Effect of years of schooling on healthcare costs: A Mendelian randomisation study

Carol Lorena Perez Romay

Supervisors:
Geir Godager and Neil M. Davies

Thesis submitted as a part of the Master of Philosophy Degree
in Health Economics, Policy and Management

University of Oslo
The Faculty of Medicine
Department of Health Management and Health Economics

November 2021

2021

Effect of years of schooling on healthcare costs: A Mendelian randomisation study

http://www.duo.uio.no/

Print: University Print Centre, University of Oslo

Acknowledgments

There are several individuals who have aided me in the process of writing this thesis and I very much appreciate all their input and assistance. First, a huge thanks to co-supervisor Neil M. Davies at the Integrative Epidemiology Unit at the University of Bristol for introducing me to Mendelian randomisation and trusting me with this challenge. Thanks also to Sean Harrison at the University of Bristol for providing data. Next, I would like to give a special thanks to my main supervisor Geir Godager at the Institute of Health and Society at the University of Oslo for his guidance, useful comments and especially for asking difficult questions. I am equally grateful to Jonas Minet Kinge at the University of Oslo and the Norwegian Institute of Public Health who made time to help me, in suggesting literature, contributing to discussions, and offering constructive feedback. Furthermore, I would like to thank my employer at the Norwegian Directorate of Health for being supportive and accommodating, and especially current and past department directors Johnny Advocaat-Vedvik and Anne-Grete Skjellanger. I am also grateful to Eva van Loenen for patiently reading my text, her constant optimism, and for helping me structure my schedule for completion. Last but not least, I would like to thank friends and family for their help, words of encouragement, and for caring about my long-lasting academic endeavours. Specially to Adrian, for all the joy and motivation.

Abstract

Cost evaluation and cost control are the main concerns of any healthcare systems since national economies, regardless of their economic power, face constrained healthcare budgets while the demand for healthcare and health costs are ever-growing. Consequently, we need to understand how healthcare costs behave as a function of patients' and health systems' characteristics. However, although the positive association between education and health has been well documented, the causal relationship between years of schooling and healthcare costs is yet to be established. The aim of this thesis is to estimate, if any, this causal effect.

To handle potential confounding and reverse causation problems in the education-healthcare costs relationship, this thesis applies a Two-sample Mendelian Randomisation (MR) approach on summary data from previous genome-wide association studies based on samples of European and white British ancestry. Given that the IV/MR-conditions discussed through this paper are fulfilled, this method might provide the causal effects of years of schooling on healthcare costs.

This thesis finds that an extra year of completed schooling would reduce total healthcare costs by approximately £61 in 2019-GBP (£219 per std in years of schooling (3.6 years); CI £267 - £168). This main result remains comparable across methods meant to uncover potential violations of the MR-assumptions. However, only using summary data from previous studies pose some limitations to the analysis. Thus, these results should be treated with care. In addition, I conclude that, contrary to earlier concerns, a Two-sample MR with summary data can be used to perform exploratory analyses of economic outcomes. Genetic variants should not be dismissed as instruments beforehand on the argument that they won't induce enough variation to the exposure, However, methodologically solid research results would require using individual data, ideally containing information on siblings and/or family trios.

There is a need to understand the biological pathways through which genetic instruments influence educational attainment. Without this knowledge it is difficult to make convincing arguments supporting the credibility of the exogenous condition, which cannot be tested otherwise. Finally, it seems of vital importance to rigorously establish which specific causal effects a MR-analysis identifies.

Table of contents

Introduction

The analysis of healthcare costs has attracted the attention of epidemiological researchers as well economist and other experimental researchers since cost evaluation and control are a main concern for healthcare systems (e.g. Gregori et al., 2011; de Meijer et al., 2011). For this reason, knowledge about causal factors of healthcare costs is needed to understand how healthcare costs behave as a function of patients' and health systems' characteristics. The aim of this thesis is to estimate the causal effects of education attainment, defined as completed years of schooling, on total healthcare costs including costs[1] in both, primary and secondary healthcare. The nature of the data at hand (described in the method section), limits the analysis to focus on healthcare costs funded by governmental schemes. The discussion will include whether our results could be generalised to overall[2] societal healthcare costs i.e., costs including governmental, and privately funded healthcare costs.

The causal relationship between completed years of schooling and healthcare costs has thus far been unclear. There is a lack of theoretical frameworks and empirical results to refer to when, for instance, considering education as strategy to reduce future healthcare costs from a policy maker's point of view. Yet, knowledge of the effects of years of schooling on healthcare costs could be relevant, for instance to national economies searching for strategies to contain ever-increasing healthcare expenditure. Estimates of the effect of years of schooling on healthcare costs, if any, would also be relevant information to cost-benefit analyses regarding interventions aimed to provide health education. For example, when evaluating the cost-benefits of an increase in years of schooling vs carrying out information campaigns targeting the general population.

Review of Relevant Literature

From a macro-perspective, previous studies of the determinants of healthcare costs in OECD countries include factors such as the share of elderly population, proximity to death of patients, technological progress, territorial decentralization, and the remuneration arrangements of healthcare suppliers (Martín et al., 2011). However, to my knowledge, educational attainment/completed years of schooling has received little attention as a determinant of

[1] In this thesis the terms cost and expenditure are used interchangeable.

[2] Overall healthcare costs refer to healthcare cost funded by governmental schemes and private funds. Total healthcare costs refer to costs including primary and secondary healthcare.

healthcare costs (i.e. there are few analyses of the effect of educational attainment on healthcare costs beyond a health effect).

At the individual level, it is well documented that there is a correlation between educational attainments i.e. years of schooling and health (e.g. Khaing et al., 2017; Xu et. al., 2015). Furthermore, there are studies on healthcare costs at the individual's level, indicating that healthcare costs are associated with several behavioural and social factors besides the individuals' health (e.g. Sterling et al., 2018). When reviewing the literature, I find that healthcare services' utilisation and health literacy emerge as potential linking paths between educational attainment and healthcare cost.

Health. A causal relation between years of schooling and health would lead us to expect, other things being equal, lower healthcare costs amongst those with more years of schooling. However, the empirical evidence of the causal relationship between education and health is still unclear. Galama et al. (2018) review evidence from experimental and quasi-experimental studies, mostly from developed countries, regarding the effect of education (years of education) on mortality and two of the most well-known risk factors for mortality and disease: obesity and smoking. The authors define mortality, obesity and smoking as expressions of people's health and conclude that the causal effect of education on health is yet to be well stablished. Specifically, they find that education's effect on mortality and smoking is context sensitive. The effect on mortality seems to depend on gender, labour market returns to education, quality of education; and whether the educational intervention affects the quality of individuals' peers. Smoking seems to be unaffected by a simple increase of schooling time, but rather to be influenced by schooling reforms that affect the individual's track or the peer group. Furthermore, Galama et al. (2018) conclude that there is no convincing evidence of a causal effect of education on obesity.

Healthcare services utilisation. Years of schooling may affect healthcare costs if there is a differential use of healthcare services depending on years of schooling. Many studies that have documented differences in healthcare services utilisation related to socioeconomic status (SES) (e.g. Jansen et al., 2018), which is closely related to education. However, I focus on literature regarding education or studies that measure SES in terms of education. The main impression after revising this literature is that observable differences in utilisation are related to the type of services being studied.

On the one hand, Fletcher and Frisvold (2009) conclude that high school graduates who later attended college, were more likely to use preventive care services such as physical examinations, dental examinations, flu shots, and cholesterol tests compared to those who did not attend college. In their study, Fletcher and Frisvold (2009) use data from a cohort of high school graduates who were followed up for nearly 50 years, from Wisconsin, USA. Furthermore, the authors point out that there is a consistent conclusion in public health and economics literature that the use of preventive healthcare for adults is correlated with education. It is difficult to assess the generalisability of Fletcher and Frisvold's (2009) results as their research suggest that these results are due to greater access to healthcare in an American context and they might not apply to, for instance, countries with universal healthcare coverage such as UK or Scandinavian countries.

Similarly, Jansen et al. (2018) find a relation between educational attainment and the likelihood of using out-of-hours primary care (OOHPC) among people with similar health status (chronic illnesses) in the Netherlands. Their results indicate that people with low education levels (no education, primary school only or lower vocational education) were more likely to use out-of-hours primary care services than people with both intermediate (intermediate or advanced general education, intermediate vocational education) and high levels (higher professional education, university) of education. Additional evidence suggesting different use of healthcare services after accounting for health status in the Netherlands, was found by Droomers and Wester (2004).

On the other hand, Dahl et al. (2014) report that there was no evidence suggesting differences in use of GPs associated with education attainment in Norway. Compatible results were found by Glazier et al. (2009) who use data from the Canadian Community Health Survey (CCHS). After adjusting for morbidity, income, urban-rural location, age and sex, Glazier et al. do not find differences in contact (measured as at least 1 visit) with GP associated with levels of education (low: not completed high school, medium: high school completion and some postsecondary education, and high: university degree). However, Glazier et al. results suggest that higher education is associated with less-frequent (fewer than 10) GP visits. Moreover based on a review of empirical results published between 2000 and 2013, Godager and Iversen (2013), and Glazier et al. (2009) observe evidence of differences in use of specialised care. Godager and Iversen's results suggest that people with higher education and higher incomes tend to utilise private specialist services outside the governmental schemes rather than visit publicly funded primary care services. Glazier et al. (2009) on their side, conclude that higher educated

people had greater contact, and higher visit frequency, with specialists. Their results also indicate that patients with higher education were more likely than those with lower education to bypass primary care to access specialists. Additionally, results of Fiva et al. (2014) suggest that education levels may also be associated with the type of treatment that individuals receive. In their study Fiva et al. (2014) investigate utilisation of highly specialised treatment and how utilisation patterns affected survival after cancer. The authors conclude that patients with higher education were more likely than other patients to be transferred to two central, highly specialized hospitals outside of their zone of residence. Furthermore, they find that differences in the probability of being transferred to the specialized hospital increased after conditioning on disease characteristics.

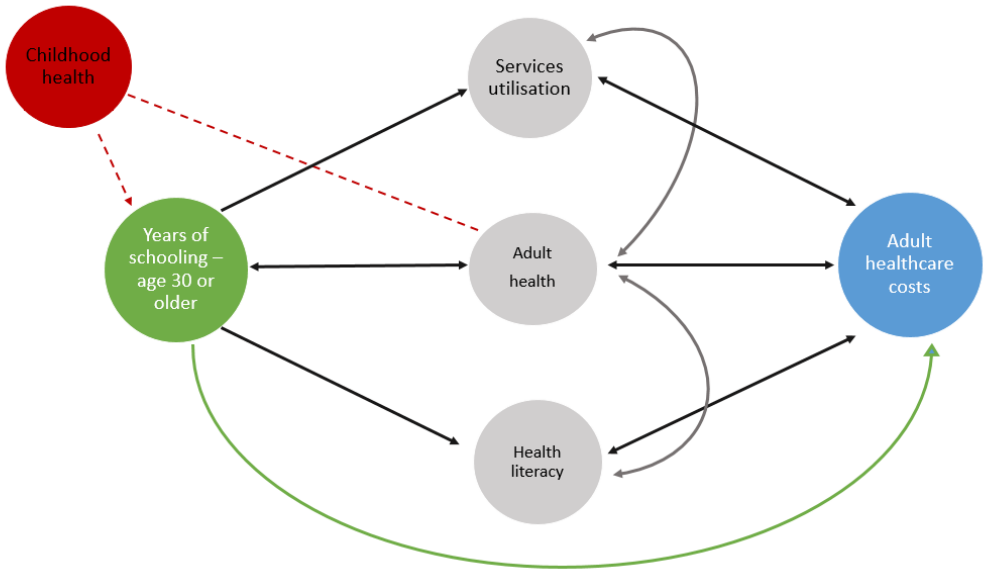
Health literacy. Health literacy has gained interest as a determinant of healthcare costs (Palumbo, 2017). At the same time, health literacy has been associated with education attainment (Jansen et al.2018). Palumbo (2017) review the literature based on USA data and concluded that limited health literacy is an important and independent predictor of, for instance patients' disengagement, health problems' exacerbation, and misuse of available health resources, which in turn will increased healthcare costs.

Causal effect of years of schooling on healthcare costs

The literature at hand may lead us to speculate that higher educational attainment will cause better health in some contexts and is associated with improved health literacy. Better health and improved health literacy would in turn mean lower healthcare costs. Furthermore, there is some reason to believe that are patterns of healthcare services utilisation depending on educational attainment. However, it is not straightforward to hypothesize the direction on which these patterns would affect costs. It is for instance not clear what differences in patterns of utilisation for different type of services would mean for costs; while highly specialised treatment could be assumed to be more costly, greater use of preventive care may lead to a reduction of healthcare costs later. Additionally, it is difficult to predict how the conclusion of Godager and Iversen's (2013), regarding the tendency of higher educated people to utilise private specialist services rather than publicly funded primary care services, would affect total healthcare costs. While it could mean a reduction in the publicly funded costs, it may lead to an increased in the total healthcare costs.

Figure 1 illustrates three examples of mechanisms by which education i.e. years of schooling may affect healthcare costs. The figure also illustrates a direct effect of years of schooling on healthcare costs.

Figure 1 Graph representing three potential pathways in which years of schooling and adult healthcare cost may be related. The graph also illustrates that childhood health is a potential confounder of this relationship. Double arrows show potential sources of reverse causation problems. The green line illustrates a direct effect of years of schooling on healthcare costs.



The literature is spare and prompts a need for more knowledge on the causal effect of years of schooling on costs. This thesis aims to estimate causal effects of years of schooling.

Methods

The estimation of the effect of years of schooling on healthcare costs is challenging, as this relationship is complex with numerous potential confounding and reverse causation problems. Examples of potential confounders (i.e. variables that both determine costs and are correlated to one or more of the included exposure variables (Stock & Watson, 2007) are cognitive ability, the patient's socioeconomic and demographic factors, such as sex, parents' economic status, place of birth, parents' education and structural conditions, for instance, public funding of health and education. Childhood health may also affect completed years of schooling and is strong correlated with adult health (Case et al., 2005).

The analysis of the effect of years of schooling on healthcare costs is further complicated by reverse causation (i.e. a situation where a change in the outcome results in a change on the

exposure of interest) because healthcare costs in adulthood affect for instance adult health (Figure 1). Adult health is highly correlated with the confounder childhood health and educational attainment (Case et al. 2005). As long as we cannot adjust for the individual's childhood health, it will not be possible to separate the effect of childhood health on earlier years of schooling (during childhood) from the effect of earlier years of schooling on adult health and thereby on adult healthcare costs. Hence, a link between adult healthcare costs and earlier years of schooling will be maintained and the reverse causation problem will persist. Additionally, the analysis of this thesis is based on data from adult individuals, and years of schooling was measured at age 30 or older. Some of those individuals might still be under education at this point. In that case, there would be a direct link between adult healthcare costs, adult health, and years of schooling at age 30 or older. Confounders and reverse causation make it difficult to disentangle the effect of the exposure (years of schooling) on the outcome (healthcare costs).

The following section makes a formal presentation of challenges that confounding factors and reverse causation pose to the analysis of the effect of the completed years of schooling on healthcare costs, and present the Mendelian Randomisation method, which aims to estimate causal relationships by avoiding bias originated by confounders and reverse causation (Lawlor et al., 2008).

Confounding and reverse causation problems

In the absence of confounding factors and reverse causation, and assuming linear relationship between the exposure and outcome, the relationship between completed years of schooling and healthcare costs can be modelled as in equation (1):

$$Y = \beta_{om} + \beta_1 X + \boldsymbol{\gamma M} + \varepsilon_m \qquad (1)$$

Y represents estimated yearly healthcare costs, β_{om} is the regression constant term, β_1X is completed years of schooling (X) times the coefficient of interest i.e. the effect of one additional year of schooling on healthcare costs (β_1). $\boldsymbol{\gamma}\mathbf{M}$ is the product of a row vector $\boldsymbol{\gamma}$ containing p coefficients corresponding to a column vector $\mathbf{M}'$ of p covariates (see appendix X for a detailed description of the covariates) and ε_m is the error term of the equation. The values of β_{om} and β_1 define the intercept and the slope of a line that describes the relationship between years of schooling and healthcare costs.

The values of βo , β_1 and $\boldsymbol{\gamma}$ can be estimated by the OLS (Ordinary Least Squares) estimators (Stock & Watson, 2007) due to their superior practical and theoretical properties. Provided that the following least squares assumptions are met the OLS estimators will be unbiased, consistent, and asymptotically normal (ibid):

A1. The error term for individual i (ε_{mi}) has conditional mean zero given the explanatory/exposure variable: $E(\varepsilon_{mi}\,|X_{i,}\mathbf{M}_i) = 0$,which implies that $corr(X_i, \varepsilon_{mi}) = 0$.

A2. The outcome, exposure, and covariate values of $(X_i, \mathbf{M_i}, Y_i)$are independent and identically distributed drawn from their joint distribution.

A3. Large outliers are unlikely: $X_i, Y_i \ and \ \mathbf{M_i}$ have nonzero finite fourth moments.

A4. There is not perfect multicollinearity i.e. no one of the regressors $(X_{i,}\mathrm{M}_{pi})$ is a perfect linear function of the other regressors.

We assume that conditions A2, A3 and A4 above are satisfied. The presence of confounders/omitted variables violates the first assumption of the OLS estimators because it correlates the exposure with the error term (Stock & Watson, 2007) so that $corr(X_i, \varepsilon_{mi}) \neq 0$ creating what is known as omitted variable/confounding bias. In practical terms, this means that a confounder, for instance cognitive ability, would lead to an under- or overestimation of the association between years of completed schooling on healthcare costs (Alexander et al., 2015).

One strategy to avoid confounding bias in observational studies using regression analysis is to include potential covariates in the regression equation (1) (Stock & Watson, 2007). In our analysis, one could consider including cognitive ability, the patient's socioeconomic/demographic factors and childhood health to avoid confounding bias, however it is resource-intensive to gather the amount of information needed to adjust for these confounders. Additionally, one cannot guaranty that the results are unaffected by unmeasured confounding factor and results should therefore not be interpreted as causal effects (Hernán & Robins, 2006).

In the case that one aimed to estimate the direct effect (green line in fig. 1) and had all required data at hand; adjusting for confounders could not solve the reverse causation problem created by adult health. This is because direct effects are estimated by conditioning for known mediators of the relationship. In our case, conditioning on adult health would be problematic as health is as a collider (i.e. a variable that is itself affected by both the exposure variable and the

outcome variable (Elwert & Winship, 2014) and conditioning on it would bias estimates of the direct effect. This bias is known as endogenous selection bias (Elwert & Winship, 2014) or simultaneous causality bias (Stock & Watson, 2007). Simultaneous causality distorts the OLS estimate in a similar way as an omitted variable, since the bias arises because the exposure of interest and covariates are correlated to the error term, violating the first OLS assumption i.e. $E(\varepsilon_{mi} | X_{i,}\mathrm{M}_i) = 0$ (Stock & Watson, 2007).

Mendelian Randomisation Method

Mendelian randomization (MR) is an epidemiological method that uses genetic variants as instruments for estimating causal relationships between a modifiable (non-genetic) exposure and an outcome by avoiding bias originated by confounding and reverse causation (Lawlor et al., 2008). Mendelian randomization (MR) is recognizable to those acquainted with instrumental IV analysis commonly used in econometrics. However, when commenting an alternative naming of the method such as Genetic Instrumental Variable Analysis, proponents of MR (Davey Smith et al., 2020; Lawlor et al., 2008) argue that, even though the name is secondary, it is important to recognise that making causal inferences based on the MR-method demands expertise beyond the conventional understanding of instrumental variable analysis. This thesis is mostly based on MR-literature and terminology, however, some references to algebraic expressions commonly used in econometrics are used whenever it seems helpful to do so.

Informally, the MR- method relies on finding a variable, the instrument, which is strongly associated with the exposure of interest (completed years of schooling) whilst the instrument can be assumed to be assigned to each individual almost at random and will therefore not be associated with confounders. Additionally, the instrument is assumed to only affect the outcome through the exposure (See conditions A5 and A6 below and appendix 3 for a detailed presentation of IV/MR conditions).

As mentioned above, the need for instrumental variable analysis arises due to, among other reasons, unmeasured confounders, or reverse causation. I limit the presentation of the estimation problem to the context of a single-equation linear model with an omitted variable problem.

The relationship of interest is as in equation 1. This means that β_1 should not be estimated by the OLS method as the resulting estimator $\widehat{\beta_{1\mathrm{OLS}}}$ would be bias due to the non-

zero correlation between the exposure and the error term of the model violating assumption A1 i.e. $E(\varepsilon_{mi}\,|X_{i,}\mathbf{M}_i)\ \neq 0$.

To estimate the MR or any IV approach it is necessary to have at least one observable variable (Z_k) for each endogenous variable (i.e. the exposure variable correlated with the error term) that, apart from not being included in equation (1), satisfies the following two conditions:

A5. Instrument relevance: This condition is often presented as $corr(Z_{ki}, X_i) \neq 0$. A more precise definition (Wooldridge, 2010) of this condition is that the coefficient of a regression/projection of X on all exogenous covariates and Z_k, is not zero i.e. $\phi \neq 0$ in eq. 2 below.

A6. Instrument exogeneity: Z_k is uncorrelated with the error term ε_{mi} i.e. $corr(Z_{ki}\,, \varepsilon_{mi}) = 0$. This mathematical expression comprises the independence and the exclusion assumption (Angrist & Pischke; 2009) presented in appendix 3.

Additional assumptions of the IV regression model are (Stock & Watson,2007):

A7. We assume that X are the only endogenous variable, while the covariates $\boldsymbol{M}$ are uncorrelated with the error term i.e. $E(\varepsilon_{mi}\,|\mathbf{M}_i) = (\varepsilon_{mi}\,|\mathrm{M}_{1i}, \dots, \mathrm{M}_{pi}) = 0$.

A8. The values $(X_i, \mathrm{M}_{1i}, \dots, \mathrm{M}_{pi}, Z_{ki}\,, Y_i)$ are independent and identically distributed (i.i.d.) drawn from their joint distribution.

A9. Large outliers are unlikely: $(X_i, \mathrm{M}_{1i}, \dots, \mathrm{M}_{pi}, Z_{ki}, Y_i)$ have nonzero finite fourth moments.

IV models also make assumptions about perfect multicollinearity of regressors (Stock & Watson, 2007), however, in the case where there is only one endogenous variable (X in our case), the condition regarding perfect multicollinearity holds whenever at least one instrument Z_k enters the regression X on all exogenous covariates and Z_k (eq. (2) below).

To obtain a point estimate of the causal effect of X on Y, it is necessary to make an additional assumption (Labrecque & Swanson, 2018)

A10. Monotonicity condition requires that the instrument affect all individuals in the same direction, in our case it implies that the allele of reference either increases or decreases years of schooling for all individuals.

In this thesis, a genetic variant (SNP_k) is an instrument for the endogenous variable years of schooling. The value of the SNP is assigned randomly to each individual from his or her

parents at conception. The SNP (instrument) of interest can, for diploids cells as the human DNA, take the values 0, 1 or 2. These numbers represents the number of alleles of reference (called the effect allele) that each individual possesses. An individual's number of effect alleles can be thought of as a randomly assigned allocation number to different groups, where each group receives a different mean level of the exposure during their life course (Bowden et al., 2016).

Equation (2) below represents the relationship of years of schooling with the instrument Z (SNPk). This equation is called first stage, a term borrowed from simultaneous equation analysis, and means that an endogenous variable is expressed only in terms of exogenous variables (Angrist & Pischke, 2015). This equation is also known as reduced-form equation (Wooldridge, 2010).

$$X = \phi_{om} + \phi_k Z_k + \boldsymbol{\psi M^x} + \varrho_m \qquad (2)$$

ψM is the product of vectors of p covariates times their coefficients.

A reduced-form equation for healthcare costs (Y) can be obtained by replacing X in equation ((1) by the reduced form (2):

$$Y = \delta_{om} + \rho_k Z_k + \boldsymbol{\upsilon M^y} + \nu_m \qquad (3)$$

$\delta_{om} = \beta_{0m} + \beta_1\phi_o, \ = \beta_1\phi$, $\boldsymbol{\upsilon} = (\beta_1\psi_1 + \gamma_1, \ \beta_1\psi_2 + \gamma_2, \dots, \ \beta_1\psi_p + \gamma_{p,})$ and $\nu_m = \beta_1\varrho_m + \ \varepsilon_m$.

(A11). Reduced-form equation 2 and 3 are assumed to satisfy the OLS conditions stated above This means among other things that instruments are chosen so that Z_k is not correlated with unmeasured confounders i.e. $cov\,(Z_{ki}, \varrho_{mi}) = 0$ and $cov\,(Z_{ki}, \upsilon_{mi}) = 0$, i.e. the independence condition is satisfied. Additionally, ϱ_{mi} and υ_{mi} are independent of each other (A12).

The causal effect of interest i.e. effect of completed years of schooling on healthcare costs is determined by the ratio of the population reduced form regression of Y on Z_k (eq. (3)) to the population reduced form regression of X on Z_k (eq. (2)) (Angrist & Pischke, 2009; Wooldridge, 2010)

$$\beta_1 = \frac{\rho}{\phi} \qquad (4)$$

Angrist & Pischke (2009) explain that a consistent estimate of β_1 can therefore be constructed by the ratio of the OLS estimators of ϕ and ρ provided that conditions A11 and A12 hold i.e. regressors (instrument and covariates) in the reduced-form equations are uncorrelated with their respective errors. Furthermore, this insight yields also when covariates are included in the reduced forms. In MR studies, the IV estimator is mostly referred to as Wald or Ratio estimator (Davies et al. 2018, Burgess et al. 2017).

Two-sample MR

There are different approaches to MR (e.g. Davey Smith & Hemani, 2014; Burgess et al. 2016). This thesis utilises the Two-sample Mendelian randomisation (Two-sample MR). The distinctive feature of the Two-sample MR is that it allows to estimate the causal effect of the exposure on the outcome by using two independent datasets from the same population. The theoretical insights that make this possible were laid out in an influential article by Angrist and Krueger (1992) where they developed a Two-sample instrumental variables (TSIV) estimator which has been applied and further developed (e.g. Inoue and Solon, 2010; Dee & Evans, 2003; Angrist & Krueger, 1995). The method relies on two important assumptions (Angrist & Pischke, 2009; Angrist & Krueger; 1992):

A13. The two datasets are drawn from the same population.

A14. There is not overlap of sample observations.

Epidemiologist have built on this insight and developed Two-sample Mendelian Randomisation (e.g. Pierce & Burgess,2013; Burgess et al.,2015). Hemani et al. (2018) explains that this method makes it possible to take advantage of the vast wealth of known genetic associations stemming from the rapid accumulation of genome-wide association studies (GWAS) during the 2010s even in the cases where the pair of traits of interest (the outcome and exposure) have not been recorded in the same sample.

An additional major advantage of the Two-sample MR is that it allows to carry out an MR analysis without having access to individual-level data, i.e. only by using Genome-wide Association Study (GWAS) summary data (Hemani et al., 2018). GWAS is an observational approach used in genetics research. The method involves scanning the genomes from big samples of people and looking for associations between specific genetic variations/genetic markers with particular traits (National Human Genome Research Institute, 2019). GWAS summary includes estimates of association (regression coefficients) between SNPs and the trait

of interest as well as standard errors of the regression coefficients. Hemani et al. (2018) point out that using summary data as the "raw data" for the analysis is a major advantage because summary statistics from GWAS are non-disclosive and are often made freely available to the public. In addition to that, the authors point out that two-samples MR adds power to the analysis as it combines the samples of two or more GWAS.

In practical terms, the estimation process is as follows: suppose that one instrument, a SNP in our case (SNP_k), is measured in two separate Genome-wide Association Studies (GWAS). Study 1 estimates the association of SNP_k with the outcome (healthcare costs), while study 2 estimates the SNPk – exposure (years of schooling) association (Bowden & Holmes, 2019). In the case that one specific SNPk from study 2 is not found in study 1, a proxy can be used (Burgess et al. 2015). The Two-sample MR estimate of β_1 using instrument Z_k (ratio estimator of the causal effect) is constructed as the ratio of the OLS- regression coefficient of the outcome-SNP_k from study 1to the OLS-regression coefficient of exposure-SNP_k from study 2 (Burgess et al. 2015; Lawlor et al. 2008). Study subscripts will be suppressed in the following for clarity.

$$\hat{\beta}_{1k} = \frac{\widehat{\rho_k}}{\widehat{\phi_k}} \qquad (5)$$

The within-ratio/sample variance is (Burgess et al., 2016):

$$Var\hat{\beta}_{1k} = var\frac{\hat{\rho}_k}{\hat{\phi}_k} = \frac{\sigma^2_{\rho k}}{\hat{\phi}^2_k} \qquad (6)$$

This sample variance it is approximate using the formula for Taylor expansions. It relays in the additional assumption that the instrument-exposure effect is estimated with negligible error ($\sigma^2_{\phi k} \approx 0$) so that $\hat{\phi} \approx \phi$ and can be therefore be treated as a constant (A15). This assumption is known as No Measurement error in the exposure (NOME) (Hemani et al., (2018); Bowden & Holmes, 2019). Furthermore, $\sigma^2_{\rho k}$ is the sample variance of $\widehat{\rho_k}$ and $\sigma^2_{\phi k}$ is the sample variance of $\widehat{\rho_\phi}$, and although commonly estimated in the data as the standard errors of $\widehat{\rho_k}$ and $\widehat{\phi_k}$, these variances are assumed to be known (A16) (Hemani et al., 2018).

Efficient combination of estimators

Using information from multiple ratio estimates may be useful to explore the possibility that MR/IV assumptions have been violated (Lawlor et al. 2008; Davey Smith & Hemani, 2014). This study utilises 71 instruments yielding 71 independent estimates of β_1 i.e. k= {1,

2,…71}.The 71 ratio estimates can be combined into an overall effect by the inverse-variance weighted (IVW) method, from the meta-analysis literature, when assuming that each ratio estimate provides independent evidence on the causal effect i.e. the SNPs are uncorrelated (A17) (Burgess & Thompson, 2017). The inverse-variance weighted estimator can be informally motivated as a weighted average of the 71 ratio estimates $\frac{\widehat{\rho_k}}{\widehat{\phi_k}}$, using weights w_k.

IVW estimator is formally defined as (Bowden & Holmes, 2019):

$$\hat{\beta}_{1IVW} = \frac{\sum_{k=1}^{K} w_k \frac{\widehat{\rho_k}}{\widehat{\phi_k}}}{\sum_{k=1}^{K} w_k} \qquad (7)$$

can accommodate either a fixed effect, an additive random effect, or a multiplicative random effects IVW model. In Mendelian randomisation studies, the fixed effect and the multiplicative are preferred (Burgess et al., 2020); both assume a single overall effect (A18) and yield the same estimate of the effect (Mawdsley et al. 2017). For comparison, the additive random effects model assumes that there is a *distribution* of true sample effects so that $\hat{\beta}_{IVW}$ is the estimate of the mean of this this distribution (Borestein et al., 2010)

Formula (8) shows the weights for the estimation of a single overall effect ($\hat{\beta}_{IVW}$) in a fixed effects model. In addition to assuming a single overall effect, the fixed effects model requires all instruments to be valid (A19) (Bowden et al. 2016). A consequence of this condition is that the fixed effects IVW model supposes that all variation between ratio estimates of β_1 ($\hat{\beta}_{1k} = \frac{\widehat{\rho_k}}{\widehat{\phi_k}}$) is exclusively due to within-ratio/sample variation i.e. there is no between-ratio variation (Borestein et al., 2010).

$$w_k FE = \frac{1}{var \frac{\hat{\rho}_k}{\hat{\phi}_k}} = \frac{\hat{\phi}_k^2}{\sigma_{\rho k}^2} \qquad (8)$$

With weights as in (8) the IVW fixed effects estimator takes the following form:

$$\hat{\beta}_{1IVW_{FE}} = \frac{\sum_{k=1}^{71} \hat{\phi}_k \hat{\rho}_k \sigma_{\rho k}^{-2}}{\sum_{k=1}^{71} \hat{\phi}_k^2 \sigma_{\rho k}^{-2}} \qquad (9)$$

The standard errors are (Mawdsley et al, 2017; Borestein et al., 2010):

$$se\left(\hat{\beta}_{1IVW\_FE}\right) = \sqrt{\frac{1}{\sum_{k=1}^{K} var\,\frac{\hat{\rho}_k}{\hat{\phi}_k}}} \qquad (10)$$

The practical estimation of the β_1 is carried out by fitting the following OLS weighted regression:

$$\hat{\rho}_k = \beta_{1ivw}\,\hat{\phi}_k + \epsilon_k;\ \ \epsilon_k \sim N\left(0, \sigma_{\rho k}^2\right) \qquad (11)$$

$$Weights = \frac{1}{\sigma_{\rho k}^2}$$

In this case, due to A19, the empirically estimated standard error of $\widehat{\rho_k}$ ($se^2(\widehat{\rho_k})$) is set equal to $\sigma_{\rho k}^2$.

Challenges to the IV assumptions in MR-studies.

The attractiveness of using genetic variants as instruments in Mendelian Randomisation stems from what is known as Mendel's second law of independent assortment. Simply put, Mendel's law declares that there is a random assortment of genes that are passed from parents to offspring during gamete formation and conception (Davey Smith & Ebrahim, 2003). Furthermore, genetic associations have the advantage that causation is unidirectional, from the genetic variation (SNP) to the trait of interest (Davey Smith & Hemani, 2014). Additionally, modern genotyping technologies and stringent quality controls, make it possible to measure genetic variants highly accurately (Pierce & WanderWeele, 2012). According to Davey Smith & Hemani (2014) the measurement of the effects of genetic variants, are also relatively free form measurement errors. However, using genetic variants as instruments comes with its own challenges. I will now describe known phenomena that call into question the assumptions when using genetic variants as instruments.

Threats to the relevance assumption

Weak instruments refer to the case when the instrument explains little variation of the exposure being investigated (Burgess & Thompson, 2011). In a two sample Mendelian randomization without individual data, it is not possible to test the relevance assumption (A5) i.e. the strength of the instrument by means of a traditional F- test. In these types of studies, researchers will typically only include independent SNPs with a significant association to the exposure at the customary genome-wide significant level of p< 5 x 10^-8 to avoid weak instruments (Zhao et al., 2020).

Bowden & Del Greco et al. (2016a) suggest, in the context of Two-sample MR with summary data, to assess the strength of each individual instrument by the F statistic and the average instrument strength by the mean value of this statistic. F and $\bar{F}$ are defined as follows:

$$F = \hat{\phi}_k^2/\sigma_{\phi k}^2 \qquad (12)$$

$$\bar{F} = \frac{\sum_{k=1}^{K}(\widehat{\phi}_k^2/\sigma_{\phi k}^2)}{K} \qquad (13)$$

Furthermore, Bowden & Holmes (2019) suggest computing a weighted average instrument ($\overline{F_w}$) using $\hat{\phi}_k^2/\sigma_{\rho k}^2$ as weights:

$$\overline{F_w} = \frac{\sum_{k=1}^{K} w_k(\widehat{\phi}_k^2/\sigma_{\phi k}^2)}{\sum_{k=1}^{K} w_k}, \qquad w_k = \frac{\hat{\phi}_k^2}{\sigma_{\rho k}^2} \qquad (14)$$

Winner's curse. Winner's curse is a type of selection/publication bias known in the MR-literature and refers to the risk that the estimated associations of the SNPs published in the GWAS-study are overestimated (Burgess et al., 2020), increasing the threat of using weak instruments in the MR-study (Lawlor, 2016). The "winner's curse" problem is especially latent when researchers estimate the SNP-exposure associations on the same dataset as the dataset where the instrumental SNPs were first discovered. The reason for this is the way leading SNPs are chosen in GWAS studies i.e. the SNP in a locus with the lowest p-value is chosen to be the leading SNP.

Threats to the independence assumption.

Lawlor et al. (2008) explains that the random assortment of genes that are passed from parents to offspring implies that the inheritance of one trait (genotype) is independent of the inheritance of other traits. This can be understood as if one trait had been randomized with respect to the other of the traits. This random assortment was originally described in Mendel's work on plants and makes the independence assumption stated above plausible.

Genetic instruments that satisfy the independence assumption will serve as unconfounded indicators of particular trait values. The independence of genetic variants is, however, not always the case and researchers applying MR should consider the presence of social confounders that may create an association between genetic variants and outcomes. Potential confounders that threat the independence assumption are dynastic effects, assortative mating and population structure (Brumpton et al., 2020).

Dynastic effects. Dynastic effects may arise if parents' phenotypes (due to genetic variants) affect their children's observable traits beyond genetic heritage (Davies & Howe et al., 2019; Kong et al.,2018). A relevant example for us can be the case where parents' healthcare costs affect the healthcare costs of their offspring because they live in an environment where it is common to overuse/underuse healthcare services. This effect will exacerbate the effect on costs due to the offspring's years of schooling and could result in a biased estimated effect.

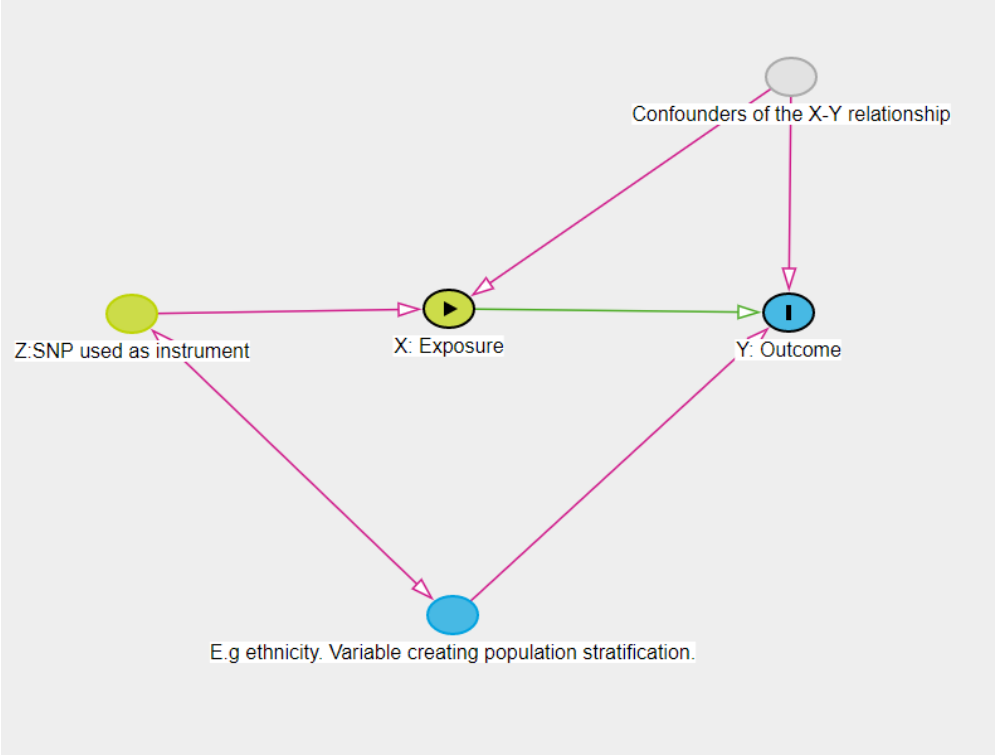
Assortative mating. is a social phenomenon consisting in people selecting partners in a non-random way, based on specific observable characteristics (phenotypes) (Brumpton et al., 2020; Hartwig et al., 2018). Hartwig et al. (2018) explain that assortative mating can become a problem if individuals with a particular genetic variant select mates with a phenotype that is genetically influenced, leading to genetic correlation between parents. The authors illustrate assortative mating by using examples of height and intelligence, both height and intelligence are known to be genetically influenced. First, tall women are more likely to select tall men, and second, women with higher intelligence select taller men. The first case is known as single-trait assortative mating. Under single-trait assortative mating, the mother's and father's exposures (or outcomes) will be correlated. This case will be unproblematic if the exposure and outcome under investigation are not genetically correlated (e.g. it will be unproblematic that taller mothers select e.g. taller fathers in a study investigating the effect of height on anything that is not genetically correlated to height).

However, as height and intelligence are highly hereditable and genetically correlated (Keller et al. 2013), the second example (i.e. women with higher intelligence select taller men) is problematic because it makes the exposure (e.g. height), the outcome (e.g. intelligence) and the genetic variants involved to be correlated, violating IV assumptions. Assortative mating as in this second example is known as is cross-trait assortative mating (Hartwig et al., 2018). Finally, Hartwig et al. (2018) argues that assertive mating depending on a third trait (other than exposure and outcome) can lead to bias if the trait under assortment is genetically correlated with exposure (e.g. height) and outcome (e.g. intelligence). For details, see Hartwig et al. (2018).

Population stratification. Population stratification refers to a situation where different population subgroups experience different outcome rates (i.e. healthcare costs in our case) and have different frequencies of the effect allele (Lawlor et al., 2008). Lawlor et al., (2008) present a graphic representation of population stratification (Figure 2) that can lead to spurious associations between the instrument (genotype) and the outcome. A hypothetical example that

could be relevant for this study is if the samples are collected from places that have systematic differences in the frequencies of the effect allele due to for example geographical distances. In that case, such differences will be problematic if there are also regional systematic differences in societal structures that affect healthcare costs, for instance difficulties in accessing healthcare services in remote regions.

Figure 2 exemplifies a situation where a variable (e.g. ethnicity) is correlated with the distribution on the instrumental SNP. This violates the IV/MR assumption if, as it is shown in the figure, there is a direct relationship between e.g. ethnicity and the outcome.



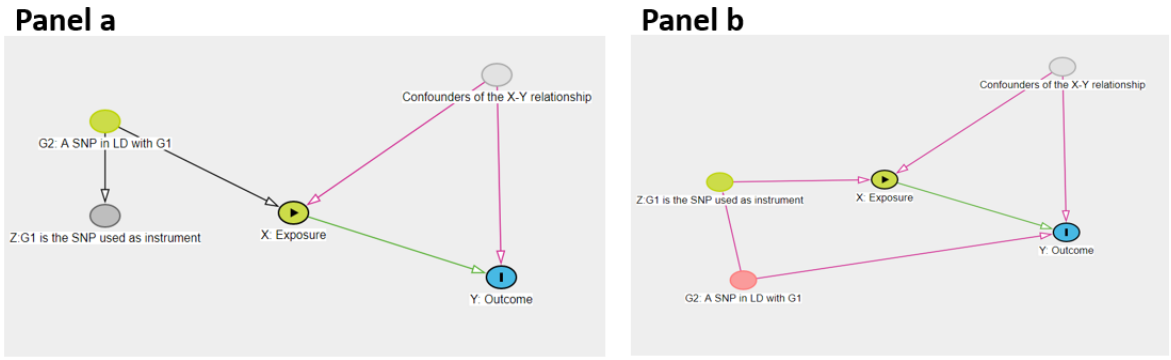
Source: Adapted from Lawlor et al. 2018

Linkage disequilibrium. Another threat to the independence assumption arises from the fact that independent assortment is not always the case (Lawlor et al., 2008). Lawlor et al. (2008) explain that work after Mendel's second law of independent assortment discovered "gene linkage" which means that not all genes are assorted independently. Furthermore, the authors explain, that it is now known that independent assortment is generally true for a specific type of chromosomes (non-homologous) but is not true for a set of genes located on a homologous chromosome, particularly if the genes are located close to each other.

The hypothetical situation in which all alleles exhibit complete independence is called linkage equilibrium (LE). A departure from this situation is referred to as linkage disequilibrium (LD) (Lawlor et al., 2008). Linkage disequilibrium means that there is correlation between

genetic variants, typically for variants physically close together on the same chromosome (Haycock et al., 2016). Lawlor et al. (2008) points out that LD is relevant to MR studies, because when an instrumental SNP is correlated (in LD) with a genetic variant, and the latter influences the outcome of interest, this may result in confounding. The authors also explain that not all correlation between genetic variants will challenge the IV assumptions and show it graphically. Figure 3 depicts examples of situations where LD is involved, as presented by the authors. Panel **a** presents a situation where LD is unproblematic. This is because the instrument (G1) is in LD with a variant (G2), which is related to the exposure, but (G2) does not have a direct effect on the outcome. However, on panel **b** LD violates IV/MR assumptions because G2 is in LD with the instrument (G1) as well as the outcome (Y).

Figure 3Examples of relationships between the instrumental SNP and other genetic variants. Panel a shows a situation where the LD between G1 and G2 do not violate the IV/MR assumptions. In panel b, on the contrary LD is a problem as there is a direct relationship from G2 to the outcome.



Source: Figure adapted from Lawlor et al. 2008

Threats to the exclusion assumption.

To consistently estimate the effect of the exposure on the outcome, the exclusion assumption must be satisfied (i.e. genetic variants used as instruments only affect the outcome via the exposure). However, it is likely that a single genetic variant influences multiple phenotypes. This phenomenon is called pleiotropy (Hemani et al., 2018; Davey Smith & Ebrahim, 2003).

There are two types of pleiotropy, vertical and horizontal pleiotropy. Vertical pleiotropy, which is also known as mediation, refers to the case were a genetic variant is associated with multiple traits (phenotypes) on the same biological pathway. This is less problematic as it does not violate the exclusion assumption (Hemani et al., 2018). Horizontal pleiotropy refers to a situation in which the instrument is associated with variables on different causal pathways to the outcome (Burgess et al., 2020).

Horizontal pleiotropy can also induce association between the instrument and a confounder. In that case, horizontal pleiotropy will violate the independence assumption as well as the exclusion assumption. In fact, linkage disequilibrium and population stratification as explained above will violate both the independence and the exclusion criteria (Hemani et al., 2018).

Horizontal pleiotropy robust methods

Horizontal pleiotropy is a major threat to MR-analysis as it contradicts the exclusion and independence assumptions, making the instruments (SNPs) invalid and thus leading to biased estimates. An algebraic expression of bias due to pleiotropy follows.

Consider the following model with one instrument and no covariates:

$$Y = Bo + \alpha_k Z_k + \beta_1 X + \Sigma \qquad (15)$$

The reduced forms are then:

$$X = \phi_{0k} + \phi_k Z_k + \varrho_k \qquad (16)$$

$$Y = \delta_{0k} + \rho_k^* Z_k + \nu_k \qquad (17)$$

Where $\delta_o = B_0 + \beta_1 \phi_{0k}$, $\rho_{\mathrm{k}}^* = \alpha_k + \beta_1 \phi_k$ and $\nu\_k = \beta\_1\ \varrho\_k + \Sigma$.

So that

$$\beta_1^* = \frac{\rho_k^*}{\phi_k} = \beta_1 + \frac{\alpha_k}{\phi_k} \qquad (18)$$

And the IVW will tent towards (Bowden et al., 2015):

$$\hat{\beta}_{1ivw}^* \approx \beta_1 + \frac{\sum_{k=1}^{K} w_k\ \alpha_j}{\sum_{k=1}^{K} w_k} = \beta_1 + Bias\ (\alpha, \phi) \qquad (19)$$

The robustness of the IVW fixed effect results to pleiotropy can be assessed by comparing them to the results of models which, contrary to the IVW fixed effects model, do not required all the SNPs to be valid instruments (Bowden et al., 2015). Similar results from pleiotropy robust methods will support our confidence in the results. Following Burgess et al. (2020), I applied a set of pleiotropy robust estimators including IVW_Multiplicative Random Effects, MR-Egger, the Median based method, and the Mode-based method. A brief description of the IVW_Multiplicative Random Effects, the Weighted Median and Mode-base estimation

models are provided below. A more detailed presentation of the Egger method is given as it allows to test for directional horizontal pleiotropy.

Multiplicative Random Effects method. The IVW Multiplicative Random Effects method (IVW_MRE), originally from meta-analyses literature, permits between studies variation steaming from the SNP-specific direct effects on healthcare costs (horizontal pleiotropic effects) to be incorporated into the model (Thompson & Sharp, 1999; Mawdsley et al., 2017). The unbiasedness of the IVW_MRE estimator relays on the assumption that the instrument specific bias term ($\frac{\alpha_k}{\phi_k}$) in equation (18) cancel each other out so that the sum of the bias term in (19) is zero as the number of instruments/SNPs increases (Burgess et al., 2020) (A20), provided that the direct effect of the instruments on healthcare costs are uncorrelated with the SNP-exposure effect (Slob & Burgess, 2020) (A21). This is known as the Instrument Strength Independent of Direct Effect (InSide) condition i.e. $corr\ (\alpha_\kappa, \phi_k) = 0$ (Bowden et al. 2015).

By the assumptions of the model, the bias derived from pleiotropic effect do not bias the overall estimator of the effect. This situation is referred to as balance horizontal pleiotropy (Burgess et al., 2020) and means that the magnitude of the multiplicative and the fixed effects IVW models are the same (Bowden et al., 2017). Burgess et al. (2020) explains that the difference between IVW_MRE and IVW_FE is in the standard errors, which will be different in the presence of excess heterogeneity (between-ratio-estimators heterogeneity). The authors also explains that in this case the multiplicative model is superior to the fixed effects as the standard errors of the latter will be misleadingly small.

$$\hat{\beta}_{1IVW_{MRE}} = \frac{\sum_{k=1}^{K} w_k^* \frac{\widehat{\rho_k^*}}{\widehat{\phi_k}}}{\sum_{k=1}^{K} w_k^*} \qquad (20)$$

In the multiplicative IVW model, weights are equal to the inverse of the ratio estimators' variance times a measure of the heterogeneity among those (ω). ω is estimated in the model (see below) and it is set to one if the estimate is lower than one. (Thompson & Sharp, 1999).

$$w_k^*(MRE) = \frac{1}{\left(var\frac{\widehat{\rho_k}}{\widehat{\phi_k}}\right) * \omega} \qquad (21)$$

The standard errors are (Mawdsley et al, 2017; Borenstein et al. 2010):

$$se\left(\hat{\beta}_{1IVW\_MRE}\right) = \sqrt{\frac{1}{\sum_{k=1}^{K}\left(var\,\frac{\widehat{\rho_k}}{\widehat{\phi_k}}\right)}} * \sqrt{\omega} \qquad (22)$$

The unbiasedness of the multiplicative IVW estimator, also requires the NOME condition (A15) condition to hold.

The practical estimation of β_1 is done by fitting the following weighted OLS-regression equation using $\frac{1}{\omega * \sigma_{\rho k}^2}$ as weights. $\sigma_{\rho k}^2$ is within-ratio/sample variance of $\widehat{\rho_k}$ while ω represents the between-ratio heterogeneity. In praxis, $\omega * \sigma_{\rho k}^2$ is set equal to the empirically estimated standard error the SNP-outcome estimator ($se^2(\widehat{\rho_k})$) (Hemani et. al., 2018a, 2918b) ω can be independently estimated as the mean standard error/deviation (MSE/MSD) of regression equation (24) .

$$\hat{\rho}_k = \beta_{1ivw}\,\hat{\phi}_k + \epsilon_k;\ \epsilon_k \sim N(0, \omega * \sigma_{\rho k}^2) \qquad (23)$$

The Egger method. This method handles the case when the horizontal pleiotropic effects do not cancel out. This situation is known as directional/unbalanced horizontal pleiotropy and will make the estimator of the causal effect of the IVW models to be biased even in large samples (Hemani et al., 2018). Hemani et al. (2018) point out that the Egger method would be consistent even if all variants were invalid instruments. The causal effect (β_{1E}) is estimated by a weighted regression equation using $\frac{1}{\sigma_{\rho k}^2}$ as weights. In this case, the empirical standard errors of the SNP-outcome estimators ($se^2(\widehat{\rho_k})$) are regard a valid expression of $\sigma_{\rho k}^2$.

$$\hat{\rho}_k = \beta_{0E} + \beta_{1E}\,\hat{\phi}_k + \epsilon_k;\ \epsilon_k \sim N(0, \sigma_{\rho k}^2) \qquad (24)$$

The Egger method also requires the InSIDE condition to hold i.e $corr\ (\alpha_\kappa, \phi_k) = 0$ (A21). The Inside condition makes the Egger method to provide a bias-reduced estimate for the true causal effect as the sample sizes and number of instruments increases (Bowden et al. 2015):

$$\widehat{\beta_{1E}} = \frac{cov\,(\rho_k^*, \hat{\phi})}{var\,(\hat{\phi})} = \hat{\beta}_1 + \frac{cov\,(\hat{\alpha}, \hat{\phi})}{var\,(\hat{\phi})};\ cov\left(\hat{\alpha}, \hat{\phi}\right) \xrightarrow{N\to\infty} cov\,(\alpha, \phi) \xrightarrow{K\to\infty} 0 \qquad (25)$$

The variance of the Egger estimate of the causal effect is inversely proportional to the weighted variance of the $\hat{\phi}$ estimates using $\sigma_{\rho k}^2$ (in praxis, $\frac{1}{se^2(\widehat{\rho_k})}$) as weights (Burgess & Thompson, 2017). Hence the standard error of the estimator is:

$$se(\hat{\beta}_{1E}) = \frac{\omega}{\left(\sum_{k=1}^{K}\left(\hat{\phi}_k - \overline{\hat{\phi}}\right)^2\right)\sigma_{\rho k}^{-2}} \quad (26)$$

X and $\overline{\hat{\phi}}$ are the weighted average SNP-exposure association using the inverse-variance weights $\sigma_{\rho k}^{-2}$ (Burgess & Thompson 2017). As before, ω is the estimated residual standard error/deviation from equation (23) and it is included to adjust the Egger standard errors for any potential between-ratio heterogeneity, while keeping the magnitude of the β_{1E} unchanged (Bowden et al., 2017; Burgess & Thompson 2017)

An advantage of the MR-Egger method is that it allows to test whether the intercept term (β_{0E} in equation (24) is different from zero (Bowden et al., 2015). Bowden et al. (2015) explain that a non-zero intercept would suggest the presence of unbalanced horizontal pleiotropy, as $\hat{\beta}_{0E}$ can be interpreted as the estimate of the average pleiotropic effects across genetic variants. This test is referred to as the MR-Eggers' test. The MR- Egger intercept test does not require the Inside assumption to be valid (Burgess & Thompson, 2017).

Bowden & Holmes, 2019 explains that the MR-Egger estimate is likely to be substantially less precise than its IVW counterpart and it is therefore suggested to use it in the context of a sensitivity analysis rather than as a replacement of the standard IVW estimators(Burgess et al., 2020). The precision of the Egger estimate, the authors explain, would depend on the amount of variation between the set of SNP-exposure associations. As in any regression model, an outlying point can influence the regression coefficients (β_{0E} , β_{1E}).

The Egger method is especially sensitive to the orientation of the SNP-exposure association. Burgess & Thompson (2017) explain that the SNP-exposure estimates of associations should be oriented in the same direction (all positive or all negative associations) before fitting the MR-regression line. This is because different orientations would change the estimate of the intercept (β_{0E}) in equation (24) as well as the sign of the pleiotropic effect of the instrument (α_k) in (15) (Bowden, Davey Smith, et al., 2016). Burgess & Thompson (2017) remark that it is possible to reorient the estimates of genetic associations as there is no specific criterion to choose the allele of reference in GWAS studies. For example, in the case that a SNP can have a C and a T allele, the association could equally well be reported as for instance 0.243 units per additional copy of the C allele, or as -0.243 units per additional copy of the T allele. SNP-exposure association are usually oriented to be positive, and, if necessary, the SNP-outcome associations should also be altered to match the orientation of the SNP-exposure estimate (Bowden, Davey Smith, et al., 2016).

Median and mode. I briefly present a category of robust methods which define the causal estimate as the median or mode of the distribution of the ratio estimates. They are usually referred to as consensus methods (Slob & Burgess, 2020) and they may be more appropriate in cases where the Egger estimator does not perform well (i.e. limited variation among the SNP-exposure estimates and/or the presence of outliers) while relaxing the IVW-fixed effects assumption that all instruments are valid (A19) (Burgess & Thompson, 2017) and the assumption of balanced horizontal pleiotropy (A20).

In contrast to the IVW method that requires all instruments to be valid, and the Egger method that allows all instruments to be invalid, the median and mode methods are based on the "majority" and "plurality" valid assumptions respectively. The former assumption (A22) ensures that as the sample size increases, the causal estimates from all valid instrumental variables will tend towards the same value, which will equal the median provided that at least 50% of the instruments are valid (Burgess & Thompson, 2017). The latter assumption (A23) implies that, in large samples, the ratio estimates for all valid instruments should be equal to the true causal effect, while ratio estimates from invalid instruments will disperse towards different values. Based on this assumption the estimate of the true causal effect is the mode value of the ratio estimates. The plurality valid assumption is also known as the Zero Modal Pleiotropy Assumption (ZEMPA) (Slob & Burgess, 2020)

Weighted variants of the median and mode estimators have been proposed, in these cases, the median and mode are taken from a distribution of the ratio estimates in which the more precise ratio estimates are given more weight (Burgess & Thompson, 2017; Hartwig et al. 2017). Additionally, there is a penalised version of the weighted median estimator which recognises and adjusts for the fact that although invalid instruments do not affect the median estimate asymptotically, they may, however, influence the position where the median value is to be found in finite samples, and bias the estimate (Bowden, Davey Smith, et al., 2016). A weakness of the mode and median estimators is that they are sensitive to the addition and removal of variants from the analysis. Additionally, these methods may be less efficient compared to methods that base their estimates on all genetic variants (Slob & Burgess, 2020). Burgess et al. (2020) point out that, the mode methods have been found to have low precision when tested in some simulated and real datasets.

Heterogeneity – Gauging violations of the MR-assumptions

Building on the assumption that there is a single overall causal effect where the only source of variation is the within-ratio sample variation (according to the IVW fixed effects model), it is expected that valid genetic instruments yield similar estimates of this effect (Bowden & Holmes 2019). Substantial heterogeneity can be interpreted as an expression of horizontal pleiotropy.

Bowden & Holmes (2019) remark that it is important to note that heterogeneity does not prove that there are horizontally pleiotropic effects; in general, it is an indication that at least one of the assumptions of the model is not met, for instance the presence of weak instruments. Furthermore, Hemani et al., 2018, point out that homogeneous results could be due to perfect confounding; a situation where all the SNP-exposure instruments arise due to another trait influencing both the exposure and the outcome.

Heterogeneity can be assessed by the Cochran's Q statistic (Bowden and Holmes 2019):

$$Q = \sum_{k=1}^{K} Q_k = \sum_{k=1}^{K} w_k \left(\hat{\beta}_{1k} - \hat{\beta}_{1IVW\_FE}\right) \sim \chi^2_{k-1} \qquad (27)$$

Where w_k are defined as in eq. (8) and $\hat{\beta}_{1IVW\_FE}$ as in eq. (9).When Q is referred against a chi^2 distribution with k-1 degrees of freedom it tests the underlying null hypothesis of the heterogeneity test that the true treatment effect is the same across studies and that variations are due to sample variation (H0: Q=0) (Bowden and Holmes 2019).

A visual assessment of the heterogeneity can be done by examining funnel plots (Walker et al. 2019). A funnel plot is a scatter plot of the effect estimates from individual studies against some measure of precision originally used to gauge for, among other issues, publication bias and heterogeneity of treatment effects in meta-analyses (Sterne et al. 2011; Sterne & Egger, 2001) and later used in MR-context by e.g. Bowden et al. (2015; 2016). In a funnel plot, the more precise results are at the top of the Y-axis on the graph. In the absence of heterogeneity/bias, the plot would resemble a symmetrical inverse funnel, centred on the overall causal effect estimate (Sterne & Egger, 2001). Asymmetry of the inverse funnel may be due to heterogeneity of the individual causal effect estimates therefore suggest violations to the IV/MR the assumptions.

Violation of the NOME assumption

The IVW and Egger methods relay on a pragmatic assumption that the variance (squared standard error) of the SNP-exposure effect estimator ($\sigma^2_{\phi k}$) is small enough to be considered equal to zero i.e. the NO Measurement Error (NOME) assumption (A15). The argument for this is that, even though, zero variance would only be true if we had infinite observations, association studies are often done in large samples often above 100 000 (approximately 300 000 in our case) (Bowden et al., 2017). The violation of this assumption results in dilution bias of the causal effect estimator (Bowden, Del Greco, et al., 2016a).

According to Hartwing (2017) dilution bias for the IVW method can be estimated by the following relations where $\bar{F}$ defined as in equation (13) and $\bar{F}_w$ defined as in equation (14)

$$\frac{(\bar{F}-1)}{\bar{F}} \qquad (28)$$

$$\frac{\bar{F}_w-1}{\bar{F}_w} \qquad (29)$$

Dilution bias of the MR-Egger estimator can be estimated by the I^2_{GX} statistic suggested by Bowden, Del Greco, et al. (2016a) and defined as follows:

$$I^2_{GX} = (Q_{GX} - (K\text{-}1))/\, Q_{GX} \qquad (30)$$

Where Q_{GX} is and the Cochrans's Q (Bowden, Del Greco, et al., 2016b) .

$$Q_{GX} = \frac{\Sigma_{k=1}^{K}(\frac{\hat{\phi}_k}{\sigma^2_{\rho k}}-\overline{\hat{\phi}})^2}{\frac{\sigma^2_{\phi k}}{\sigma^2_{\rho k}}}, \overline{\hat{\phi}} = \frac{\Sigma_{k=1}^{K}\hat{\phi}_k/\sigma^2_{\phi k}}{\Sigma_{k=1}^{K}1/\sigma^2_{\phi k}} \qquad (31)$$

Note that Bowden and Del Greco et al.'s (2016b) definition of the Cochrans's Q in equation (31) is an extended version of equation (27) above. In the extended version, the weight term can incorporate uncertainty in the denominator of the ratio estimates ($\frac{\rho}{\phi}$) i.e. does not require the NOME condition to be hold.

Statistical analyses

Data

This study follows a liberal strategy for choosing instruments, which means that one uses relevant SNPs found along the genome. This is opposed to using SNPs for which the

researcher has a clear understanding of the biological path between the SNP and the trait of interest.

SNP-years of schooling associations. The estimates of the SNP-years of schooling association ($\widehat{\phi_k}$) were obtained from Okbay et al. (2016a). The authors performed 64 cohort specific association analyses comprising a total sample of 293,723 individuals of European descent (Okbay et al., 2016b) which were thereafter meta-analysed using simple sizes as weights. Okbay et al. (2016a) defined years of schooling as a continuous variable with the numbers of years of schooling completed (mean=14.3 and std. = 3.6) and identified 74 loci associated with the number of years of schooling completed at the genome-wide significant level. A summary of sample descriptive statistics and meta-analysis results are presented in Table 2.

In their analysis, Okbay et al. (2016b) included demographic control variables such as year of birth, sex and combination of those. A set of dummy variables to mark significant societal events that affected part of the cohort observations such as World War II, or changes in the educational system in the individual's country of origin were also included, finally, ten principal components indicating genetic ancestors were included as covariates (see Table 1). These principal components are estimated by a Principal components analysis (PCA) (Price et al., 2006 in Okbay et al., 2016b). Intuitively, the PCA recognises clusters of individuals in the sample by identifying genetic relationships among observations. In a GWAS with *J* SNPs, PCA analysis yields *J*-principal components (PC), each PC is interpreted as a marker of a specific ancestry. Including principal components as covariates is conceptually similar to using a set of dummies to mark the individual's ancestral group

Spelling out each of the *p* covariates in equation (2) in Okbay et al. (2016a) yields the following equation for each instrument (Rietveld, 2013):

$$X = \phi_{om} + \phi_k Z_k + \psi_{1-10} m_{1-10}^x + \psi_{11} m_{11}^x + \psi_{12} m_{12}^x + \psi_{13} m_{13}^x + \psi_{14} m_{14}^x + \psi_{15} m_{15}^x + \psi_{16} m_{16}^x + \psi_{17} m_{17}^x + \psi_{18-p} m_{18-p}^x + \varrho_m$$

Table 1 Description of variables included in regressions models by Okbay et al. (2016a)

X	*US years of schooling equivalent*
Z_k	*One instrument –Number of effect alleles at the specific SNP. Can take values 0, 1 or 2.*
m_{1-10}^x	*Ten principal components*
m_{11}^x	*birth year*

m_{12}^{x}	*birth year^2*
m_{13}^{x}	*birth year^3*
m_{14}^{x}	*Sex*
m_{15}^{x}	*sex*birth year*
m_{16}^{x}	*sex*birth year^2*
m_{17}^{x}	*sex*birth year^3*
$m_{18-p^{x}}^{x}$	*dummies marking a significant societal affecting only parts of the cohorts individuals*
ϱ_{m}	*Error term*

Source: Okbay et al. (2016b)

Depending on the values of association, the authors defined a lead SNP in each of the loci being investigated, as the SNP in the area with the smallest P-value. The estimated effect sizes of the lead SNPs in Okbay et al.'s (2016a) study ranged from 0.014 to 0.048 standard deviations of completed schooling years corresponding to 2.7 and 9 weeks of schooling respectively (Table 2). The number of years of schooling completed was assessed at or above age 30 (Okbay et al., 2016a). This study was chosen, as it does not contain data from the UK biobank dataset avoiding sample overlap.

Table 2 Sample descriptive information of Okbay et al.'s (2016) sample and results.

Meta-Analysis **Variable**	
Total Sample size N	293,723
Cohort sample size (min-max)	(318 – 76,155)
Share of females (min-max; mean)	(0 – 1; 0.55)
Mean year of birth (min-max)	(1921-1971)
Year of birth (min-max)	(1893-1985)
Overall years of schooling completed. Mean (std)	14.3 (3.6)
Number of lead SNPs	74
Estimated lead SNP effect (min max)	(0.014 std.- 0.048 std) (2.9 to 9 weeks)
R2 value of lead SNPs (min-max)	(0.01 % – 0.035 %)

Source: Okbay et al., 2016a; Okbay et al, .2016b

Summary data from the Okbay et al. (2016a) study was retrieved from the IEU GWAS database (Hemani et al., 2018a) by applying the extract_instruments command from the MR

Instruments R-package (Hemani et al., 2018a; Hemani et al., 2018b). To increase the probability of using independent instruments, the available SNPs were revised to identify pairs of SNPs in linkage disequilibrium. Linkage disequilibrium is here defined by customary parameters used in MR-studies i.e. a SNP is in linkage disequilibrium if it shows a pairwise R2 greater than 0.001 with another SNP within a distance of 10,000 kb. The R2 is a measure of association between allele frequencies that range from 0 to 1 where 0 shows perfect equilibrium (Slatkin,2008) i.e. random association. For each pair of SNPs in linkage disequilibrium, only one of them, the one with the strongest evidence of association with the exposure, was kept. This procedure is known as clumping (Hemani et al., 2018a) and led to the exclusion of one of the 74 SNPs identified by Okbay et al. (2016a).

Two more SNPs were excluded for being palindromic with intermediate allele frequencies. Intuitively, a palindromic SNP makes it difficult to assess the direction in which the gene sequence is being read. This leads to technical challenges when coding the value of the instrument variable (i.e. the number of effect alleles that a person has at a given SNP). The ambiguity can be solved by looking at differences in frequencies of the variant/allele that is less common in the population (minor allele frequency), however, when there are no substantial differences between the minor and the major allele frequencies, the problem cannot be resolved, and the SNP should not be used (see Hartwig et al., 2016).

SNP-health care costs associations. Summary data including coefficients of association between SNPs and healthcare costs were kindly provided by Senior research associate Sean Harrison at the University of Bristol. Harrison et al. (2021a) estimate healthcare costs as part of a Mendelian Randomisation study on the long-term cost-effectiveness of interventions for obesity in the UK based on a data set extracted from the UK biobank database (for an introduction to of the UK biobank, see e.g. Bycroft et al., 2018). The UK biobank contains information of approximately 500 000 people recruited among individuals who were registered at the National Health Service (NHS) and who lived within reasonable traveling distance of a total of 22 assessment centres across the UK between 2006 and 2010, moreover data from UK biobank has been linked to medical data from hospital episodes and primary care data (Harrison et al, 2021a).

In their main analysis, Harrison et al. (2021a) use a restricted sample from UK biobank consisting of unrelated individuals of white British ancestry recruited at centres in England or Wales, as costs data were not available for other centres (Harrison et al., 2021b). British ancestry was defined as participants who self-reported being “White British” and who had very

similar ancestral backgrounds according to the principal component analysis, additionally, participants without a BMI measure were also excluded from the analysis and are therefore not part of dataset on which my analysis is based. After exclusions, 310,913 participants aged between 39 and 72 years remained in the main dataset (Harrison et al., 2021b).

Harrison et al. (2021a) use available follow-up data including primary healthcare costs, on average for a period of 8.1 years. Follow-up period for secondary healthcare costs was on average 6.1 years. For each participant, primary care costs were calculated between recruitment and 31 March 2017 or death in the case that it occurred before this date. Costs were estimated using 2019-tariffs (November). Those costs were then averaged over years of follow-up to get individual average yearly primary care healthcare costs. Average yearly secondary healthcare costs for all participants in Harrison et al.'s study were available from Dixon et al. (2020). Dixon et al. (2020) estimates covered the period from recruitment to 31 March 2015, death, or emigration, whichever occurred first, and were originally reported in 2016/17 pounds sterling. Harrison et al. adjusted costs to 2019-pounds sterling (GBP) (Harrison et al., 2021a).

Average yearly total healthcare costs were estimated by combining the estimated average yearly primary and secondary care healthcare costs for each person yielding an average yearly estimate of total NHS-based healthcare including inpatient hospital care episodes, primary care appointments and primary care drug prescriptions and private healthcare received in NHS hospitals. The estimate excludes emergency care, outpatient appointments, private healthcare undertaken in private facilities as well as, diagnostic test costs in primary care (Harrison et al. 2021a). Estimated average yearly total health care costs as well as other relevant demographics of Harrison's et al. dataset are presented in Table 3.**Feil! Fant ikke referansekilden.**

Table 3 Summary demographics of sample used in Harrisons et al. (2021)

Variable	**All**	**Men**	**Women**
N	310 913	144 032	166 881
Age at recruitment, years [Mean (SD)]	56.9 (7.99)	57.1 (8.10)	56.7 (7.90)
Years of follow-up primary s [Mean (SD)]	8.1 (0.80)	8.1 (0.80)	8.1 (0.80)
Years of follow-up secondary costs [Mean (Median)]	6.1 (6.1)		
Average Total Healthcare costs per year [Median (IQR)]*	£601 (£212 to £1,217)	£605 (£206 to £1,240)	£596 (£216 to £1,199)

*Results from imputed data, median & IQR are the medians of 100 medians & IQRs

Source: (Harrison et al., 2021a.; Dixon et al., 2020)

Harrison et al.(2021a) included age at baseline assessment, sex, 40 principal components and UK Biobank recruitment centre location were included as covariates in their regression models (see Table 4). Spelling out the each of the *p* covariates in equation (3) yields the following equation for each instrument:

$$Y = \delta_{om} + \rho_k Z_k + \upsilon_1 m_1^y + \upsilon_2 m_2^y + \upsilon_{3-43} m_{3-43}^y + \upsilon_{44-p} m_{44-p^y}^y + \nu_m$$

Table 4 Description of variables included in regressions models by Harrison et al., (2021a)

Y	*Average yearly total healthcare costs in 2019 pounds sterling*
Z_k	*One instrument –Number of effect alleles at the specific SNP. Can take values 0, 1 or 2.*
m_1^y	*Age at baseline assessment*
m_2^y	*Sex*
m_{3-43}^y	*40 principal components*
$m_{44-p^y}^y$	*A set of dummies indicating UK Biobank recruitment centre location*
ν_m	*Error term*

Source: (Harrison et al., 2021a)

The estimated effect sizes per allele of the 71 SNPs being used as instruments, when SNPs-Exposure are positively oriented, the association range between, approximately, a reduction of £16 in 2019-GBP and an increment of £5 in 2019-GBP.

Primary analysis

The fixed effects model was in practice done by performing regression equation (15) using $\frac{1}{se^2(\widehat{\rho_k})}$ as weights. Primary and sensitivity analysis were carried out by using the R package for performing 2-sample MR (Hemani et al., 2018a, 2918b).

Assessment of assumptions validity

Weak instruments. I only included SNPs with a significant association to the exposure at the customary genome-wide significant level of p< 5 x 10^-8 to avoid weak instruments (Zhao et al., 2020). Following Bowden et al. (2017), Bowden & Holmes (2019), and Bowden et al. (2016a), the strength of each individual instrument was assessed by the F statistic (eq.(12)) additional to the simple average instrument strength (eq.(13)) and the weighted average

instrument (eq.(14)). In the practical calculations $se^2(\widehat{\rho_k})$ was used instead of $\sigma^2_{\phi k}$ by virtue of A16.

Okbay et al. (2016a) was the first study to discover most of the SNPs used as instruments in this thesis, there was therefore reason to fear the winner's curse problem in this thesis. In their GWAS, the authors examine the out-of-sample replicability of their results by using a subsample of the UK Biobank (n=11,349) released in 2015. Additionally, the SNP's effect adjusted for the winner's course was estimated by applying the methods described in a Rietveld et al.'s paper from 2014 (Rietveld et al., (2014) in Okbay et al. 2016b). Based on corrected effects, Okbay et al. find that 71.4 of the 74 SNPs were expected to have matching signs, 40.3 SNPs were expected to be significant at the 5% level, and 0.6 SNPs were expected to be genome-wide significant. The observed numbers are, respectively, 72, 51 and 7. Okbay et al. (2016a) conclude that the observed associations between SNPs and the number of years of schooling completed observed in their primary analysis, is not driven by an overestimation of the genetic associations in their original dataset. Okbay et al. (2016a) results support the assumption that the instruments used in our analysis are of sufficient strength.

Exogeneous condition. The risk of including instruments in linkage disequilibrium is handle by pruning the SNP (i.e. applying the clumping procedure describe in the methods chapter). I concentrate next on assessing population stratification and horizontal pleiotropy. Potential bias from assortative mating and dynastic effects can be avoided by within-family analysis (Brumpton et al., 2020). However, as this type of analysis requires individual data, it was not possible to carry out for this thesis.

The risk of population stratification is reduced by using summary data from studies on assumed homogeneous populations i.e. individuals of European and white British ancestry our case. In addition, Okbay et al.(2016a) and Harrison et al.(2021a) took additional steps to assess and correct for population stratification in their analysis. Both studies carry out Principal Components Analysis (PCA) and used the *t*-th principal components as covariates in their regression equations. Furthermore, Okbay et al. (2016a) assessed remaining effects due to confounding factors, such as population stratification by performing a linkage disequilibrium score regression (Sullivan et al., 2015 in Okbay et al., 2016a), a common technique in GWAS studies, and concluded that the effects of population stratification were small.

I estimate the overall causal effect by applying horizontal pleiotropy robust methods. Following Burgess et al. (2020), I applied the IVW-MRE, MR-Egger, the median based method

and the mode-based method. The IVW-MRE estimated was obtained by carrying out equation (23) using $\frac{1}{se^2(\widehat{\rho_k})}$ as weights while the Egger-estimator was estimated by regression equation (24) with the same weights. The necessary reorientation of the SNP-exposure estimates was done automatically when applying the "mr_egger_regression" option the R-MR package ((Hemani et al., 2018a, 2918b)). The Egger test of the intercept was performed by the "mr_pleiotropy_test" command in same package. Additionally, I performed a leave-one-out analysis (Burgess& Thompson, 2017; Hemani et al., 2018a, 2018b) to detect influential data points.

Heterogeneity. I assessed the risk for horizontal pleiotropy by evaluating heterogeneity among the ratio estimate of the causal effects (Hemani et al., 2018).In this study, the heterogeneity measure Q is in practice estimated by rearranging the Higgins' metric $H^2 = Q/(k-1)$ (Higgins & Thompson, 2002) to $Q = H^2(k-1)$ and taking advantage of that H^2 is equivalent to the mean standard error/deviation MSE (ω) from the regression equation (23) so that $Q = \omega(k-1)$ (Mawdsley et al. 2017; Higgins & Thompson, 2002; Hemani et al., 2018b).

Additionally, heterogeneity was assessed by examining funnel plots (Walker et al. 2019). A funnel plot (Figure 5) were created by the "mr_funnel_plot"-command included in the Two-sample MR R-package (Hemani et al., 2018b) and depict the effect estimates of each individual SNP as a separate instrument SNP on the X-axis against the inverse of the standard error of the each of the causal estimates. The overall causal effect in depicted in the graph as the vertical line from which symmetry is assessed (Hemani et al., 2018a, 2018b; Walker et al. 2019).

NOME Assumption. The magnitude of a potential dilution bias for the IVW estimates due to a violation of the NOME assumption (A15) was estimated by the relations described in eq. (28) and (29) (Hartwig et al., 2017). Dilution bias of the MR-Egger estimator was estimated by the I_{GX}^2 (eq. (30)) statistic (Bowden, Del Greco, et al. (2016a).

Results

Under the assumption that changes in genetic variants influence years of schooling in the same direction for all individuals (A10), our results suggest that there a negative effect of years of schooling on total healthcare costs later in life.

The standard Inverse Variance Weighted (IVW) method shows a reduction in yearly total healthcare costs of £219 per additional standard deviation of years of schooling (3.6 years) corresponding to £61 per additional year of schooling (Table 5and Figure 4). The effect of schooling on total healthcare costs remains negative across all additional methods, ranging between a reduction of £58 per additional years of schooling for the Weighted Median method and 29.4 for the MR-Egger method (£207 and £106 per standard deviation).

Results are not highly sensitive to the varying assumptions regarding the validity of instruments across methods. Under the scenario where up to 50 % of the instruments are invalid, the effect of years of schoolings on healthcare costs remains negative, of similar magnitude to the IVW estimator (which assumes that all instruments are valid) and highly significant as shown by the Weighted Median estimator.

The effect of years of schooling on healthcare costs also remains negative and significant under a scenario where the majority of instruments are invalid (Mode estimates). In the case that the majority of instruments were, in fact, invalid, we expect the Mode estimates of £-207 and £-186 (weighted and simple respectively) to be less biased than the Weighted Median and the IVW estimators (Hartwig et al., 2017). The weighted method being the less biased among the two mode estimators. The confidence intervals of the Mode estimators are, as expected (Hartwig et al., 2017), wider than those from the IVW and Weighted Median methods. The similarity in point estimates and the fact that they persist significantly across IVW, Weighted Median and Mode (weighted and simple) support the results of the conventional IVW-results showing a significant reduction in healthcare costs for an extra year of schooling.

The Egger-estimate is not significant. However, under the assumption of no directional horizontal pleiotropy, simulation studies (Hartwig et al., 2017; Bowden et al.,2015) have shown that the MR Egger is a considerably less powered method to detect a causal effect compared to the IVW-method, especially when the SNP-exposure effect sizes are relative homogeneous. Therefore, the considerable wider confidence intervals of the MR-Egger could be compatible with a scenario with no horizontal pleiotropy. For comparison, the Simple Mode method was found the second less powered method, followed, in increasing order by the Weighted Mode, Weighted Median and the IVW methods. The MR- Egger intercept test show that the intercept was not significant different from zero (value: -2; Std error: 2.76; P-value: 0.48) supporting the assumption of no directional horizontal pleiotropy. The MR- Egger intercept test does not require the Inside assumption to be valid (Burgess & Thompson, 2017).

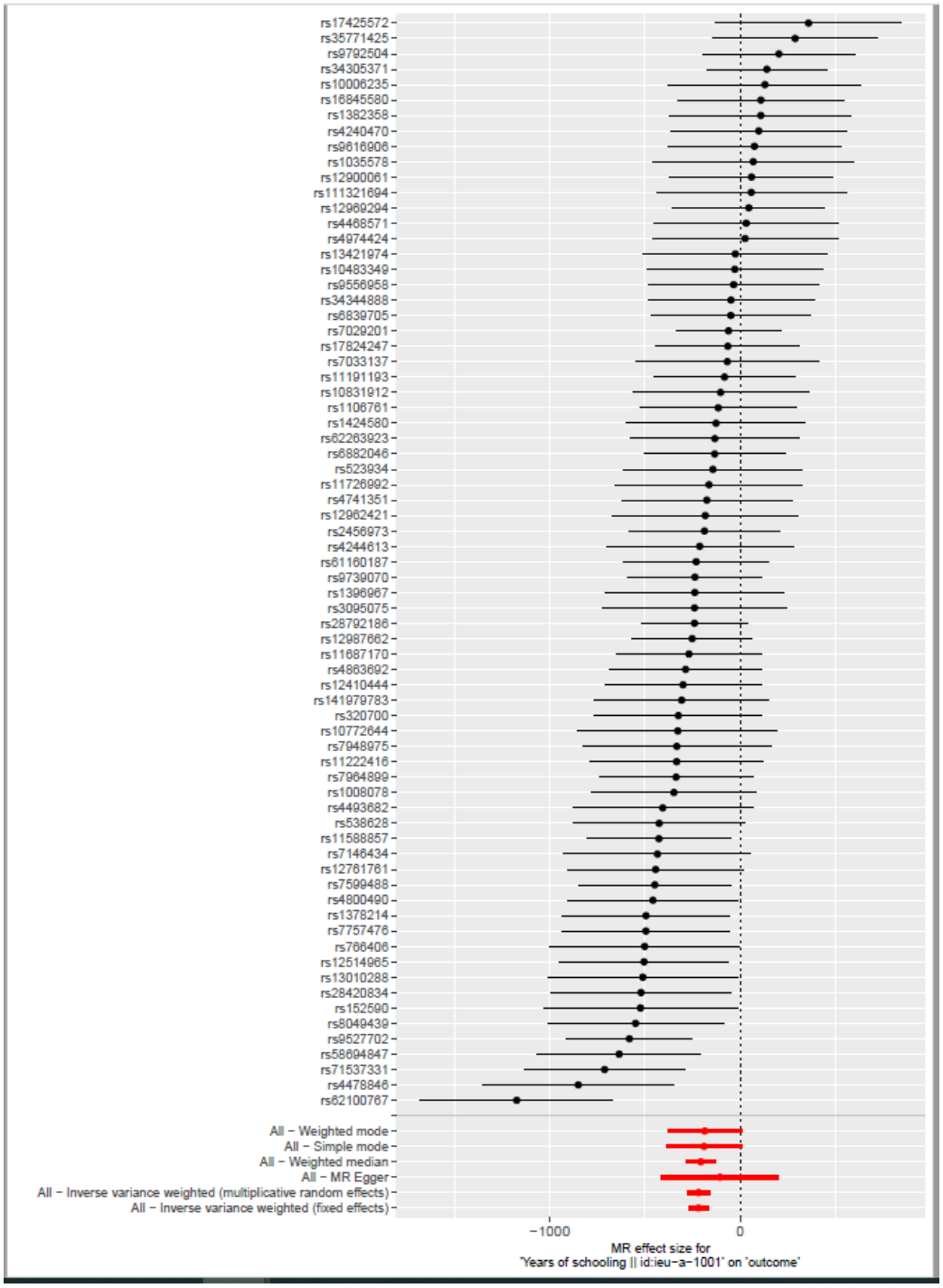
The strength of each individual instrument (F) ranges from 22 to 81. The simple average instrument strength ($\bar{F}$) is 38 while the weighted average instrument ($\overline{F_w}$) is 41. Potential dilution bias of the causal effect estimator due to violations of the No Measurement Error (NOME) for the IVW estimator was calculated by eq. (28) and estimated to be 3 %. Dilution bias of the MR-Egger estimator (I^2_{GX} = 0.97) indicate a small dilution bias attenuation of approximate 3 %.

Table 5 Results of the analysis. Estimates reflect changes in healthcare costs per one extra unit of years of schooling, standard errors, and confidence intervals of the estimates. Table shows changes for one additional year of schooling and for one extra standard deviation of years of schooling.

Method	*2019 GBP/year*	*2019 GBP/std. Years of schooling (3.6 years)*			*P Value*	
		Beta_1	*SE*	*95 % CI*		
IVW-fixed effects	-60.8	-218.9	25.8	(-269.5, -168.3)	2.22E-17	***
IVW-multiplicative random effects	-60.8	-218.9	30.6	(-278.8, -159)	8.10E-13	***
Weighted median	-57.6	-207.3	39.6	(-284.9, -129.7)	1.64E-07	***
Weighted mode	-51.6	-185.9	87.6	(-357.7, -14.1)	0.037446	*
Simple mode	-52.8	-190.1	96.8	(-379.9, -0.34)	0.053554	*
MR Egger	-29.6	-106.5	158	(-416.4, 203.1)	0.502559	ns

*Note: All methods are calculated using N= 71 Snps. Abbreviations: IVW, inverse variance weighted; MR, Mendelian randomisation. ns= P> 0.05, * = P ≤ 0.05, **= P ≤ 0.01, *** P ≤ 0.001*

Figure 4 Forest plot. The plot depicts the 71 ratio estimates of the effect size (β_{1k}) and their 95% confidence intervals used in this analysis. The six red dots diamonds at the bottom of the plot represents the overall effect, using different estimation methods. All point estimates are negative and of similar size, but MR-Egger is not significant and associated with much greater uncertainty compared to the alternative methods. The overall impression suggests a reduction of total (average yearly) healthcare costs per additional standard deviation of years of schooling (3.6 years).



I assess heterogeneity among ratio estimates to explore potential violations of the IV/MR assumptions, including, but not limited to, horizontal pleiotropy (Bowden & Holmes 2019; Hemani et al., 2018). The Cochran's Q statistic of the IVW and MR-Egger estimates were calculated (Q_{ivw}:98; Q_{ivw}_df: 70; Q_{Egger}:97; Q_{Egger}_df: 69) and test for a null of no

heterogeneity. The tests results show no indication of heterogeneity as both yielded P-values of approximately 0,014. Heterogeneity was additionally assessed by visual inspection of funnel plots (Figure 5) which suggest no signs of considerable heterogeneity, supporting the impression that heterogeneity and directional pleiotropy are not a considerable issue for this study as suggested by the Cochran's Q statistic and the MR- Egger test.

Figure 5 Funnel Plot. X-axis shows the estimated overall effect of an additional unit years of schooling on yearly healthcare costs per effect allele in 2019-punds sterling against the inverse of its standard error. The more precise results are at the top of the Y-axis. In the absence of heterogeneity, the plot would resemble a symmetrical inverse funnel, centred at the overall causal effect estimate (solid lines).

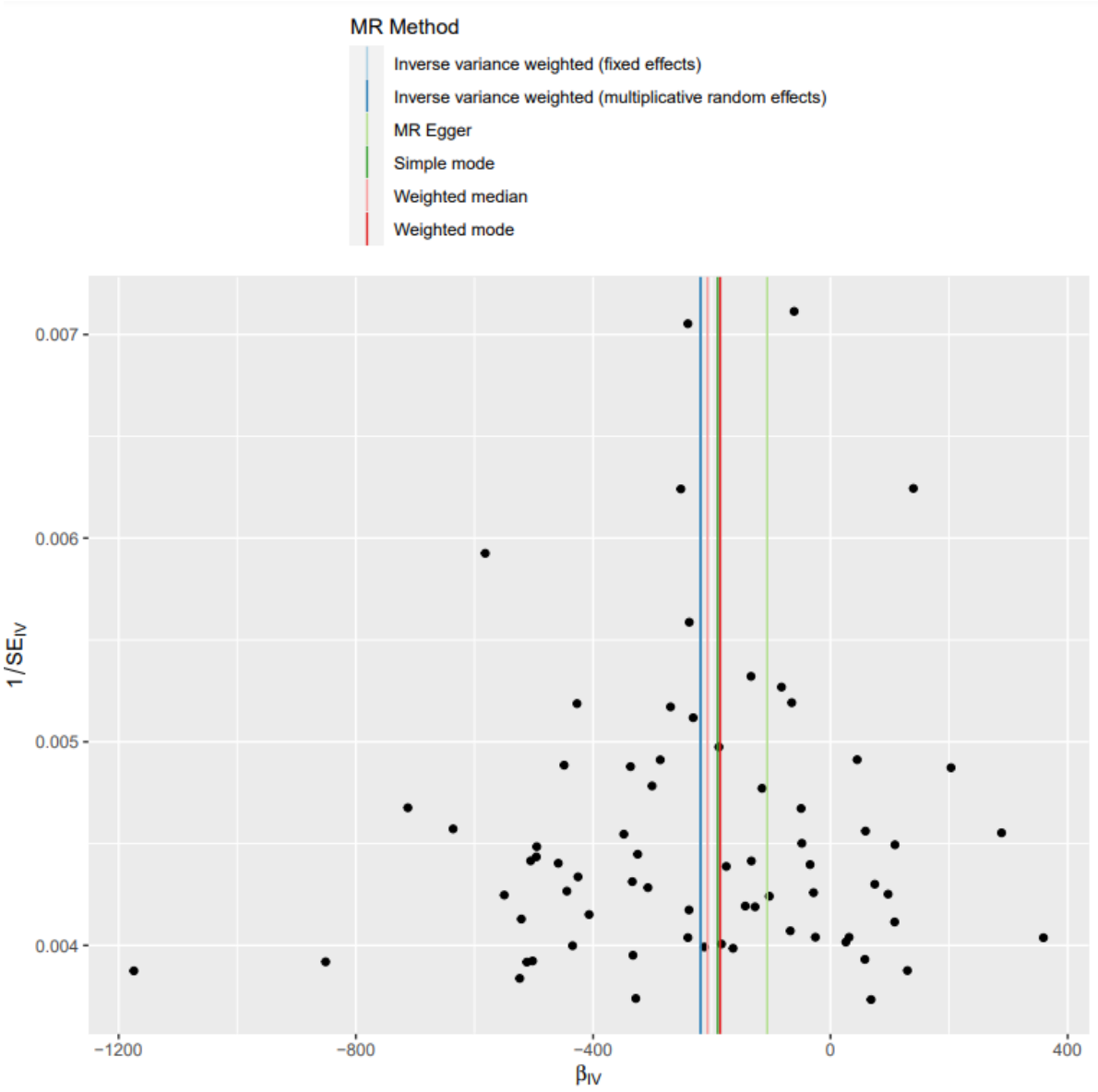
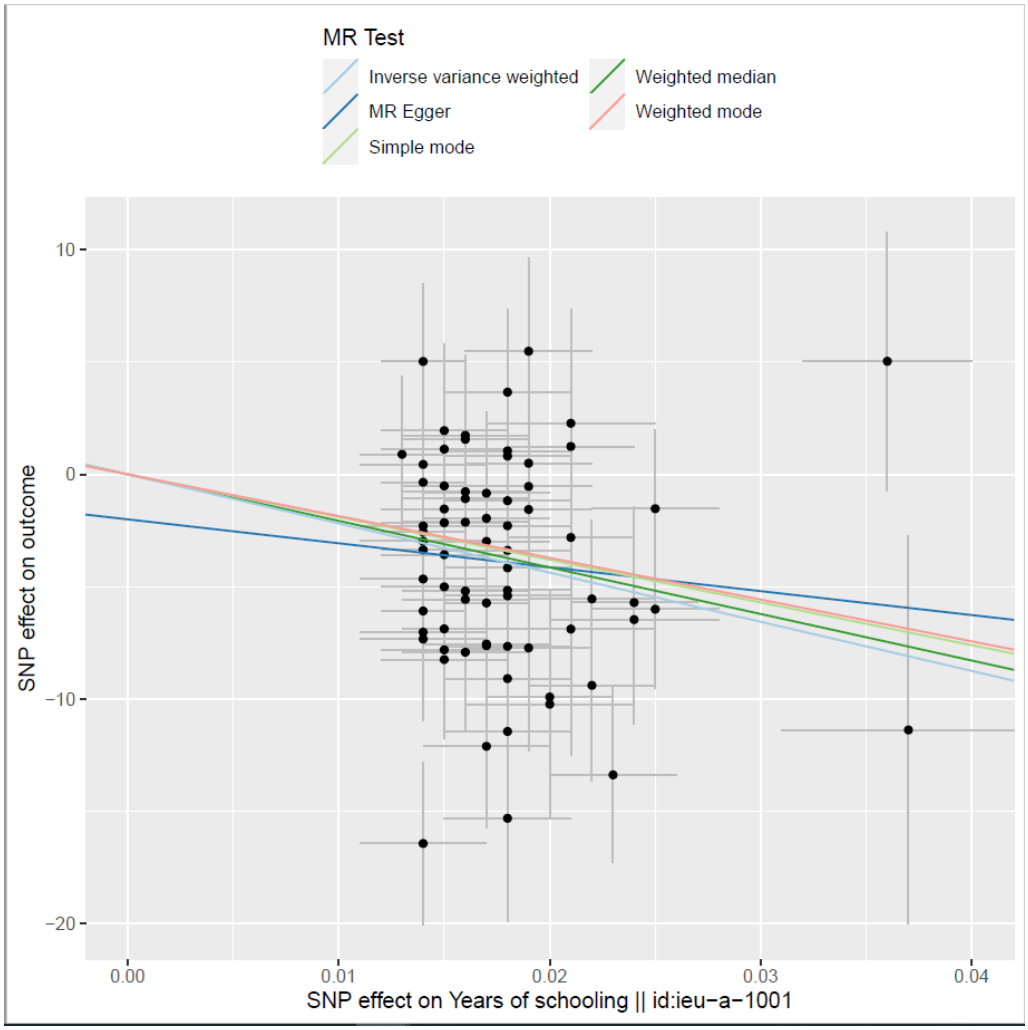


Figure 6 is a graphical summary of the analysis. The estimates of an increase of one standard deviation of years of schooling per effect allele at the instrumental SNP ($\hat{\phi}_k$) are depicted on the X-axis. The estimates of association between an effect allele at the instrumental

SNP and yearly healthcare costs measured in 2019-GBP ($\hat{\rho}_k$). The gradient from the origo to each point is the ratio estimate ($\frac{\hat{\rho}_k}{\hat{\phi}_k}$) of the effect of years of schooling on healthcare costs. The slope of the lines represents the overall estimates of effect of years of schooling on healthcare costs according to the different methods. The scatter plot shows graphically that there are no extreme discrepancies among the ratio estimates and the overall estimates. We see however that there are some ratio estimates on the periphery. Those ratio estimates influence the overall results (except for the simple mode estimate) depending on their precision.

Figure 6. Scatter Plot. X-axis shows an increase of one standard deviation of years of schooling per effect allele. Y-axis shows the increase in yearly healthcare costs per effect allele in 2019-punds sterling. The slope of each line correspond to the overall effect estimate under each method.



I investigate whether the results are dominated by a single variant/SNP, by carrying out a leave-one-out analyses. The results are summarised on Table 6 suggesting that the primary/original results were not driven by a single SNP. Figure 7 is a visual presentation of the results of leave-one-out analysis for the IVW-Multiplicative estimate. It presents the original overall effect IVW-Multiplicative estimate with confidence intervals (in red) along the

estimates and confidence intervals obtained by a series of iterations where one SNP was left out at the time. This plot shows that, for instance, the IVW-Multiplicative overall estimate becomes 209.24 when SNP rs62100767 at the top of panel is excluded from the analysis. Results indicated that the IVW-Multiplicative stimator is not driven by individual SNPS. Figures 8 to 11 in appendix 4 visualises the results for the leave-one-out analysis of the Egger, Weighted Median and Mode, and Simple Mode estimates

Figure 7. Forest plot of the leave-one-analysis of the IVW-Multiplicative effects method. Each black dot represents the overall estimate using IVW multiplicative method when the analysis is replicated by leaving out one SNP. The overall IVW-multiplicative estimate become, for instance, 209.24 when SNP rs62100767 at the top of panel is excluded from the analysis. Results suggest that the IVW-Multiplicative estimator is not driven by individual SNPS.

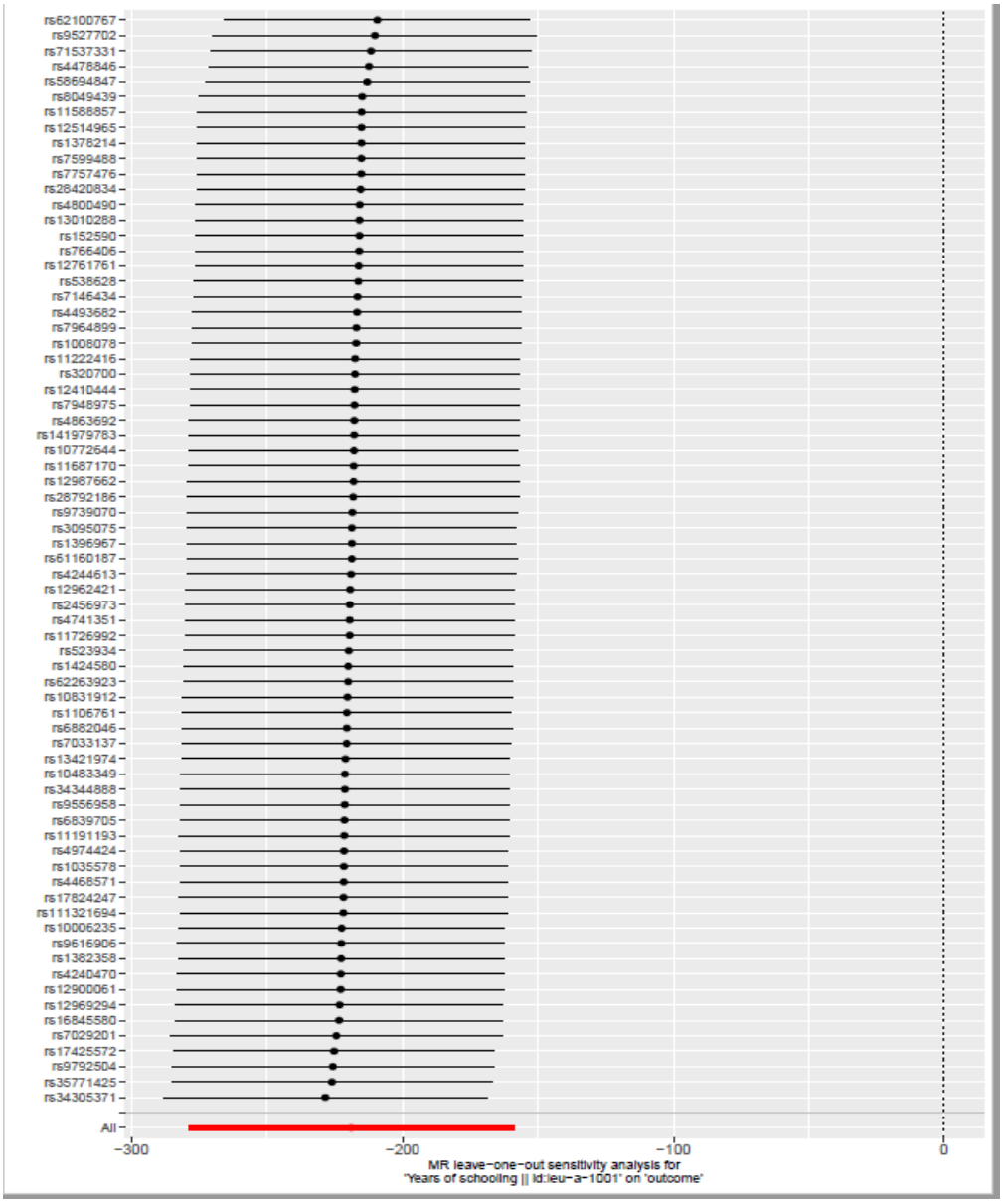


Table 6 Summary of leave-one-out analyses. For each method, 71 regressions were run. Each regression left out one SNP and recovered an overall estimate of the causal effect. The original estimate (including 71 SNPs) is showed in column 2. Column 3 presents the range of estimates produced by the series of 71 regressions.

Estimator	Original estimate	Range of estimates from leave-one-out analyses
Inverse variance weighted (fixed effects)	-218.9	(-228.5,-209.3)
Inverse variance weighted (multiplicative random effects)	-218.9	(-228.5,-209.3)
MR Egger	-106.5	(-241.7, -42.12)
Weighted median	-207.3	(-235.5, -186.18)
Simple mode	-190.1	(-215, - 159.9)
Weighted mode	-185.9	(-215.1, - 143)

Discussion

In this thesis, I have estimated the effect of a marginal increase in completed years of schooling on yearly total healthcare costs (primary and secondary healthcare). The aim of this thesis is to estimate a causal effect. However, this is challenging as the relationship involves potential confounding and reverse causation problems. Confounding and reverse causation problems in a naïve regression model would hamper a causal interpretation of the coefficients. Hence, instrumental variables in a Two-sample Mendelian Randomisation analysis were applied. This method might provide causal effects of education on health care costs, given that certain conditions are fulfilled.

This thesis has two main empirical findings and provides some significant methodological insights. Firstly, my results show a reduction of healthcare costs of approximately £61 in 2019-GBP for an extra year of completed schooling (£219 per std in years of schooling (3.6 years); CI £267 - £168).). Secondly, estimates are not highly sensitive to the varying assumptions regarding the validity of instruments across the various methods applied, neither show signs of considerable heterogeneity or potential dilution bias, supporting our main results. The last conclusion that I draw from my experiences working on this thesis is that

genetic variants as instruments in a Two sample-MR with publicly available summary data from well powered GWAS can be used to carry out exploratory analyses of economic outcomes.

To the best of my knowledge, this is the first paper that has studied the impact of education attainment/years of schooling for the general population on health care costs. However, the literature on the association between education and health, education and healthcare services utilisation, and education and health literacy reviewed earlier might provide some insight on the relationship between years of schooling and healthcare costs. Even though the causal effect of educational attainment on health is yet to be consistently stablished (Galama et al., 2018; Clark & Royer, 2013), a large and robust association between education and health have been documented, also after controlling for several background characteristics. My results are in line with studies that find a positive effect of education on indicators of health such as mortality by applied instrumental variable methods3 (e.g. Lleras-Muney (2006); Davies, Dickson, et al., 2019). However, my estimates would reflect potential additional mechanism highlighted in healthcare costs literature, which may work in opposite directions. There is, for instance, evidence suggesting that there are differences in the use of healthcare services, and types of treatment received related to education levels, even when comparing individuals with similar health status (e.g. Droomers & Wester, 2004; Fiva et al., 2014), and there are also empirical results showing that health literacy appears to be a predictor of factors influencing healthcare costs e.g. patients' disengagement, and misuse of available health resources (Palumbo, 2017). Overall, my results suggest that a reduction in costs related to gains in health and health literacy associated with additional years of schooling may surpass potential costs' increments linked to differences in services utilisation i.e., for given health, better educated individual may make more use of expensive services compared to less educated patients.

Strengths. Under a set of conditions (summarised in appendix 2), Mendelian randomisation (MR)/IV methods allow to avoid endogeneity problems such as biases due to confounding and reverse causation and make it possible to make causal inferences from my results (Lawlor et al., 2008). In this thesis, I reviewed a set of circumstances that may challenge the plausibility of the assumptions required by MR. Whenever possible, the likelihood of these assumptions was assessed by statistical means. Additionally, to reduce the threat of weak instruments, only SNPs associated with years of schooling at the genome-wide significant level of p<5x10^-8 were included. To increase the possibilities of using exogenous variation, SNPs

[3] It is worth mentioning that not all IV-analysis that have found such effect (e.g Mazumder (2008) or Fletcher (2015) in Galama et al., 2018).

were pruned to only include independent instruments (not in Linkage Disequilibrium) and I used summary data from studies based on individuals of European and white British ancestry with non-overlapping samples. Moreover, I relay in Okbay et al.'s (2016a) and Harrison et al.'s (2021) efforts to produce valid estimates of the SNP-years of schooling, and SNP-costs associations.

Provided that the MR/IV-assumptions hold, my results shed light on the powerful effect of education attainment on healthcare costs but, equally important, they provide some methodological insights about the use of genetic instruments in the analysis of economic outcomes. Von Hinke et al. (2016) pointed out that MR might not be a compelling method for the analysis effects of endogenous behavioural phenotypes (e.g. smoking or drinking) on economic outcomes. The distinctive characteristic of behavioural phenotypes is that much of the observed variation may stem from environmental factors, von Hinke et al.(2016) presented a series of arguments including the following: 1) genetic variants, while not weak in a statistical sense, will generally, only create small shifts in the exposure, this will be problematic as economic outcomes often require large changes in the exposure to impact the outcome, 2) the sample sizes of data sets containing data on relevant exposures and outcomes were too small (about 4000-5000 observations) to produce precise estimates.

With these arguments in mind, it was therefore a somewhat unexpected results that the small amount of exogeneous variation in years of schooling introduced by the instruments (2.9 to 9 weeks of additional education per effect allele; R-squared values ranged between 0.01 % and 0.035 %) lead to statistically significant changes in healthcare costs. This point is particular important when using a two sample MR approach with summary data. The reason behind it is that the use of aggregate data limits the possibilities to aggregate the variance induced by each SNPs into a single, more powerful instrument. For context, Davies, Dickson, et al. (2019) construct a single, combined instrument based on same SNPs as I use. Davies, Dickson, et al. (2019) observe that a unit increase in their combined instrument, is associated with 1.45 additional years of education (std. 0.05), while the variation induces by each SNPs used individually in this study range between 2.9 and 9 weeks of additional education. My results make the case that one should not discard using genetic variants solely on the argument that they won't induce enough variation to the exposure. The usefulness of SNPs as instruments, as von Hinke et al. (2016) also suggested, would depend on how much of a change in the exposure is needed to produce a change in the outcome. In our case, had healthcare costs only be altered by large shifts in completed years of education, using SNPs as instruments may not detect the

causal effect. Thus, two-sample MR with free available summary data and statistical software (Hemani et al. 2018a), might therefore be an affordable method to analyse economic outcomes. Furthermore, my results are relative precise thanks to the rapid increase in sample sizes of the datasets, and the implementation of two-sample methods, which relax the need of having a single dataset containing genetic markers, information on the exposure, and information on the outcome. However, it is important to keep in mind is that my estimates may come to better use when analysing relatively modest changes in years of schooling as inferring the effects of a small change onto a larger one (e.g. one whole additional year of education) would require making assumptions on the way effects accumulate.

Weaknesses. There are some concerning issues regarding the internal and external validity of my estimates. Firstly, it has been documented that participants UK biobank study are among other things, healthier, wealthier and are better educated than the general British population (e.g. Dixon et al., 2020). Education and healthcare costs seems likely to have influenced their participation in the study, creating a selection bias, which will mean that my estimates could be distorted. Secondly, as I did not have access to individual data, it was not possible to carry out analyses to assess potential bias from assortative mating and dynastic effects, or to test for non-linearity of the effects of years of schooling on healthcare costs.

Furthermore, to my understanding, it has not been rigorously established which particular causal effect is identified when using genetic variants as instruments. Imbens & Angrist (1994) showed that, under the monocitcy assumption, IV-methods identify the *local effect treatment effect* (LATE) which is the causal effect for those individuals[4] who received treatment only because they were induced by the instrument (Aronow & Carnegie, 2013). In many cases, however, the estimate of interest is the average treatment effect for the whole population of interest (ATE).

On the one hand, some authors have proposed that, in MR studies, the LATE and ATE effects are equivalents (Davies, Dickson, et al., 2019; Dixon et al., 2020). Davies, Dickson, et al. (2019) compare using genetic variants to other types of instruments, specifically, an education reform in Britain, and suggest that while the education reform would only affect those at the end of the education distribution, genetic instruments, would affect individuals

[4] These individuals are known as *compliers.* For reference, Aronow & Carnegie (2013) explain that in a setting with a binary instrument and a binary treatment, population is made of three subpopulations additional to compliers: 1) *defiers* i.e. individuals who do the opposite of what is expected by the value of their instrumental variable/assignment, 2) *never-takers* who do not take the treatment regardless of the value of their instrumental variable and 3) *always takers*, those who take the treatment, whatever their assignment.

independently of observed levels of education i.e. across the entire education distribution. Dixon et al. (2020) on their side, justify their conclusion by referring to the fact that the exposure in question (BMI) is a continuous variable. A supportive argument for assuming the LATE and ATE effects to be similar might be that, when using summary data, each ratio/IV estimate is generated from a single SNP, thus, each ratio estimate is associated with a different group of *compliers.* Therefore, as the set of ratio estimates in this study show no signs of considerable heterogeneity, we might imply that the estimated causal effects of years of schooling may be similar across the UK-biobank sample. This could in turn be interpreted as an indication of the similarity between the LATE and ATE effects of years of schooling on costs.

On the other hand, the potential selection bias mentioned above may imply that the sample being studied only covers part of the distribution of educational attainment from the general population. Additionally, as, to my awareness, the knowledge on the biological paths making the effect alleles to result in more years of completed schooling is not extensive, we cannot rule out that various of SNPs being used work through the same biological mechanism. Thus, that the set of compliers across SNPs might, in fact, be substantially overlapping. This is particularly relevant to my results since Okbay et al. (2016a) concluded that the 74 SNPs affecting educational attainment were disproportionally found in genomic regions related to neural development and showed that they were genetically correlated to cognitive performance and intercranial volume. It may therefore be that the genetic variants do not affect the exposure independently of education levels as suggested by Davies, Dickson, et al. (2019) but rather, mostly affect those at the top of the education distribution, given the positive correlation between cognitive performance and education levels (Deary & Johnson, 2010). This would in turn weaken the belief that the LATE effects estimated by my MR-analysis are equivalent to the average treatment effect (ATE) for the population.

Moreover, the genetic correlation between cognitive ability and completed years of schooling discovered by Okbay et al. (2016a) could potentially violate the exclusion criteria if there was any direct association between cognitive performance and health/healthcare unrelated to education attainment. A way to assess the likelihood of (horizontal) pleiotropy due to e.g. cognitive performance is to carry out a multivariable MR-analysis. With the data at hand, it was not possible to carry out this type of analysis.

Furthermore, my estimates of healthcare costs do not include, among other things, costs of emergency care, outpatient care or diagnostic tests' costs in primary care (Harrison et al., 2021a), if those unobserved where to increase with a marginal increase in education, the effect

size could be closer to zero or even result in an increase in costs following an increment in education. However, to my knowledge, there is not conclusive evidence of an education gradient in the use of primary, emergency or outpatient healthcare. As seen earlier, existing studies have demonstrated either no association, a positive association between education and the use of preventive care, or a causal effect between education and the use of out-of-hours-primary care (i.e., less educated individuals are more likely to use out- of-hours primary care services). While it is difficult to assess the way in which including costs of emergency care or diagnostic test costs in primary care would change my estimate, it seems rather likely that including those costs would increase the size of the effect and not change the direction

Additionally, it may be argued that, from a societal point of view, the policy relevant question is whether there is an effect of years of schooling on the overall healthcare expenditure/costs[5] instead of the governmental financed expenditure/costs that I have used in this thesis. Analyses of the overall healthcare expenditure might, for instance, yield effect estimates which take potential changes in financing patterns associated with higher levels of educational attainment into account (Cf. results suggesting that higher educated people and higher incomes tend to prefer private specialist services outside the governmental schemes). If the foregoing mechanisms were strong enough, they could lead to contradictory results when analysing overall healthcare costs, regardless of the financing source, and governmental financed healthcare costs. However, potential discrepancies would probably depend on the financing structure of the healthcare system in question. In our case, costs' estimates are based on NHS data from the UK, which is characterised by universal access to healthcare and where governmental financed expenditure stands for most of the overall healthcare costs. For instance, in 2019, it was calculated that governmental financed expenditure corresponded to 79 % of the overall health care expenditure in the UK. The same year, out of pocket, private insurance and other financing sources shares were recorded to be 16 %, 3 % and 2 % respectively (Office for National Statistics, 2020). These figures might allow us to assume governmental financed costs' estimates used here are proxies of the overall healthcare costs and, thus, one could expect to obtain a negative effect of an increment in years of schooling on healthcare costs, even when using costs estimated including both governmental and private expenditure.

It is a methodological concern that the knowledge about the functions of genetic variants and the pathways in which they affect education attainment are, to the best of my knowledge,

[555] That is, adding expenditure financed by out-of-pocket expenditure, private health insurance and other sources of financing to the government-financed expenditure.

still not well known. This is important because, if as von Hinke et al. (2016) argues, MR is a controversial approach within economics, because it is not possible to test the exogenous criteria. In that case, it is vital to gain that knowledge, so that one can make compelling theoretical arguments that support the exogenous assumption.

Policy relevance and further research. Under the set of assumptions discussed through this paper, and if reading my results as estimates of the average treatment effect in the population, the findings of this thesis would be relevant for any policy maker facing a constrained healthcare budget, regardless of the economic power of their countries. Several examples of policy relevant questions that causal estimates of the effect of years of schooling on healthcare costs could inform come to mind. One being cost-benefits analysis of alternative interventions aiming to increase health literacy or to reduce healthcare costs, for instance, when comparing increasing years of formal schooling to information campaigns targeting the general population. Another example could be policy makers considering using education as a strategy to reduce future healthcare costs.

However, further work is needed before it would be possible to make more confident use of the findings of this thesis. A natural next step in future research would be to replicate my analysis using a random sample of the population to avoid selection bias. Testing for non-linear effects and adjusting for e.g. cognitive performance by applying a multivariable MR-method would be a significant improvement on my analysis. Ideally, one would estimate the effect of years of schooling on healthcare costs using individual data on siblings and/or family trios to assess the risk of assortative mating and dynastic effects.

In more general terms, there is a need to understand the biological pathways through which genetic instruments influence the exposure of interest. It would provide insights on the specific type of causal effects being estimated in Mendelian Randomisation studies. On a methodological note, it seems imperative to fully understand which causal effect a MR-analysis identifies. In that endeavour, it could be useful to explore the methods proposed by Aronow & Carnegie (2013) to deduce ATE estimates from the LATE ones.

Conclusion

This thesis finds that an extra year of completed schooling would reduce total healthcare costs (from primary and secondary healthcare) by approximately £61 in 2019-GBP (£219 per std in years of schooling (3.6 years); CI £267 - £168).). Furthermore, estimates are not highly

sensitive to the varying assumptions regarding the validity of instruments across the various methods applied, neither show signs of considerable heterogeneity or potential dilution bias, supporting our main results. However, only using summary data from previous studies pose some limitations to the analysis. Thus, these results should be treated with care. Given the financing system in Britain, from which the data on costs was collected, the reduction in costs seems to yield the overall healthcare costs, including both governmental and private expenditure.

Additionally, I conclude that, contrary to earlier concerns, a Two-sample MR with summary data can be used to perform exploratory analyses of economic outcomes. Genetic variants should not be dismissed as instruments beforehand on the argument that each SNP provides very little exogenous variation to heavily environmental-influenced exposures. This is because the usefulness of genetic instruments to identify a causal effect would depend on the strength of the relation between the exposure and the outcome, the mechanisms that mediate this relationship, and the size of the samples being used.

Moreover, conducting this research has led to the conclusion that a Two sample-MR with non-disclosive and easily accessible summary data from well powered GWAS can be used as an affordable method to carry out exploratory analyses of economic outcomes. However, methodologically solid research results would require using individual data from a random sample. Ideally containing information on siblings and/or family trios.

An important additional note regarding the MR-method is that there is a need to understand the biological pathways through which genetic instruments influence the exposure of interest. Without this knowledge it is difficult to make convincing arguments supporting the credibility of the exogenous condition, which cannot be tested otherwise. Finally, it seems of vital importance to fully understand and rigorously establish which specific causal effects a MR-analysis identifies.

References

Angrist, J. D., & Krueger, A. B. (1992). The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples. *Journal of the American Statistical Association*, *87*(418), 328-336. https://doi.org/10.2307/2290263

Angrist, J. D., & Krueger, A. B. (1995). Split-sample instrumental variables estimates off the return to schooling. *Journal of business & economic statistics*, *13*(2), 225-235. https://doi.org/10.1080/07350015.1995.10524597

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press.

Angrist, J. D., & Pischke, J.-S. (2015). *Mastering metrics: the path from cause to effect*. Princeton University Press.

Aronow, P. M., & Carnegie, A. (2013). Beyond LATE: Estimation of the Average Treatment Effect with an Instrumental Variable. *Polit. anal*, *21*(4), 492-506. https://doi.org/10.1093/pan/mpt013

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97-111. https://doi.org/https://doi.org/10.1002/jrsm.12

Bowden, J., Davey Smith, G., & Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology*, *44*(2), 512-525. https://doi.org/10.1093/ije/dyv080

Bowden, J., Davey Smith, G., Haycock, P. C., & Burgess, S. (2016). Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic Epidemiology*, *40*(4), 304-314. https://doi.org/10.1002/gepi.21965

Bowden, J., Del Greco M, F., Minelli, C., Davey Smith, G., Sheehan, N. A., & Thompson, J. R. (2016). Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *International Journal of Epidemiology*, dyw220. https://doi.org/10.1093/ije/dyw220

Bowden, J., Del Greco M, F., Minelli, C., Davey Smith, G., Sheehan, N. A., & Thompson, J. R. (2016). Technical appendix to accompany "Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic". *International Journal of Epidemiology*, *Oxford University Press (OUP)*. https://doi.org/10.1093/ije/dyw220

Bowden, J., Fabiola, Minelli, C., Davey Smith, G., Sheehan, N., & Thompson, J. (2017). A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Statistics in Medicine*, *36*(11), 1783-1802. https://doi.org/10.1002/sim.7221

Bowden, J., & Holmes, M. V. (2019). Meta-analysis and Mendelian randomization: A review. *Research Synthesis Methods*, *10*(4), 486-496. https://doi.org/10.1002/jrsm.1346

Brumpton, B., Sanderson, E., Heilbron, K., Hartwig, F. P., Harrison, S., Vie, G. Å., Cho, Y., Howe, L. D., Hughes, A., Boomsma, D. I., Havdahl, A., Hopper, J., Neale, M., Nivard, M. G., Pedersen, N. L., Reynolds, C. A., Tucker-Drob, E. M., Grotzinger, A., Howe, L., Morris, T., Li, S., Auton, A., Windmeijer, F., Chen, W.-M., Bjørngaard, J. H., Hveem, K., Willer, C., Evans, D. M., Kaprio, J., Davey Smith, G., Åsvold, B. O., Hemani, G., & Davies, N. M. (2020). Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-17117-4

Burgess, S., Davey Smith, G., Davies, N. M., Dudbridge, F., Gill, D., Glymour, M. M., Hartwig, F. P., Holmes, M. V., Minelli, C., Relton, C. L., & Theodoratou, E. (2020). Guidelines for performing Mendelian randomization investigations. *Wellcome Open Research*, *4*, 186. https://doi.org/10.12688/wellcomeopenres.15555.2

Burgess, S., Dudbridge, F., & Thompson, S. G. (2016). Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and

summarized data methods. *Statistics in Medicine*, *35*(11), 1880-1906. https://doi.org/10.1002/sim.6835

Burgess, S., Scott, R. A., Timpson, N. J., Davey Smith, G., Thompson, S. G., & Consortium, E.-I. (2015). Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *European Journal of Epidemiology*, *30*(7), 543-552. https://doi.org/10.1007/s10654-015-0011-z

Burgess, S., Small, D. S., & Thompson, S. G. (2017). A review of instrumental variable estimators for Mendelian randomization. *Statistical Methods in Medical Research*, *26*(5), 2333-2355. https://doi.org/10.1177/0962280215597579

Burgess, S., & Thompson, S. G. (2011). Avoiding bias from weak instruments in Mendelian randomization studies. *International Journal of Epidemiology*, *40*(3), 755-764. https://doi.org/10.1093/ije/dyr036

Burgess, S., & Thompson, S. G. (2017). Interpreting findings from Mendelian randomization using the MR-Egger method. *European Journal of Epidemiology*, *32*(5), 377-389. https://doi.org/10.1007/s10654-017-0255-x

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203-209. https://doi.org/10.1038/s41586-018-0579-z

Case, A., Fertig, A., & Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics*, *24*(2), 365-389. https://doi.org/10.1016/j.jhealeco.2004.09.008

Clark, D., & Royer, H. (2013). The effect of education on adult mortality and health: Evidence from Britain. *Am Econ Rev*, *103*(6), 2087-2120. https://doi.org/10.1257/aer.103.6.2087

Dahl, E., Bergsli, H., & van der Wel, K. (2014). *Sosial ulikhet i helse: en norsk kunnskapsoversikt* (SAM - Department of Social Work, Child Welfare and Social Policy 296, Issue. https://hdl.handle.net/20.500.12199/738

Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?. *International Journal of Epidemiology*, *32*(1), 1-22. https://doi.org/10.1093/ije/dyg070

Davey Smith, G., & Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics*, *23*(R1), R89-R98. https://doi.org/10.1093/hmg/ddu328

Davey Smith, G., Holmes, M. V., Davies, N. M., & Ebrahim, S. (2020). Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues. *European Journal of Epidemiology*, *35*(2), 99-111. https://doi.org/10.1007/s10654-020-00622-7

Davies, N. M., Dickson, M., Davey Smith, G., Windmeijer, F., & van den Berg, G. J. (2019). *The Causal Effects of Education on Adult Health, Mortality and Income: Evidence from Mendelian Randomization and the Raising of the School Leaving Age.* (No. 12192). Institute of Labor Economics (IZA), Bonn. https://www.econstor.eu/handle/10419/196690

Davies, N. M., Holmes, M. V., & Davey Smith, G. (2018). Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*, k601. https://doi.org/10.1136/bmj.k601

Davies, N. M., Howe, L. J., Brumpton, B., Havdahl, A., Evans, D. M., & Davey Smith, G. (2019). Within family Mendelian randomization studies. *Human Molecular Genetics*, *28*(R2), R170-R179. https://doi.org/10.1093/hmg/ddz204

de Meijer, C., Koopmanschap, M., d' Uva, T. B., & van Doorslaer, E. (2011). Determinants of long-term care spending: Age, time to death or disability? *Journal of Health Economics*, *30*(2), 425-438. https://doi.org/https://doi.org/10.1016/j.jhealeco.2010.12.010

Deary, I. J., & Johnson, W. (2010). Intelligence and education: causal perceptions drive analytic processes and therefore conclusions. *International Journal of Epidemiology*, *39*(5), 1362-1369. https://doi.org/10.1093/ije/dyq072

Dee, T. S., & Evans, W. N. (2003). Teen drinking and educational attainment: Evidence from two-sample instrumental variables estimates. *Journal of labor economics*, *21*(1), 178-209. https://doi.org/10.1086/344127

Dixon, P., Hollingworth, W., Harrison, S., Davies, N. M., & Davey Smith, G. (2020). Mendelian Randomization analysis of the causal effect of adiposity on hospital costs. *Journal of Health Economics*, *70*, 102300-102300. https://doi.org/10.1016/j.jhealeco.2020.102300

Droomers, M., & Westert, G. P. (2004). Do lower socioeconomic groups use more health services, because they suffer from more illnesses? *European Journal of Public Health*, *14*(3), 311-313. https://doi.org/10.1093/eurpub/14.3.311

Elwert, F., & Winship, C. (2014). Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable. *Annual review of sociology*, *40*, 31-53. https://doi.org/10.1146/annurev-soc-071913-043455

Fiva, J. H., Hægeland, T., Rønning, M., & Syse, A. (2014). Access to treatment and educational inequalities in cancer survival. *Journal of Health Economics*, *36*, 98-111. https://doi.org/10.1016/j.jhealeco.2014.04.001

Fletcher, J. M., & Frisvold, D. E. (2009). Higher Education and Health Investments: Does More Schooling Affect Preventive Health Care Use? *Journal of human capital*, *3*(2), 144-176. https://doi.org/10.1086/645090

Galama, T., Lleras-Muney, A., & Kippersluis, H. v. (2018). The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence. https://doi.org/10.1093/acrefore/9780190625979.013.7

Glazier, R. H., Agha, M. M., Moineddin, R., & Sibley, L. M. (2009). Universal Health Insurance and Equity in Primary Care and Specialist Office Visits: A Population-Based Study. *The Annals of Family Medicine*, *7*(5), 396-405. https://doi.org/10.1370/afm.994

Godager, G., & Iversen, T. (2013). *Empirisk litteratur om sosial ulikhet i bruk av helsetjenester i Norge. Underlagsrapport til Sosial ulikhet i helse: En norsk kunnskapsoversikt*. Elektronisk publisering. Retrieved 14 November 2021 from https://docplayer.me/2685872-Empirisk-litteratur-om-sosial-ulikhet-i-bruk-av-helsetjenester-i-norge-underlagsrapport-til-sosial-ulikhet-i-helse-en-norsk-kunnskapsoversikt.html

Gregori, D., Petrinco, M., Bo, S., Desideri, A., Merletti, F., & Pagano, E. (2011). Regression models for analyzing costs and their determinants in health care: an introductory review. *International Journal for Quality in Health Care*, *23*(3), 331-341. https://doi.org/10.1093/intqhc/mzr010

Harrison, S., Dixon, P., Jones, H. E., Davies, A. R., Howe, L. D., & Davies, N. M. (2021). Long-term cost-effectiveness of interventions for obesity: A mendelian randomisation study. *PLOS Medicine*, *18*(8), e1003725. https://doi.org/10.1371/journal.pmed.1003725

Harrison, S., Dixon, P., Jones, H. E., Davies, A. R., Howe, L. D., & Davies, N. M. (2021). S2 Text. Inclusion criteria an genotyping. Supporting information to Long-term cost-effectiveness of interventions for obesity: A mendelian randomisation study. *PLOS Medicine*, *18*(8). https://doi.org/https://doi.org/10.1371/journal.pmed.1003725.s002

Hartwig, F. P., Davey Smith, G., & Bowden, J. (2017). Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *International Journal of Epidemiology*, *46*(6), 1985-1998. https://doi.org/10.1093/ije/dyx102

Hartwig, F. P., Davies, N. M., & Davey Smith, G. (2018). Bias in Mendelian randomization due to assortative mating. *Genetic Epidemiology*, *42*(7), 608-620. https://doi.org/10.1002/gepi.22138

Hartwig, F. P., Davies, N. M., Hemani, G., & Davey Smith, G. (2016). Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int J Epidemiol*, *45*(6), 1717-1726. https://doi.org/10.1093/ije/dyx028

Haycock, P. C., Burgess, S., Wade, K. H., Bowden, J., Relton, C., & Davey Smith, G. (2016). Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *The American Journal of Clinical Nutrition*, *103*(4), 965-978. https://doi.org/10.3945/ajcn.115.118216

Hemani, G., Bowden, J., & Davey Smith, G. (2018). Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Human Molecular Genetics*, *27*(R2), R195-R208. https://doi.org/10.1093/hmg/ddy163

Hemani, G., Zheng, J., Elsworth, B., Wade, K., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., Tan, V. Y., Yarmolinsky, J., Shibab, H., Timpson, N., Evans, D., Relton, C., Martin, R., Davey Smith, G., Gaunt, T., Haycock, P., & Collaboration, T. M.-B. (2018a). The MR-Base platform supports systematic causal inference across the human phenome. *eLife*, *7(), e34408.* https://doi.org/https://doi.org/10.7554/eLife.34408

Hemani, G., Zheng, J., Elsworth, B., Wade, K., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., Tan, V. Y., Yarmolinsky, J., Shibab, H., Timpson, N., Evans, D., Relton, C., Martin, R., Davey Smith, G., Gaunt, T., Haycock, P., & Collaboration, T. M.-B. (2018b). Detailed code for The MR-Base platform supports systematic causal inference across the human phenome. *eLife*, *7(), e34408*. https://doi.org/ https://rdrr.io/github/MRCIEU/TwoSampleMR/f/

Hernán, M. A., & Robins, J. M. (2006). Instruments for Causal Inference: An Epidemiologist's Dream? *Epidemiology*, *17*(4), 360-372. https://doi.org/10.1097/01.ede.0000222409.00878.37

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539-1558. https://doi.org/10.1002/sim.1186

Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, *62*(2), 467-475. https://doi.org/10.2307/2951620

Inoue, A., & Solon, G. (2010). Two-sample instrumental variables estimators. *The review of economics and statistics*, *92*(3), 557-561. https://doi.org/10.1162/REST_a_00011

Jansen, T., Rademakers, J., Waverijn, G., Verheij, R., Osborne, R., & Heijmans, M. (2018). The role of health literacy in explaining the association between educational attainment and the use of out-of-hours primary care services in chronically ill people: a survey study. *BMC health services research*, *18*(1), 394-394. https://doi.org/10.1186/s12913-018-3197-4

Keller, M. C., Garver-Apgar, C. E., Wright, M. J., Martin, N. G., Corley, R. P., Stallings, M. C., Hewitt, J. K., & Zietsch, B. P. (2013). The genetic correlation between height and IQ: shared genes or assortative mating? *PLoS genetics*, *9*(4), e1003451. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003451

Khaing, W., Vallibhakara, S. A., Attia, J., McEvoy, M., & Thakkinstian, A. (2017). Effects of education and income on cardiovascular outcomes: A systematic review and meta-analysis. *European Journal of Preventive Cardiology*, *24*(10), 1032-1042. https://doi.org/10.1177/2047487317705916

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., Benonisdottir, S., Oddsson, A., Halldorsson, B. V., Masson, G., Gudbjartsson, D. F., Helgason, A., Bjornsdottir, G., Thorsteinsdottir, U., & Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*(6374), 424-428. https://doi.org/doi:10.1126/science.aan6877

Labrecque, J., & Swanson, S. A. (2018). Understanding the Assumptions Underlying Instrumental Variable Analyses: a Brief Review of Falsification Strategies and Related Tools. *Curr Epidemiol Rep*, *5*(3), 214-220. https://doi.org/10.1007/s40471-018-0152-1

Lawlor, D. A. (2016). Commentary: Two-sample Mendelian randomization: opportunities and challenges. *International Journal of Epidemiology*, *45*(3), 908-915. https://doi.org/10.1093/ije/dyw127

Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N., & Davey Smith, G. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, *27*(8), 1133-1163. https://doi.org/10.1002/sim.3034

Lleras-Muney, A. (2006). The Relationship between Education and Adult Mortality in the United States Adriana Lleras-Muney. *The Review of Economic Studies*, *72,73*(3), 189,847-847. https://doi.org/10.1111/j.1467-937X.2006.00398.x

Martín, J. J. M., Puerto López Del Amo González, M., & Dolores Cano García, M. (2011). Review of the literature on the determinants of healthcare expenditure. *Applied Economics*, *43*(1), 19-46. https://doi.org/10.1080/00036841003689754

Mawdsley, D., Higgins, J. P. T., Sutton, A. J., & Abrams, K. R. (2017). Accounting for heterogeneity in meta-analysis using a multiplicative model-an empirical study. *Research Synthesis Methods*, *8*(1), 43-52. https://doi.org/10.1002/jrsm.1216

Office for National Statistics (ONS), (2020, 14 October 2020). *Healthcare expenditure, UK Health Accounts: 2019*. Retrieved 14 November 2021 from https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/ukhealthaccounts/2019

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., Turley, P., Chen, G.-B., Emilsson, V., Meddens, S. F. W., Oskarsson, S., Pickrell, J. K., Thom, K., Timshel, P., De Vlaming, R., Abdellaoui, A., Ahluwalia, T. S., Bacelis, J., Baumbach, C., Bjornsdottir, G., Brandsma, J. H., Pina Concas, M., Derringer, J., Furlotte, N. A., Galesloot, T. E., Girotto, G., Gupta, R., Hall, L. M., Harris, S. E., Hofer, E., Horikoshi, M., Huffman, J. E., Kaasik, K., Kalafati, I. P., Karlsson, R., Kong, A., Lahti, J., Lee, S. J. V. D., Deleeuw, C., Lind, P. A., Lindgren, K.-O., Liu, T., Mangino, M., Marten, J., Mihailov, E., Miller, M. B., Van Der Most, P. J., Oldmeadow, C., Payton, A., Pervjakova, N., Peyrot, W. J., Qian, Y., Raitakari, O., Rueedi, R., Salvi, E., Schmidt, B., Schraut, K. E., Shi, J., Smith, A. V., Poot, R. A., St Pourcain, B., Teumer, A., Thorleifsson, G., Verweij, N., Vuckovic, D., Wellmann, J., Westra, H.-J., Yang, J., Zhao, W., Zhu, Z., Alizadeh, B. Z., Amin, N., Bakshi, A., Baumeister, S. E., Biino, G., Bønnelykke, K., Boyle, P. A., Campbell, H., Cappuccio, F. P., Davies, G., De Neve, J.-E., Deloukas, P., Demuth, I., Ding, J., Eibich, P., Eisele, L., Eklund, N., Evans, D. M., Faul, J. D., Feitosa, M. F., Forstner, A. J., Gandin, I., Gunnarsson, B., Halldórsson, B. V., Harris, T. B., Heath, A. C., Hocking, L. J., Holliday, E. G., Homuth, G., Horan, M. A., Hottenga, J.-J., De Jager, P. L., Joshi, P. K., Jugessur, A., Kaakinen, M. A., Kähönen, M., Kanoni, S., Keltigangas-Järvinen, L., Kiemeney, L. A. L. M., Kolcic, I., Koskinen, S., Kraja, A. T., Kroh, M., Kutalik, Z., Latvala, A., Launer, L. J., Lebreton, M. P., Levinson, D. F., Lichtenstein, P., Lichtner, P., Liewald, D. C. M., Cohort Study, L., Loukola, A., Madden, P. A., Mägi, R., Mäki-Opas, T., Marioni, R. E., Marques-Vidal, P., Meddens, G. A., McMahon, G., Meisinger, C., Meitinger, T., Milaneschi, Y., Milani, L., Montgomery, G. W., Myhre, R., Nelson, C. P., Nyholt, D. R., Ollier, W. E. R., Palotie, A., Paternoster, L., Pedersen, N. L., Petrovic, K. E., Porteous, D. J., Räikkönen, K., Ring, S. M., Robino, A., Rostapshova, O., Rudan, I., Rustichini, A., Salomaa, V., Sanders, A. R., Sarin, A.-P., Schmidt, H., Scott, R. J., Smith, B. H., Smith, J. A., Staessen, J. A., Steinhagen-Thiessen, E., Strauch, K., Terracciano, A., Tobin, M. D., Ulivi, S., Vaccargiu, S., Quaye, L., Van Rooij, F. J. A., Venturini, C., Vinkhuyzen, A. A. E., Völker, U., Völzke, H., Vonk, J. M., Vozzi, D., Waage, J., Ware, E. B., Willemsen, G., Attia, J. R., Bennett, D. A., Berger, K., Bertram, L., Bisgaard, H., Boomsma, D. I., Borecki, I. B., Bültmann, U., Chabris, C. F., Cucca, F., Cusi, D., Deary, I. J., Dedoussis, G. V., Van Duijn, C. M., Eriksson, J. G., Franke, B., Franke, L., Gasparini, P., Gejman, P. V., Gieger, C., Grabe, H.-J., Gratten, J., Groenen, P. J. F., Gudnason, V., Van Der Harst, P., Hayward, C., Hinds, D. A., Hoffmann, W., Hyppönen, E., Iacono, W. G., Jacobsson, B., Järvelin, M.-R., Jöckel, K.-H., Kaprio, J., Kardia, S. L. R., Lehtimäki, T., Lehrer, S. F., Magnusson, P. K. E., Martin, N. G., McGue, M., Metspalu, A., Pendleton, N., Penninx, B. W. J. H., Perola, M., Pirastu, N., Pirastu, M., Polasek, O., Posthuma, D., Power, C., Province, M. A., Samani, N. J., Schlessinger, D., Schmidt, R., Sørensen, T. I. A., Spector, T. D., Stefansson, K., Thorsteinsdottir, U., Thurik, A. R., Timpson, N. J., Tiemeier, H., Tung, J. Y., Uitterlinden, A. G., Vitart, V., Vollenweider, P., Weir, D. R., Wilson, J. F., Wright, A. F., Conley, D. C., Krueger, R. F., Davey Smith, G., Hofman, A., Laibson, D. I., Medland, S. E., Meyer, M. N., Yang, J., Johannesson, M., Visscher, P. M., Esko, T., Koellinger, P. D., Cesarini, D., & Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604), 539-542. https://doi.org/10.1038/nature17671

Okbay, A., et al. (2016b). Supplementary information to Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604). https://doi.org/10.1038/nature17671

National Human Genome Research Institute, (2019, 26.03.2019). *Genome-Wide Association Studies (GWAS)*. Retrieved 28.04.2021 from https://www.genome.gov/genetics-glossary/Genome-Wide-Association-Studies

Palumbo, R. (2017). Examining the impacts of health literacy on healthcare costs. An evidence synthesis. *Health Services Management Research*, *30*(4), 197-212. https://doi.org/10.1177/0951484817733366

Pierce, B. L., & Burgess, S. (2013). Efficient Design for Mendelian Randomization Studies: Subsample and 2-Sample Instrumental Variable Estimators. *American Journal of Epidemiology*, *178*(7), 1177-1184. https://doi.org/10.1093/aje/kwt084

Pierce, B. L., & VanderWeele, T. J. (2012). The effect of non-differential measurement error on bias, precision and power in Mendelian randomization studies. *International Journal of Epidemiology*, *41*(5), 1383-1393. https://doi.org/10.1093/ije/dys141

Rietveld, C. A (2013). *Educational Attainment Analysis Plan–Phase 2 to Genome-wide association study identifies 74 loci associated with educational attainment*. Center for open science. https://osf.io/paj9m/

Slatkin, M. (2008). Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nature reviews. Genetics*, *9*(6), 477-485. https://doi.org/10.1038/nrg2361

Slob, E. A. W., & Burgess, S. (2020). A comparison of robust Mendelian randomization methods using summary data. *Genetic Epidemiology*, *44*(4), 313-329. https://doi.org/10.1002/gepi.22295

Sterling, S., Chi, F., Weisner, C., Grant, R., Pruzansky, A., Bui, S., Madvig, P., & Pearl, R. (2018). Association of behavioral health factors and social determinants of health with high and persistently high healthcare costs. *Preventive Medicine Reports*, *11*, 154-159. https://doi.org/https://doi.org/10.1016/j.pmedr.2018.06.017

Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *J Clin Epidemiol*, *54*(10), 1046-1055. https://doi.org/10.1016/S0895-4356(01)00377-8

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, *343*, d4002. https://doi.org/10.1136/bmj.d4002

Stock, J. H., & Watson, M. W. (2007). *Introduction to econometrics* ([2nd ed.]. ed.). Pearson Addison Wesley.

Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, *18*(20), 2693-2708. https://doi.org/https://doi.org/10.1002/(SICI)1097-0258(19991030)18:20<2693::AID-SIM235>3.0.CO;2-V

von Hinke, S., Davey Smith, G., Lawlor, D. A., Propper, C., & Windmeijer, F. (2016). Genetic markers as instrumental variables. *J Health Econ*, *45*, 131-148. https://doi.org/10.1016/j.jhealeco.2015.10.007

Walker, V. M., Davies, N. M., Hemani, G., Zheng, J., Haycock, P. C., Gaunt, T. R., Davey Smith, G., & Martin, R. M. (2019). Using the MR-Base platform to investigate risk factors and drug targets for thousands of phenotypes. *Wellcome Open Research*, *4*, 113. https://doi.org/10.12688/wellcomeopenres.15334.2

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed. ed.). MIT Press.

Xu, W., Tan, L., Wang, H.-F., Jiang, T., Tan, M.-S., Tan, L., Zhao, Q.-F., Li, J.-Q., Wang, J., & Yu, J.-T. (2015). Meta-analysis of modifiable risk factors for Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, jnnp-2015-31054. https://doi.org/10.1136/jnnp-2015-310548

Zhao, Q., Wang, J., Hemani, G., Bowden, J., & Small, D. S. (2020). Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Annals of Statistics*, *48*(3), 1742-1769. https://doi.org/10.1214/19-aos1866

Appendix A. Overview of equations

Formula numbering refers to equation numbers in main text. References for definitions and theoretical results are to be found in the main text.

Variables

Y	Total healthcare costs (Average yearly total healthcare costs in 2019-pounds sterling)
X	US years of schooling equivalent.
$\boldsymbol{\gamma}\mathbf{M}$	Row vector with the product of a P covariates times their corresponding coefficients in structural equation 1
Z_k	Instrument k. Number of effect alleles at the specific SNP. Each individual is, at conception, assigned a number of effect alleles (Z_{ki}). Z_{ki} can take values 0, 1 or 2.
$\boldsymbol{\psi}\mathbf{M}^{\mathbf{x}}$	Row vector with the product of p^x covariates times their corresponding coefficients in reduced-form equation 2 *Covariates:* m^x_{1-10} *Ten principal components* m^x_{11} *Birth year* m^x_{12} *Birth year^2* m^x_{13} *Birth year^3* m^x_{14} *Sex* m^x_{15} *Sex*birth year* m^x_{16} *Sex*birth year^2* m^x_{17} *Sex*birth year^3* $m^x_{18-p^x}$ *Dummies marking a significant societal event affecting only parts of the cohort's individuals*
$\boldsymbol{\upsilon}\mathbf{M}^{\mathbf{y}}$	Row vector with the product of p^y covariates times their corresponding coefficients in reduced-form equation 3 *Covariates:* m^y_1 *Age* m^y_2 *Sex* m^y_{3-43} *40 principal components* $m^y_{44-p^y}$ *A set of dummies indicating UK Biobank recruitment centre location*

ϱ_m , ν_m	*Error terms*

Structural equation:

$$Y = \beta_{om} + \beta_1 X + \boldsymbol{\gamma M} + \varepsilon_m \qquad (1)$$

Where the years of schooling of individual *i* is correlated with unmeasured factors included in the error term *i.e.* $corr(X_i, \varepsilon_i) \neq 0$.

There is one instrument (Z_k) which ideally satisfies conditions A5-A9 yielding the following reduced forms. Each reduced-form equation was estimated in a separate sample:

$$X = \phi_{om} + \phi_k Z_k + \boldsymbol{\psi M^x} + \varrho_m \qquad (2)$$

$$Y = \delta_{om} + \rho_k Z_k + \boldsymbol{\upsilon M^y} + \nu_m \qquad (3)$$

Notice that the covariates in the structural equation are not well/unambiguously defined but represent a similar collection of demographic variables.

The magnitude of a potential causal effect can, provided that the monotonicity assumption A10 holds, be estimated by |the ratio of the OLS-estimator of association between the instrument and healthcare cost from reduced-form equation (3) to the OLS-estimator of association between the instrument and years of schooling from equation (2). Given conditions A11-A12

The OLS estimators $\widehat{\rho_k}$ and $\widehat{\phi_k}$ used as input data for my study were calculated in to separated samples assumed to satisfy conditions A13 and A14 (Okbay et al., 2016; Harrison et al., 2021). In total, this study utilises 71 instruments assumed to be independent. The resulting 71 ratio estimates of β_1 are assumed to provide independent evidence (A17) of a common overall estimate (A18). Each ratio estimate is represented as follows:

$$\hat{\beta}_{1k} = \frac{\widehat{\rho_k}}{\widehat{\phi_k}}, \ k=(1,\ldots,71) \qquad (5)$$

The standard errors and variance of the ratio is presented below. Those are estimated by the Taylor approximations formula while requiring conditions A15 and A16 to hold.

$$Var\hat{\beta}_{1k} = var\frac{\hat{\rho}_k}{\hat{\phi}_k} = \frac{\sigma^2_{\rho k}}{\hat{\phi}^2_k} \qquad (6)$$

$\sigma^2_{\rho k}$ is the variance of $\widehat{\rho_k}$.

Information from the 71 estimates is combined by using the Fixed-effects IVW method, originally from meta-analysis literature. The Fixed-effects IVW requires all included SNP's to be valid instruments (A19):

$$\hat{\beta}_{1IVW} = \frac{\sum_{k=1}^{K} w_k \frac{\widehat{\rho_k}}{\widehat{\phi_k}}}{\sum_{k=1}^{K} w_k} \quad (7)$$

In equation *(7)*, weights are set equal to the inverse of the ratio estimators variance (eq. *(6)*): *(8)*

$$w_k FE = \frac{1}{var \frac{\hat{\rho}_k}{\hat{\phi}_k}} = \frac{\hat{\phi}_k^2}{\sigma_{\rho k}^2}$$

The IVW_FE becomes: *(9)*

$$\hat{\beta}_{1IVW_{FE}} = \frac{\sum_{k=1}^{71} \hat{\phi}_k \hat{\rho}_k \sigma_{\rho k}^{-2}}{\sum_{k=1}^{71} \hat{\phi}_k^2 \sigma_{\rho k}^{-2}}$$

And the standard errors are: *(10)*

$$se(\hat{\beta}_{1IVW\_FE}) = \sqrt{\frac{1}{\sum_{k=1}^{K} var \frac{\hat{\rho}_k}{\hat{\phi}_k}}}$$

The practical estimation of the B1 is carried out by fitting the following OLS weighted regression:

$$\hat{\rho}_k = \beta_{ivw} \hat{\phi}_k + \epsilon_k; \ \ \epsilon_k \sim N(0, se^2(\widehat{\rho_k}))$$

$$Weights = \frac{1}{se^2(\widehat{\rho_k})} \quad (11)$$

Under the assumption (A16) that $se^2(\widehat{\rho_k}) = \sigma_{\rho k}^2$

Sensitivity analyses

A major threat to instrument exogeneity when using genetic variants is that it might exist a direct effect from the SNP to the outcome. This is known as horizontal pleiotropy.

Consider the following model with one instrument and no covariates for simplicity

$$Y = \mathrm{B}o + \alpha_k Z_k + \beta_1 \mathrm{X} + \Sigma \quad (15)$$

The reduced forms are then

$$X = \phi_{0k} + \phi_k Z_k + \varrho_k \tag{16}$$

$$Y = \delta_{0k} + \rho_k^* Z_k + \nu_k,\ \rho^* = \alpha_k + \beta_1 \phi_k \tag{17}$$

So that

$$\beta_1^* = \frac{\rho_k^*}{\phi_k} = \beta_1 + \frac{\alpha_k}{\phi_k} \tag{18}$$

And the overall IVW-estimate will tent towards:

$$\hat{\beta}_{1ivw}^* \approx \beta_1 + \frac{\sum_{k=1}^{K} w_k\ \alpha_j}{\sum_{k=1}^{K} w_k} = \beta_1 + Bias\ (\alpha, \phi) \tag{19}$$

Multiplicative random effects

$$\hat{\beta}_{1IVW_{MRE}} = \frac{\sum_{k=1}^{K} w_k^* \frac{\widehat{\rho_k^*}}{\widehat{\phi_k}}}{\sum_{k=1}^{K} w_k^*} \tag{20}$$

By the additional assumptions of the model (A20 and A21), the pleiotropic effects of each instrument, cancel each other, eliminating the bias of the estimator. The magnitude of the multiplicative and the fixed effects IVW models are thus the same.

In the multiplicative IVW model, weights are equal to the inverse of the ratio estimators' variance times a measure of the heterogeneity among those (ω).

$$w_k^*\_MRE = \frac{1}{\left(var \frac{\widehat{\rho_k}}{\widehat{\phi_k}}\right) * \omega} \tag{21}$$

With the following standard error:

$$se\left(\hat{\beta}_{IVW\_MRE}\right) = \sqrt{\frac{1}{\sum_{k=1}^{K}\left(var \frac{\widehat{\rho_k}}{\widehat{\phi_k}}\right)}} * \sqrt{\omega} \tag{22}$$

The practical estimation of B1 is done by fitting the following weighted OLS-regression equation using $\frac{1}{se^2(\widehat{\rho_k})}$ as weights. In this case, the $se^2(\widehat{\rho_k})$ is used as the empirical estimator of $\omega * \sigma_{\rho k}^2$. ω can be independently estimated as the mean standard error/deviation (MSE/MSD) of regression equation (24) .

$$\hat{\rho}_k = \beta_{1ivw}\hat{\phi}_k + \epsilon_k;\ \ \epsilon_k \sim N(0, \omega * \sigma^2_{\rho k}) \qquad (23)$$

The Egger estimator

The Egger method handle the case when the pleiotropic effects do not cancel out (directional unbalanced horizontal pleiotropy). The Egger estimator would yield a consistent estimate of β_1 even in the caser where all instruments were invalid instruments. The causal effect (β_{1E}) is estimated by a weighted regression equation using $\frac{1}{se^2(\widehat{\rho_k})}$ as weights. In this case, the empirical $se^2(\widehat{\rho_k})$ are consider to represent $\sigma^2_{\rho k}$.

$$\hat{\rho}_k = \beta_{0E} + \beta_{1E}\hat{\phi}_k + \epsilon_k;\ \ \epsilon_k \sim N(0, \sigma^2_{\rho k}) \qquad (24)$$

The Egger method also depends on the Inside condition (A21) i.e. $corr\ (\alpha_\kappa, \phi_k) = 0$

$$\widehat{\beta_{1E}} = \frac{cov\ (\widehat{\rho^*}, \hat{\phi})}{var\ (\hat{\phi})} = \hat{\beta}_1 + \frac{cov\ (\hat{\alpha}, \hat{\phi})}{var\ (\hat{\phi})} \text{ and } cov\ (\hat{\alpha}, \hat{\phi}) \xrightarrow{N\to\infty} cov\ (\alpha, \phi) \xrightarrow{K\to\infty} 0 \qquad (25)$$

With the following standard error:

$$se(\hat{\beta}_{1E}) = \frac{\omega}{\left(\sum_{k=1}^{K}\left(\hat{\phi}_k - \overline{\hat{\phi}}\right)^2\right)\sigma_{\rho k}^{-2}} \qquad (26)$$

$\overline{\hat{\phi}}$ is the weighted average SNP-exposure association using the inverse-variance weights $\frac{1}{\sigma^2_{\rho k}}$, $\frac{1}{se^2(\widehat{\rho_k})}$ in praxis. ω is, as above, estimated as residual standard error/deviation of regression equation (). It is included to adjust the standard errors of the Egger estimator for potential between-ratio heterogeneity.

Other metrics used in the analysis:

Formula	**Eq. number in main text**
$F = \hat{\phi}_k^2/\sigma_{\phi k}^2$	*(12)*
$\bar{F} = \frac{\sum_{k=1}^{K}(\widehat{\phi}_k^2/\sigma_{\phi k}^2)}{K}$	*(13)*
$\overline{F_w} = \frac{\sum_{k=1}^{K} w_k(\widehat{\phi}_k^2/\sigma_{\phi k}^2)}{\sum_{k=1}^{K} w_k}$, $w_k = \frac{\hat{\phi}_k^2}{\sigma_{\rho k}^2}$	*(14)*
$Q = \sum_{k=1}^{K} Q_k = \sum_{k=1}^{K} w_k \left(\hat{\beta}_{1k} - \hat{\beta}_{1IVW\_FE}\right) \sim \chi_{k-1}^2$	*(27)*
$\frac{(\bar{F}-1)}{\bar{F}}$ $\bar{\mathrm{F}}$ defined as in equation (13)	*(28)*
$\frac{\bar{F}_w-1}{\bar{F}_w}$, $\bar{\mathrm{F}}_\mathrm{w}$ defined as in equation (14)	*(29)*
$I_{GX}^2 = (Q_{GX} - (K\text{-}1))/\, Q_{GX}$	*(30)*
$Q_{GX} = \frac{\Sigma_{k=1}^{K}(\frac{\hat{\phi}_k}{\sigma_{\rho k}^2} - \overline{\hat{\phi}})^2}{\frac{\sigma_{\phi k}^2}{\sigma_{\rho k}^2}}$, $\overline{\hat{\phi}} = \frac{\sum_{k=1}^{K}\hat{\phi}_k/\sigma_{\phi k}^2}{\sum_{k=1}^{K}1/\sigma_{\phi k}^2}$	*(31)*

Appendix B. Overview of assumptions/conditions

References in main text

A1	The error term for individual i (ε_{mi}) has conditional mean zero given the explanatory/exposure variable: $E(\varepsilon_{mi}\,|X_{i,}\mathbf{M_i}) = 0$,which implies that $corr(X_i, \varepsilon_{mi}) = 0$	Multivariable OLS
A2	The outcome, exposure and covariate values of $(X_i, \mathbf{M_i}, Y_i)$are independent and identically distributed (i.i.d) drawn from their joint distribution.	Multivariable OLS
A3	Large outliers are unlikely: X_i, Y_i and $\mathbf{M_i}$ have nonzero finite fourth moments.	Multivariable OLS
A4	There is not perfect multicollinearity i.e. no one of the regressors ($X_{i,}\mathrm{M}_{pi}$) is a perfect linear function of the other regressors.	Multivariable OLS
A5	Instrument relevance: This condition is often presented as $corr(Z_{ki}, X_i) \neq 0$. A more precise definition (Wooldridge 2010) of this condition is that the coefficient of a regression/projection of X on all exogenous covariates and Z_k, is not zero i.e $\phi \neq 0$	IV/MR (1)
A6	Instrument exogeneity: Z_k is uncorrelated with the error term ε_{mi} i.e $corr(Z_{ki}\,, \varepsilon_{mi}) = 0$. This mathematical expression comprises the independence and the exclusion assumption (Angrist & Pischke; 2009) presented in appendix X	IV/MR (2,3)
A7	We assume that X are the only endogenous variable, while the covariates $\boldsymbol{M}$ are uncorrelated with the error term i.e. $E(\varepsilon_{mi}\,|\mathbf{M}_i) = (\varepsilon_{mi}\,|\mathrm{M}_{1i}, \ldots, \mathrm{M}_{pi}) = 0.$	IV/MR

A8	The values $(X_i, \mathrm{M}_{1i}, \dots, \mathrm{M}_{pi}, Z_i, Y_i)$ are independent and identically distributed (i.i.d) drawn from their joint distribution.	IV/MR
A9	Large outliers are unlikely: $(X_i, \mathrm{M}_{1i}, \dots, \mathrm{M}_{pi}, Z_i, Y_i)$ have nonzero finite fourth moments.	IV/MR
A10	Monotonicity condition requires that the instrument affect all individuals in the same direction, in our case it implies that the allele of reference either increases or decreases years of schooling for all individuals.	LATE effect
A11	Reduced-form equation 2 and 3 are assumed to satisfy the OLS conditions stated above This means among other things that instruments are chosen so that $\mathrm{Z_k}$ is not correlated with unmeasured confounders i.e $cov\,(Z_{ki}, \varrho_{mi}) = 0$ and $cov\,(Z_{ki}, \upsilon_{mi}) = 0$, i.e. the independence condition is satisfied.	Two sample MR
A12	$\varrho_{mi}\ and\ \upsilon_{mi}$ are independent of each other	Two sample MR
A13	The two datasets are drawn samples from the same population	Two sample MR
A14	The two samples are independent i.e. there is not overlap of sample observations.	Two sample MR
A15	The instrument-exposure effect is estimated with negligible error $(\sigma^2_{\phi k} \approx 0)$ so that $\hat{\phi} \approx \phi$ and can be therefore be treated as a constant. This assumption is known as No Measurement error in the exposure (NOME)	IVW, Egger, (median and median kan do it, but can also not do it)
A16	$\sigma^2_{\rho k}$ is the sample variance of $\widehat{\rho_k}$ and $\sigma^2_{\phi k}$ is the sample variance of $\widehat{\phi_k}$, although commonly estimated in the data as the standard errors of $\widehat{\rho_k}$ and $\widehat{\phi_k}$, are assumed to be known.	Two sample MR
A17	SNPs are uncorrelated	Two sample MR with summary data

A18	Single overall effect i.e. all of the incorporated genetic variables provide evidence of the same causal effects.	IVW Fixed effects, Multiplicative effect, Egger
A19	All instruments are valid. A consequence of this condition is that the fixed effects IVW model supposes that all variation between ratio estimates of is exclusively due to within-ratio/sample variation i.e. there is no between-ratio variation (Borestein et al. 2010).	IVW Fixed effects
A20	Balanced horizontal pleiotropy. The instrument specific bias term (horizontal pleiotropic effects) $\frac{\alpha_k}{\phi_k}$ in equation (16) cancel each other out so that the sum of the bias term in (17) is zero.	IVW multiplicative random effects
A21	Instrument Strength Independent of Direct Effect (InSide) : the direct effect of the instruments on healthcare costs are uncorrelated with the SNP-exposure effect i.e. $corr\ (\alpha_k, \phi_k) = 0$	IVW multiplicative, Egger
A22	Majority valid assumption. As the sample size increases, the causal estimates from all valid instrumental variables will tend towards the same value, which will equal the median provided that at least 50% of the instruments are valid	Median
A24	ZEMPA Zero Modal Pleiotropy Assumption or Plurality valid assumption. In large samples, the ratio estimates for all valid instruments should equal to the true causal effect, while ratio estimates from invalid instruments will disperse towards different values	Mode

Appendix C. Alternative presentation of the IV/MR assumptions

It is customary, in epidemiological literature, to present the relevance and exogeneity conditions (A5 and A6) with a slightly different terminology than the one used in my main text .I use this opportunity to add some informative remarks about these assumptions.

A. *Relevance assumption:* The instrument Z_k is associated with the exposure of interest X in a known direction (Smith & Hemani, 2014) some authors, however, sustain that there is a need for a causal relationship (Angrist & Pischke, 2015). Lawlor et al. (2008) points out that instruments used in many Mendelian Randomisation studies are not necessary causal but only correlated to the causal genetic variant affecting the exposure.

B. *Independence assumption:* The instrument Z_k is independent of the confounding factors (U) that confound the association of the exposure and the outcome (Lawlor et al., 2008) i.e the instrument being randomly assigned or as "good as randomly assigned" in the sense that it is unrelated to potential omitted variables/confounders (Angrist & Pischke,2015). This assumption is known as the *independence assumption* (Ibid), *exchangeability assumption*, *ignorable treatment assignment* or described as *no confounding for the effect of* Z_k *on Y (*Labrecque & Swanson 2018).

C. *Exclusion restriction:* The instrument Z_k is independent of the outcome Y given X and the confounding factors U (Lawlor et al., 2008). This is known as an exclusion restriction (Angrist & Pischke, 2015) and means that there is a single channel through which the instrument affects outcomes.

Appendix D. Leave-one-out analysis

Figure 8-11 present the original overall effect estimates with confidence intervals (in red), along the estimates and confidence intervals obtained by a series of iterations where SNP was left out at the time.

Figure 8 MR Egger

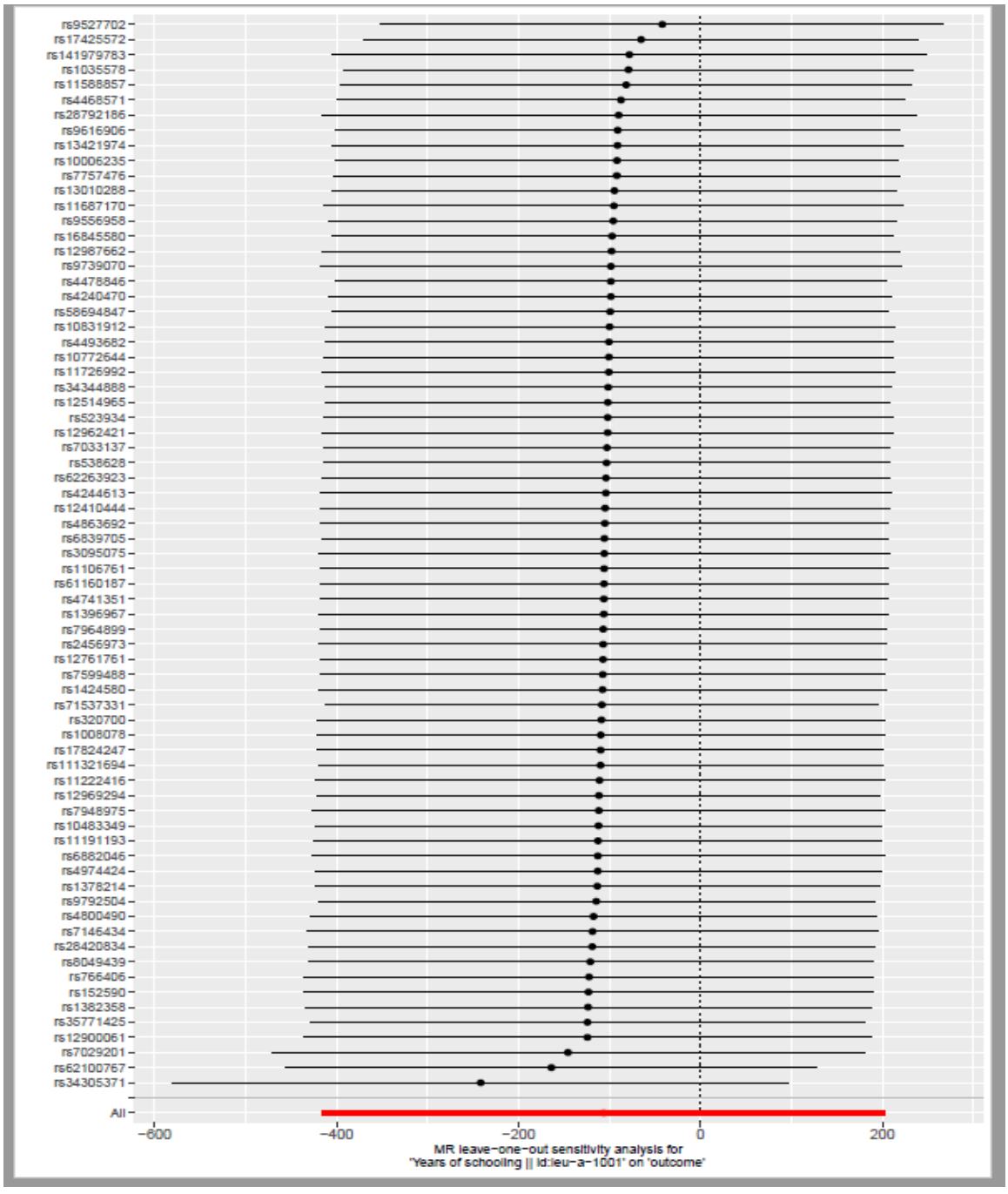


Figure 9 Weighted Median

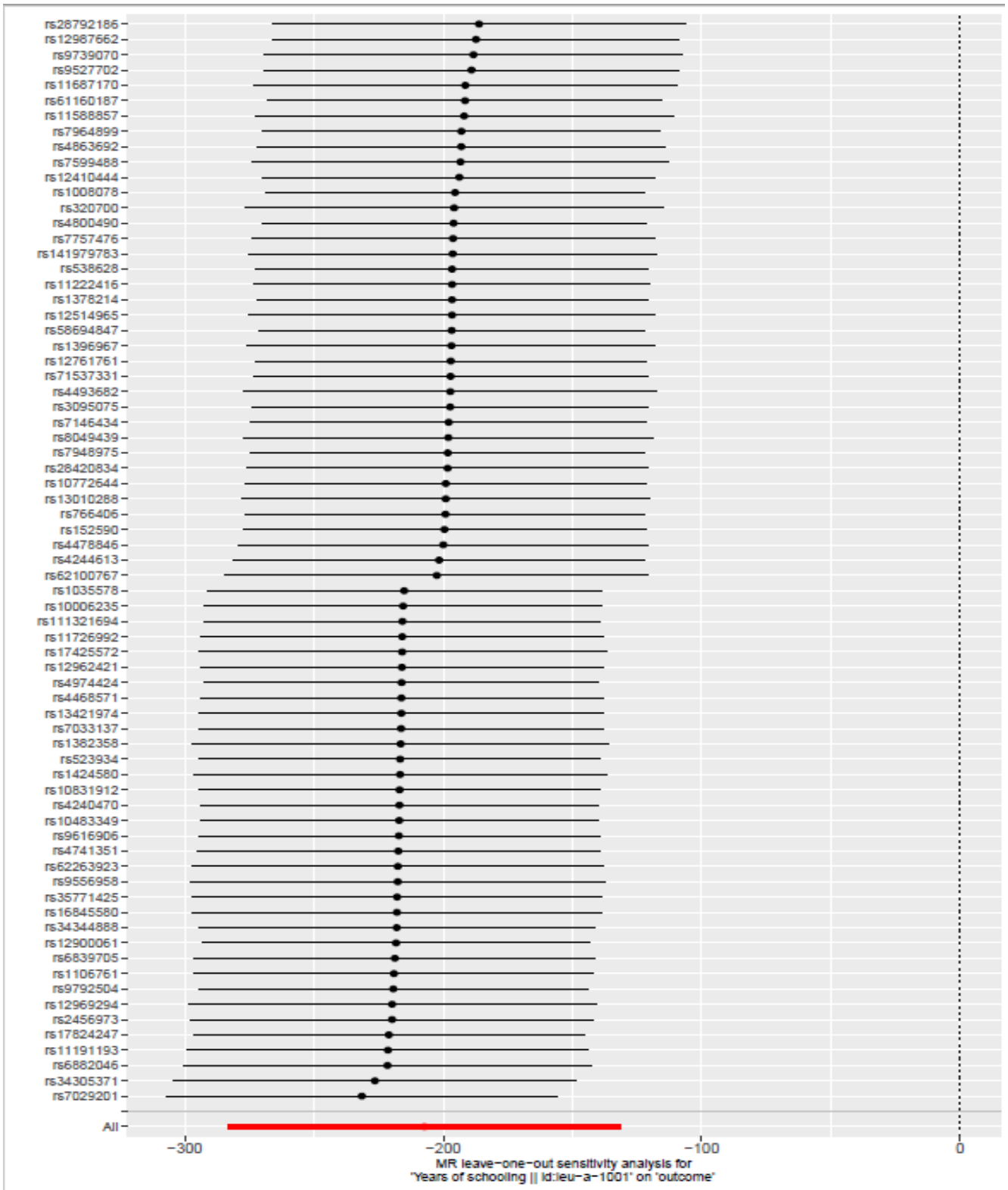


Figure 10 Weighted Mode

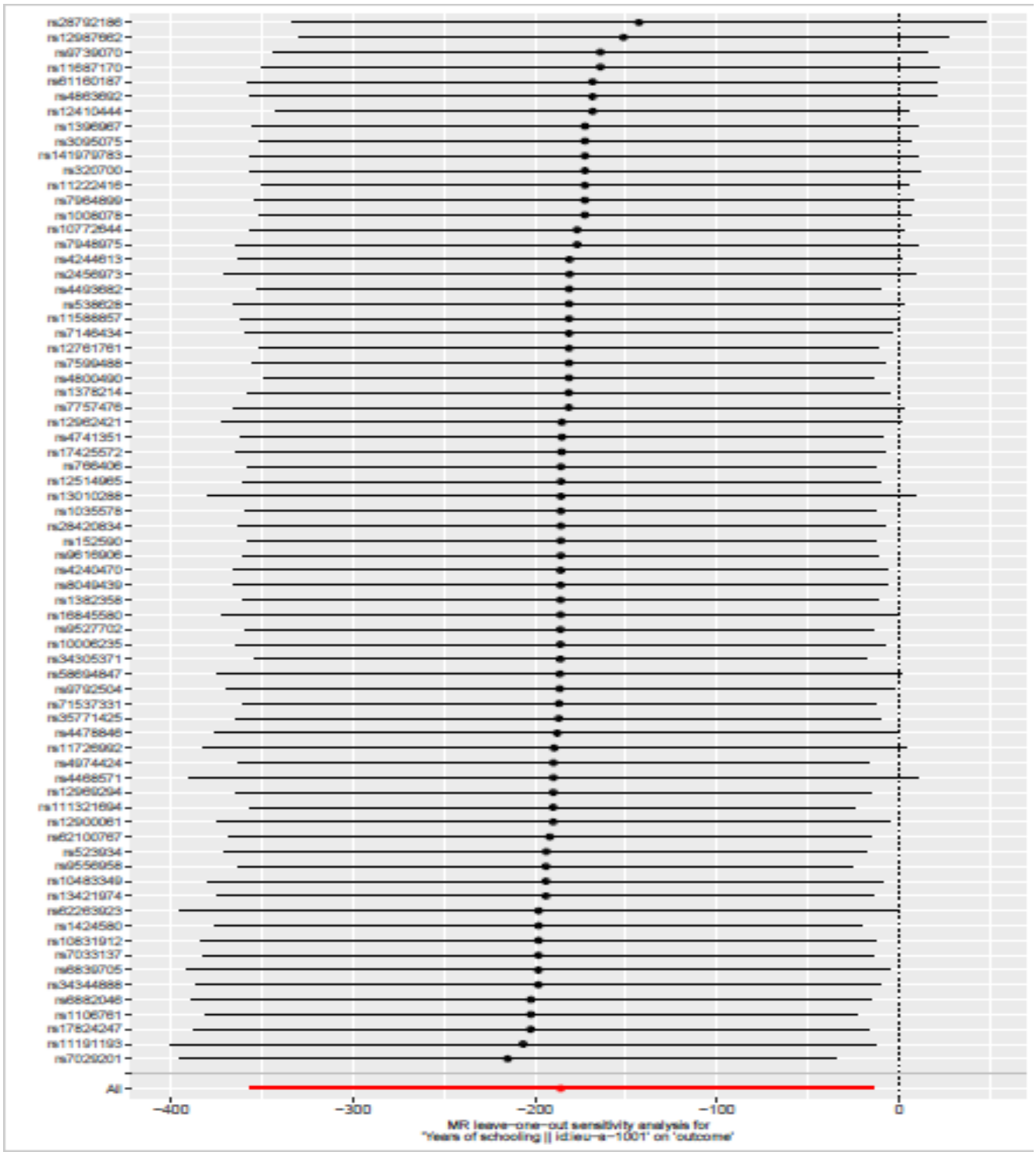


Figure 11 Simple mode