Supplementary Material

A.1 Technical discussion about estimation of indirect and direct effects

There are several versions of direct and indirect effects. We present definitions using counterfactual terminology, using potential values of the outcome $Y(x, m)$, representing the outcome which would be observed if the risk factor (or exposure) X were set (by intervention) to x and M were set to m, and potential values of the mediator $M(x)$, the value taken by the mediator if X were set to x. All effects are given on the difference scale; with a binary outcome, effects on a relative risk or odds ratio scale can also be defined, but the decomposition is more complex [VanderWeele and Vansteelandt, 2010; Kaufman, 2010]. This text is adapted from Burgess et al. [2015].

A total effect is defined as the effect of a change in the exposure from, say, $X = x$ to $X = x + 1$. It comprises the effects of the change in the exposure, and the change in the mediator as a result of the change in the exposure:

$$TE(x, x + 1) = Y(x + 1, M(x + 1)) - Y(x, M(x)) \tag{A1}$$

A controlled direct effect is defined as the effect of a change in the exposure keeping the mediator fixed at a given level, say $M = m$ [Robins and Greenland, 1992; Pearl, 2001]. The controlled direct effect may depend on the choice of m:

$$CDE(m; x, x + 1) = Y(x + 1, m) - Y(x, m) \tag{A2}$$

A natural direct effect is defined as the effect of a change in the exposure with the mediator fixed at the level it would naturally take if the exposure were fixed at a given level, say $X = x$:

$$NDE(x; x, x + 1) = Y(x + 1, M(x)) - Y(x, M(x)) \tag{A3}$$

A natural indirect effect is defined as the effect of a change in the mediator from the value it would naturally take if the exposure were unchanged to the level it would take if the exposure were changed. The exposure itself is kept fixed at a given level, say $X = x + 1$:

$$NIE(x + 1; x, x + 1) = Y(x + 1, M(x + 1)) - Y(x + 1, M(x)) \tag{A4}$$

In the linear case, the natural direct and indirect effects represent a decomposition of the total effect, in that $TE(x, x+1) = NDE(x; x, x+1) + NIE(x+1; x, x+1)$ (or alternatively $TE(x, x+1) = NDE(x+1; x, x+1) + NIE(x; x, x+1)$). Under the condition:

$$Y(x+1, m_1) - Y(x, m_1) = Y(x+1, m_2) - Y(x, m_2) \tag{A5}$$

for all values of $M = m_1, m_2$, and for all individuals, the controlled direct effect is equal to the natural direct effect [Robins and Greenland, 1992]. The natural direct effect has a clearer intuitive interpretation as a measure of mediation than the controlled direct effect. However, it is not possible to conceive of an experiment which would produce the natural direct effect, as the quantity requires the outcome if the exposure were set at two different levels (for example, in $NDE(x; x, x+1)$, $Y(x+1, M(x))$ requires $X = x+1$ for Y, but $X = x$ for M). This is known as a "cross-world" quantity, as setting the exposure to two different values is only possible in two different worlds [Richardson and Robins, 2013].

As we argue in Burgess et al. [2015], we would regard the controlled direct effect as the quantity that is targeted by mediation analysis with instrumental variables, as this is what would be obtained if we were to intervene separately on the risk factor and mediator. As we assume that all relationships between variables are linear and there is no effect heterogeneity, the natural and controlled direct effects are equal, and hence we refer to a 'direct effect' throughout this manuscript without further qualification.

A.2 Software code

We provide R code to implement the methods discussed in this paper. The associations of the genetic variants with the risk factor are denoted `betaXG` with standard errors `sebetaXG`. The associations of the genetic variants with the mediator are denoted `betaMG` with standard errors `sebetaMG`. The associations of the genetic variants with the outcome are denoted `betaYG` with standard errors `sebetaYG`. When variables are continuous, these associations are typically estimated using linear regression.

Estimation of the total causal effect using summarized data:

```
   total.effect =         lm(betaYG ~ betaXG - 1, weights = sebetaYG^-2)$coef[1]
resid.std.error = summary(lm(betaYG ~ betaXG - 1, weights = sebetaYG^-2))$sigma
se.total.effect = summary(lm(betaYG ~ betaXG - 1, weights = sebetaYG^-2))$coef[1,2]
ci.upper.total  = max(total.effect + qnorm(0.975) * se.total.effect / resid.std.error,
                      total.effect + qt(0.975, df=length(betaXG)-1) * se.total.effect)
ci.lower.total  = min(total.effect - qnorm(0.975) * se.total.effect / resid.std.error,
                      total.effect - qt(0.975, df=length(betaXG)-1) * se.total.effect)
```

The weighted regression model for estimating the total effect is equivalent to a meta-analysis of the variant-specific causal estimates. Setting the residual standard error as 1 is equivalent to a fixed-effect assumption in the meta-analysis formula [Thompson and Sharp, 1999]. If there is no heterogeneity between the causal estimates identified by the individual variants, then the residual standard error should tend to 1 asymptotically. If the estimate of the residual standard error is greater than 1 (overdispersion), then we do not correct for this; this is equivalent to a (multiplicative) random-effects meta-analysis [Burgess and Thompson, 2017]. This would occur if different genetic variants identify different causal estimates (say, different variants influence the risk factor via different mechanisms). However, there is no biological rationale for underdispersion (residual standard error estimate is less than 1). Hence, we correct for underdispersion by dividing the standard error for the total effect by the residual standard error.

The multiplicative random-effects analysis fits the following model, with ϕ representing the residual standard error:

$$\hat{\beta}_{Yj} = \theta_T \hat{\beta}_{Xj} + \epsilon_{Tj}, \quad \epsilon_{Tj} \sim \mathcal{N}(0, \phi^2 \, \mathrm{se}(\hat{\beta}_{Yj})^2). \tag{A6}$$

For a fixed-effect analysis, the residual standard error is assumed to be known; hence it is appropriate to use a normal distribution for inferences. For a random-effect analysis, as the residual standard error (the overdispersion parameter ϕ) is estimated rather than

known, a t-distribution should be used for making inferences. In the confidence intervals, we take the upper bound to be the maximum of the bounds based on the fixed-effect and random-effect analyses; similarly for the lower bound as the minimum. This ensures that confidence intervals are no wider than they would be from a fixed-effect analysis, but that under-precision is not doubly penalized (by setting the residual standard error to be 1, and then using a t-distribution for inferences).

Estimation of the direct causal effect using summarized data:

```
   direct.effect =         lm(betaYG ~ betaXG + betaMG - 1, weights = sebetaYG^-2)$coef[1]
se.direct.effect = summary(lm(betaYG ~ betaXG + betaMG - 1, weights = sebetaYG^-2))$coef[1,2]/
               min(summary(lm(betaYG ~ betaXG + betaMG - 1, weights = sebetaYG^-2))$sigma, 1)
ci.upper.direct  = max(direct.effect + qnorm(0.975) * se.direct.effect / resid.std.error,
                       direct.effect + qt(0.975, df=length(betaXG)-1) * se.direct.effect)
ci.lower.direct  = min(direct.effect - qnorm(0.975) * se.direct.effect / resid.std.error,
                        direct.effect - qt(0.975, df=length(betaXG)-1) * se.direct.effect)
```

As the additional term in the regression analysis for the estimate of the direct effect lowers the residual standard error, we take the estimated residual standard error from the regression model for the total causal effect. This is because we want this term to represent overdispersion in the genetic associations with the outcome, not the residual associations after adjustment. Hence the t-distribution for making inferences is still on $J - 1$ degrees of freedom.

If the outcome is binary, then genetic associations with the outcome are typically estimated using logistic regression. Beta-coefficients from logistic regression can be used in the estimation of direct and indirect effects, but the precise magnitude of effect estimates should not be over-interpreted, as odds ratios suffer from non-collapsibility when the rare disease assumption is not applicable (instrumental variable estimates represent population-averaged causal effects, which are not the same as subject-specific causal effects on the odds ratio scale, hence the indirect and direct effects may not precisely sum to give the total effect). Therefore in the applied example in this paper, we do not report an indirect effect.

With correlated variants, this correlation can be accounted for by generalized weighted linear regression [Burgess et al., 2016]. We assume that `rho` is the matrix of correlations between genetic variants:

```
Omega                  = sebetaYG%o%sebetaYG*rho
total.effect.correl    = solve(t(betaXG)%*%solve(Omega)%*%betaXG)*t(betaXG)%*%solve(Omega)%*%betaYG
se.total.effect.fixed  = sqrt(solve(t(betaXG)%*%solve(Omega)%*%betaXG))
resid.total            = betaYG-total.effect.correl*betaXG
se.total.effect.random = sqrt(solve(t(betaXG)%*%solve(Omega)%*%betaXG))*
```

```
                      max(sqrt(t(resid.total)%*%solve(Omega)%*%resid.total/(length(betaXG)-1)),1)
direct.effect.correl     = solve(t(cbind(betaXG, betaMG))%*%solve(Omega)%*%
                      cbind(betaXG, betaMG))%*%t(cbind(betaXG, betaMG))%*%solve(Omega)%*%betaYG
se.direct.effect.fixed  = sqrt(solve(t(cbind(betaXG, betaMG))%*%solve(Omega)%*%cbind(betaXG, betaMG))[1,1])
resid.direct            = betaYG-direct.effect.correl[1]*betaXG-direct.effect.correl[2]*betaMG
se.direct.effect.random = sqrt(solve(t(cbind(betaXG, betaMG))%*%solve(Omega)%*%cbind(betaXG, betaMG))[1,1])*
                      max(sqrt(t(resid.direct)%*%solve(Omega)%*%resid.direct/(length(betaXG)-2)),1)
```

Standard errors are given corresponding to both fixed-effect and random-effects assumptions.

Two different approaches for calculating the indirect effect are provided below. We recall that the indirect effect only has a clear interpretation when all variables are continuous and all relationships are linear. The linear assumptions are particularly crucial, as the indirect effect is calculated based on the total effect minus the direct effect (difference method) or the effect of the risk factor on the mediator multiplied by the effect of the mediator on the outcome (product method).

```
   indirect.effect.difference = total.effect - direct.effect
se.indirect.effect.difference = sqrt(se.total.effect^2 + se.direct.effect^2)
   indirect.product.1  = summary(lm(betaYG~betaXG+betaMG-1, weights = sebetaYG^-2))$coef[2]
se.indirect.product.1  = summary(lm(betaYG~betaXG+betaMG-1, weights = sebetaYG^-2))$coef[2,2]/
                     min(summary(lm(betaYG~betaXG-1, weights = sebetaYG^-2))$sigma, 1)
   indirect.product.2  = summary(lm(betaMG~betaXG-1, weights = sebetaMG^-2))$coef[1]
se.indirect.product.2  = summary(lm(betaMG~betaXG-1, weights = sebetaMG^-2))$coef[1,2]/
                     min(summary(lm(betaMG~betaXG-1, weights = sebetaMG^-2))$sigma, 1)
 indirect.boot = NULL; straps = 1000
 for (k in 1:straps) {
  indirect.boot[k] = rnorm(1, indirect.product.1, se.indirect.product.1)*
                     rnorm(1, indirect.product.2, se.indirect.product.2)
  }
      indirect.effect.product = indirect.product.1 * indirect.product.2
lower.indirect.effect.product = sort(indirect.boot)[0.025*straps+1]
upper.indirect.effect.product = sort(indirect.boot)[0.975*straps]
```

These two methods for calculating and performing inferences on the indirect effect are compared in the simulation study in Supplementary Material A.3.

A.3 Additional details of simulation study

For the simulation study in the paper, the risk factor X was generated as:

$$X_i = \sum_{j=1}^{10} \alpha_j G_{ij} + U_i + \epsilon_{Xi}$$

where G_{ij} is the number of variant alleles for genetic variant j, U is a confounder, ϵ_{Xi} is an independent error term. The number of variant alleles for each variant was drawn from a binomial distribution with 2 trials and probability 0.3, representing a single nucleotide polymorphism with minor allele frequency 0.3. The genetic effects on the risk factor α_j were generated from a normal distribution with mean 0.2 and variance 0.1^2. The variants in total explained on average 5.1% of the variance in the risk factor, corresponding to an average F statistic of 53.5 with a sample size of 10 000. The confounder U and all error terms $(\epsilon_X, \epsilon_M, \epsilon_Y)$ were drawn from independent standard normal distributions. The mediator M was generated as:

$$M_i = \theta_1 X_i + U_i + \epsilon_{Mi} + \sum_{j=1}^{10} \phi_j G_{ij}$$

where θ_1 is the causal effect of X on M, and ϕ_j are direct effects of the genetic variants on the mediator. These effects are included in the simulation model to ensure that the direct effect is identified, as otherwise genetic associations with the risk factor and mediator would be perfectly correlated for large sample sizes, leading to unstable estimates of the direct effect. The ϕ_j parameters were generated from a normal distribution with mean zero and variance 0.1^2. The outcome Y was generated as:

$$Y_i = \theta_2 X_i + \theta_3 M_i + U_i + \epsilon_{Yi}$$

where θ_2 is the direct effect of X on Y, and θ_3 is the effect of M on Y. The indirect effect of X on Y via M is $\theta_1\theta_3$, and the total effect of X on Y is $\theta_2 + \theta_1\theta_3$. In total, 10 000 simulated datasets were generated for each choice of parameter values.

Inferences for the indirect effect

We explored two different approaches to make inferences for the indirect effect. Briefly, a number of different approaches have been proposed in the literature, including resampling (bootstrap) methods and approaches based on the indirect effect being the product of the effect of X on M and the effect of M on Y [MacKinnon et al., 2004]. As we only have summarized data, resampling methods are limited in their applicability here, as we cannot explore genetic associations with the risk factor, mediator, or outcome derived from bootstrapped samples. While parametric bootstrap approaches appear attractive, correlations between the genetic associations with the risk factor and the mediator are typically unknown. It would be crucial to correctly specify these correlations in order to estimate the direct effect of X on Y or the effect of M on Y. Hence we consider difference and product methods for estimating the indirect effect.

The difference indirect effect is calculated as the total effect minus the direct effect, with the standard error of the indirect effect as the square root of the sum of the squared standard errors for the total effect and the direct effect:

$$\hat{\theta}_{I1} = \hat{\theta}_T - \hat{\theta}_D$$
$$\text{se}(\hat{\theta}_{I1}) = \sqrt{\text{se}(\hat{\theta}_T)^2 + \text{se}(\hat{\theta}_D)^2}.$$

Confidence intervals for the difference indirect effect are calculated using a normal approximation: lower and upper bounds of the 95% confidence interval are $\hat{\theta}_{I1} \pm 1.96 \times \text{se}(\hat{\theta}_{I1})$.

For the product indirect effect, we first estimate the effect of X on M (θ_A):

$$\hat{\beta}_{Mj} = \theta_A \hat{\beta}_{Xj} + \epsilon_{Aj}, \quad \epsilon_{Aj} \sim \mathcal{N}(0, \text{se}(\hat{\beta}_{Mj})^2)$$

and then estimate the effect of M on Y (θ_M) using the same regression model as used to estimate the direct effect:

$$\hat{\beta}_{Yj} = \theta_D \hat{\beta}_{Xj} + \theta_M \hat{\beta}_{Mj} + \epsilon_{Dj}, \quad \epsilon_{Dj} \sim \mathcal{N}(0, \text{se}(\hat{\beta}_{Yj})^2).$$

The product indirect effect estimate is calculated as:

$$\hat{\theta}_{I2} = \hat{\theta}_A \times \hat{\theta}_M.$$

Confidence intervals for the product indirect effect are obtained by a Monte Carlo method: we calculate the standard errors for $\hat{\theta}_A$ and $\hat{\theta}_M$ using similar formulae as for the total effect and direct effect of X on Y respectively. We then obtain a Monte Carlo distribution by drawing from normal distributions with mean $\hat{\theta}_A$ and variance $\text{se}(\hat{\theta}_A)^2$ and with $\hat{\theta}_M$ and variance $\text{se}(\hat{\theta}_M)^2$, and multiplying the results. This process was repeated 1000 times, and the 2.5th and 97.5th percentiles of this Monte Carlo distribution were taken as the lower and upper bounds of the 95% confidence interval.

Supplementary Table A1 provides means and standard deviations of estimates of the indirect effect across the same simulated datasets as in the main simulation study of the paper, coverage of the 95% confidence interval (that is, the proportion of confidence intervals including the true value of the indirect effect), empirical power to detect a non-null indirect effect (that is, the proportion of confidence intervals excluding zero), and median width of the 95% confidence interval. We see that means and standard deviations of estimates are almost identical between the two methods. Estimates are close to unbiased, with some evidence of weak instrument bias. Confidence intervals from the difference method are on average wider, and have better coverage and Type 1 error properties (coverage close to 95% in all scenarios except Scenario 6 – see below). Confidence intervals from the product method have worse coverage properties, but greater power. It may be that the poor coverage properties are due to weak instrument bias, however coverage is even underestimated in Scenario 5 when the true indirect effect is zero. Of note is Scenario 6, in which the mediator has no effect on the outcome. In this scenario, confidence intervals for the difference method are much wider than those from the product method, and coverage rates are much higher than the nominal 95% level. The reason for this is that the correlation between estimates of the total and direct effects is ignored in the calculation of the standard error. This is not unreasonable in the other scenarios, where the correlations across the 10 000 simulations are around 0.2 to 0.3. However in Scenario 6, this correlation is 0.76, leading to highly conservative inferences. We note that if the estimates of total and direct effect are positively correlated (as in all the simulation scenarios), then failure to account for this correlation leads to conservative estimates of the difference standard error.

Overall, on this basis of this simulation study, the difference method seems to be the preferable approach for constructing confidence intervals as it does not suffer from inflated Type 1 error rates.

			Indirect	Difference method					Product method				
θ_1	θ_2	θ_3	effect	Mean	SD	Coverage	Power	CI width	Mean	SD	Coverage	Power	CI width
0.3	0.2	1	0.3	0.324	0.163	95.1	50.1	0.650	0.324	0.163	92.5	57.1	0.606
0.3	0.2	−1	−0.3	−0.295	0.150	94.8	48.4	0.601	−0.295	0.150	92.5	57.1	0.555
0.3	−0.2	1	0.3	0.322	0.166	94.7	49.7	0.648	0.322	0.166	91.9	57.3	0.604
−0.3	−0.2	1	−0.3	−0.304	0.163	94.7	45.2	0.644	−0.304	0.164	92.6	52.6	0.601
0.0	0.2	1	0.0	0.008	0.160	94.8	5.2	0.634	0.008	0.160	91.8	8.2	0.591
0.3	0.2	0	0.0	0.014	0.038	99.5	0.5	0.265	0.014	0.038	96.4	3.4	0.172
0.3	0.0	1	0.3	0.323	0.164	94.7	49.5	0.646	0.324	0.164	92.5	56.9	0.602
−0.2	0.2	1	−0.2	−0.198	0.161	94.7	22.6	0.638	−0.199	0.161	92.1	29.3	0.594

Supplementary Table A1: Mean, standard deviation (SD), coverage of 95% confidence interval (%), empirical power (%), and median width of the 95% confidence interval (CI) for Mendelian randomization estimates of the indirect effect from two methods across 10 000 simulated datasets for different mediation scenarios (X = risk factor, M = mediator, Y = outcome).

Varying the degree of heterogeneity

We also experimented with different values of the variance of the ϕ_j parameters in the data-generating model. Results are shown in Supplementary Table A2. When there was low heterogeneity, estimates were more variable and bias from weak instruments was more pronounced. This is expected, as the associations with the risk factor and mediator are increasingly collinear as the heterogeneity decreases. To demonstrate that the bias is an artifact of limited sample size (so called 'weak instrument bias'), we repeated the simulation with 100 000 participants (100 iterations per scenario only). As expected, bias did not decrease when there was no heterogeneity, as the collinearity problem does not disappear with increasing sample sizes in this case. However, in all other cases, increasing the sample size decreased bias sharply.

Sample size: 10 000						
θ_1	θ_2	θ_3	$\text{var}(\phi) = 0$	$\text{var}(\phi) = 0.05^2$	$\text{var}(\phi) = 0.1^2$	$\text{var}(\phi) = 0.2^2$
0.3	0.2	1	0.054 (0.110)	0.165 (0.077)	0.196 (0.055)	0.203 (0.050)
0.3	0.2	−1	0.052 (0.115)	0.165 (0.076)	0.195 (0.056)	0.204 (0.050)
0.3	−0.2	1	−0.343 (0.113)	−0.235 (0.071)	−0.205 (0.059)	−0.194 (0.050)
−0.3	−0.2	1	−0.049 (0.101)	−0.153 (0.073)	−0.181 (0.058)	−0.187 (0.052)
0.0	0.2	1	0.205 (0.041)	0.207 (0.048)	0.208 (0.048)	0.207 (0.048)
0.3	0.2	0	0.053 (0.106)	0.168 (0.074)	0.196 (0.058)	0.203 (0.053)
0.3	0.0	1	−0.146 (0.113)	−0.035 (0.071)	−0.004 (0.059)	0.003 (0.050)
−0.2	0.2	1	0.302 (0.076)	0.235 (0.057)	0.213 (0.050)	0.210 (0.048)
Sample size: 100 000						
θ_1	θ_2	θ_3	$\text{var}(\phi) = 0$	$\text{var}(\phi) = 0.05^2$	$\text{var}(\phi) = 0.1^2$	$\text{var}(\phi) = 0.2^2$
0.3	0.2	1	0.053 (0.092)	0.191 (0.027)	0.200 (0.019)	0.201 (0.016)
0.3	0.2	−1	0.051 (0.114)	0.194 (0.030)	0.195 (0.019)	0.202 (0.016)
0.3	−0.2	1	−0.341 (0.098)	−0.206 (0.028)	−0.197 (0.016)	−0.198 (0.015)
−0.3	−0.2	1	−0.049 (0.087)	−0.191 (0.027)	−0.197 (0.020)	−0.202 (0.017)
0.0	0.2	1	0.199 (0.012)	0.202 (0.016)	0.199 (0.016)	0.200 (0.016)
0.3	0.2	0	0.055 (0.106)	0.196 (0.027)	0.199 (0.019)	0.200 (0.017)
0.3	0.0	1	−0.136 (0.099)	−0.005 (0.027)	0.003 (0.018)	0.000 (0.017)
−0.2	0.2	1	0.296 (0.072)	0.206 (0.018)	0.200 (0.018)	0.200 (0.016)

Supplementary Table A2: Mean (standard deviation) of multivariable Mendelian randomization estimates of the direct effect θ_2 across 10 000 simulated datasets (100 datasets for larger sample size) for different values of the variance of the heterogeneity parameters ϕ.

A.4 Additional simulation scenario: bidirectional causal effects between risk factor and mediator

In the applied example, it may be that as well as the risk factor having a causal effect on the mediator, that the mediator also has a causal effect on the risk factor. To consider this scenario, we simulate causal effects in both directions and consider Mendelian randomization and multivariable Mendelian randomization estimates. The data-generating model is:

$$
\begin{aligned}
X_{0i} &= \sum_{j=1}^{10} \alpha_j G_{ij} + U_i + \epsilon_{Xi} \\
M_i &= \theta_1 X_{0i} + U_i + \epsilon_{Mi} + \sum_{j=1}^{10} \phi_j G_{ij} \\
X_{1i} &= X_{0i} \pm M_i \\
Y_i &= \theta_2 X_{1i} + \theta_3 M_i + U_i + \epsilon_{Yi}
\end{aligned}
$$

This is the same as the previous data-generating model, except that we first generate X_{0i} and then generate a second risk factor variable X_{1i} that has a causal effect from the mediator. These could be thought of as values of the risk factor at different time points. We consider cases where the mediator has a positive and a negative effect on the risk factor. All other aspects of this simulation are the same as the original.

Results are shown in Supplementary Table A3. The total effect varies depending on whether the effect of the mediator on the risk factor is positive or negative, and is not simply an estimate of $\theta_2 + \theta_1\theta_3$ (as there are additional components of the total effect via the effect of the mediator on the risk factor). However, the direct effect as estimated by multivariable Mendelian randomization is invariant to any bidirectional effect. Therefore the direct effect of age at menarche on breast cancer risk not via BMI can be estimated using multivariable Mendelian randomization whether or not there is a bidirectional relationship between age at menarche and BMI.

			Positive effect		Negative effect	
θ_1	θ_2	θ_3	Univariable	Multivariable	Univariable	Multivariable
0.3	0.2	1	0.525	0.195	0.222	0.195
0.3	0.2	−1	−0.103	0.194	0.173	0.194
0.3	−0.2	1	0.123	−0.204	−0.169	−0.204
−0.3	−0.2	1	−0.195	−0.180	−0.504	−0.180
0.0	0.2	1	0.381	0.208	0.045	0.208
0.3	0.2	0	0.209	0.195	0.197	0.195
0.3	0.0	1	0.323	−0.005	0.017	−0.005
−0.2	0.2	1	0.273	0.217	−0.060	0.217

Supplementary Table A3: Mean of univariable and multivariable Mendelian randomization estimates across 10 000 simulated datasets for different mediation scenarios with positive and negative bidirectional effect of the mediator on the risk factor.

Supplementary References

Burgess, S., Daniel, R., Butterworth, A., Thompson, S., and EPIC-InterAct Consortium 2015. Network Mendelian randomization: extending instrumental variable techniques. *International Journal of Epidemiology*, 44(2):484–495.

Burgess, S., Dudbridge, F., and Thompson, S. G. 2016. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Statistics in Medicine*, 35(11):1880–1906.

Burgess, S. and Thompson, S. G. 2017. Interpreting findings from Mendelian randomization using the MR-Egger method. *European Journal of Epidemiology*. Available online before print. doi: 10.1007/s10654-017-0255-x.

Kaufman, J. S. 2010. Invited commentary: Decomposing with a lot of supposing. *American Journal of Epidemiology*, 172(12):1349–1351.

MacKinnon, D. P., Lockwood, C. M., and Williams, J. 2004. Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1):99–128.

Pearl, J. 2001. Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pages 411–420.

Richardson, T. and Robins, J. 2013. Single World Intervention Graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. Technical Report 128, Center for Statistics and the Social Sciences, University of Washington.

Robins, J. and Greenland, S. 1992. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2):143–155.

Thompson, S. and Sharp, S. 1999. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, 18(20):2693–2708.

VanderWeele, T. and Vansteelandt, S. 2010. Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology*, 172(12):1339–1348.