Supplementary Material for: Fulfilling the promise of Mendelian randomization

Joseph K. Pickrell[†1,2]

[1] New York Genome Center, New York, NY
[2] Department of Biological Sciences, Columbia University, New York, NY
[†] Correspondence to: jkpickrell@nygenome.org

April 15, 2015

Table S1: List of causal relationships evaluated by Mendelian randomization, compiled from Table 1 of Davey Smith and Hemani [11] and Table 1 of Burgess and Thompson [5]. In the column for effect sizes "+" denotes a positive correlation between the trait and the outcome, and "−" denotes a negative correlation between the trait and the outcome. MR = Mendelian randomization, RCT = randomized controlled trial. For one study listed in Burgess and Thompson [5] we were unable to locate the listed Mendelian randomization study, this is listed as a study of folate and blood pressure by Thompson et al. (2005).

Trait	Outcome	MR effect	Reference	RCT?	RCT effect	Reference
cholesterol levels	cancer	None	[33]	Yes	None	[9]
C-reactive protein levels	insulin resistance	None	[32]	No	NA	NA
C-reactive protein levels	cartoid thickness	None	[20]	No	NA	NA
C-reactive protein levels	cancer	None	[1]	No	NA	NA
HDL cholesterol	myocardial infarction	None	[34]	Yes	None	[3]
homocysteine	stroke	+	[7]	Yes	Conflicting	[16, 31]
lipoprotein(a)	myocardial infarction	+	[19]	No	NA	NA
sex hormone binding globulin	type 2 diabetes	-	[12]	No	NA	NA
BMI	cartoid thickness	+	[21]	Yes	+	[15, 29]
BMI	age at menarche	+	[25]	No	NA	NA
BMI	employment	None	[26]	No	NA	NA
fat mass	academic achievement	None	[35]	No	NA	NA
alcohol intake	blood pressure	+	[8]	No	NA	NA
caffeine intake	stillbirth	None	[4]	No	NA	NA
milk intake	metabolic syndrome	+	[2]	No	NA	NA
alcohol abuse	drug abuse	None	[18]	No	NA	NA
ADHD	education	None	[13]	No	NA	NA
depression	education	-	[13]	No	NA	NA
interuterine folate	neural tube defects	-	[10]	Yes	-	[24]
C-reactive protein levels	coronary heart disease	None	[6]	No	NA	NA
serum iron	Parkinson's disease	-	[28]	No	NA	NA
uric acid	coronary heart disease	None	[27]	No	NA	NA
macrophage migration inhibitory factor	type 2 diabetes	+	[14]	No	NA	NA
interleukin 6 levels	coronary heart disease	-	[17]	No	NA	NA
smoking	anxiety	None	[23]	No	NA	NA
alcohol consumption	blood pressure	+	[8]	No	NA	NA
BMI	gallstone disease	+	[30]	No	NA	NA
maternal alcohol consumption	childhood school performance	-	[36]	No	NA	NA
maternal BMI	childhood fat mass	None	[22]	No	NA	NA

1 Simulation details

If a genetic variant influences two traits by different mechanisms, the expected "causal effect" of the first trait on the second (as estimated by Mendelian randomization) is non-zero. This is true simply from first principles, but to illustrate this problem and determine its severity, we simulated Mendelian randomization studies. Specifically, we simulated Mendelian randomization studies that use a genetic score as an "instrument". Our overall strategy was to simulate 100 genetic variants that influence a trait and a disease, then to simulate phenotypes for individuals with those variants, and then to test for a causal relationship between the quantitative trait and the disease using standard methods. In all simulations, there was no causal relationship between the two, so all positive results were false positives. All simulation code is available at `https://github.com/joepickrell/mr_paper`. The details were:

Simulate 100 variants that influence a trait. Let f_i be the allele frequency of variant i, β_{1i} be the effect size of variant i on the quantitative trait, and β_{2i} be the effect size of variant i on liability to disease (measured

in arbitrary units). Let p be the proportion of variants that influence both the quantitative trait and the disease (see Figure 1 in the main text). To simulate a set of 100 variants, we:

1. Draw an allele frequency from the distribution $f_i \sim Beta(0.5, 0.5)$
2. Draw effect sizes from the distribution $[\beta_{1i}, \beta_{2i}] \sim MVN([0,0], \Sigma)$, where the diagonal entries of Σ are set to 0.005, and the off-diagonal entries to 0.005ρ. For Figure 1 in the main text, $\rho = 0$, while in Figure S1 we also show results for $\rho = 0.4$.
3. With probability $1 - p$, set $\beta_{2i} = 0$.
4. Compute the expected proportion of variance in the quantitative trait explained by the variant as $2\beta_{1i}^2 f_i(1 - f_i)$. If this is less than 0.0003 (for example, the allele frequency or effect size is very small), return to step 1.
5. Repeat until there are 100 genetic variants.

On average, this resulted in a genetic score that explained around 20% of the variance in the quantitative trait.

Simulate case/control status for individuals. We then simulated genotypes for N individuals (where N varied across simulation settings, see Figure 1 in the main text) from the 100 variants as follows:

1. For each of the 100 variants in each of the N individuals, draw $g_{ij} \sim Binom(2, f_i)$, where j indexes individuals and i indexes variants.
2. For each individual, simulate the disease liability of the individual as $y_j \sim N(\sum_i g_{ij}\beta_{2i}, 1 - \sum_i 2\beta_{2i}^2 f_i[1 - f_i])$.
3. Convert liability to case/control status by defining a disease prevalence K and setting all individuals that fall in the top K quantile of disease liability as cases and the rest as controls. In all simulations, $K = 0.1$.

Perform Mendelian randomization to test for a causal effect of the quantitative trait on the disease. We then tested for a "causal effect" of the quantitative trait on the disease:

1. Define the genetic score for individual as $s_j = \sum_i \beta_{1i} g_{ij}$
2. Define the case/control status of each individual as z_j.
3. Use logistic regression of z_j on s_j to test for a causal relationship the quantitative trait and disease, using a P-value threshold of 0.05.

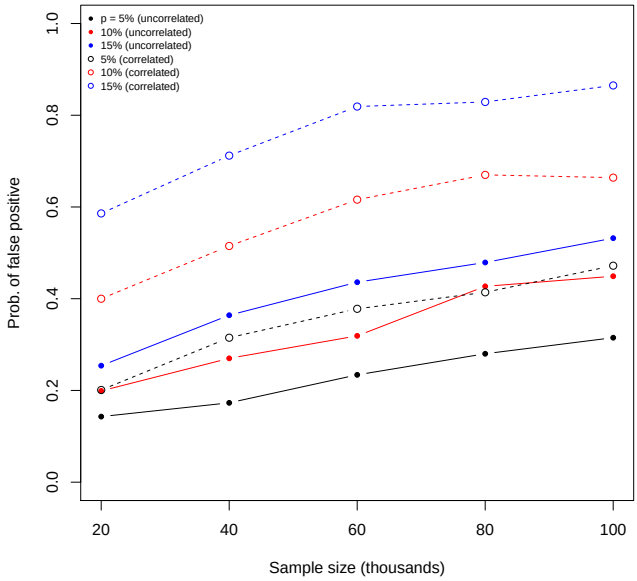

Figure S1. **Simulations of Mendelian randomization in the absence of a causal relationship between a trait and a disease.** In addition to the simulations from Figure 1 in the main text, we also performed simulations where the effects of a variant on the trait and on liability to disease were correlated, with a correlation coefficient of 0.4 (see simulation details in the Supplementary Information). Each point is an average over 1,000 simulations.

References

[1] Allin, K. H., Nordestgaard, B. G., Zacho, J., Tybjaerg-Hansen, A., and Bojesen, S. E., 2010. C-reactive protein and the risk of cancer: a mendelian randomization study. *J Natl Cancer Inst*, **102**(3):202–6.

[2] Almon, R., Alvarez-Leon, E. E., Engfeldt, P., Serra-Majem, L., Magnuson, A., and Nilsson, T. K., 2010. Associations between lactase persistence and the metabolic syndrome in a cross-sectional study in the Canary Islands. *European journal of nutrition*, **49**(3):141–146.

[3] Barter, P. J., Caulfield, M., Eriksson, M., Grundy, S. M., Kastelein, J. J. P., Komajda, M., Lopez-Sendon, J., Mosca, L., Tardif, J.-C., Waters, D. D., *et al.*, 2007. Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med*, **357**(21):2109–22.

[4] Bech, B. H., Autrup, H., Nohr, E. A., Henriksen, T. B., and Olsen, J., 2006. Stillbirth and slow metabolizers of caffeine: comparison by genotypes. *Int J Epidemiol*, **35**(4):948–53.

[5] Burgess, S. and Thompson, S. G., 2015. *Mendelian Randomization: Methods for using Genetic Variants in Causal Estimation*. CRC Press.

[6] C Reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC), Wensley, F., Gao, P., Burgess, S., Kaptoge, S., Di Angelantonio, E., Shah, T., Engert, J. C., Clarke, R., Davey-Smith, G., *et al.*, 2011. Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. *BMJ*, **342**:d548.

[7] Casas, J. P., Bautista, L. E., Smeeth, L., Sharma, P., and Hingorani, A. D., 2005. Homocysteine and stroke: evidence on a causal link from mendelian randomisation. *Lancet*, **365**(9455):224–32.

[8] Chen, L., Smith, G. D., Harbord, R. M., and Lewis, S. J., 2008. Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach. *PLoS Med*, **5**(3):e52.

[9] Dale, K. M., Coleman, C. I., Henyan, N. N., Kluger, J., and White, C. M., 2006. Statins and cancer risk: a meta-analysis. *JAMA*, **295**(1):74–80.

[10] Davey Smith, G. and Ebrahim, S., 2003. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*, **32**(1):1–22.

[11] Davey Smith, G. and Hemani, G., 2014. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*, **23**(R1):R89–98.

[12] Ding, E. L., Song, Y., Manson, J. E., Hunter, D. J., Lee, C. C., Rifai, N., Buring, J. E., Gaziano, J. M., and Liu, S., 2009a. Sex hormone-binding globulin and risk of type 2 diabetes in women and men. *N Engl J Med*, **361**(12):1152–63.

[13] Ding, W., Lehrer, S. F., Rosenquist, J. N., and Audrain-McGovern, J., 2009b. The impact of poor health on academic performance: New evidence using genetic markers. *Journal of health economics*, **28**(3):578–597.

[14] Herder, C., Klopp, N., Baumert, J., Müller, M., Khuseyinova, N., Meisinger, C., Martin, S., Illig, T., Koenig, W., and Thorand, B., *et al.*, 2008. Effect of macrophage migration inhibitory factor (MIF) gene variants and MIF serum concentrations on the risk of type 2 diabetes: results from the MONICA/KORA Augsburg Case-Cohort Study, 1984-2002. *Diabetologia*, **51**(2):276–84.

[15] Hjerkinn, E. M., Abdelnoor, M., Breivik, L., Bergengen, L., Ellingsen, I., Seljeflot, I., Aase, O., Ole Klemsdal, T., Hjermann, I., and Arnesen, H., *et al.*, 2006. Effect of diet or very long chain omega-3 fatty acids on progression of atherosclerosis, evaluated by carotid plaques, intima-media thickness and by pulse wave propagation in elderly men with hypercholesterolaemia. *Eur J Cardiovasc Prev Rehabil*, **13**(3):325–33.

[16] Huo, Y., Li, J., Qin, X., Huang, Y., Wang, X., Gottesman, R. F., Tang, G., Wang, B., Chen, D., He, M., *et al.*, 2015. Efficacy of Folic Acid Therapy in Primary Prevention of Stroke Among Adults With Hypertension in China: The CSPPT Randomized Clinical Trial. *JAMA*, .

[17] IL6R Genetics Consortium Emerging Risk Factors Collaboration, Sarwar, N., Butterworth, A. S., Freitag, D. F., Gregson, J., Willeit, P., Gorman, D. N., Gao, P., Saleheen, D., Rendon, A., *et al.*, 2012. Interleukin-6 receptor pathways in coronary heart disease: a collaborative meta-analysis of 82 studies. *Lancet*, **379**(9822):1205–13.

[18] Irons, D. E., McGue, M., Iacono, W. G., and Oetting, W. S., 2007. Mendelian randomization: a novel test of the gateway hypothesis and models of gene-environment interplay. *Dev Psychopathol*, **19**(4):1181–95.

[19] Kamstrup, P. R., Tybjaerg-Hansen, A., Steffensen, R., and Nordestgaard, B. G., 2009. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA*, **301**(22):2331–9.

[20] Kivimäki, M., Lawlor, D. A., Smith, G. D., Kumari, M., Donald, A., Britton, A., Casas, J. P., Shah, T., Brunner, E., Timpson, N. J., *et al.*, 2008a. Does high C-reactive protein concentration increase atherosclerosis? The Whitehall II Study. *PLoS One*, **3**(8):e3013.

[21] Kivimäki, M., Smith, G. D., Timpson, N. J., Lawlor, D. A., Batty, G. D., Kähönen, M., Juonala, M., Rönnemaa, T., Viikari, J. S. A., Lehtimäki, T., *et al.*, 2008b. Lifetime body mass index and later atherosclerosis risk in young adults: examining causal links using Mendelian randomization in the Cardiovascular Risk in Young Finns study. *Eur Heart J*, **29**(20):2552–60.

[22] Lawlor, D. A., Timpson, N. J., Harbord, R. M., Leary, S., Ness, A., McCarthy, M. I., Frayling, T. M., Hattersley, A. T., and Smith, G. D., 2008. Exploring the developmental overnutrition hypothesis using parental-offspring associations and FTO as an instrumental variable. *PLoS Med*, **5**(3):e33.

[23] Lewis, S. J., Araya, R., Smith, G. D., Freathy, R., Gunnell, D., Palmer, T., and Munafò, M., 2011. Smoking is associated with, but does not cause, depressed mood in pregnancy–a mendelian randomization study. *PLoS One*, **6**(7):e21689.

[24] MRC Vitamin Study Research Group et al., 1991. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. *The lancet*, **338**(8760):131–137.

[25] Mumby, H. S., Elks, C. E., Li, S., Sharp, S. J., Khaw, K.-T., Luben, R. N., Wareham, N. J., Loos, R. J. F., and Ong, K. K., 2011. Mendelian Randomisation Study of Childhood BMI and Early Menarche. *J Obes*, **2011**:180729.

[26] Norton, E. C. and Han, E., 2008. Genetic information, obesity, and labor market outcomes. *Health Economics*, **17**(9):1089–1104.

[27] Palmer, T. M., Nordestgaard, B. G., Benn, M., Tybjærg-Hansen, A., Davey Smith, G., Lawlor, D. A., and Timpson, N. J., 2013. Association of plasma uric acid with ischaemic heart disease and blood pressure: mendelian randomisation analysis of two large cohorts. *BMJ*, **347**:f4262.

[28] Pichler, I., Del Greco M, F., Gögele, M., Lill, C. M., Bertram, L., Do, C. B., Eriksson, N., Foroud, T., Myers, R. H., PD GWAS Consortium, *et al.*, 2013. Serum iron levels and the risk of Parkinson disease: a Mendelian randomization study. *PLoS Med*, **10**(6):e1001462.

[29] Shai, I., Spence, J. D., Schwarzfuchs, D., Henkin, Y., Parraga, G., Rudich, A., Fenster, A., Mallett, C., Liel-Cohen, N., Tirosh, A., *et al.*, 2010. Dietary intervention to reverse carotid atherosclerosis. *Circulation*, **121**(10):1200–8.

[30] Stender, S., Nordestgaard, B. G., and Tybjærg-Hansen, A., 2013. Elevated body mass index as a causal risk factor for symptomatic gallstone disease: a Mendelian randomization study. *Hepatology*, **58**(6):2133–2141.

[31] Study of the Effectiveness of Additional Reductions in Cholesterol and Homocysteine (SEARCH) Collaborative Group, Armitage, J. M., Bowman, L., Clarke, R. J., Wallendszus, K., Bulbulia, R., Rahimi, K., Haynes, R., Parish, S., Sleight, P., *et al.*, 2010. Effects of homocysteine-lowering with folic acid plus vitamin B12 vs placebo on mortality and major morbidity in myocardial infarction survivors: a randomized trial. *JAMA*, **303**(24):2486–94.

[32] Timpson, N. J., Lawlor, D. A., Harbord, R. M., Gaunt, T. R., Day, I. N. M., Palmer, L. J., Hattersley, A. T., Ebrahim, S., Lowe, G. D. O., Rumley, A., *et al.*, 2005. C-reactive protein and its role in metabolic syndrome: mendelian randomisation study. *Lancet*, **366**(9501):1954–9.

[33] Trompet, S., Jukema, J. W., Katan, M. B., Blauw, G. J., Sattar, N., Buckley, B., Caslake, M., Ford, I., Shepherd, J., Westendorp, R. G. J., *et al.*, 2009. Apolipoprotein e genotype, plasma cholesterol, and cancer: a Mendelian randomization study. *Am J Epidemiol*, **170**(11):1415–21.

[34] Voight, B. F., Peloso, G. M., Orho-Melander, M., Frikke-Schmidt, R., Barbalic, M., Jensen, M. K., Hindy, G., Hólm, H., Ding, E. L., Johnson, T., *et al.*, 2012. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet*, **380**(9841):572–80.

[35] von Hinke Kessler Scholder, S., Smith, G. D., Lawlor, D. A., Propper, C., and Windmeijer, F., 2010. Genetic markers as instrumental variables: An application to child fat mass and academic achievement. Technical report, cemmap working paper.

[36] Zuccolo, L., Lewis, S. J., Smith, G. D., Sayal, K., Draper, E. S., Fraser, R., Barrow, M., Alati, R., Ring, S., Macleod, J., *et al.*, 2013. Prenatal alcohol exposure and offspring cognition and school performance. A 'Mendelian randomization' natural experiment. *Int J Epidemiol*, **42**(5):1358–70.