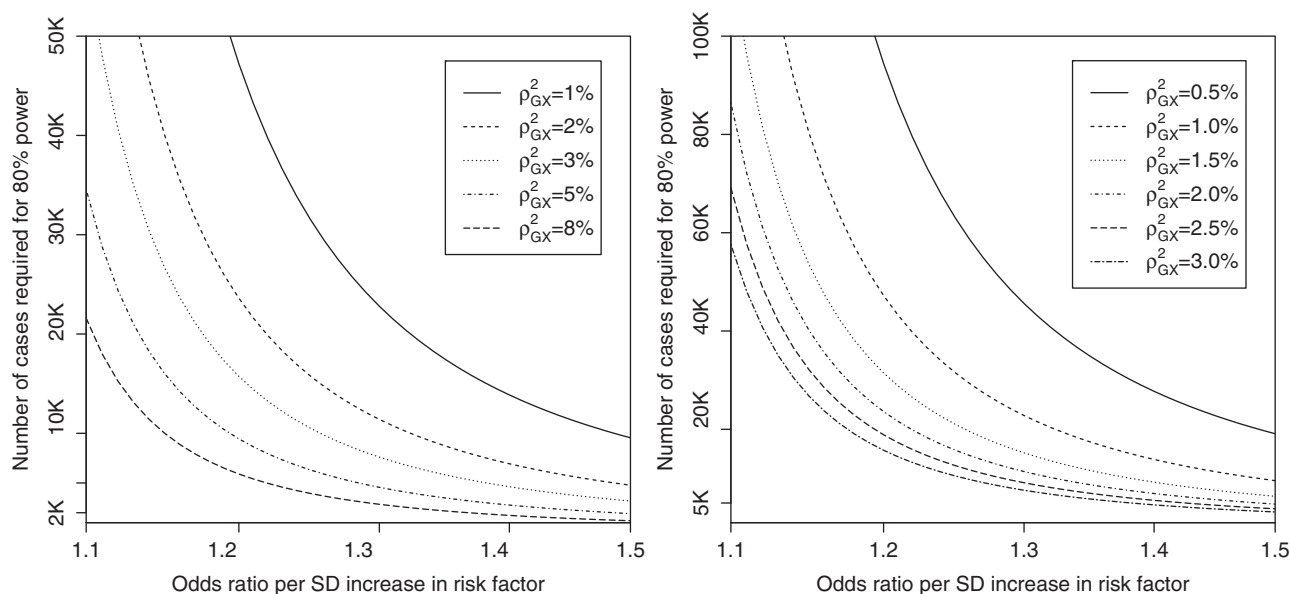
Figure 2. Number of cases required in a Mendelian randomization analysis with a binary outcome and a single instrumental variable for 80% power with a 5% significance level and 1:1 ratio of cases:controls varying the size of causal effect [odds ratio per standard deviation (SD) increase in risk factor, $\exp(\beta_1)$] for different values of IV strength. Left panel: $\rho^2_{GX} = 1\%$–8%. Right panel: $\rho^2_{GX} = 0.5\%$–3.0%

constructed using Fieller's method,[23] and so does not rely on the same asymptotic assumption as the analytical method for estimating the power. Previous simulations have shown that confidence intervals from Fieller's method maintain nominal coverage levels even with weak instruments.[16] To obtain a case-control sample of the necessary size, we initially simulate data for a large number of individuals, and then take the required number of cases and controls from this population.

Simulation results

Results from the validation simulation are given in Table 1. The Monte Carlo standard error (the expected variation from the true value due to the limited number of simulations) in the simulation estimates of power is at most 0.5%. The coverage levels of the 95% confidence interval from Fieller's method are close to 95% throughout (between 94.8 and 95.9 for the 54 scenarios).

We note that estimates of power from the formula of equation (12) are similar to those from the simulation approach. There is no apparent systematic bias in the estimates from the analytical formula, with simulation estimates being greater and less than those from the formula a similar number of times (when rounded to nearest 0.1%, the estimate from the simulation was less 24 times and greater 19 times). Estimates from both approaches are no more different than would be expected due to chance alone. Similar results are obtained with a dichotomous risk

factor; details are given in Web Table A1 (available as Supplementary data at *IJE* online). In comparing estimates of power with equal numbers of cases, greater power is achieved when there is a case:control ratio of 1:2 than with a ratio of 1:1. However, when the total sample size is fixed, the estimate of power is greatest when the numbers of cases and controls are equal. This can be seen by comparing estimates with 30 000 cases and a ratio of 1:1, and with 20 000 cases and a ratio of 1:2.

In response to concerns from a reviewer that the power estimates may not be valid with a discrete instrumental variable (such as a single nucleotide polymorphism) or when there is confounding, additional validation simulations were performed in these scenarios. Results are given in the Web Appendix (Web Tables A2–A4, available as Supplementary data at *IJE* online). No substantial differences were observed from the validation simulation in the main paper when the instrumental variable was discrete. When there was confounding, estimates from the analytical formula slightly overestimated power, particularly when the confounding was in the same direction as the causal effect. However, this overestimation was slight (on average less than 1% when the confounding was in the opposite direction, and less than 2% when the confounding was in the same direction). As the magnitude of confounding is not possible to estimate in applied practice, conservative estimates of the correlation and causal effect parameters used in power calculations are recommended, particularly if confounding is thought to be substantial.

Table 1. Validation simulation to compare estimates of power in a Mendelian randomization analysis with a continuous risk factor and a binary outcome from analytical formula and simulation study with a 5% significance level varying the size of causal effect (β_1), the IV strength (ρ_{GX}^2), the sample size and the ratio of cases to controls

Case:control ratio = 1:1		10000 cases		20000 cases		30000 cases	
		Formula	Simulation	Formula	Simulation	Formula	Simulation
$\rho_{GX}^2 = 0.01$	$\beta_1 = 0.1$	10.5%	10.2%	16.9%	16.6%	23.1%	22.4%
	$\beta_1 = 0.2$	29.3%	28.4%	51.6%	51.2%	68.8%	69.5%
	$\beta_1 = 0.3$	56.4%	56.4%	85.1%	85.0%	95.7%	95.7%
$\rho_{GX}^2 = 0.02$	$\beta_1 = 0.1$	16.9%	17.2%	29.3%	28.9%	41.0%	41.1%
	$\beta_1 = 0.2$	51.6%	51.0%	80.7%	80.2%	93.4%	93.6%
	$\beta_1 = 0.3$	85.1%	84.9%	98.9%	98.9%	99.9%	100.0%
$\rho_{GX}^2 = 0.03$	$\beta_1 = 0.1$	23.1%	22.9%	41.0%	40.8%	56.4%	57.0%
	$\beta_1 = 0.2$	68.8%	68.5%	93.4%	93.3%	98.9%	99.0%
	$\beta_1 = 0.3$	95.7%	95.5%	99.9%	99.9%	100.0%	100.0%

Case:control ratio = 1:2		10000 cases		20000 cases		30000 cases	
		Formula	Simulation	Formula	Simulation	Formula	Simulation
$\rho_{GX}^2 = 0.01$	$\beta_1 = 0.1$	12.6%	12.9%	21.0%	21.4%	29.3%	28.9%
	$\beta_1 = 0.2$	37.2%	37.5%	63.7%	64.4%	80.7%	81.1%
	$\beta_1 = 0.3$	68.8%	68.2%	93.4%	93.3%	98.9%	98.8%
$\rho_{GX}^2 = 0.02$	$\beta_1 = 0.1$	21.0%	21.2%	37.2%	37.8%	51.6%	51.6%
	$\beta_1 = 0.2$	63.7%	63.9%	90.4%	90.7%	97.9%	97.9%
	$\beta_1 = 0.3$	93.4%	93.2%	99.8%	99.8%	100.0%	100.0%
$\rho_{GX}^2 = 0.03$	$\beta_1 = 0.1$	29.3%	29.0%	51.6%	51.4%	68.8%	68.8%
	$\beta_1 = 0.2$	80.7%	80.8%	97.9%	97.7%	99.8%	99.9%
	$\beta_1 = 0.3$	98.9%	98.9%	100.0%	100.0%	100.0%	100.0%

Discussion

In this paper, we have provided information on sample sizes and power calculations in a Mendelian randomization analysis with a single IV and a binary outcome. We have shown in the continuous setting how the power depends on the magnitude of causal effect and the proportion of variance in the risk factor explained by the IV. With a binary outcome, the precision of the coefficient in the regression of the outcome on the IV is reduced compared with a continuous outcome, as the outcome can only take two values. As a result, the required sample sizes to obtain 80% power are much larger.

For a given applied example, the magnitude of the causal effect of a risk factor is fixed, as is the expected proportion of variance in the risk factor explained by each variant. However, the expected proportion of variance in the risk factor explained by the IV depends on the choice of IV. The required sample size for a given power level can be reduced (or equivalently the expected power at a given sample size can be increased) by including more genetic variants into the IV. This can be achieved by using multiple variants as separate IVs,[13] or as a single IV using an allele score approach. With an allele score, power can be further increased by the use of relevant weights for the variants.[10] Provided that weights are not derived naively from the data under analysis, the allele score approach avoids some of the problems of bias from weak instruments resulting from using many IVs.[24] A disadvantage of the inclusion of many variants in an IV analysis, whether in a multiple IV or an allele score model, is that one or more of the variants may not be a valid IV. If a variant is associated with a confounder of the risk factor–outcome association, or with the outcome through a pathway not via the risk factor of interest, then the estimate associated with this IV may be biased. If the function and relevance of some variants as IVs are uncertain, investigators will have to balance the risk of a biased analysis against the risk of an underpowered analysis. Sensitivity analysis may be a valuable tool to assess the homogeneity of IV estimates using different sets of variants.

If there are missing data, this may adversely impact the power of an analysis. When there are multiple genetic variants, individuals with sporadic missing genetic data can be included in an analysis using an imputation approach.[25]

This can minimize the impact of missing data on the power of the analysis, particularly if the distributions of genetic variants are correlated (the variants are in linkage disequilibrium).

The calculations in this paper make several assumptions. The distribution of the IV estimator is assumed to be well approximated by a normal distribution. This is known to be a poor approximation when the IV is weak;[26] however, if the IV is weak, then the power will usually be low. The standard deviation of this normal distribution is assumed to be close to the first-order term from the delta expansion. This term only involves the uncertainty in the coefficient from the genetic association with the outcome. The uncertainty in the estimate of the genetic association with the risk factor is not accounted for. Typically, this uncertainty will be small in comparison as the genetic association with the outcome is assumed to be mediated through the risk factor. Again, if this uncertainty is large, then the power of the analysis will usually be low. If a more precise estimate of the power is required, either further terms from the delta expansion could be used, or a direct simulation approach could be undertaken. The model of the logistic-transformed probability of an outcome event is assumed to be linear in the risk factor. As the power is very sensitive to the squared correlation term ρ_{GX}^2, it is advisable to take a conservative estimate of this parameter, or to perform a sensitivity analysis for a range of values of ρ_{GX}^2. Despite these approximations, the validation simulation suggests that estimates of sample size and power from the formulae in this paper will be close to the true values for a range of realistic values of the parameters involved.

The ratio method used in this paper has been criticized for use with binary outcomes to estimate an odds ratio.[27,28] This is due to the non-collapsibility of the odds ratio, meaning that the parameter estimate depends on the choice of covariate adjustment.[29] This is a general property of odds ratios, and not a specific feature of the ratio method. The estimate from the ratio method approximates a population averaged odds ratio,[15] and is close to a conditional odds ratio under certain specific circumstances.[30] The choice of odds ratio estimate does not affect the consistency of the estimator under the null.[31] As effect estimation is usually secondary to the demonstration of a causal effect, the precise identification of the parameter estimated by the ratio method is not of particular importance in Mendelian randomization analyses, and over-literal interpretation of Mendelian randomization estimates should be avoided even outside the odds ratio case.[32]

Although the sample sizes required in Mendelian randomization experiments are often large, it is not always necessary to measure the risk factor on all of the participants in a study. Simulations have shown that, in some cases,

90% of the power of the complete-data analysis can be obtained while only measuring the risk factor for 10% of participants.[33] This means that obtaining measurements of the risk factor, which may be expensive or impractical for a large sample, should not be the prohibitive factor for a Mendelian randomization investigation.

Supplementary Data

Supplementary data are available at *IJE* online.

Conflict of interest: None declared.

References

1. Davey Smith G, Ebrahim S. Data dredging, bias, or confounding. *BMJ* 2002;**325**:1437.
2. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32**:1–22.
3. Lawlor D, Harbord R, Sterne J, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008;**27**: 1133–63.
4. Nitsch D, Molokhia M, Smeeth L, DeStavola B, Whittaker J, Leon D. Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. *Am J Epidemiol* 2006;**163**:397–403.
5. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000;**29**:722–29.
6. Martens E, Pestman W, de Boer A, Belitser S, Klungel O. Instrumental variables: application and limitations. *Epidemiology* 2006;**17**:260–67.
7. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 2007;**16**(4):309–30.
8. Schatzkin A, Abnet C, Cross A. Mendelian randomization: how it can – and cannot – help confirm causal relations between nutrition and cancer. *Cancer Prev Res* 2009;**2**:104–13.
9. Freeman G, Cowling B, Schooling M. Power and sample size calculations for Mendelian randomization studies. *Int J Epidemiol* 2013;doi:10.1093/ije/dyt110.
10. Burgess S, Thompson S. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol* 2013;**42**: 1134–44.
11. Pierce B, Ahsan H, VanderWeele T. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol* 2011;**40**: 740–52.
12. Didelez V, Meng S, Sheehan N. Assumptions of IV methods for observational epidemiology. *Stat Sci* 2010;**25**:22–40.
13. Palmer T, Lawlor D, Harbord R *et al*. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res* 2011;**21**:223–42.
14. Cai B, Small D, Ten Have T. Two-stage instrumental variable methods for estimating the causal odds ratio: Analysis of bias. *Stat Med* 2011;**30**:1809–24.

15. Burgess S; CHD CRP Genetics Collaboration. Identifying the odds ratio estimated by a two-stage instrumental variable analysis with a logistic regression model. *Stat Med* 2013;**32**: 4726–47.

16. Burgess S, Thompson S. Improvement of bias and coverage in instrumental variable analysis with weak instruments for continuous and binary outcomes. *Stat Med* 15 2012;**31**:1582–600.

17. Nelson C, Startz R. The distribution of the instrumental variables estimator and its *t*-ratio when the instrument is a poor one. *J Business* 1990;**63**:125–40.

18. Wooldridge J. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2002.

19. Shakhbazov K. *mRnd: Power Calculations for Mendelian Randomization*. http://glimmer.rstudio.com/kn3in/mRnd (27 January 2014, date last accessed).

20. Thomas D, Conti D. Commentary: the concept of 'Mendelian Randomization'. *Int J Epidemiol* 2004;**33**:21–25.

21. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2011.

22. Burgess S. *Online Sample Size and Power Calculator for Mendelian Randomization With a Binary Outcome*. http://spark.rstudio.com/sb452/power/ (27 January 2014, date last accessed).

23. Buonaccorsi J. Fieller's theorem. In: Armitage P, Colton T (eds). *Encyclopedia of Biostatistics.*Hoboken NJ: Wiley, 2005.

24. Burgess S, Thompson S; CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* 2011;**40**:755–64.

25. Burgess S, Seaman S, Lawlor D, Casas J, Thompson S. Missing data methods in Mendelian randomization studies with multiple instruments. *Am J Epidemiol* 2011;**174**:1069–76.

26. Burgess S, Thompson S. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med* 2011;**30**:1312–23.

27. Palmer T, Thompson J, Tobin M, Sheehan N, Burton P. Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. *Int J Epidemiol* 2008;**37**:1161–68.

28. Palmer T, Sterne J, Harbord R *et al*. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. *Am J Epidemiol* 2011;**173**: 1392–403.

29. Greenland S, Robins J, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci* 1999;**14**:29–46.

30. Harbord R, Didelez V, Palmer T, Meng S, Sterne J, Sheehan N. Severity of bias of a simple estimator of the causal odds ratio in Mendelian randomization studies. *Stat Med* 2013;**32**: 1246–58.

31. Vansteelandt S, Bowden J, Babanezhad M, Goetghebeur E. On instrumental variables estimation of causal odds ratios. *Stat Sci* 2011;**26**:403–22.

32. Burgess S, Butterworth A, Malarstig A, Thompson S. Use of Mendelian randomisation to assess potential benefit of clinical intervention. *BMJ* 2012;**345**:e7325.

33. Pierce B, Burgess S. Efficient design for Mendelian randomization studies: subsample and two-sample instrumental variable estimators. *Am J Epidemiol* 2013;**178**:1177–84.