**SUPPLEMENTARY MATERIAL TO: Estimating the causal effect of genetic liability to prevalent disease on hospital costs using Mendelian Randomization**

Padraig Dixon1,2\*, Sean Harrison1,2, William Hollingworth1, Neil M Davies1,2, George Davey Smith1,2,3

1: Population Health Sciences, University of Bristol

2: MRC Integrative Epidemiology Unit, University of Bristol

3: NIHR Biomedical Research Centre, University of Bristol

30 June 2020

# Directed acyclic graph summary of the instrumental variable assumptions

Figure A1 summarises the first three instrumental variable assumptions using a directed acyclic graph (DAG). Valid instrumental variables are related to the treatment variable, indicated by the directed line emanating from G (the genetic variants) to X (the treatment variable – disease status in this case). G is independent of all known and unknown omitted variables U (indicated by the absence of any directed line between U and G. Finally, G affects the outcome Y (costs in this case) only via the treatment variable – there are no paths from G to Y that are not via X.

**Figure A1 Instrumental variable**



# Intuition for different estimators

This section provides simple intuition for the overidentified instrumental variable models: the inverse variance weighted (IVW) model, the median estimator, the mode estimator and the MR-Egger estimator.

Consider the following example. Twenty hypothetical SNPs are represented in scatter plot below (Figure A1) using made-up data, with their position on the plane being fixed by their association (on the horizontal axis) with disease liability, and on the vertical axis by their association with the cost outcome. The line from the origin to any point is the causal effect of that variant, since the slope of such a line is measured by the ratio of the effect of the variant on the outcome divided by the effect of the SNP on disease liability.

**Figure A2 Scatter plot of hypothetical causal estimates from twenty genetic variants**



A useful starting point in approaching estimation of these relationships is to interpret conditional random allocation of genetic variation at conception as a type of randomized controlled trial, in which the individual is allocated to higher or lower susceptibility to disease. When many such variants are available, we may proceed, as in a conventional meta-analysis, to summarize their effects on the outcome in a manner that is potentially more informative than scrutiny of any one variant.

To calculate the overall effect of all variants, we may weight all 20 estimates by their precision and then calculate an overall inverse variance weighted effect. An IVW estimator is shown in the second panel of Figure A1. Note that line summarizing the IVW effect of all twenty estimators must pass through the origin, since summary causal effects are determined only by the ratio of the effect of variants on the outcome to effects of variants on disease liability. This is a direct consequence of the exclusion restriction: there should be no effect of any variant on the outcome that does not also affect liability to disease.

A random-effects meta-analysis allows some or all variants to have pleiotropic effects in violation of the exclusion restriction, but assumes that the effect of this pleiotropy “balances out” so that pleiotropy that increases the causal effect estimate is matched by pleiotropy that reduces the causal effect estimate. Thus, the effect size is estimated without bias, although with a greater standard error than would be the case without variation induced by horizontal pleiotropy.

Cochran’s Q, defined in the main text, may indicate the presence of heterogeneity caused by horizontal pleiotropy. One mechanism to generate heterogeneity of this type would be the influence of a variant on the outcome through multiple independent channels. If so, alternatives to the IVW estimator may be analyzed. Note that the vertical distance from any one point to the IVW line reflects the contribution of the variant represented by that point to the Q statistic. This motivates the use of the Radial MR technique described in the main text.

Median-based estimators (1) assume that the median estimate, constructed by forming an empirical density of all estimates, is unbiased. Excluding pleiotropic variants will mean that they offer no weight to the calculation of the overall summary effect. For example, it might be the case the variants highlighted below are excluded by a median-based estimator, with the consequence that the estimate of causal effect (i.e. the slope of the line) is different than the IVW estimator which included all SNPs. Note that the set of excluded variants in this and subsequent figures is not intended to be realistic, but instead simply to illustrate the point that include different sets of variants in an analysis may lead to a change in the causal effect estimate.

**Figure A3 Scatter plot of hypothetical causal estimates and a median causal estimate**



A related method makes an assumption that modal values of individual variants are not in violation of the exclusion restriction (2). This permits some SNPs, possibly most SNPs, to be pleiotropic and therefore in violation of the exclusion restriction, provided that the modal variants, or modal set of variants, are unbiased (Figure A3).

**Figure A4 Scatter plot of hypothetical causal estimates and a modal causal estimate**



The MR Egger estimator does not constrain the intercept in the scatter plot to be zero. The intercept represents the average pleiotropic effect from all SNPs. MR Egger may produce asymptotically unbiased estimates even if all include SNPs are pleiotropic (Figure A4).

**Figure A5 Scatter plot of hypothetical causal estimates and a MR Egger causal estimate**



The next section presents scatter plots for all six disease phenotypes using the analysis sample described in the main text.

# Scatter plots for all six disease phenotypes

This section presents scatter plots of the SNP-disease and SNP-outcome associations for all six disease phenotypes.

**Figure A6 Asthma scatter plot**



**Figure A7 Eczema scatter plot**



**Figure A8 Migraine scatter plot**



**Figure A9 Coronary heart disease scatter plot**



**Figure A10 Type 2 diabetes scatter plot**



**Figure A11 Depression scatter plot**



# Results from inverse weighted models using summary data for a 10% relative increase in genetic liability

Figure A12 presents results from inverse variance weighted models for a 10% relative increase in genetic liability.

**Figure A12 Inverse variance weighted effect estimates for all conditions for a 10% relative increase in genetic liability**



References

1. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304-14.

2. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985-98.