Research Article

Selection bias in instrumental variable analyses

Rachael A. Hughes*,[1] Neil M. Davies,[2] George Davey Smith,[1,2] and Kate Tilling[1,2]

[1] *Bristol Medical School, University of Bristol, Bristol, United Kingdom*
[2] *MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom*

Correspondence: *Rachael A. Hughes, Bristol Medical School, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, United Kingdom. Email: rachael.hughes@bristol.ac.uk

Received xxxxx; Revised xxxxx; Accepted xxxxx

Summary

Participants in epidemiological and genetic studies are rarely truly random samples of the populations they are intended to represent, and both known and unknown factors can influence participation in a study (also known as selection into a study). The circumstances in which selection causes bias in an instrumental variable (IV) analysis are not well understood. We use directed acyclic graphs (DAGs) to depict assumptions about the selection mechanism (i.e., the factors affecting selection into the study), and show how DAGs can be used to determine when a two stage least squares (2SLS) IV analysis is biased by selection. For a range of selection mechanisms we explain the structure of the selection bias and, via simulations, we illustrate the potential bias caused by selection in an IV analysis. We show that selection can result in a biased 2SLS estimate of the causal exposure effect, substantial undercoverage of its confidence interval, and the chance of reaching an incorrect conclusion about the causal exposure effect. We consider whether the bias caused by selection differ according to different instrument strengths, between a linear and nonlinear exposure-instrument association, and for a causal and non-causal exposure effect. In addition, we present the results of a real data example where nonrandom selection into the study was suspected. We conclude that selection bias can have a major effect on an IV analysis and that statistical methods for estimating causal effects using data from nonrandom samples are needed.

Keywords: collider stratification bias, instrumental variable, selection bias, two stage least squares estimate

1 Introduction

The main aim of many epidemiological studies is to estimate the causal effect of an exposure (or treatment) on an outcome; that is, the effect on the outcome of changing the exposure. A major challenge for observational studies is unmeasured confounding of the outcome and exposure but standard analyses, such as ordinary least squares

regression, make the untestable assumption that there is no unmeasured confounding. One solution is an instrumental variable (IV) analysis which can control for unmeasured confounders. An IV analysis requires a variable, known as the instrument, to satisfy three assumptions: (1) the instrument is associated with the exposure, (2) the instrument only causes the outcome to change via its impact on the exposure, and (3) there is no confounding between the instrument and the outcome (1–3). Based on the observed data, the first IV assumption can be tested, but the latter two IV assumptions are untestable.

As with any statistical analysis, to obtain a valid inference about the causal exposure effect the participants of the analysed data should be a representative sample of the population under study. There are several reasons why the participants of the analysed data (i.e., our selected sample) may not be a representative sample, including: nonrandom sampling, nonrandom participant refusal, loss to follow-up, subgroups analysis, and missing data (i.e., restricting the analysis to those participants with complete data). In all of these cases there may be both known and unknown factors that influence the "selection" of participants for analysis. Following Hernán and Robins (4), we consider selection bias to be distinct from confounding; that is, for an exposure-outcome association, selection bias is due to conditioning on common effects (or common causes) of the outcome and the exposure, whilst confounding is due to the presence of common causes of the outcome and exposure. We consider our IV estimate of the causal exposure effect, based on our selected sample for analysis, to be biased by selection when it systematically differs to the value of the causal exposure effect in the population of interest (5),.

Although there is extensive literature on IV analysis, there is relatively limited information on the effects of selection. Recent exceptions include examples where selection depends on the: exposure (6), instrument plus measured and unmeasured confounders (of the outcome-exposure association) (7), exposure and measured variable (which causes the outcome) (8), missing values of measured covariates (9, 10), and unmeasured confounder plus measured covariates (11). Typically, an IV analysis ignores how participants were selected for analysis, and so implicitly assumes that the sample selected for analysis are representative of the population. However, ignoring the selection process in an IV analysis can lead to bias (6, 8). In addition to the limited information on the consequences of nonrandom selection, the structure of the selection bias is not always fully understood. For example, Canan et al (8) describe an IV analysis where the outcome is associated with a measured variable and selection depends on the exposure and the measured variable. The authors state that in the selected sample the outcome-instrument association is confounded by the measured variable. Although correct, in this example, selection also results in unmeasured confounding of the outcome-instrument association; hence adjusting for the measured variable would not eliminate selection bias.

In this paper, we review the circumstances in which an IV analysis is biased by selection for eight examples of selection which can occur in practice. We use directed acyclic graphs (DAGs) to jointly depict the IV analysis and selection process, and show how DAGs can be used to determine when selection violates the IV assumptions. Also,

we illustrate the consequences of ignoring selection in an IV analysis via simulations. In Section 2, we introduce our IV analysis, and describe the structure of the selection bias for the eight examples of selection considered. Section 3 presents our simulation study, and section 4 presents the results of our data example. We conclude with a general discussion in Section 5.

2 When does selection lead to bias?

Whether an IV analysis is biased by selection depends on the causes of selection and the causal relationships among the variables of the IV analysis. We begin with a description of our IV analysis, and then discuss, for a range of examples of selection, whether our analysis would be biased by selection.

2.1 Description of our IV analysis

We want to estimate the effect of a continuous exposure X on a continuous outcome Y and we shall denote this causal exposure effect by β_X. The $Y-X$ association is confounded by unmeasured variables U and measured variables C. In the full sample (i.e., the selected and unselected participants), the instrument Z satisfies the three core assumptions of an IV analysis (without conditioning on C). Figure 1a is a directed acyclic graph (DAG) showing the causal relationships among these variables.

To identify β_X we assume homogeneous exposure effects; that is, β_X is the same for all individuals (12). We estimate β_X using the two stage least squares (2SLS) method (13) and denote its 2SLS estimate by $\hat{\beta}_X^{2SLS}$. In the first stage of 2SLS, X is regressed on Z to give fitted values $\hat{X}$. In the second stage, the regression coefficient of Y on fitted values $\hat{X}$ is the 2SLS estimate, $\hat{\beta}_X^{2SLS}$. When Z is a single instrument, $\hat{\beta}_X^{2SLS}$ is equivalently estimated using the ratio of coefficients method (14, 15)

$$\hat{\beta}_X^{2SLS} = \frac{\hat{E}(Y|Z)}{\hat{E}(X|Z)}, \tag{1}$$

where the numerator, $\hat{E}(Y|Z)$, is the estimated coefficient from the regression of Y on Z, and the denominator, $\hat{E}(X|Z)$, is the estimated coefficient from the regression of X on Z. We also estimate the causal exposure effect conditional on (or adjusted for) measured confounders C, and denote this conditional 2SLS estimate by $\hat{\beta}_{X|C}^{2SLS}$.

2.2 Selection mechanisms

Whether IV estimate, $\hat{\beta}_X^{2SLS}$, is biased by selection depends on the reasons for selection, which we shall call the "selection mechanism". Figures 1b to 1i depict 8 different selection mechanisms, where S is a binary variable indicating whether a participant is selected or not selected. Restricting the analysis to the selected sample implies conditioning on S which is represented by a box around S. Among the selection mechanisms we consider, $\hat{\beta}_X^{2SLS}$ is biased by

selection when selection implies conditioning on a collider in a non-causal pathway between Z and Y via C or U, which results in violating the IV assumption that there is no confounding between Z and Y (16, 17). We can ignore confounding of the $X - Z$ association because outcome Y is continuous, and so $\hat{\beta}_X^{2SLS}$ is consistently estimated even when the first stage regression (X on Z) is misspecified (18).

Figure 1b depicts selection mechanism "completely at random", where selection does not depend on any observed or unobserved data, and so the selected participants are a representative sample of the unselected participants. For the remaining selection mechanisms, there are systematic differences between the selected and unselected participants. Below, we explain when selection does and does not bias $\hat{\beta}_X^{2SLS}$, and table 1 provides a summary of this discussion. Unless otherwise stated, whether $\hat{\beta}_X^{2SLS}$ is biased by selection equally applies when X does and does not cause Y (i.e., true causal effect is not null and null). Also note, in our example all variables are measured without error since our concern is selection bias and not measurement bias. However, whether $\hat{\beta}_X^{2SLS}$ is biased by selection equally applies when selection depends on the variables measured with error (4).

When selection is completely at random (Figure 1b) or selection depends on Z (Figure 1c), $\hat{\beta}_X^{2SLS}$ is not biased by selection. The $Y - Z$ association is not confounded in the selected sample because all non-causal pathways between Z and Y (i.e., via C and U) remain blocked by colliders X and Y.

When selection depends on $Z + C$ (i.e., Z and C; Figure 1d), $\hat{\beta}_X^{2SLS}$ is biased by selection because conditioning on S opens the non-causal pathway $Z \rightarrow \boxed{\text{S}} \leftarrow C \rightarrow Y$, resulting in confounding of $Y - Z$ by C. However, conditioning on C re-blocks this pathway, thus eliminating the selection bias. Therefore, whilst $\hat{\beta}_X^{2SLS}$ is biased by selection, the 2SLS estimate conditional on C, $\hat{\beta}_{X|C}^{2SLS}$, is not biased by selection. We note one exception. Suppose selection depended on $Z + C$ but we have measured confounder C with error, denoted by C^* (DAG not shown). Conditioning on this mismeasured variable C^* would not block the non-causal pathway between Z and Y via S and C, and so selection bias would not be eliminated. Furthermore, conditioning on C^* would result in confounding the $Y - Z$ association by the measurement error of C, known as "residual confounding". However, if selection depended on Z plus the mismeasured variable C^* then conditioning on C^* would now block the non-causal pathway between Z and Y via S and C, thus eliminating the selection bias, and it would not induce residual confounding; hence the 2SLS estimate conditional on C^* would be unbiased.

When selection depends on X, $X + Z$, $X + C$ or $X + Y$ (Figures 1e, 1f, 1g and 1h respectively), $\hat{\beta}_X^{2SLS}$ is biased by selection due to confounding of the $Y - Z$ association by C and U. Because S is a descendant of X, conditioning on S unblocks any pathways which include X as a collider; that is, the non-causal pathway between Z and Y via C, $Z \rightarrow \boxed{\text{X}} \leftarrow C \rightarrow Y$, and the non-causal pathway via U, $Z \rightarrow \boxed{\text{X}} \leftarrow U \rightarrow Y$. Note, conditioning on a descendant of X is not the same as directly conditioning on X (such as in the regression of Y on X). Because we are not directly conditioning on X, pathways between Z and Y which include X as a partial mediator are not closed by selection.

For selection mechanisms $X + Z$ and $X + C$ there are also non-causal pathways between Z and Y via S and C or U, which are unblocked by conditioning on S; for example, $Z \rightarrow \boxed{\text{S}} \leftarrow X \leftarrow U \rightarrow Y$ for mechanism $X + Z$, and $Z \rightarrow X \rightarrow \boxed{\text{S}} \leftarrow C \rightarrow Y$ for mechanism $X + C$.

Selection depending on Y (Figure 1i) has the special property that $\hat{\beta}_X^{2SLS}$ is only biased by selection when X causes Y because all sources of selection bias, for this mechanism, require an edge between X and Y. When X causes Y, S is a descendant of X and so conditioning on S unblocks any non-causal pathway that includes collider X (i.e., $Z \rightarrow \boxed{\text{X}} \leftarrow C \rightarrow Y$ and $Z \rightarrow \boxed{\text{X}} \leftarrow U \rightarrow Y$); thus, the $Y - Z$ association is confounded by C and U in the selected sample. When X does not cause Y, X is not a descendant of S and so these non-causal pathways remain blocked by collider X. Consequently, the $Y - Z$ association remains unconfounded in the selected sample, and $\hat{\beta}_X^{2SLS}$ is not biased by selection.

For selection mechanisms X, $X + C$, $X + Z$, $X + Y$, and Y, estimating the causal exposure effect conditional on C, $\hat{\beta}_{X|C}^{2SLS}$, reduces the magnitude of the selection bias because conditioning on C eliminates confounding by C. However, $\hat{\beta}_{X|C}^{2SLS}$ remains biased by selection because of unmeasured confounding of the $Y - Z$ association.

3 Simulation study

3.1 Methods

For the IV analysis of Section 2.1, we investigated the effects of the 8 selection mechanisms described in Section 2.2, on the 2SLS estimate of the causal exposure effect. The simulated data were generated under the following model

$$
\begin{aligned}
z_i \sim N(0,1),\ c_i \sim N(0,1),\ u_i \sim N(0,1),\ \epsilon_{xi} \sim N(0,1),\ \epsilon_{yi} \sim N(0,1) \\
x_i = \alpha_0 + \alpha_Z z_i + \alpha_{Z^3} z_i^3 + \alpha_C c_i + \alpha_U u_i + \epsilon_{xi}, \\
y_i = \beta_0 + \beta_X x_i + \beta_C c_i + \beta_U u_i + \epsilon_{yi},
\end{aligned}
\tag{2}
$$

where, y_i, x_i, z_i, c_i and u_i respectively denote realizations of variables Y, X, Z, C and U for participant i, and ϵ_{xi} and ϵ_{yi} are normally distributed errors. The values of the data model parameters are listed in Supplementary Table S1. For simplicity, we set the constant terms, α_0 and β_0, in the equations to zero. We considered a causal exposure effect of 1 (i.e., $\beta_X = 1$) and a null causal exposure effect (i.e., $\beta_X = 0$). We set the values of $\alpha_C, \alpha_U, \beta_C$ and β_U such that, in the full sample (i.e., the selected and unselected participants), the ordinary least squares estimate of β_X (i.e., estimated by regressing Y on X) was biased due to confounding and the magnitude of this bias was close to 0.22. We considered different instrument strengths (i.e., the amount of variation of the exposure explained by the instrument) by setting the value of α_Z. To ensure consistency across all simulation settings, we measured instrument

strength using the partial $R^2_{X|Z}$ statistic rather than the F statistic since the latter is subject to large sampling variability, and so is an unreliable measure of instrument strength (19). Note, we report the population F statistic of our simulations in Section 3.2. We chose values of α_Z such that, in the full sample, partial $R^2_{X|Z}$ was close to 0.39 for the strong instrument, and 0.045 for the moderate instrument. As a sensitivity analysis, we also considered an instrument strength of partial $R^2_{X|Z}$ close to 0.15. We ensured that all instruments had sufficient strength in order to avoid weak instrument bias (i.e., bias due to finite sample size) in the full and selected samples. Lastly, we investigated whether the effects of selection differed between a linear ($\alpha_{Z^3} = 0$) and a nonlinear ($\alpha_{Z^3} = 1$) $X - Z$ association. For each combination of true value of exposure effect, form of the $X - Z$ association, and instrument strength we generated $3,000$ simulated datasets, each with $20,000$ participants for the full sample.

Following generation of the full sample datasets, participants were independently selected using the following selection model

$$Pr(\text{participant } i \text{ is selected for analysis}) = \text{expit}(\eta_0 + \eta_Z z_i + \eta_C c_i + \eta_X x_i + \eta_Y y_i),$$

where different selection mechanisms were created by setting certain parameters to 0. For example, when selection depended on Z, parameters η_C, η_X and η_Y were set to 0. For all selection mechanisms, we set the value of η_0 such that the probability of selection was close to 0.6; that is, close to 60% of the participants were selected. Supplementary Table S2 shows the values of the selection model parameters.

We used Stata command *ivregress* to perform 2SLS estimation. Of interest was the bias of $\hat{\beta}^{2SLS}_X$, the relative error of its standard error compared to the empirical standard deviation of $\hat{\beta}^{2SLS}_X$, and the coverage of the 95% confidence interval (CI) for $\hat{\beta}^{2SLS}_X$. Similarly, for the conditional estimate, $\hat{\beta}^{2SLS}_{X|C}$. We analysed the simulation results using the Stata command *simsum* (20).

3.2 Results

When there was no selection (i.e., all $20,000$ observations were analysed), $\hat{\beta}^{2SLS}_X$ was unbiased and CI coverage was nominal for all simulation settings (causal and non-causal exposure effect, linear and nonlinear $X - Z$ association, and strong and moderate instrument). (Results available on request). The F-statistics of X on Z were large, in part, due to the large sample size of $20,000$ observations: median F-statistic of 946 (interquartile range (IQR) 84) for linear $X - Z$ and moderate instrument, 12784 (IQR 386) for linear $X - Z$ and strong instrument, 936 (IQR 146) for nonlinear $X - Z$ and moderate instrument, and 12958 (IQR 470) for nonlinear $X - Z$ and strong instrument.

Table 2 presents the results of the simulation study according to the moderate and strong instrument, when the true causal exposure effect was 1, and the $X - Z$ association was linear. When selection was completely at random or

depended on Z only, $\hat{\beta}_X^{2SLS}$ was unbiased and CI coverage was nominal. Because this finding applied to all simulation settings we shall not discuss these two selection mechanisms further. For the remaining selection mechanisms, $\hat{\beta}_X^{2SLS}$ was negatively biased with poor (88%) to severe (0%) CI undercoverage. For selection mechanisms $Z + C$ and $X + Z$ the level of bias increased with decreasing instrument strength: in comparison to the strong instrument ($R^2_{X|Z} = 0.39$), the level of bias was 3 to 4 times larger for the moderate instrument ($R^2_{X|Z} = 0.045$) and close to double for instrument strength $R^2_{X|Z} = 0.15$ (results available on request). When selection did not depend on Z, there were only small differences in the level of bias between the instrument strengths. For all selection mechanisms, decreasing the instrument strength resulted in larger standard errors and higher CI coverage: in comparison to the strong instrument, the mean standard errors were 3 to 4 times larger for the moderate instrument, and close to double for instrument strength $R^2_{X|Z} = 0.15$ (results available on request). The standard errors of $\hat{\beta}_X^{2SLS}$ were neither underestimated nor overestimated (results of the relative error statistic not shown). Therefore, CI undercoverage was only due to a biased 2SLS estimate.

Despite a biased 2SLS estimate and poor CI coverage, 100% of the CIs showed evidence of a positive exposure effect for all simulation settings where the true casual effect was 1. This was in part due to the large sample size and relatively large effect size of 1 (i.e., relatively far from 0). When we lowered the effect size to 0.25, and used the moderate instrument strength, a small (1%) to large percentage (76%) of the CIs showed evidence of a null exposure effect (i.e., included the null), where the percentage of CIs including the null effect increased with larger selection bias. (Results available on request).

In general, the results for a nonlinear $X-Z$ association (Supplementary Table S3) follow the same pattern noted for a linear $X-Z$ association. However, certain differences were noted between a linear and nonlinear $X-Z$ association. For the moderate instrument the level of bias was slightly higher for nonlinear $X - Z$ when selection depended on $Z+C$, whilst for the remaining selection mechanisms it was slightly lower for nonlinear $X-Z$. To understand further we shall compare the effect of a nonlinear $X - Z$ association on selection mechanisms $Z+C$ and X for the moderate instrument. For selection mechanism $Z + C$, a nonlinear $X - Z$ association resulted in greater selection bias (i.e., stronger confounding of $Y - Z$ by C) because the $C - Y$ association was stronger for nonlinear $X - Z$. Note, the $Z-C$ association was comparable between linear and nonlinear $X-Z$ because its magnitude depended on quantities (i.e., magnitudes of the $Z - S$ and $C - S$ associations, and the distribution of S (16)) that were independent of X. For selection mechanism X, a nonlinear $X - Z$ association resulted in lower selection bias (i.e., weaker confounding of $Y - Z$ confounding by C and U) because the induced $Z-C$ and $Z-U$ associations were noticeably weaker when $X - Z$ was nonlinear. The strength of these induced associations depended on the strengths of the $Z - X$, $X - C$ and $X - U$ associations, and the distribution of X. Whilst the former were similar between linear and nonlinear $X - Z$, the distribution of X differed: for nonlinear $X - Z$, the spread of the distribution was much wider and it

was more positively skewed. The effect of this change in the distribution of X was a larger proportion of selected observations with a positive value of X, and the induced $Z - C$ and $Z - U$ associations were weak in those with a positive value of X. Consequently, among all selected values of X these induced associations were weak. When the instrument was strong, the levels of selection bias were more comparable between linear and nonlinear $X - Z$ because the nonlinearity had less of an effect on features, such as the $C - Y$ association and the distribution of X, that determined the magnitude of the selection bias.

For the moderate and strong instruments, the standard errors of $\hat{\beta}_X^{2SLS}$ were smaller for nonlinear $X - Z$ compared to linear $X - Z$, with larger differences for the moderate instrument. Note, the asymptotic variance of $\hat{\beta}_X^{2SLS}$ is $\frac{\hat{\sigma}^2}{SST_X R^2_{X|Z}}$, where $\hat{\sigma}^2$ is the estimate of the variance of the residuals from the second stage regression of 2SLS (i.e., Y on fitted values $\hat{X}$), and SST_X is the total sum of squares for X (21). The smaller standard errors for nonlinear $X - Z$ were primarily due to a substantial increase in $\hat{\sigma}^2$ (results not shown). When the level of bias was comparable between linear and nonlinear $X - Z$, CI coverages were poorer for nonlinear $X - Z$ due to smaller standard errors. However, in situations where a nonlinear $X - Z$ lowered the level of bias (e.g., when the instrument was moderate and selection depended on X only) then CI coverages were substantially higher, despite smaller standard errors.

In keeping with our discussion in Section 2.2, selection depending on Y did not bias $\hat{\beta}_X^{2SLS}$ when the causal exposure effect was null (Supplementary Table S4 for linear $X - Z$ and Supplementary Table S5 for nonlinear $X - Z$). For the remaining selection mechanisms, the results for a causal and non-causal exposure effect were very similar.

Since selecting on $Z + C$ resulted in confounding $Y - Z$ by C only, then estimating the causal exposure effect conditional on C eliminated the selection bias; that is, $\hat{\beta}_{X|C}^{2SLS}$ was unbiased and CI coverage was nominal (Supplementary Table S6). For the remaining mechanisms, whilst conditioning on C reduced the level of selection bias it did not eliminate all of the bias because the $Y - Z$ association was also confounded by U.

4 Data example

We now consider an IV analysis examining the causal effect of education on the decision to smoke using data from the UK Biobank study (22) where nonrandom selection of participants into the analysis dataset was suspected. Full details of the IV analysis are provided in (23).

Smoking is still a major avoidable risk factor for ill health and premature mortality worldwide (24). Higher educational achievement predicts lower levels of smoking uptake and, among those who started smoking, higher levels of quitting smoking (25, 26). However, we do not know whether this is due to a causal effect of education on the decision to smoke, or if this association is due to unmeasured confounding factors. To counteract the possible effects of unmeasured confounding an IV analysis was conducted where the instrument was a policy reform, introduced on 1st September 1972, which raised the school leaving age from 15 to 16 years. Individuals who turned 15 before

this date could leave school at age 15 whilst younger individuals, who turned 15 after 1st September 1972, had to remain in school at least until age 16. This example uses the policy reform (often referred to as ROSLA, Raising of School Leaving Age) as an instrument for time spent in education. The ROSLA has been used as an instrument in previous examples (23, 27, 28) and for the purposes of this example we assume that the instrument is valid. Note, the example is intended to demonstrate the impact of sample selection on linear regression and IV analyses, and not as a comprehensive analysis of the effect of education on smoking in adult life.

The UK Biobank study is a sample of $502,644$ individuals who attended 23 study clinics across the UK, and did not subsequently withdraw from the study, between 2006 and 2010 (22). At the time of the study the invited participants were aged between 40 and 69 years old and so would have turned age 15 some time between 1952 and 1985. The study achieved a response rate of 5.5% after inviting 9.2 million individuals to participate in the study (29). Higher levels of educational achievement predicted attendance at a study clinic, and so sampling by UK Biobank (29). Therefore, it is plausibe that study participants were nonrandomly selected depending on the exposure of interest, educational attainment. As discussed in Section 2.2 and shown in Section 3.2, selection depending on the exposure can bias an IV analysis. We considered it reasonable to assume that the sampling of study participants by UK Biobank was completely at random, and so this selection process was ignored. Because the ROSLA only affected England and Wales, we restricted our sample to those participants born in England or Wales. In addition, among those born in England or Wales, we only included participants who answered questions about the age at which they left school or their highest level of qualification. For the purposes of this example we have assumed that there were no systematic differences between the UK Biobank participants selected for analysis (i.e., born in England or Wales and information on the exposure) and those UK Biobank participants not selected for analysis (i.e., born outside of England and Wales, or no information on the exposure) with respect to the exposure, outcome, or any confounders of the exposure-outcome association. Under this assumption the selection process would not bias our analysis; therefore, we have ignored selection on place of birth and whether information on the exposure was observed or missing.

Using the notation of Section 2.1, the outcome Y was a binary variable, equal to one if the participant had ever smoked (i.e., includes ex-smokers and current smokers), and equal to zero if the participant had never smoked. We also considered a second outcome which was also a binary variable, equal to one if, at the time of the study clinic, the participant was a current smoker, and equal to zero if the participant did not smoke at that time (i.e., included ex-smokers and never smokers). Separate analyses were performed on each outcome using the same exposure and instrument. The exposure X was a binary variable, equal to one if the participant had left school age 16 or older, and equal to zero otherwise. The instrument, Z, was also a binary variable, equal to one if the participant turned 15 before the policy reform was introduced, and equal to zero otherwise. There were no measured confounders, C, of the exposure-outcome association but we suspect there were unmeasured confounders, U. Although the validity

of the instrument did not require conditioning on any measured variables, following Clark and Royer (27) we have adjusted for sex and month of birth.

We conducted the IV analysis using the linear probability model. This model is a form of 2SLS estimation in which the outcome, exposure and instrument are binary, and the causal exposure effect is on the risk difference scale (30). For a single binary instrument the causal exposure effect estimated under the linear probability model can be expressed as in equation 1, where the numerator, $\hat{E}(Y|Z)$, and denominator, $\hat{E}(X|Z)$, are linear regression coefficients. We note, for both of these linear regressions, Y given Z and X given Z, the dependent variable is binary which may produce predicted values outside of the 0 to 1 range (31). Robust standard errors were calculated to account for assumptions about homogeneous exposure effects and the outcome distributions. For comparison, we also considered the equivalent standard analysis; that is, the linear regression of Y on X (e.g., current smoking on "left school age 16 or older"), also with robust standard errors. Although the standard analysis may be biased by unmeasured confounding of the exposure-outcome association, we know from the missing data literature (e.g., (32)) that the results of the standard analysis are not biased by selection on the exposure, X.

A weighted IV analysis, in which the weights to try to make the selected participants representative of all invited participants, (e.g., (7)) can account for unmeasured confounding and selection bias. We used the weighting method known as inverse probability weighting (33). Under the assumption that selection only depended on the exposure X, the weights were the inverse of the conditional probabilities of selection given the exposure. The selected participants who left school age 15 were weighted by 34.29, and the selected participants who left school age 16 or older were weighted by 12.37 (weights rounded to 2 decimal places). So, those participants suspected to be under-represented in the selected sample (i.e., left school age 15) contributed more to the analysis than those suspected to be over-represented in the selected sample (i.e. left school age 16 or older). For comparison, we carried out a weighted linear regression analysis using the same weights. (See the supplementary material for information on the calculation of the weights.) For both weighted analyses the weights were treated as known (i.e., uncertainty in the estimated weights were ignored), which generally leads to conservative inference (33).

Table 3 presents the results for the exposure effect estimated using unweighted and weighted versions of the linear regression and IV analyses. For the IV analysis there were noticeable differences between the unweighted and weighted analyses. For outcome ever smoker, the weighted IV estimate was more than double that of the unweighted IV estimate, although there was some overlap between the corresponding 95% confidence intervals. Both analyses suggested staying in school at least one extra year (i.e., leaving at age 16 or older) decreased the likelihood of being an ever smoker compared to those who left school at age 15, although the confidence interval for the unweighted analysis included all 3 possible effects: risk decrease, no effect, and risk increase. For outcome current smoker, the results of the unweighted IV analysis suggested staying in school at least one extra year increased the likelihood of

being a current smoker compared to those who left school at age 15, whilst the results of the weighted IV suggested the opposite effect. Similarly, the confidence interval for the unweighted analysis included all 3 possible effects. We note, weighting also improved efficiency resulting in narrower confidence intervals.

As expected, the unweighted and weighted results of the linear regression analyses were identical. Weighting under the assumption of selection depending on the exposure will not affect the estimated value because selection under this mechanism does not bias the exposure effect estimate of a linear regression.

Comparing the (unweighted or weighted) linear regression results to the weighted IV, both of which should not be biased by selection on the exposure. Although the direction of the exposure effects were the same, the linear regression exposure effect estimates were about 2 to 3 times larger than those of the weighted IV, and there was no overlap in the 95% confidence intervals. Note, the confidence intervals of the IV analyses were noticeably wider than those of the linear regression analyses because standard errors of a 2SLS estimate are always larger than those of an OLS estimate (21).

5 Discussion

For eight different selection mechanisms, we have explained the structure of the selection bias and showed how DAGs can be used to determine if selection violates any of the IV assumptions. Using simulations, we then illustrated the effects of the 8 selection mechanisms on the 2SLS estimate of the causal exposure effect. For a causal and null exposure effect, the 2SLS estimate was not biased when the selected sample was representative of the population under study or when selection depended on the instrument only. However, it was biased when selection depended on the instrument plus measured confounder, or depended (in part or entirely) on the exposure, or depended on the outcome plus exposure. A special case was selection depending on the outcome only, where the 2SLS estimate was only biased when X truly caused Y. Decreasing the instrument strength resulted in an increase in the level of bias for selection mechanisms partly depending on the instrument, but had little effect on the other selection mechanisms. However, weakening the instrument increased the standard errors for all selection mechanisms. Whilst the larger standard errors improved CI coverage when the 2SLS estimate was biased, there was still substantial CI undercoverage. Estimating the conditional 2SLS estimate eliminated selection bias when caused by measured confounding, but only reduced the level of bias when selection resulted in measured and unmeasured confounding. Changing the exposure-instrument association from linear to nonlinear (i.e., cubic) reduced the size of the standard errors, but its effect on bias depended on the structure of the selection bias.

Similarly to our study, other authors have used DAGs to illustrate when the instrument-outcome association becomes confounded due to selection (6–8). However, two of these papers (7, 8) only discuss selection causing confounding by either measured or unmeasured variables when in fact there was confounding by both types of

variables. In keeping with the results of our simulation study, non-trivial levels of selection bias were demonstrated via simulations (6, 8). However, these papers only considered a single selection mechanism, and did not investigate if the effects of selection differed according to a null and non-null causal exposure effects, or a linear and nonlinear exposure-instrument association. Differently from our study, others have examined the effects of selection when it depends on unmeasured or partially observed data (e.g., (9–11)).

There is evidence that nonrandom selection can occur in practice with large differences in the characteristics of the selected and unselected participants (i.e., indicating a strong selection mechanism). For example, consider selection into the UK Biobank study. The percentage of subjects who owned their property outright was 56.7% in the UK Biobank study (i.e., the selected sample) and 40.6% in the 2001 UK census (i.e., the study population) (29). Since recruitment to the UK Biobank study was rare we shall assume that the percentage of subjects who owned their property outright was 40.6% in those who declined to participate in the UK Biobank study (i.e., the not selected sample). Therefore, using this approximation and the symmetry property of the odds ratio, the odds of selection among outright property owners was almost double that of those who were not outright property owners. Another example is selection into the Avon Longitudinal Study of Parents and Children (ALSPAC) study (34). Using similar calculations, where the study population is the Avon population and not the whole of the UK, the odds of selection among households with a car was almost double the odds of section among households without a car.

In our simulation study the direction of the selection bias was negative for a causal and non-causal exposure effect. However, we would have obtained positive bias if we had set the direction of the exposure's associations with the confounders to be in the opposite direction to that of the outcome's associations with the confounders; that is, set the exposure to be positively associated with the confounders and the outcome to be negatively associated with the confounders, or vice versa. Also, we note that the magnitude of the selection bias depends on several features of the data (e.g., distribution of the exposure, strength of the instrument-exposure, magnitude of the confounder-exposure and confounder-outcome associations) (6), as well as the strength of the selection mechanism (e.g., how strongly the exposure is associated with being selected). Regarding testing for the existence of a causal exposure effect, the sample size and true effect size (i.e., distance from the null) also affect how likely the results will lead to an incorrect conclusion.

Our simulation study has several limitations. First, whilst we considered eight plausible selection mechanisms, it was not possible to investigate all possible selection mechanisms even for a single IV analysis example. Second, in practice an IV analysis may use weaker instruments than we considered. We chose a sample size that was typical of an IV analysis so that even for an partial $R^2_{X|Z}$ of 0.045 the instrument would not be considered weak. However, for the purposes of our study, we wanted to ensure that any bias was attributable only to selection and not to weak instrument bias (15). Third, our simulation study was designed to show the effects of different selection mechanisms on an IV

analysis and not an exhaustive investigation of the magnitudes of selection bias that could occur in practice. Such an investigation would be best limited to a single selection mechanism where different values for the parameters, of the data and selection models, that can affect the magnitude of selection bias are considered, and choosing parameters values based on real IV analyses. Fourth, our use of non-parametric DAGs, to determine if selection would violate one of the core IV assumptions, are not suitable for all types of selection mechanisms (e.g., selection depending on the parameterization of the IV analysis (5)). However, for non-parametric selection mechanisms, DAGs are easier to understand than mathematical statements about the validity of an IV analysis under selection.

For some selection mechanisms, selection biases the 2SLS estimate but not the ordinary least squares estimate; for example, when selection depends on the exposure. Although, unlike the IV estimate, the ordinary least squares estimate may be biased by confounding. In such cases the magnitude of the selection bias (for the 2SLS estimate) may exceed the bias due to confounding of the ordinary least squares estimate. Ideally, estimation of the causal exposure effect should correctly account for unmeasured confounding and nonrandom selection. Inverse probability weighting (8) and multiple imputation (10) have been used to appropriately account for selection in an IV analysis. Commercially available implementations of these methods usually assume the chance of selection depends on measured variables that are fully observed for all participants of the full sample. When selection depends on unmeasured variables or partially observed variables, MI and IPW may not fully account for selection and give a biased 2SLS estimate (7, 10). There are a limited number of approaches that try to account for selection depending on unmeasured or partially observed data (7, 9, 10), but they tend to be less straightforward, make untestable assumptions or are specific to a particular type of IV analysis.

With individual level data on the selected and unselected participants, an IV analyst can investigate possible factors that influence selection. However, this is not possible when the IV analyst only has summary level data. Providers of summary level data should indicate the chance that the summary level data is based on a nonrandom sample of the population of interest, and posit possible selection mechanisms. Where possible, these providers could generate summary level data accounting for nonrandom selection (e.g., summary level data from a weighted analysis, or summary level data adjusted for known factors associated with selection). Two sample IV analyses tend to be conducted using summary level data, and these analyses are further complicated because there are two opportunities for nonrandom selection to occur, and possibly two different selection mechanisms to take into account. Future work could develop doubly robust methods to account for selection, which would give valid results even when one of the samples has been incorrectly adjusted for selection.

There is current debate regarding representativeness of a study's results and findings to populations that differ from the study population (e.g., (35)). However, nonrandom selection is concerned with the accuracy of a study's results for the study population; that is, whether the selected sample is unrepresentative of the study population

leading to study results that systematically differ from their true values in the study population. Therefore, we consider nonrandom selection to be a separate issue from generalisability to other populations.

In summary, ignoring how participants are selected for analysis can result in a biased 2SLS estimate, substantial CI undercoverage, and lead to an incorrect conclusion that an exposure is not causal. This limitation should be noted in guidelines for IV analyses. DAGs can be used to assess if the IV analysis may be biased by the assumed selection mechanism. Future work could provide researchers guidance on statistical methods, diagnostic tools and sensitivity analyses for estimating causal effects from nonrandom samples.

Acknowledgments

We thank Professor Frank Windmeijer, of the Department of Economics, University of Bristol, for his helpful comments.

The Medical Research Council (MRC) and the University of Bristol support the MRC Integrative Epidemiology Unit [MC_UU_12013/1, MC_UU_12013/9]. The Economics and Social Research Council (ESRC) support Neil Davies via a Future Research Leaders grant [ES/N000757/1]. No funding body has influenced data collection, analysis or its interpretations. This publication is the work of the authors, who serve as the guarantors for the contents of this paper. This work was carried out using the computational facilities of the Advanced Computing Research Centre - http://www.bris.ac.uk/acrc/ and the Research Data Storage Facility of the University of Bristol - http://www.bris.ac.uk/acrc/storage/.

This research has been conducted using the UK Biobank Resource under Application Number 8786.

References

1. Angrist J.D., Imbens G.W., Rubin D.B. Identification of causal effects using instrumental variables. *J Am Stat Assoc.* 1996;91:444–455.

2. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol.* 2000;29:722–729.

3. Hernán M.A., Robins J.M. Instruments for causal inference. An epidemiologist's dream?. *Epidemiology.* 2006;17:360–372.

4. Hernán M.A., Robins J.M. *Causal Inference.*: Chapman & Hall/CRC, forthcoming; 2017.

5. Hernán M.A. Selection bias without colliders. *American Journal of Epidemiology.* 2017;185:1048–1050.

6. Swanson S.A., Robins J.M., Miller M., Hernán M.A. Selecting on treatment: A pervasive form of bias in instrumental variable analyses. *American Journal of Epidemiology.* 2015;181:191–197.

7. Ertefaie A., Small D., Flory J., Hennessy S. Selection bias when using instrumental variable methods to compare two treatments but more than two treatments are available. *The International Journal of Biostatistics.* 2016;12:219–232.

8. Canan C., Lesko C., Lau B. Instrumental variable analyses and selection bias. *Epidemiology*. 2017;28:396–398.

9. Ertefaie A., Flory J.H., Hennessy S., Small D.S. Instrumental variable methods for continuous outcomes that accommodate nonignorable missing baseline values. *American Journal of Epidemiology*. 2017.

10. Yang F., Lorch S.A., Small D.S. Estimation of causal effects using instrumental variables with nonignorable missing covariates: application to effect of type of delivery NICU on premature infants. *The Annals of Applied Statistics*. 2014;8:48–73.

11. Mogstad M., Wiswall M. Instrumental variables estimation with partially missing instruments. *Economics Letters 114*. 2012;114:186–189.

12. Baiocchi M., Cheng J., Small D.S. Instrumental variable methods for causal inference. *Statistics in Medicine*. 2014;33:2297–2340.

13. Baum C., Schaffer M., Stillman S. Instrumental variables and gmm: estimation and testing. *Stata Journal*. 2003;3(1):1–31.

14. Lawlor D., Harbord R., Sterne J., Timpson N., G. Davey Smith. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*. 2008;27:1133–1163.

15. Burgess S., Thompson S.G. Bias in causal estimates from mendelian randomization studies with weak instruments. *Statistics in Medicine*. 2011;30:1312–1323.

16. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300Ũ306.

17. Hernán M.A., Hernández-Diaz S., Robins J.M. A structural approach to selection bias. *Epidemiology*. 2004;15:615–625.

18. Wooldridge J.M. *Econometric analysis of cross section and panel data.*: Cambridge, MA: MIT Press; 2002.

19. Burgess S., Thompson S.G. *Mendelian randomization. methods for using genetic variants in causal estimation.*: Chapman & Hall/CRC; 2015.

20. White I.R. simsum: Analyses of simulation studies including Monte Carlo error. *The Stata Journal*. 2010;10:369–385.

21. Wooldridge J.M. *Introductory econometrics. a modern approach*. Sixth Edition: Cengage Learning; 2016.

22. Sudlow C., Gallacher J., Allen N., Beral V., Burton P., Danesh J., Downey P., Elliott P., Green J., Landray M., Liu B., Matthews P., Ong G., Pell J., Silman A., Young A., Sprosen T., Peakman T., Collins R. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*. 2015;12, DOI doi: 10.1371/journal.pmed.1001779.

23. Davies N.M., Dickson M., Davey Smith G., van den Berg G., Windmeijer F. *The Causal Effects of Education on Health, Mortality, Cognition, Well-being, and Income in the UK Biobank*. Article submitted for peer review.

24. Jha P. Avoidable global cancer deaths and total deaths from smoking. *Nature Reviews Cancer*. 2009;9:655–664.

25. Gage S.H., Bowden J., Davey Smith G., Munafo M.R. *Investigating causality in associations between education and smoking: A two-sample Mendelian randomization study*. Article submitted for peer review.

26. Jeffris B.J., Power C., Graham H., Manor O. Changing social gradients in cigarette smoking and cessation over two decades of adult follow-up in a British birth cohort. *Journal of Public Health*. 2004;26:13–18.

27. Clark D., Royer H. The Effect of Education on Adult Mortality and Health: Evidence from Britain. *American Economic Review*. 2013;103:2087–2120.

28. Harmon C., Walker I. Estimates of the Economic Return to Schooling for the United Kingdom. *The American Economic Review.* 1995;85:1278–1286.

29. Fry A., Littlejohns T.J., Sudlow C., Doherty N., Adamska L., Sprosen T., Collins R., Allen N.E. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology.* 2017, DOI doi: 10.1093/aje/kwx246.

30. Klungel O.H., Jamal Uddin M., de Boer A., Belitser S.V., Groenwold R.H., Roes K.C. Instrumental Variable Analysis in Epidemiologic Studies: An Overview of the Estimation Methods. *Pharmaceutica Analytica Acta.* 2015;353, DOI 10.4172/2153-2435.1000353.

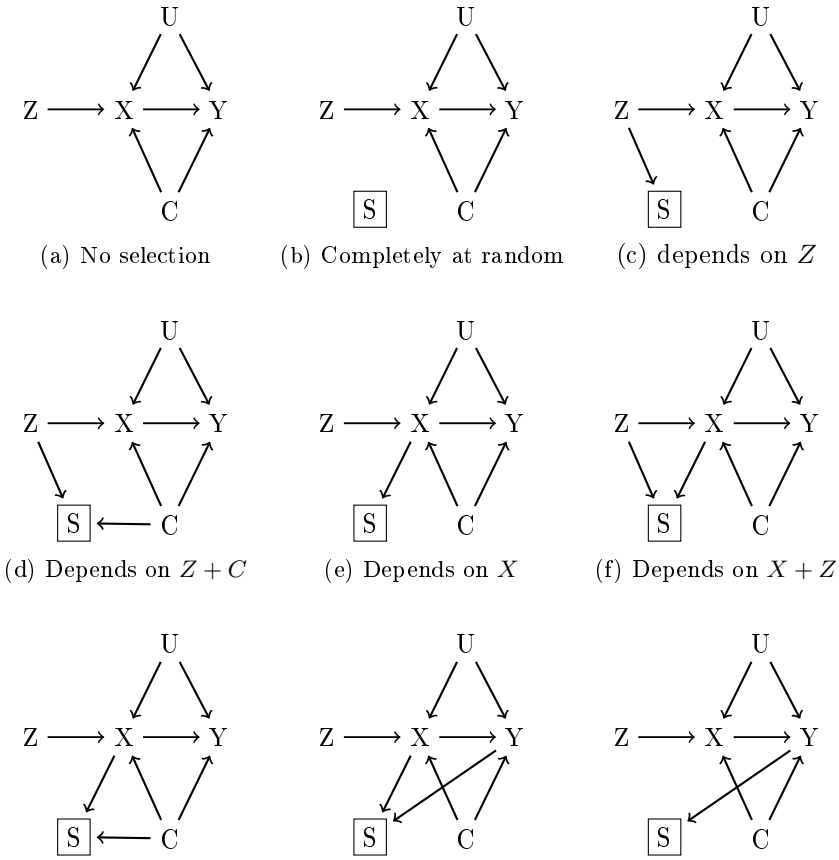
31. Cameron A.C. *Microeconometrics: methods and applications.*: Cambridge University Press; 2005.

32. Bartlett J.W., Harel O., Carpenter J.R. Asymptotically unbiased estimation of exposure odds ratios in complete records logistic regression. *American Journal of Epidemiology.* 2015;182:730–736.

33. Seaman S.R., White I.R. Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research.* 2013;22:278–295.

34. Fraser A., Macdonald-Wallis C., Tilling K., Boyd A., Golding J., Davey Smith G., Henderson J., Macleod J., Molloy L., Ness A., Ring S., Nelson S.M., Lawlor D.A. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology.* 2013;42:97–110.

35. Rothman K.J., Gallacher J.E.J., Hatch E.E. Why representativeness should be avoided. *International Journal of Epidemiology.* 2013;42:1012–1014.



(a) No selection (b) Completely at random (c) depends on Z

(d) Depends on $Z + C$ (e) Depends on X (f) Depends on $X + Z$

(g) Depends on $X + C$ (h) Depends on $X + Y$ (i) Depends on Y

Figure 1: Directed acyclic graphs of an instrumental variable analysis under no selection and eight different selection mechanisms.

Table 1: Potential bias of the two stage least squares (2SLS) estimate of the casual exposure effect, $\hat{\beta}_X^{2SLS}$, and the corresponding 2SLS estimate conditional on C, $\hat{\beta}_{X|C}^{2SLS}$, according to different selection mechanisms.

Selection is/ depends on	$Y - Z$ association$^\sharp$	$\hat{\beta}_X^{2SLS}$	$\hat{\beta}_{X\|C}^{2SLS}$
Completely at random	Unconfounded	Unbiased	Unbiased
Z	Unconfounded	Unbiased	Unbiased
$Z + C$	Confounded by C	Biased	Unbiased
X	Confounded by C and U	Biased	Biased
$X + C$	Confounded by C and U	Biased	Biased
$X + Z$	Confounded by C and U	Biased	Biased
$X + Y$	Confounded by C and U	Biased	Biased
Y	Confounded by C and U	Biased$^\flat$	Biased$^\flat$

$^\sharp$ $Y - Z$ association in the selected sample.

$^\flat$ Not biased by selection when X does not cause Y.

Table 2: Summary of the simulation results of the two stage least squares (2SLS) estimate of the casual exposure effect, $\hat{\beta}_X^{2SLS}$, for true value of 1 and linear $X - Z$ association: Median (interquartile range) of the partial $R^2_{X|Z}$ of the $X - Z$ association, bias of $\hat{\beta}_X^{2SLS}$, mean of the estimated standard errors (SE) of $\hat{\beta}_X^{2SLS}$, and coverage percentage of the 95% confidence interval for $\hat{\beta}_X^{2SLS}$. Monte Carlo errors in brackets, except for partial $R^2_{X|Z}$.

Selection is/	Moderate Instrument				Strong Instrument			
depends on	partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage	partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage
Completely at random	0.0452	0.000840	0.0420	95.4	0.390	0.000286	0.0114	95.3
	(0.00370)	(0.000758)	(0.0000360)	(0.382)	(0.00695)	(0.000205)	(0.00000341)	(0.385)
Z	0.0384	0.000144	0.0460	95.2	0.350	0.000121	0.0125	94.8
	(0.00338)	(0.000832)	(0.0000419)	(0.390)	(0.00693)	(0.000226)	(0.00000393)	(0.404)
Z + C	0.0322	−0.153	0.0525	15.2	0.353	−0.0470	0.0125	3.33
	(0.00317)	(0.000951)	(0.0000591)	(0.655)	(0.00698)	(0.000227)	(0.00000406)	(0.328)
X	0.0382	−0.0423	0.0502	87.4	0.348	−0.0429	0.0133	9.67
	(0.00345)	(0.000915)	(0.0000483)	(0.606)	(0.00691)	(0.000243)	(0.00000426)	(0.540)
X + C	0.0425	−0.0384	0.0461	87.7	0.370	−0.0432	0.0126	6.90
	(0.00364)	(0.000854)	(0.0000422)	(0.600)	(0.00690)	(0.000229)	(0.00000391)	(0.463)
X + Z	0.0163	−0.159	0.0781	46.1	0.325	−0.0404	0.0136	15.9
	(0.00229)	(0.00142)	(0.000125)	(0.910)	(0.00696)	(0.000249)	(0.00000458)	(0.668)
Y	0.0425	−0.0833	0.0441	52.7	0.373	−0.0704	0.0121	0.00
	(0.00366)	(0.000812)	(0.0000397)	(0.912)	(0.00693)	(0.000223)	(0.00000374)	(0.00)
X + Y	0.0406	−0.0857	0.0472	55.1	0.362	−0.0688	0.0127	0.00
	(0.00355)	(0.000866)	(0.0000436)	(0.908)	(0.00692)	(0.000231)	(0.00000398)	(0.00)

Table 3: Risk difference %, of ever smoker or current smoker, for leaving school at age 16 or older compared to leaving school at age 15 using unweighted and weighted versions of linear regression and instrumental variable (IV) analysis. 95% confidence intervals displayed within brackets.

	Linear regression		IV analysis	
Outcome	Unweighted	Weighted	Unweighted	Weighted
Ever smoker	-20.5 $(-22.8, -18.3)$	-20.5 $(-22.8, -18.3)$	-4.8 $(-11.6, 1.9)$	-11.6 $(-15.4, -7.8)$
Current smoker	-14.1 $(-15.5, -12.7)$	-14.1 $(-15.5, -12.7)$	1.8 $(-1.5, 5.0)$	-5.6 $(-7.6, -3.6)$

6 Supplementary Material

6.1 Calculation of the weights

In our data example, Section 4, we conducted a weighted IV analysis and a weighted standard analysis. To calculate the weights we have assumed that the exposure (an extra year of compulsory education) was the only factor causing participation into Biobank. Also, we have assumed that there was no effect of cigarette smoking on selection into UK Biobank (although Fry et al, (29), would suggest otherwise). This example is for illustrative purposes only; in practice a thorough investigation of the factors likely to affect selection would be necessary.

Under the assumption that selection S only depended on the exposure X, the weights were the inverse of the conditional probabilities of selection given the exposure, $Pr(S = 1|X)$. We can see from figure 1e that under this selection mechanism, selection is independent of outcome Y and instrument Z after conditioning on X (i.e., pathways between S and Z, and S and Y are closed after conditioning on X). To calculate $Pr(S = 1|X)$ we use Bayes theorem

$$Pr(S|X) = \frac{Pr(X|S)Pr(S)}{Pr(X)} \tag{3}$$

.

First, let us calculate the conditional probability of selection for those participants who left school at age 15 (i.e., $Pr(S = 1|X = 0)$). We used information from the 2011 UK census to approximate information on all individuals invited to participate in the study (i.e., the selected and unselected). Prior to the introduction of the policy reform (i.e., $Z = 0$), September 1972, 33% of individuals left school at age 15, and so we set $Pr(X = 0|Z = 0) = 0.33$. Among the selected participants, 17.5% left school at age 15 prior to September 1972; that is, $Pr(X = 0|S = 1, Z = 0) = 0.175$. After the introduction of the policy reform, a small number of individuals did leave school at age 15. However, for the purposes of our illustrative example we have assumed from September 1972 onwards that all individuals left school aged 16 or older; that is, $Pr(X = 0|Z = 1) = 0$ and $Pr(X = 0|S = 1, Z = 1) = 0$. Therefore, $Pr(X = 0) = Pr(X = 0|Z = 0)Pr(Z = 0)$ and $Pr(X = 0|S = 1) = Pr(X = 0|S = 1, Z = 0)Pr(Z = 0)$. For the probability of selection, we know that $502,644$ out of 9.2 million invited participants attended the study clinic, and

so $Pr(S = 1) = 502644/9200000 = 0.055$ (2 decimal places). Putting everything together we have:

$$\begin{aligned} Pr(S = 1|X = 0) &= \frac{Pr(X = 0|S = 1)Pr(S = 1)}{Pr(X = 0)} \\ &= \frac{Pr(X = 0|S = 1, Z = 0)Pr(Z = 0)Pr(S = 1)}{Pr(X = 0|Z = 0)Pr(Z = 0)} \\ &= \frac{0.175 \times 0.055}{0.33} \\ &= 0.02916667 \end{aligned} \tag{4}$$

.

Similarly, we can calculate the conditional probability of selection for those participants who left school age 16 or older (i.e., $Pr(S = 1|X = 1)$). From the UK census we have $Pr(X = 1|Z = 0) = 1 - 0.33$, and from the selected sample we have $Pr(X = 1|S = 1, Z = 0) = 1 - 0.175$. Given our assumption that post September 1972 all individuals left school aged 16 or older, then $Pr(X = 1|Z = 1) = 1$ and $Pr(X = 1|S = 1, Z = 1) = 1$. Since individuals could leave school age 16 or older before and after September 1972 then $Pr(X = 1) = 0.67 \times Pr(Z = 0) + Pr(Z = 1) = 1 - 0.33 \times Pr(Z = 0)$, and $Pr(X = 1|S = 1) = 0.825 \times Pr(Z = 0) + Pr(Z = 1) = 1 - 0.175 \times Pr(Z = 0)$. The probability $Pr(Z = 0)$ is the proportion of the 9.2 million invited participants to which the introduction of the policy reform did not apply; that is, the proportion that turned 15 before September 1972. The recruitment period for the UK Biobank study was between 2006 and 2010, and so the mid-point was 2008. Participants unaffected by the policy reform had to have turned 51 years before September 2008. We approximate $Pr(Z = 0)$ using population estimates for England and Wales, Scotland and Northern Ireland supplied by the Office for National Statistics. Based on these population estimates for June 2008, among those aged between 40 to 70 years (UK Biobank's sampling ages) the proportion over 50 was 0.58. So, we estimated that 58% of the invited UK Biobank participants had turned 51 years before September 2008 (i.e., $Pr(Z = 0) = 0.58$). Putting everything together we have:

$$\begin{aligned} Pr(S = 1|X = 1) &= \frac{Pr(X = 1|S = 1)Pr(S = 1)}{Pr(X = 1)} \\ &= \frac{[1 - 0.175 \times Pr(Z = 0)]Pr(S = 1)}{1 - 0.33 \times Pr(Z = 0)} \\ &= \frac{[1 - 0.175 \times 0.58] \times 0.055}{1 - 0.67 \times 0.58} \\ &= 0.08082679 \end{aligned} \tag{5}$$

.

Therefore, the selected participants who left school age 15 were weighted by 34.29 and the selected participants who left school age 16 or older were weighted by 12.37 (weights rounded to 2 decimal places). So, those participants

suspected to be under-represented in the selected sample (i.e., left school age 15) contributed more to the analysis than those suspected to be over-represented in the selected sample (i.e. left school age 16 or older).

Table S1: Values of the data model's parameters according to whether the exposure−instrument $(X - Z)$ association was linear or nonlinear, instrument strength was moderate or strong, and true causal exposure effect was 1 or 0. For all simulation settings, constant coefficients $\alpha_0 = \beta_0 = 0$, and confounders' coefficients $\alpha_C = \beta_U = \beta_C = \alpha_U$.

$X - Z$ association	Instrument Strength	Causal exposure effect	α_Z	α_Z^3	α_U
Linear	Moderate	0 or 1	0.248	0	0.380
Linear	Strong	0 or 1	1.00	0	0.531
Nonlinear	Moderate	0	−2.34	1	1.04
Nonlinear	Moderate	1	−2.35	1	1.04
Nonlinear	Strong	0 or 1	−0.320	1	1.43

Table S2: Values of the selection model's parameters for the 8 selection mechanisms, including selection completely at random (SCAR), according to whether the exposure−instrument ($X - Z$) association was linear or nonlinear, instrument strength was moderate or strong, and true causal exposure effect was 1 or 0.

Selection is/ depends on	$X - Z$ association	Instrument	Causal exposure effect	η_0	η_Z	η_C	η_X	η_Y
SCAR	Linear or nonlinear	Moderate or strong	1 or 0	0.425	0	0	0	0
Z	Linear or nonlinear	Moderate or strong	1 or 0	0.490	1.00	0	0	0
$Z + C$	Linear or nonlinear	Moderate or strong	1 or 0	0.500	0.700	0.700	0	0
X	Linear	Moderate	1 or 0	0.517	0	0	0.875	0
	Linear	Strong	1 or 0	0.517	0	0	0.640	0
	Nonlinear	Moderate	0	0.510	0	0	0.435	0
	Nonlinear	Moderate	1	0.500	0	0	0.425	0
	Nonlinear	Strong	0	0.500	0	0	0.325	0
	Nonlinear	Strong	1	0.490	0	0	0.320	0
$X + C$	Linear	Moderate	1 or 0	0.500	0	0.575	0.575	0
	Linear	Strong	1 or 0	0.500	0	0.475	0.475	0
	Nonlinear	Moderate	0	0.500	0	0.340	0.340	0
	Nonlinear	Moderate	1	0.500	0	0.340	0.340	0
	Nonlinear	Strong	0	0.500	0	0.260	0.260	0
	Nonlinear	Strong	1	0.500	0	0.265	0.265	0
$X + Z$	Linear	Moderate	1 or 0	0.500	0.600	0	0.600	0
	Linear	Strong	1 or 0	0.500	0.425	0	0.425	0
	Nonlinear	Moderate	0	0.500	0.425	0	0.425	0
	Nonlinear	Strong	0	0.500	0.270	0	0.270	0
	Nonlinear	Moderate	1	0.500	0.425	0	0.425	0
	Nonlinear	Strong	1	0.500	0.275	0	0.275	0
Y	Linear	Moderate	0	0.500	0	0	0	0.880
	Linear	Moderate	1	0.500	0	0	0	0.550
	Linear	Strong	0	0.500	0	0	0	0.800
	Linear	Strong	1	0.500	0	0	0	0.435
	Nonlinear	Moderate	0	0.500	0	0	0	0.570
	Nonlinear	Moderate	1	0.500	0	0	0	0.270
	Nonlinear	Strong	0	0.500	0	0	0	0.450
	Nonlinear	Strong	1	0.500	0	0	0	0.200
$X + Y$	Linear	Moderate	0	0.500	0	0	0.550	0.550
	Linear	Moderate	1	0.500	0	0	0.355	0.355
	Linear	Strong	0	0.500	0	0	0.440	0.440
	Linear	Strong	1	0.500	0	0	0.265	0.265
	Nonlinear	Moderate	0	0.505	0	0	0.275	0.275
	Nonlinear	Moderate	1	0.500	0	0	0.170	0.170
	Nonlinear	Strong	0	0.500	0	0	0.200	0.200
	Nonlinear	Strong	1	0.500	0	0	0.122	0.122

Table S3: Summary of the simulation results of the two stage least squares (2SLS) estimate of the casual exposure effect, $\hat{\beta}_X^{2SLS}$, for true value of 1, and nonlinear $X - Z$ association: Median (interquartile range) of the partial $R^2_{X|Z}$ of the $X - Z$ association, bias of $\hat{\beta}_X^{2SLS}$, mean of the estimated standard errors (SE) of $\hat{\beta}_X^{2SLS}$, and coverage percentage of the 95% confidence interval for $\hat{\beta}_X^{2SLS}$. Monte Carlo errors in brackets, except for partial $R^2_{X|Z}$.

Selection is/	Moderate Instrument				Strong Instrument			
depends on	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage
Completely at random	0.0449	0.000179	0.0250	95.0	0.393	0.000130	0.00766	95.1
	(0.00652)	(0.000451)	(0.0000486)	(0.398)	(0.00866)	(0.000138)	(0.00000426)	(0.394)
Z	0.0314	−0.000305	0.0314	95.1	0.350	0.0000716	0.00865	95.1
	(0.00552)	(0.000567)	(0.0000736)	(0.395)	(0.00829)	(0.000157)	(0.00000506)	(0.394)
Z + C	0.0267	−0.176	0.0388	0.00	0.352	−0.0485	0.00877	0.00
	(0.00520)	(0.000794)	(0.000118)	(0.00)	(0.00840)	(0.000162)	(0.00000529)	(0.00)
X	0.0330	0.00117	0.0319	94.9	0.339	−0.0519	0.00936	0.00
	(0.00571)	(0.000574)	(0.0000732)	(0.400)	(0.00826)	(0.000173)	(0.00000574)	(0.00)
X + C	0.0335	−0.00919	0.0315	94.6	0.350	−0.0501	0.00908	0.00
	(0.00569)	(0.000569)	(0.0000709)	(0.414)	(0.00833)	(0.000168)	(0.00000545)	(0.00)
X + Z	0.0177	−0.132	0.0492	18.4	0.322	−0.0599	0.00974	0.00
	(0.00442)	(0.000959)	(0.000189)	(0.708)	(0.00822)	(0.000181)	(0.00000622)	(0.00)
Y	0.0343	−0.0195	0.0302	90.7	0.365	−0.0503	0.00862	0.00
	(0.00571)	(0.000555)	(0.0000666)	(0.531)	(0.00840)	(0.000159)	(0.00000499)	(0.00)
X + Y	0.0334	−0.0107	0.0312	94.5	0.354	−0.0526	0.00894	0.00
	(0.00568)	(0.000569)	(0.0000696)	(0.417)	(0.00836)	(0.000165)	(0.00000531)	(0.00)

Table S4: Summary of the simulation results of the two stage least squares (2SLS) estimate of the casual exposure effect, $\hat{\beta}_X^{2SLS}$, for true value of 0, and linear $X - Z$ association: Median (interquartile range) of the partial $R^2_{X|Z}$ of the $X - Z$ association, bias of $\hat{\beta}_X^{2SLS}$, mean of the estimated standard errors (SE) of $\hat{\beta}_X^{2SLS}$, and coverage percentage of the 95% confidence interval for $\hat{\beta}_X^{2SLS}$. Monte Carlo errors in brackets, except for partial $R^2_{X|Z}$.

Selection is/	Moderate Instrument				Strong Instrument			
depends on	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage
Completely at random	0.0452	0.000840	0.0420	95.4	0.390	0.000286	0.0114	95.3
	(0.00370)	(0.000758)	(0.0000360)	(0.382)	(0.00695)	(0.000205)	(0.00000341)	(0.385)
Z	0.0384	0.000144	0.0460	95.2	0.350	0.000121	0.0125	94.8
	(0.00338)	(0.000832)	(0.0000419)	(0.390)	(0.00693)	(0.000226)	(0.00000393)	(0.404)
Z + C	0.0322	−0.153	0.0525	15.2	0.353	−0.0470	0.0125	3.33
	(0.00317)	(0.000951)	(0.0000591)	(0.655)	(0.00698)	(0.000227)	(0.00000406)	(0.328)
X	0.0382	−0.0423	0.0502	87.4	0.348	−0.0429	0.0133	9.67
	(0.00345)	(0.000915)	(0.0000483)	(0.606)	(0.00691)	(0.000243)	(0.00000426)	(0.540)
X + C	0.0425	−0.0384	0.0461	87.7	0.370	−0.0432	0.0126	6.90
	(0.00364)	(0.000854)	(0.0000422)	(0.600)	(0.00690)	(0.000229)	(0.00000391)	(0.463)
X + Z	0.0163	−0.159	0.0781	46.1	0.325	−0.0404	0.0136	15.9
	(0.00229)	(0.00142)	(0.000125)	(0.910)	(0.00696)	(0.000249)	(0.00000458)	(0.668)
Y	0.0457	0.00108	0.0385	95.6	0.395	0.000284	0.0105	95.7
	(0.00370)	(0.000693)	(0.0000323)	(0.373)	(0.00695)	(0.000189)	(0.00000309)	(0.369)
X + Y	0.0425	−0.0833	0.0441	52.7	0.373	−0.0716	0.0121	0.00
	(0.00366)	(0.000812)	(0.0000397)	(0.912)	(0.00693)	(0.000224)	(0.00000374)	(0.00)

Table S5: Summary of the simulation results of the two stage least squares (2SLS) estimate of the casual exposure effect, $\hat{\beta}_X^{2SLS}$, for true value of 0 and nonlinear $X - Z$ association: Median (interquartile range) of the partial $R^2_{X|Z}$ of the $X - Z$ association, bias of $\hat{\beta}_X^{2SLS}$, mean of the estimated standard errors (SE) of $\hat{\beta}_X^{2SLS}$, and coverage percentage of the 95% confidence interval for $\hat{\beta}_X^{2SLS}$. Monte Carlo errors in brackets, except for partial $R^2_{X|Z}$.

Selection is/	Moderate Instrument				Strong Instrument			
depends on	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage
Completely at random	0.0455	0.000180	0.0249	95.0	0.393	0.000130	0.00766	95.1
	(0.00655)	(0.000448)	(0.0000479)	(0.397)	(0.00866)	(0.000138)	(0.00000426)	(0.394)
Z	0.0319	−0.000298	0.0312	95.1	0.350	0.0000716	0.00865	95.1
	(0.00555)	(0.000563)	(0.0000724)	(0.395)	(0.00829)	(0.000157)	(0.00000506)	(0.394)
Z + C	0.0272	−0.174	0.0384	0.00	0.352	−0.0485	0.00877	0.00
	(0.00523)	(0.000785)	(0.000115)	(0.00)	(0.00840)	(0.000162)	(0.00000529)	(0.00)
X	0.0335	0.00156	0.0317	95.0	0.339	−0.0528	0.00938	0.00
	(0.00573)	(0.000570)	(0.0000719)	(0.397)	(0.00824)	(0.000174)	(0.00000575)	(0.00)
X + C	0.0340	−0.00962	0.0312	94.4	0.351	−0.0488	0.00906	0.0333
	(0.00572)	(0.000565)	(0.0000698)	(0.421)	(0.00832)	(0.000168)	(0.00000541)	(0.0333)
X + Z	0.0180	−0.132	0.0486	17.9	0.323	−0.0583	0.00970	0.00
	(0.00445)	(0.000947)	(0.000184)	(0.699)	(0.00822)	(0.000180)	(0.00000618)	(0.00)
Y	0.0466	0.000692	0.0228	95.2	0.405	0.000225	0.00703	94.7
	(0.00643)	(0.000419)	(0.0000424)	(0.392)	(0.00855)	(0.000129)	(0.00000381)	(0.409)
X + Y	0.0347	−0.0200	0.0301	90.4	0.365	−0.0503	0.00862	0.00
	(0.00574)	(0.000550)	(0.0000657)	(0.537)	(0.00840)	(0.000159)	(0.00000499)	(0.00)

Table S6: Summary of the simulation results of the two stage least squares (2SLS) estimator of the casual exposure effect conditional on C, $\beta^{2SLS}_{X,C}$, for a linear $X - Z$ association and a causal exposure effect of 1: Median (interquartile range) of the partial $R^2_{X|Z}$ of the $X - Z$ association, bias of $\beta^{2SLS}_{X,C}$, mean of the estimated standard errors (SE) of $\beta^{2SLS}_{X,C}$, and coverage percentage of the 95% confidence interval for $\beta^{2SLS}_{X,C}$. Monte Carlo errors in brackets, except for partial $R^2_{X|Z}$.

Selection is/	Moderate Instrument				Strong Instrument			
depends on	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage
Completely at random	0.0508	0.000937	0.0394	95.1	0.438	0.000274	0.0103	94.8
	(0.00390)	(0.000714)	(0.0000304)	(0.393)	(0.00678)	(0.000186)	(0.00000274)	(0.405)
Z	0.0431	0.0000655	0.0432	95.2	0.396	0.0000484	0.0113	95.2
	(0.00356)	(0.000791)	(0.0000357)	(0.392)	(0.00679)	(0.000207)	(0.00000315)	(0.390)
Z + C	0.0468	0.000468	0.0413	95.5	0.417	0.000156	0.0108	95.4
	(0.00376)	(0.000745)	(0.0000333)	(0.378)	(0.00678)	(0.000195)	(0.00000296)	(0.382)
X	0.0437	−0.0208	0.0462	93.3	0.399	−0.0215	0.0117	54.6
	(0.00368)	(0.000846)	(0.0000395)	(0.456)	(0.00681)	(0.000213)	(0.00000326)	(0.909)
X + C	0.0474	−0.00859	0.0425	94.8	0.416	−0.0118	0.0111	82.7
	(0.00382)	(0.000778)	(0.0000347)	(0.407)	(0.00677)	(0.000202)	(0.00000307)	(0.691)
X + Z	0.0209	−0.0743	0.0662	80.3	0.377	−0.0200	0.0120	60.7
	(0.00259)	(0.00120)	(0.0000829)	(0.727)	(0.00693)	(0.000221)	(0.00000347)	(0.892)
Y	0.0479	−0.0687	0.0411	60.9	0.421	−0.0527	0.0108	0.333
	(0.00384)	(0.000760)	(0.0000332)	(0.891)	(0.00677)	(0.000199)	(0.00000297)	(0.105)
X + Y	0.0460	−0.0659	0.0436	67.8	0.411	−0.0477	0.0113	1.20
	(0.00374)	(0.000801)	(0.0000358)	(0.853)	(0.00680)	(0.000205)	(0.00000311)	(0.199)

Table S7: Summary of the simulation results of the two stage least squares (2SLS) estimator of the casual exposure effect conditional on C, $\beta^{2SLS}_{X,C}$, for a linear $X - Z$ association and a null causal exposure effect: Median (interquartile range) of the partial $R^2_{X|Z}$ of the $X - Z$ association, bias of $\beta^{2SLS}_{X,C}$, mean of the estimated standard errors (SE) of $\beta^{2SLS}_{X,C}$, and coverage percentage of the 95% confidence interval for $\beta^{2SLS}_{X,C}$. Monte Carlo errors in brackets, except for partial $R^2_{X|Z}$.

Selection is/	Moderate Instrument				Strong Instrument			
depends on	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage
Completely at random	0.0508	0.000937	0.0394	95.1	0.438	0.000274	0.0103	94.8
	(0.00390)	(0.000714)	(0.0000304)	(0.393)	(0.00678)	(0.000186)	(0.00000274)	(0.405)
Z	0.0431	0.0000655	0.0432	95.2	0.396	0.0000484	0.0113	95.2
	(0.00356)	(0.000791)	(0.0000357)	(0.392)	(0.00679)	(0.000207)	(0.00000315)	(0.390)
Z + C	0.0468	0.000468	0.0413	95.5	0.417	0.000156	0.0108	95.4
	(0.00376)	(0.000745)	(0.0000333)	(0.378)	(0.00678)	(0.000195)	(0.00000296)	(0.382)
X	0.0437	−0.0208	0.0462	93.3	0.399	−0.0215	0.0117	54.6
	(0.00368)	(0.000846)	(0.0000395)	(0.456)	(0.00681)	(0.000213)	(0.00000326)	(0.909)
X + C	0.0474	−0.00859	0.0425	94.8	0.416	−0.0118	0.0111	82.7
	(0.00382)	(0.000778)	(0.0000347)	(0.407)	(0.00677)	(0.000202)	(0.00000307)	(0.691)
X + Z	0.0209	−0.0743	0.0662	80.3	0.377	−0.0200	0.0120	60.7
	(0.00259)	(0.00120)	(0.0000829)	(0.727)	(0.00693)	(0.000221)	(0.00000347)	(0.892)
Y	0.0510	0.00101	0.0366	95.6	0.440	0.000235	0.00961	95.0
	(0.00388)	(0.000659)	(0.0000279)	(0.376)	(0.00675)	(0.000174)	(0.00000256)	(0.397)
X + Y	0.0479	−0.0687	0.0411	60.9	0.420	−0.0537	0.0108	0.267
	(0.00384)	(0.000760)	(0.0000332)	(0.891)	(0.00677)	(0.000199)	(0.00000297)	(0.0942)

Table S8: Summary of the simulation results of the two stage least squares (2SLS) estimator of the casual exposure effect conditional on C, $\beta^{2SLS}_{X,C}$, for a nonlinear $X - Z$ association and a causal exposure effect of 1: Median (interquartile range) of the partial $R^2_{X|Z}$ of the $X - Z$ association, bias of $\beta^{2SLS}_{X,C}$, mean of the estimated standard errors (SE) of $\beta^{2SLS}_{X,C}$, and coverage percentage of the 95% confidence interval for $\beta^{2SLS}_{X,C}$. Monte Carlo errors in brackets, except for partial $R^2_{X|Z}$.

Selection is/	Moderate Instrument				Strong Instrument			
depends on	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage
Completely at random	0.0505	0.000300	0.0203	95.0	0.443	0.000119	0.00593	95.0
	(0.00691)	(0.000367)	(0.0000377)	(0.399)	(0.00914)	(0.000108)	(0.00000313)	(0.397)
Z	0.0357	−0.000303	0.0255	94.9	0.401	0.00000144	0.00669	94.8
	(0.00591)	(0.000468)	(0.0000572)	(0.402)	(0.00873)	(0.000123)	(0.00000370)	(0.405)
Z + C	0.0407	0.0000338	0.0234	95.1	0.420	0.0000624	0.00634	95.0
	(0.00615)	(0.000424)	(0.0000478)	(0.394)	(0.00887)	(0.000116)	(0.00000338)	(0.397)
X	0.0365	0.000362	0.0263	95.5	0.394	−0.0263	0.00704	3.23
	(0.00605)	(0.000476)	(0.0000590)	(0.377)	(0.00882)	(0.000130)	(0.00000395)	(0.323)
X + C	0.0368	−0.00329	0.0260	95.2	0.403	−0.0195	0.00685	18.9
	(0.00603)	(0.000467)	(0.0000574)	(0.390)	(0.00888)	(0.000127)	(0.00000379)	(0.715)
X + Z	0.0218	−0.0651	0.0358	55.2	0.379	−0.0300	0.00723	1.23
	(0.00493)	(0.000680)	(0.000116)	(0.908)	(0.00869)	(0.000133)	(0.00000418)	(0.202)
Y	0.0381	−0.0122	0.0248	92.2	0.416	−0.0289	0.00659	0.667
	(0.00608)	(0.000458)	(0.0000534)	(0.491)	(0.00899)	(0.000121)	(0.00000355)	(0.149)
X + Y	0.0371	−0.00649	0.0256	94.3	0.407	−0.0291	0.00678	1.10
	(0.00602)	(0.000471)	(0.0000561)	(0.424)	(0.00891)	(0.000126)	(0.00000372)	(0.190)

Table S9: Summary of the simulation results of the two stage least squares (2SLS) estimator of the casual exposure effect conditional on C, $\beta^{2SLS}_{X,C}$, for a linear $X - Z$ association and a null causal exposure effect: Median (interquartile range) of the partial $R^2_{X|Z}$ of the $X - Z$ association, bias of $\beta^{2SLS}_{X,C}$, mean of the estimated standard errors (SE) of $\beta^{2SLS}_{X,C}$, and coverage percentage of the 95% confidence interval for $\beta^{2SLS}_{X,C}$. Monte Carlo errors in brackets, except for partial $R^2_{X|Z}$.

Selection is/	Moderate Instrument				Strong Instrument			
depends on	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage	Partial $R^2_{X\|Z}$	Bias	Mean SE	Coverage
Completely at random	0.0512	0.000299	0.0202	95.0	0.443	0.000119	0.00593	95.0
	(0.00694)	(0.000365)	(0.0000372)	(0.399)	(0.00914)	(0.000108)	(0.00000313)	(0.397)
Z	0.0363	−0.000298	0.0253	94.9	0.401	0.00000144	0.00669	94.8
	(0.00594)	(0.000464)	(0.0000562)	(0.403)	(0.00873)	(0.000123)	(0.00000370)	(0.405)
Z + C	0.0413	0.0000353	0.0232	95.1	0.420	0.0000624	0.00634	95.0
	(0.00618)	(0.000421)	(0.0000470)	(0.394)	(0.00887)	(0.000116)	(0.00000338)	(0.397)
X	0.0370	0.000546	0.0262	95.5	0.393	−0.0268	0.00705	3.03
	(0.00607)	(0.000473)	(0.0000581)	(0.378)	(0.00881)	(0.000130)	(0.00000396)	(0.313)
X + C	0.0373	−0.00344	0.0258	95.0	0.404	−0.0189	0.00684	20.5
	(0.00606)	(0.000463)	(0.0000565)	(0.397)	(0.00887)	(0.000127)	(0.00000377)	(0.737)
X + Z	0.0223	−0.0649	0.0355	54.7	0.380	−0.0292	0.00720	1.60
	(0.00496)	(0.000672)	(0.000113)	(0.909)	(0.00871)	(0.000133)	(0.00000415)	(0.229)
Y	0.0515	0.000532	0.0191	95.6	0.447	0.000113	0.00563	94.7
	(0.00674)	(0.000348)	(0.0000341)	(0.374)	(0.00901)	(0.000103)	(0.00000291)	(0.409)
X + Y	0.0386	−0.0125	0.0247	92.0	0.416	−0.0289	0.00659	0.667
	(0.00611)	(0.000454)	(0.0000526)	(0.494)	(0.00899)	(0.000121)	(0.00000355)	(0.149)

□