


RESEARCH ARTICLE OPEN ACCESS

Exploring and Accounting for Genetically Driven Effect Heterogeneity in Mendelian Randomization

Annika Jaitner¹  | Krasimira Tsaneva-Atanasova^{2,3} | Rachel M. Freathy¹ | Jack Bowden^{1,4}

¹Department of Clinical and Biomedical Sciences, Faculty of Health and Life Sciences, University of Exeter, Exeter, UK | ²Department of Mathematics and Statistics, Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK | ³EPSRC Hub for Quantitative Modelling in Healthcare University of Exeter, Exeter, UK | ⁴Novo Nordisk Research Centre Oxford, Oxford, UK

Correspondence: Annika Jaitner (a.jaitner@exeter.ac.uk)

Received: 9 May 2024 | **Revised:** 15 August 2024 | **Accepted:** 2 September 2024

Funding: This study was funded by the Wellcome Trust (WT088806, WT220390), US National Institutes of Health (R01 DK077659), Wellcome Trust (WT087997MA), EPSRC via grant EP/T017856/1, Wellcome Senior Research Fellowship (WT220390).

Keywords: ALSPAC | birth weight | causal inference | heterogeneity | mendelian randomization

ABSTRACT

Mendelian randomization (MR) is a framework to estimate the causal effect of a modifiable health exposure, drug target or pharmaceutical intervention on a downstream outcome by using genetic variants as instrumental variables. A crucial assumption allowing estimation of the average causal effect in MR, termed *homogeneity*, is that the causal effect does not vary across levels of any instrument used in the analysis. In contrast, the science of pharmacogenetics seeks to actively uncover and exploit genetically driven effect heterogeneity for the purposes of precision medicine. In this study, we consider a recently proposed method for performing pharmacogenetic analysis on observational data—the Triangulation Within a Study (TWIST) framework—and explore how it can be combined with traditional MR approaches to properly characterise average causal effects and genetically driven effect heterogeneity. We propose two new methods which not only estimate the genetically driven effect heterogeneity but also enable the estimation of a causal effect in the genetic group with and without the risk allele separately. Both methods utilise homogeneity-respecting and homogeneity-violating genetic variants and rely on a different set of assumptions. Using data from the ALSPAC study, we apply our new methods to estimate the causal effect of smoking before and during pregnancy on offspring birth weight in mothers whose genetics mean they find it (relatively) easier or harder to quit smoking.

1 | Introduction

Confirming or refuting causal relationships is difficult in observational study settings as one can never be sure if all confounders have been identified, appropriately measured and adjusted for. However, one can take advantage of random genetic inheritance from parents to offspring in an observational analysis to help uncover true causal mechanisms and estimate the causal effect of health interventions (Davey Smith, and Ebrahim 2003). Mendelian randomization (MR) is the formal science of using genetic variants as instrumental variables (IVs) for this purpose (Bowden

and Holmes 2019). Rather than testing the direct association between an exposure and outcome, a genetically predicted exposure is used instead. Under the assumption of random distribution of genetic variants from parents to offspring at conception, an individual's genetically predicted exposure should be far less susceptible to confounding bias. MR requires three core assumptions to hold for a genetic variant, G , to be valid instrument to test for a causal relationship between a modifiable exposure and health outcome (Lawlor et al. 2008). These are termed the relevance assumption, the independence assumption and the exclusion restriction. To go beyond testing for causality,

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Genetic Epidemiology* published by Wiley Periodicals LLC.

an additional assumption is required to estimate (or ‘point identify’) the causal effect. The most commonly used fourth assumption is homogeneity. It states that the causal effect an individual experiences is not affected by the value of their genetic instrument. When this is satisfied, an IV analysis can in theory estimate the average causal effect (ACE) of an intervention on the exposure for the entire study population. However, for continuous outcomes, this assumption is often biologically implausible unless a suitable ‘typical’ range for the exposure is defined (Hernán and Robins 2006). In cases where homogeneity is deemed implausible, an alternative assumption termed monotonicity can instead be applied to enable causal estimation (Bowden et al. 2021). In the context of an MR study using a genetic variant, G , monotonicity means that there is no individual whose exposure would be higher if they did not carry the exposure raising allele of G than if they did. Such individuals would be ‘Defiers’, and assuming that none exist allows the estimation of the causal effect in the subset of ‘Compliers’—defined as the group of individuals whose exposure level would always be greater with the exposure-raising allele of G than without.

Although homogeneity is typically invoked for interpretation of causal estimates in MR studies, in pharmacogenetic investigations genetic variants are explicitly sought to explain apparent heterogeneity in a treatment’s effectiveness. For example, many pharmaceutical interventions are pro-drugs, which require a specific metabolic process to occur for the patient to experience the full treatment effect. If the patient has a genetic variant that hinders the drug’s metabolism (e.g., a ‘Loss-of-function’ [LoF] mutation), the treatment effect may be less pronounced in individuals who carry it. For example, Pilling et al. (2021) showed that *CYP2C19* LoF alleles were associated with higher incidence of ischaemic events among those taking the commonly prescribed anti-stroke drug, Clopidogrel. National Institute of Health and Care Excellence guidance now recommends genotyping individuals on Clopidogrel who experience an ischaemic event, with a view to altering their medication if the LoF variant is found (Advisory Committee of NICE n.d.).

Observational data can be used to quantify the extent of genetically driven treatment effect heterogeneity, but the analysis can be compromised by strong confounding by indication and off-target genetic effects on the outcome of interest that are independent of any gene–drug interaction. A recently proposed method of pharmacogenetic causal inference using observational data—Triangulation Within a Study (TWIST) (Bowden et al. 2021)—defined the assumptions required to estimate the difference in treatment effect estimates between those with and without a pharmacogenetic variant, as a measure of genetically driven effect heterogeneity. A range of different methods were proposed to estimate this quantity as well as a framework for combining them if sufficiently similar. Although it is a useful tool for estimating this difference, in its most basic form it cannot estimate the causal effect of treatment on the outcome in each genetic group, which is a limitation.

Instances of genetically driven effect heterogeneity do exist in mainstream epidemiological investigations of nonpharmaceutical interventions. For example, smoking in pregnancy has been shown to have measurable consequences on offspring birth weight, which is an important marker of long-term health (Pereira et al. 2022). Specifically, Freathy et al. (2009) show that single-nucleotide

polymorphism (SNP) rs1051730 on chromosome 15 is associated with smoking cessation during pregnancy as well as smoking quantity. However, the same SNP is not associated with smoking initiation. Therefore, mothers with the rs1051730 risk allele are not more likely to smoke than mothers without, but if they do smoke they tend to smoke more heavily than non-carriers and find it harder to quit, meaning the effect of smoking on birth weight could easily be moderated by rs1051730.

In this study, we review the standard MR method, which utilises homogeneity-respecting genetic instruments, and the TWIST method, which utilises homogeneity-violating instruments. We highlight the different conceptual starting points for each approach, in terms of their modelling assumptions, and how estimates are biased if these assumptions are violated. Subsequently, we explore the integration of both sets of instruments into a unified analysis to properly characterise the ACEs and genetically driven effect heterogeneity. Using data from the ALSPAC study, we apply our new method to estimate the causal effect of smoking on offspring birth weight in distinct genetic subgroups of pregnant mothers; the magnitude of the effect heterogeneity; and the potential public health impact of genetically targeted treatment going forward.

2 | Methods

Let S and G be binary variables capturing the exposure and genetic variant of interest. In our applied example, S reflects the smoking status of the mother. We allow for the effect of the exposure on the outcome, Y , to be altered through an interaction with G , denoted as $S^* = G \times S$. To motivate the method, we assume the following linear interaction model for the mean outcome Y given S , G , and additionally measured (Z) and unmeasured (U) confounders of S and Y respectively:

$$\begin{aligned} E[Y|S, G, Z, U] &= \gamma_0 + \beta_1 SG + \beta_0 S(1 - G) + \gamma_{YG} G \\ &\quad + \gamma_{YZ} Z + \gamma_{YU} U \\ &= \gamma_0 + \beta_0 S + (\beta_1 - \beta_0) S^* + \gamma_{YG} G \\ &\quad + \gamma_{YZ} Z + \gamma_{YU} U \end{aligned} \quad (1)$$

Figure 1a depicts the directed acyclic graph consistent with the model described in Equation (1) and highlights various key assumptions using coloured arrows. We first consider the traditional set of assumptions required to estimate the ACE of the exposure on the outcome. We can express the ACE as the expected contrast between the potential outcomes of all mothers if they smoked during pregnancy, $Y(S = 1)$, and if they did not, $Y(S = 0)$:

$$ACE = E[Y(S = 1) - Y(S = 0)].$$

These assumptions are (Hernán and Robins 2020):

- IV1 (relevance) : The genetic instrument G predicts the exposure S (orange arrow);
- IV2 (independence) : The genetic instrument G is independent of any confounders U (no yellow arrow);

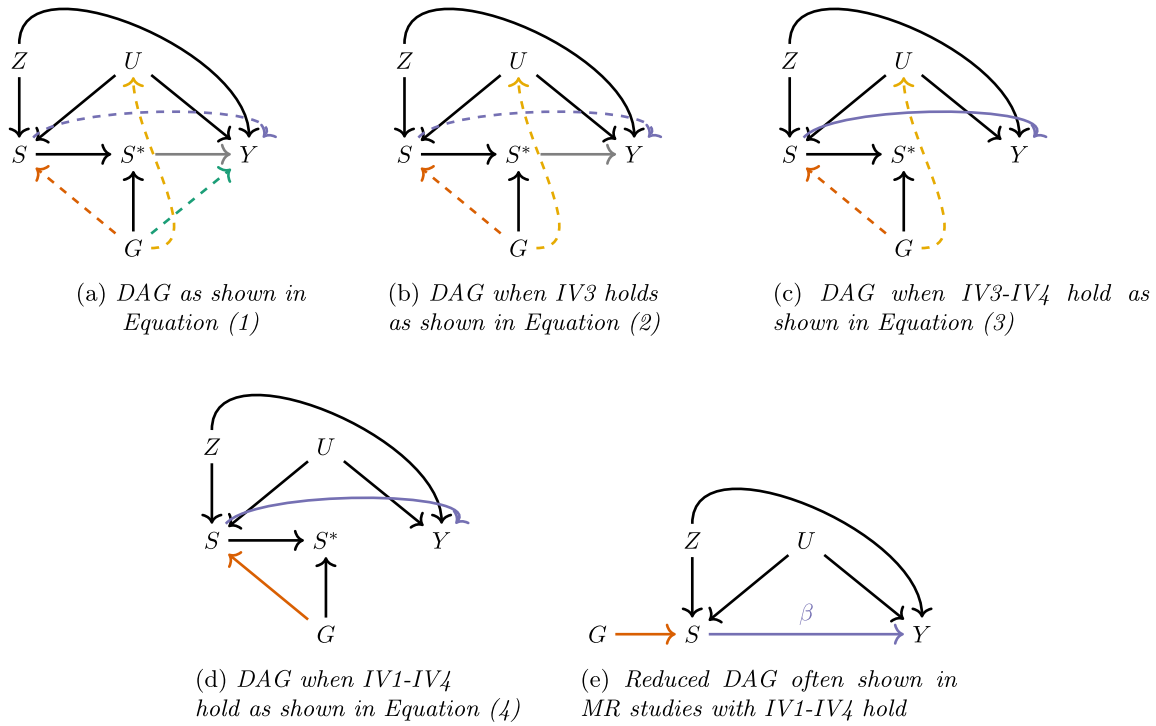


FIGURE 1 | Causal diagrams (a–e) illustrating key assumptions in the paper.

IV3 (exclusion) : The genetic instrument G is independent of the outcome Y given the exposure S and any confounders U (no green arrow).

IV4 (homogeneity) : The effect of the exposure S on the outcome Y is independent of the genetic instrument G (no grey arrow).

Assumptions IV1–4 enable us to extract the ACE via an IV analysis, by turning the general model and causal diagram in Figure 1 into the reduced model and causal diagram in Figure 1, through the following steps:

$$E[Y | S, G, Z, U] = \gamma_0 + \beta_0 S + (\beta_1 - \beta_0) S^* + \gamma_{YG} G + \gamma_{YZ} Z + \gamma_{YU} U,$$

$$\text{(from IV3)} = \gamma_0 + \beta_0 S + (\beta_1 - \beta_0) S^* + \gamma_{YZ} Z + \gamma_{YU} U, \quad (2)$$

$$\text{(from IV4)} = \gamma_0 + \beta S + \gamma_{YZ} Z + \gamma_{YU} U, \quad (3)$$

$$\text{(from IV1 and IV2)} = \gamma_0 + \beta \hat{S} + \gamma_{YZ} Z + \epsilon_Y, \quad (4)$$

where $\hat{S} = E[S|G]$ and ϵ_Y is a residual error term that is crucially independent of \hat{S} . The reduced causal diagram in Figure 1e is often shown in MR studies.

2.1 | What Does an MR Analysis Estimate Under Violation of IV2–4?

We now consider what is targeted by the Wald ratio estimate for the causal effect in an MR analysis, assuming the data model

described in Equation (1), when IV1 holds but initially, assumptions IV2–V4 do not. Note that we do not include the measured confounder Z in the following derivations as any bias through Z can be adjusted for. Under our assumed model as described in Equation (1), the Wald ratio estimate can be expressed as:

$$\begin{aligned} \frac{\text{Cov}(G, Y)}{\text{Cov}(G, S)} &= \frac{\beta_1 E[S | G = 1] - \beta_0 E[S | G = 0]}{E[S | G = 1] - E[S | G = 0]} \\ &+ \frac{\gamma_{YG} + \gamma_{YU}(E[U | G = 1] - E[U | G = 0])}{E[S | G = 1] - E[S | G = 0]} \quad (5) \\ &= \frac{\beta_1 E[S | G = 1] - \beta_0 E[S | G = 0]}{E[S | G = 1] - E[S | G = 0]} + B. \end{aligned}$$

IV1 guarantees that the denominator of Equation (5) is nonzero and so the ratio terms are well defined. If homogeneity is violated, but monotonicity holds, we show in Supporting Information that Equation (5) equals the Complier Average Causal Effect (CACE) plus any bias due to violation of IV2 and IV3 (B term). Compliers are defined as individuals that smoke if they have the risk allele ($G = 1$) and do not smoke if they do not have it ($G = 0$).

2.2 | Genetically Moderated Exposure Effect (GMEE)

Genetic instruments that satisfy the homogeneity assumption enable estimation of the ACE. However, in studies into the consequences of smoking versus not smoking, this assumption will be demonstrably false if attempting the analysis with a SNP like rs1051730, since the smoking patterns of people with and without this variant are likely to be different. In this case, a more practical starting point would be to assume the underlying DAG structure in Figure 1a and aim to quantify the magnitude of homogeneity violation as the difference in smoking effects

between the two genetic sub-groups. This ‘genetically moderated exposure effect’ (GMEE) is represented by arrow between S^* and the outcome Y . From Equation (1) this is equal to $\beta_1 - \beta_0$.

Bowden et al. (2021) discuss various methods for estimating this quantity, which we refer to as the GMEE, but which they referred to as the GMTE (T being for treatment). Each of the methods presented in Bowden et al. (2021) relies on a different set of assumptions. For example, when the genetic instrument G is independent of the exposure (i.e., no orange arrow in Figure 1 due to violation of IV1), is independent of any unmeasured confounder (i.e., no yellow arrow in Figure 1 and IV2 satisfied), and only affects the outcome through the moderated exposure variable (i.e., no green arrow in Figure 1 and IV3 satisfied), the GMEE can be estimated in the exposed population only. In our setting, this would be estimated by the difference in mean outcomes across the genetic groups among the population of smokers only:

$$GMEE(1) = \hat{E}[Y|S = 1, G = 1] - \hat{E}[Y|S = 1, G = 0].$$

Here the ‘(1)’ notation reminds the analyst that only smoker’s data is used and $G = 1/0$ refers to the presence/absence of at least one risk allele of SNP rs1051730. A more robust estimate of the GMEE is the

$$RGMEE = GMEE(1) - GMEE(0),$$

where

$$GMEE(0) = \hat{E}[Y|S = 0, G = 1] - \hat{E}[Y|S = 0, G = 0].$$

Here, the R prefix in RGMEE stands for ‘robust’, since it can estimate the GMEE without bias even if IV3 is violated (i.e., G affects the outcome directly as indicated by the green arrow in Figure 1). Indeed it is this bias term that is estimated by GMEE (0) before being subtracted out.

Bowden et al. (2021) state that the RGMEE is unbiased even if the genetic instrument violates IV2, by being associated with the outcome through the unmeasured confounder (yellow arrow in Figure 1). Our investigations in this paper have shown this to be incorrect (see Supporting Information: Section 6). Nevertheless, this actually makes it more straightforward to verify if the assumptions for the RGEE hold (i.e., a desired

violation of IV1 but no violation of IV2), since they imply that G and S are independent. Testing for an association between G and S is therefore an important prerequisite for its use.

3 | Enhancing Robustness Through the Integration of MR and GMEE Methods

The genetically moderated exposure effect introduced in the previous section proposes an array of methods for estimating the difference $\beta_1 - \beta_0$ under different assumptions, but not the individual values β_1 and β_0 . To address this, we now formally extend the previous framework by incorporating a second variant, G_2 , that is a ‘standard’ instrument for the exposure satisfying assumptions IV1–4. In our case, it therefore influences smoking initiation directly, but does not moderate an individual’s smoking habits, thereby violating homogeneity. We now explore two scenarios that expand upon the standard TWIST approach, utilising novel methods that leverage the two available genetic instruments. The DAGs for these two separate methods are shown in Figure 2.

3.1 | Method 1: (G, G_2) are Jointly Valid Instruments for (S, S^*)

We first consider estimation of β_1 and β_0 using genetic instruments G and G_2 within a multivariable model. This can be enacted in a two-step procedure by using G and G_2 to predict S in stage 1, and then plugging in the predicted values into the linear interaction model in stage 2:

1. Stage 1: Estimate \hat{S} with a consistent estimate of $\hat{E}[S|G, G_2]$;
2. Stage 2: $Y = \gamma_0 + \beta_1 \hat{S} G + \beta_0 \hat{S}(1 - G)$, with \hat{S} being the fitted values from stage 1.

This approach is robust to the case where the (assumed) effect modifying variant G violates IV1 but satisfies IV2 and IV3 (Figure 2, Left). In Supporting Information, we show through simulation that the true values of β_1 and β_0 can be recovered when these assumptions are satisfied, but violation of the assumptions lead to bias. The standard error for β_1 and β_0 can be obtained directly from the linear model output. As both parameters are estimated in the same model we can use the

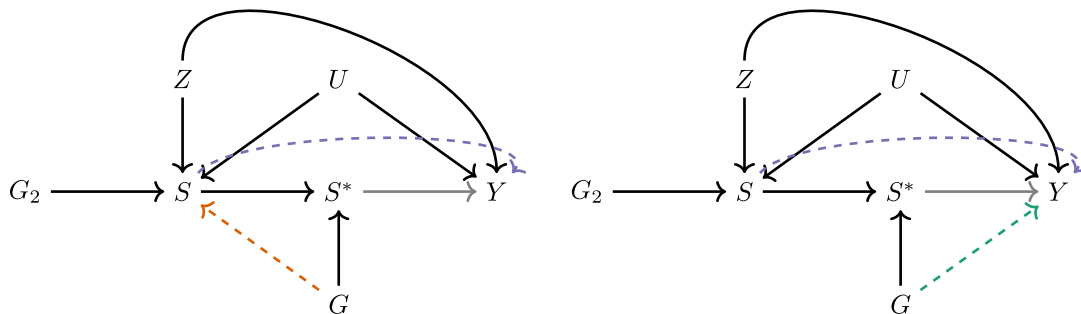


FIGURE 2 | Left: DAG of linear interaction model from Equation (1) highlighting in colour which assumptions need to hold for a consistent (i.e., asymptotically unbiased) estimate of β_1 and β_0 when using Method 1. Right: DAG of linear interaction model (1) highlighting in colour which assumptions need to hold for a consistent estimate of β_1 and β_0 when using Method 2.

covariance matrix of β_1 and β_0 to derive the variance of $\beta_1 - \beta_0$ and hence can estimate the standard error for $\beta_1 - \beta_0$.

3.2 | Method 2: Allowing for a Pleiotropic Effect of G on Y

We now propose a robust procedure that combines the general MR approach with the RGMEE given in Bowden et al. (2021). We first apply the RGMEE method to consistently estimate the genetically moderated effect $\beta_1 - \beta_0$. We then define a new variable $Y(S^* = 0)$ created by subtracting the genetically moderated effect times the moderated exposure from the original outcome Y . More formally, $Y(S^* = 0)$ is a potential outcome in which the treatment effect of S^* on Y has been set to zero. It is equal to Y (and therefore observed) for individuals with an $S^* = 0$, but is unobserved for those with $S^* = 1$. Finally, we perform an MR analysis using the genetic instrument G_2 , the exposure S and $Y(S^* = 0)$. This enables estimation of β_0 , which can then be used in combination with the RMGEE to estimate β_1 :

1. Estimate the RGMEE $\widehat{(\beta_1 - \beta_0)}$ using G ;
2. Estimate \hat{S} with a consistent estimate of $\hat{E}[S|G_2]$;
3. Estimate β_0 from model $E[Y(S^* = 0)|\hat{S}] = \gamma_0 + \beta_0\hat{S}$, where $Y(S^* = 0) = Y - \widehat{(\beta_1 - \beta_0)}S^*$.

Method 2 delivers consistent estimates if the RGMEE estimate can be consistently estimated using G and β_0 can be consistently determined using G_2 once the GMEE effect has been removed. Compared to Method 1, it allows a direct pleiotropic effect of G on Y (IV3 violation) but requires G to be independent of S (IV1 violated, but IV2 satisfied). When these assumptions are not met, our simulations show that it leads to bias (see Supporting Information). The standard error for β_0 and $\beta_1 - \beta_0$ can be directly taken from the respective model output. We make the assumption that $\widehat{(\beta_1 - \beta_0)}$ is independent of $\hat{\beta}_0$, so that $S.E(\beta_1) \approx \sqrt{Var(\widehat{(\beta_1 - \beta_0)}) + Var(\hat{\beta}_0)}$. In simulations we show that it leads to confidence intervals (CI) with only a slightly conservative coverage.

3.3 | What Does the Standard MR Estimate Using G_2 as the IV Target?

When including a homogeneity respecting instrument as shown in Figure 2, a standard MR analysis with G_2 as the IV is possible. Using the two-stage regression approach means:

1. Stage 1: Estimate \hat{S} with a consistent estimate of $\hat{E}[S|G_2]$;
2. Stage 2: $Y = \alpha_0 + \alpha_1\hat{S} + \alpha_2Z + \epsilon_Y$.

Here, α_1 is the ACE on the outcome Y if all mothers were exposed compared to if all mothers were not exposed: $E[Y(S = 1)] - E[Y(S = 0)]$. It can be shown that under the model described in Figure 2 and Equation (1):

$$\alpha_1 = \beta_0 + (\beta_1 - \beta_0)E[G|S = 1]. \quad (6)$$

4 | Simulation Results

4.1 | Data Generation

We simulated data consistent with Figure 2 in the following manner:

$$\begin{aligned} G &\sim \mathcal{B}(0.55), \\ G_2 &\sim \mathcal{B}(0.4), \\ U &= \gamma_{UG}G + \mathcal{N}(0, 1), \\ \eta &= -2 + \gamma_{SG}G + \gamma_{SG_2}G_2 + \gamma_{SU}U + \mathcal{N}(0, 0.5), \\ p_S &= \frac{\exp(\eta)}{1 + \exp(\eta)}, S \sim \mathcal{B}(p_S), \\ Y &= 3500 + \beta_1SG + \beta_0S(1 - G) + \gamma_{YG}G + \gamma_{YU}U \\ &\quad + \mathcal{N}(0, 470). \end{aligned} \quad (7)$$

The outcome (Y) model in (7) was chosen so that simulated data closely matched real birth weight data (in grams) for mothers with a history of smoking in the Avon Longitudinal Study of Parents and Children (ALSPAC) (Boyd et al. 2013; Fraser et al. 2013; Northstone et al. 2023) which we will subsequently use in our applied analysis. By choosing zero and nonzero values for the parameters γ_{UG} , γ_{SG} and γ_{YG} , we were able to explore the performance of Methods 1 and 2 in estimating the causal effect parameters β_1 and β_0 . We choose to set $\beta_1 = -200$ and $\beta_0 = -100$, which assumes a genetically moderated effect of $\beta_1 - \beta_0 = -100$ g. For all simulations, we made sure that the assumptions of Methods 1 and 2 held. Each simulation was repeated $N = 20,000$ times, which enabled the calculation of bias, coverage and statistical power. For further details, Supporting Information Table S2 provides a summary of the simulated data under all of the explored scenarios.

4.2 | Estimation Accuracy With Increasing Sample Size

We investigated the sample size needed to unbiasedly estimate β_1 and β_0 using each approach when their respective assumptions were satisfied. Data sets were generated with sample sizes between 100 and 80,000 individuals. Figure 3 shows the mean values of β_1 , β_0 and $\beta_1 - \beta_0$ using Methods 1 and 2 across 20,000 simulations. Shaded areas reflect, for each mean parameter estimate obtained from a given sample size, a 95% CI calculated as $\pm 1.96 \times SD(\cdot)$ around the mean, $SD(\cdot)$ being the standard deviation of the 20,000 estimates (Morris et al. 2019). Note that for each estimate we display three subfigures (per column) with a different range on the y-axis: one for small sample sizes, one for medium size sample sizes and one for large sample sizes. Details on the parameter values used for each simulation are described in Supporting Information. Estimation of β_1 and β_0 become more precise with shrinking CIs as the sample size increases. Both, Methods 1 and 2 lead to similar results. However, for small sample sizes (below 2000), CIs for Method 1 estimates are wider. The third column of Figure 3 shows the mean estimates for $\beta_1 - \beta_0$ using Methods 1 and 2. Here we can see a distinction in the performance of the methods across all

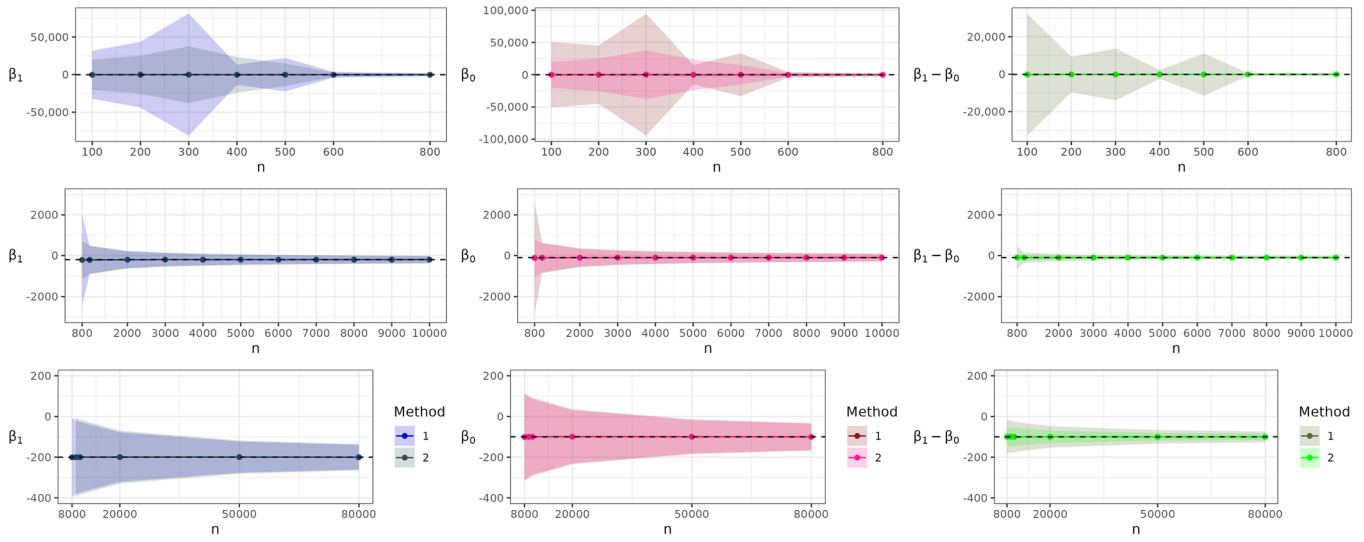


FIGURE 3 | Mean estimates for β_1, β_0 and $\beta_1 - \beta_0$ over 20,000 simulations for different sample sizes. The shaded area shows the 95% CI interval derived with the Monte Carlo standard error. The dashed line indicates the true value.

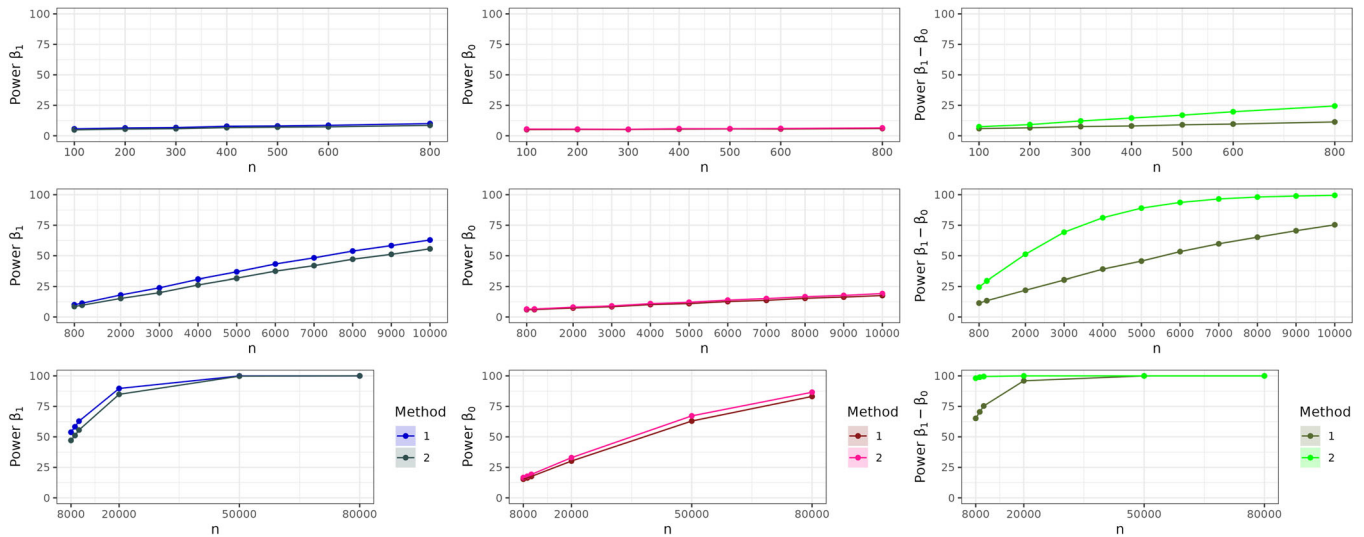


FIGURE 4 | Power for β_1, β_0 and $\beta_1 - \beta_0$ over 20,000 simulations for different sample sizes. The shaded area shows the 95% CI interval for the power derived with the Monte Carlo standard error.

sample sizes. Method 2, which uses the RGMEE method, yields narrower CIs for $\beta_1 - \beta_0$ even for small sample sizes.

4.3 | Power and the coverage

We estimated the power to reject the null hypothesis that β_1, β_0 and $\beta_1 - \beta_0$ were statistically different from zero at the 5% significance level, using Methods 1 and 2. For each simulation, we also calculated CIs for the parameter estimates based on estimated standard errors, and reported the coverage of 95% CIs across the 20,000 simulations. The results for power and coverage are shown in Figures 4 and 5 respectively, along with their Monte-Carlo standard errors (Morris et al. 2019). Our results show that Method 2 results in a higher power when estimating β_0 and $\beta_1 - \beta_0$ than Method 1 for a given sample size. However, when considering estimation of β_1 , this is

reversed. The power to detect β_0 is lower than the power to detect β_1 due to its lower effect size of -100 (compared to $\beta_1 = -200$). Figure 5 reveals a near nominal coverage for both methods close to 95%. Crucially, our assumption that $\beta_1 - \beta_0$ and $\hat{\beta}_0$ are independent leads only to a slightly conservative coverage when estimating a CIs for β_1 with Method 2.

5 | Applied Example

5.1 | Biological Example for Genetically Driven Exposure Effects

Research into the adverse consequences of smoking has been ongoing since the 1950s, up until the present day (Doll and Hill 1950; U.S. Department of Health and Human Services 2014). In the specific context of maternal health, it is well established

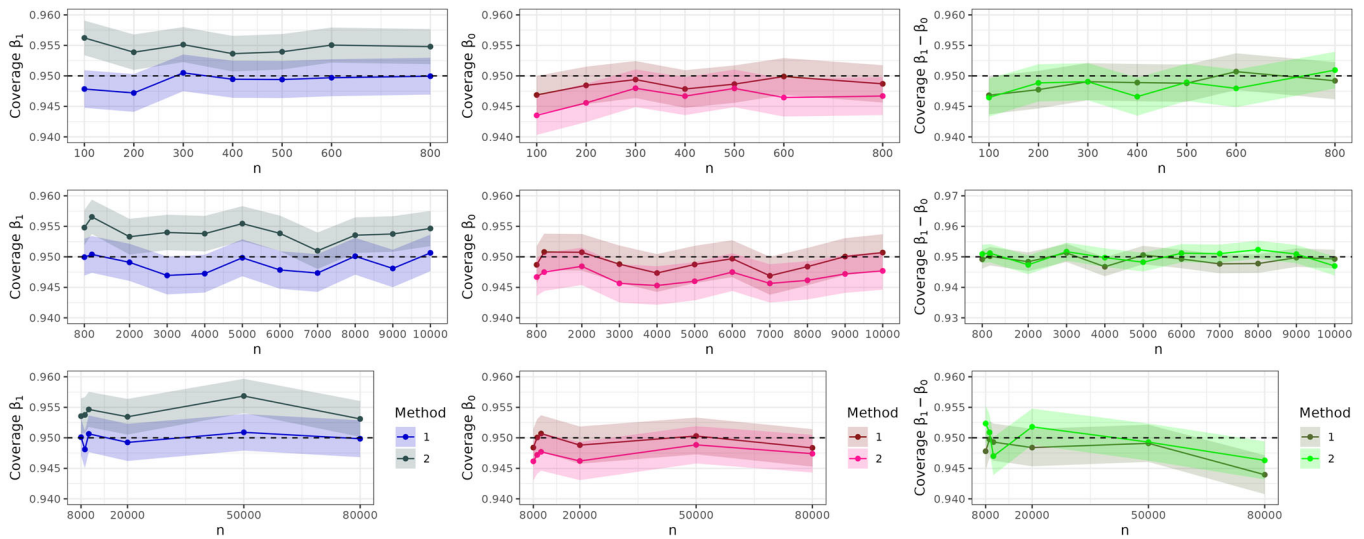


FIGURE 5 | Coverage for β_1 , β_0 and $\beta_1 - \beta_0$ over 20,000 simulations for different sample sizes. The shaded area shows the 95% CI interval for the coverage derived with the Monte Carlo standard error.

that smoking during pregnancy is associated with lower offspring birth weight, which is itself an important predictor of infant mortality and many later life health outcomes, such as cardiovascular disease, high blood pressure, coronary heart disease and type 2 diabetes (Moen et al. 2020; Tyrrell et al. 2012; Warrington et al. 2019). Attributing the correct proportion of these estimated associations that are due to the causal consequences of smoking is not straightforward, due to strong confounding between smoking and later life outcomes by socioeconomic factors which are very hard to completely control for. Despite this, smoking is viewed as a key modifiable risk factor, and reducing its prevalence during pregnancy remains an important public health target (Cnattingius 2004; U.S. Department of Health and Human Services 2014). Unfortunately, NHS digital service statistics indicate that approximately 8.6% of UK mothers were known smokers at the time of delivery in the first half of 2023 (Population Health, Clinical Audit, Team, Specialist Care, & Lead Analyst: Walt Treloar 2023). Identifying which individuals are at a higher risk of not giving up smoking and therefore might face more severe pregnancy outcomes can be crucial when targeting smoking cessation programmes, to provide support as well as closer monitoring during pregnancy.

Recently, genome-wide association studies (GWAS) have identified genetic variants that are associated with smoking initiation, smoking cessation, the age of starting smoking and smoking quantity (Liu et al. 2019). Freathy et al. (2009) show that rs1051730 on chromosome 15 is associated with smoking cessation during pregnancy as well as smoking quantity. A strong biological rationale for this exists as rs1051730 is in the nicotine acetylcholine receptor gene cluster *CHRNA5-CHRNA3-CHRNA4*. Rare variant burden associations have implicated all three of these genes as important in influencing smoking quantity (Rajagopal et al. 2023). However, it has also been shown that rs1051730 is *not* associated with smoking initiation (Freathy et al. 2009). The methods we have introduced thus far appear well suited to estimating the causal effect of smoking on birth weight using traditional genetic instruments for smoking initiation, whilst at the same time, quantifying the genetically moderated smoking effect via rs1051730.

5.2 | The Effect of Smoking on Birth Weight in the ALSPAC Study

The Avon Longitudinal Study of Parents and Children (ALSPAC) (Boyd et al. 2013; Fraser et al. 2013; Northstone et al. 2023) invited pregnant women resident in Avon, UK with expected dates of delivery between 1 April 1991 and 31 December 1992, to take part in the study. The initial number of pregnancies enrolled was 14,541. Of the initial pregnancies, there was a total of 14,676 fetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year. We restricted our analysis to unrelated mothers with available genetic information. Additionally, we excluded multiple births and preterm births (pregnancy duration ≤ 37 weeks) (Jaitner et al. 2024). The analysis data set had a sample size of 7752 individual mothers. For the traditional genetic instrument ‘ G_2 ’ we created a weighted genetic risk score (GRS) among the smoking initiation SNPs identified by the latest GWAS (Liu et al. 2019). The effect sizes from the same GWAS were used as weights. We used rs1051730 as genetic effect-modifying instrument ‘ G ’, coded as 0 and 1 corresponding to no and at least one risk allele respectively. Various different smoking definitions were used for the exposure outlined in the following sections. The ALSPAC study website contains details of all the data that are available through a fully searchable data dictionary and variable search tool (<http://www.bristol.ac.uk/alspac/researchers/our-data/>).

5.2.1 | Exposure S is Smoking Before Pregnancy

Each mother was asked at 16–18 weeks of gestation whether she smoked before pregnancy. We coded mothers that reported ‘yes’ as $S = 1$ and mothers who reported ‘no’ as $S = 0$. Figure 6a displays the assumed DAG for our analysis. We aimed to apply Methods 1 and 2 to estimate the causal effect of pre-pregnancy smoking on birth weight in the $G = 1$ group, β_1 , the $G=0$ group, β_0 , and also the genetically moderated exposure effect $\beta_1 - \beta_0$. We would expect this latter quantity to be nonzero if the pre-pregnancy smoking effect persisted

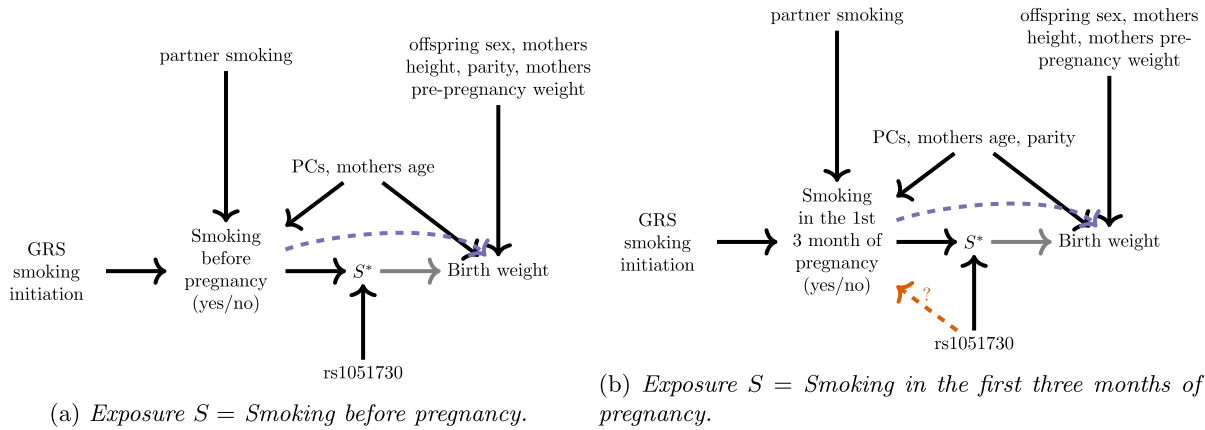


FIGURE 6 | DAG shows the relationships between the genetic instrument, the exposure, the outcome and how those are affected by different confounders. $S^* = G \times S$ represents the interaction between the genetic variant rs1051730 and the exposure S .

TABLE 1 | Logistic regression model results for $\text{logit}(Pr(S = 1)) = \alpha_0 + \alpha_1 G + \alpha_2 G_2 + \delta^T Z$ as the first stage of Method 1 using pre-pregnancy smoking as the outcome variable S . $G_2 =$ smoking initiation GRS, $G =$ rs1051730 and Z is a vector of covariates.

$S =$		Est.	Std. Error	p -value	F statistic
Smoking before pregnancy	rs1051730 (G)	$\hat{\alpha}_1 = 0.02$	0.06	0.73	0.12
	GRS smoking initiation (G_2)	$\hat{\alpha}_2 = 1.46$	0.16	4.85e-20	84
Smoking in the first 3 months	rs1051730 (G)	$\hat{\alpha}_1 = 0.13$	0.07	0.06	3.55
	GRS smoking initiation (G_2)	$\hat{\alpha}_2 = 1.39$	0.18	2.58e-15	62.6

differently throughout pregnancy across the two genetic groups. For the first stage of Method 1, we perform a logistic regression of S on the GRS of smoking initiation (G_2) and rs1051730 (G). The results are shown in Table 1. Variant rs1051730 was not associated with smoking before pregnancy, which helpfully means that Method 2 is not ruled out as an analysis option. The GRS is also associated with smoking before pregnancy and we assume it acts as a true IV for this exposure. Two crucial assumptions are that the GRS of smoking initiation has no pleiotropic effect on birth weight and it does not modify the causal effect between smoking and birth weight in the exposed and the unexposed. To apply Method 1, rs1051730 cannot have a pleiotropic effect on birth weight either but, for Method 2, this assumption is relaxed. The results from applying both methods are shown in Figure 7. To increase the precision of the estimates we adjust our regression models for different sets of covariates. The model for the genetic prediction of smoking is adjusted for whether the partner of the mother smoked, the mothers' age and the first 10 genetic principal components. The model predicting birth weight is adjusted for offspring sex, mother age, mothers' height, parity, mothers' prepregnancy weight and the first 10 genetic principal components. We viewed these variables as confounders for either smoking before pregnancy or birth weight or both. For mothers who have at least one G risk allele, the ACE of smoking before pregnancy, β_1 , was estimated to be a 168 and 169 g reduction in birth weight using Methods 1 and 2 respectively. On the other hand, the corresponding causal effect (β_0) in smoking mothers without a rs1051730 risk allele is a 159 and 161 g birth weight reduction for Methods 1 and 2 respectively compared to nonsmoking mothers without the risk allele.

5.2.2 | Exposure S is Smoking in the First 3 Months of Pregnancy

Mothers were asked at 16–18 weeks of gestation whether they smoked in the first 3 months of pregnancy. We coded those that reported 'yes' as $S = 1$ and mothers who reported 'no' as $S = 0$ and proceeded to estimate the causal effect of this exposure at the two levels (no and at least one risk allele) of rs1051730 and the genetically moderated exposure effect. In this analysis, no information about smoking before pregnancy was taken into account and therefore the mothers reporting to be non-smokers either gave up smoking when getting pregnant or did not smoke before pregnancy. Figure 6b displays the assumed DAG for our analysis. Methods 1 and 2 were applied to derive estimates for β_1 , β_0 and $\beta_1 - \beta_0$, with results shown in Table 1 and Figure 7. For these analyses, the logistic model for S given G and G_2 was adjusted for whether the partner of the mother smokes, the mothers' age, parity and the first 10 genetic principal components. The model predicting birth weight was adjusted for offspring sex, mothers' age, mothers' height, parity, mothers' prepregnancy weight and the first 10 genetic principal components. We viewed these variables as confounders for either smoking before pregnancy or birth weight or both. For this analysis, a stronger association was observed between rs1051730 and S , meaning that we were cautious with the interpretation of Method 2 results, given its crucial role in the estimation of the genetically moderated exposure effect. The causal effect of smoking during pregnancy on birth weight in those with a rs1051730 risk allele, β_1 , was estimated to be -212 and -222 g using Methods 1 and 2 respectively, whereas the effect of smoking in the individuals without the genetic variant on birth weight is -220 and -205 g for Method 1 and 2 respectively.

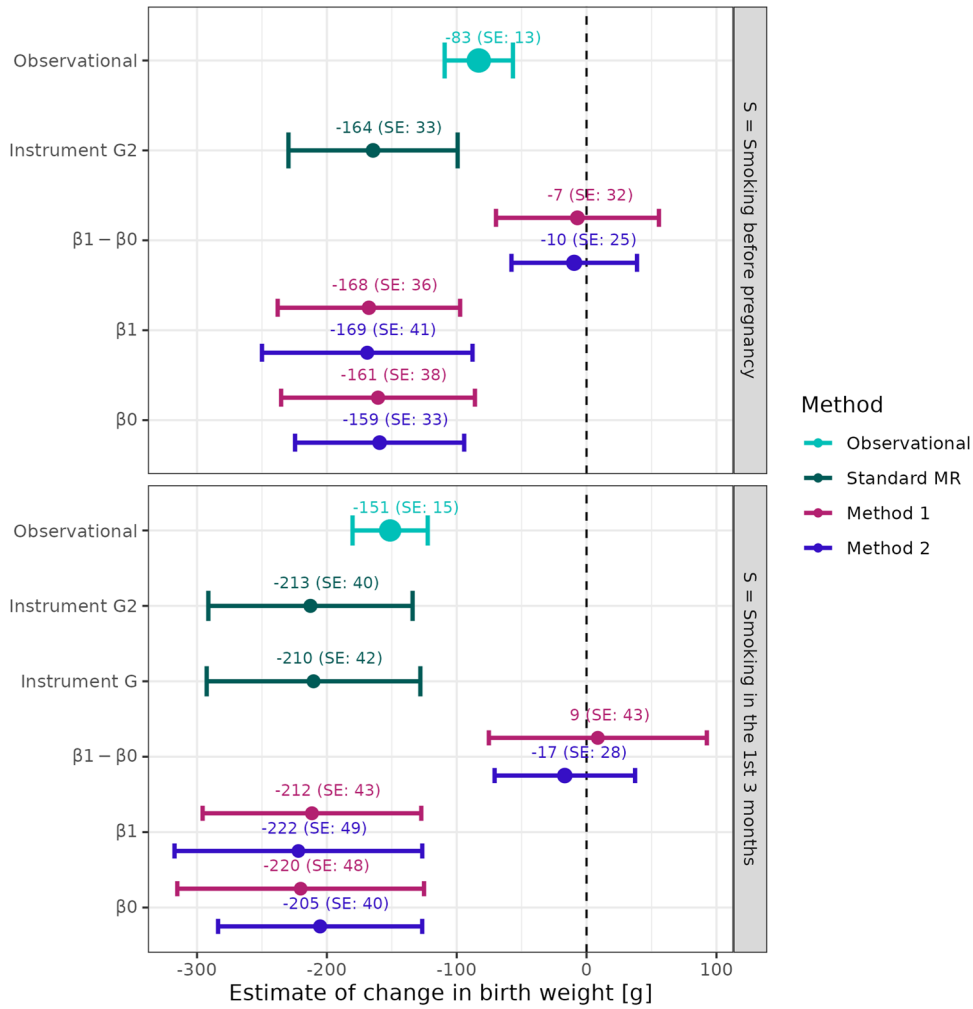


FIGURE 7 | $\beta_1 - \beta_0$, β_1 and β_0 estimates from applying Methods 1 and 2 to the ALSPAC data set. Bars indicate the 95% confidence interval. Observational analysis results arise from a linear regression of the smoking variable on birth weight adjusted for partner smoking, mothers' age, mothers' height, mothers' pre-pregnancy weight, parity, and offspring sex. Dark green are the results from the individual level data MR analysis using the two-stage least squares method using G or G_2 as the instrument, S as the exposure and birth weight is the outcome. G has no association with $S =$ smoking before pregnancy, hence we did not perform an MR analysis. The results for applying Methods 1 and 2 to estimate $\beta_1 - \beta_0$, β_1 and β_0 are shown in red and blue respectively.

For both smoking exposures, we were unable to identify a difference between the two genetic groups. This could be because the effect of pre-pregnancy smoking and smoking in the first 3 months does not truly differ in people with or without variant rs1051730. However, a simulation investigation showed that large numbers of individuals would be required to identify a genetically driven effect of smoking with the magnitudes we observed in our analysis. Specifically, we simulated data with sample sizes from 7000 to 500,000 and a true difference between the genetic groups of $\beta_1 - \beta_0$ between -20 and -5 g. To reach 80% power in estimating $\beta_1 - \beta_0 = -15$ a sample size of 500,000 individuals is required in our simulation when using Method 1. The RGMEE (Method 2), which is more robust to pleiotropy compared to Method 1, is able to estimate $\beta_1 - \beta_0 = -15$ with a power of over 80% with 200,000 individuals. More details on the results of this are shown in Supporting Information: Section 5. This provides important guidance on the much larger sample size, way beyond the 7752 mothers in the ALSPAC study, that would be required to detect a difference between the genetic groups in the region of what we observe here.

Despite not being able to detect a statistically meaningful genetically moderated exposure effect, overall our results suggest that smoking before pregnancy or smoking in the first 3 months of pregnancy results in a lower birth weight compared to not smoking. This is in line with previous publications (Larsen et al. 2018; Tyrrell et al. 012; Pereira et al. 2022).

5.3 | Observational Analysis and 'Standard' MR

In addition to applying the new proposed methods to the ALSPAC data set, we also looked at the observational association between smoking and birth weight. A linear regression of S (using both smoking definitions) on birth weight adjusted for partner smoking, mothers' age, mothers' height, mothers' pre-pregnancy weight, parity and offspring sex yielded negative associational estimates. Although the observational analysis likely suffers from residual confounding, and cannot be interpreted as a causal effect, the direction of effect remains the

same compared to estimating β_1 and β_0 (Figure 7), albeit of a smaller magnitude.

As indicated in Table 1, rs1051730 (G) is not associated with smoking before pregnancy. Therefore, we are unable to perform a standard MR analysis using G as the genetic instrument for $S =$ smoking before pregnancy. However, we did perform a standard MR analysis using individual level data and the two-stage least squares approach with $S =$ smoking in the first 3 months of pregnancy. We explain in Section 2.1 that the standard MR with a homogeneity violating instrument like rs1051730 estimates the CACE (while the monotonicity assumption holds). The CACE of smoking in the first 3 months of pregnancy on birth weight is -210 g (95% CI: $[-293, -128]$). Note that these results potentially suffer from weak instrument bias.

Additionally, and as a confirmatory test of our previous genetic analyses, we calculated a standard MR estimate using the GRS for smoking initiation to instrument smoking before and smoking in the first 3 months of pregnancy. For these analyses rs1051730 is not considered. The methodology is described in Section 3.3. Using Equation (6), the estimated β -values obtained from applying Method 1 and 2 to the ALSPAC data we derive an ACE of -165 g for smoking before pregnancy. This compares to the ACE of -164 g (Figure 7) estimated with standard MR approach. Similarly, for smoking in the first 3 months of pregnancy we obtain an estimate of -213 g using the β -values from Methods 1 or 2 outputs and the formula provided in Equation (6).

6 | Discussion

In this study, we propose a general framework for MR that allows the inclusion of traditional genetic instruments, as well as those that violate the key homogeneity assumption. This enables an analysis that goes beyond estimation of the ACE to consider estimation within specific genetic sub-groups, with a view to quantifying genetically driven effect heterogeneity. Our approach builds on ideas from the pharmacogenetic TWIST framework proposed by Bowden et al. (2021) to a more mainstream epidemiological setting, as well as incorporating the technique of multivariable MR (Sanderson et al. 2019). Specifically, Method 1 offers a new approach to estimating the genetically moderated exposure effect, which could be triangulated with existing methods. Furthermore, Method 1 allows for a direct effect between the homogeneity-violating instrument and the exposure. In the presence of unmeasured confounding, the methods proposed in the paper by Bowden et al. (2021) require the instrument to be independent of the exposure. To allow for a direct pleiotropic effect between the homogeneity-violating instrument and the outcome we proposed Method 2. Simulation studies revealed the necessary sample sizes to detect an effect with sufficient power under Methods 1 and 2, considering plausible genetically moderated exposure effect sizes. To motivate the methods, we applied them to data from the ALSPAC cohort to investigate the effect of smoking before and in pregnancy on birth weight using a traditional GRS for smoking and rs1051730 as an effect modifier.

Our work could be further extended by considering the incorporation of additional methods to allow for the relaxation of

further key assumptions. For example, allowing a genetic variant with a pleiotropic effect that acts through an unmeasured confounder (i.e., correlated pleiotropy). We assumed that the underlying data structure follows a linear interaction model. Future work could explore different data structures and non-linear effects. For simplicity, and to naturally follow on from the approach proposed in Bowden et al. (2021), we assumed a binary effect modifying instrument through the dichotomisation of the genetic instrument rather than using the number of risk alleles. Future work could relax this assumption.

Despite not being able to show a difference between the two genetic groups in our applied example due to a limitation in sample size, our investigation clarifies how large future cohort study samples need to be to estimate effects of the magnitude we observed. We believe our framework is a useful methodological extension to investigate genetically driven heterogeneity. Our methods could, for example, be applied in other setting where larger sample sizes are available, or by meta-analysing results with additional cohorts. Settings outside of pregnancy research are also possible and would not require mother and child pair information. For example, investigating the genetically driven effect of continuing smoking on lung cancer. Other examples could be using genetic variants associated with reducing alcohol consumption and the effects on various health outcomes. Data sets like the UK Biobank with genetic information available for 500,000 individuals could be used. R code for implementing the methods can be found at https://github.com/AJaitner/paper_heterogeneity_MR as well as code to implement the simulation studies and applied analysis.

Acknowledgements

The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and Annika Jaitner and Jack Bowden will serve as guarantors for the contents of this study.

Genotyping of the ALSPAC maternal samples were funded by the Wellcome Trust (WT088806). Specific funds for recent detailed data collection on the mothers were obtained from the US National Institutes of Health (R01 DK077659) and Wellcome Trust (WT087997MA) for completion of selected items of obstetric data extraction, including placental weights. A comprehensive list of grants funding is available on the ALSPAC website (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>).

Annika Jaitner received funding for her PhD studentship from the Faculty of Health and Life Sciences at the University of Exeter.

Jack Bowden is funded by research grant MR/X011372/1.

Krasimira Tsaneva-Atanasova gratefully acknowledges the financial support of the EPSRC via grant EP/T017856/1.

Rachel M. Freathy is supported by a Wellcome Senior Research Fellowship (WT220390).

This project utilised high-performance computing funded by the UK Medical Research Council (MRC) Clinical Research Infrastructure Initiative (award number MR/M008924/1). This study was supported by the National Institute for Health and Care Research Exeter Biomedical Research Centre. The views expressed are those of the authors and not

necessarily those of the NIHR or the Department of Health and Social Care.

This research was funded in part, by the Wellcome Trust (Grant number: WT220390). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

The authors would like to acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in carrying out this work.

We would also like to thank Xiaoran Liang for her feedback and useful discussion which improved the quality of this work.

Ethics Statement

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time.

Conflicts of Interest

Jack Bowden is a part-time employee of Novo Nordisk, however, this work is unrelated to his role at the company. The remaining authors declare no conflicts of interest.

Data Availability Statement

The data in ALSPAC is fully available, via managed systems, to any researchers. The managed system is a requirement of the study funders, but access is not restricted on the basis of overlap with other applications to use the data or on the basis of peer review of the proposed science. The ALSPAC data management plan describes in detail the policy regarding data sharing, which is through a system of managed open access. The following steps highlight how to apply for access to the data included in this paper and all other ALSPAC data. (1) Please read the ALSPAC access policy, which describes the process of accessing the data and samples in detail and outlines the costs associated with doing so. (2) You may also find it useful to browse the fully searchable ALSPAC research proposals database, which lists all research projects that have been approved since April 2011. (3) Please submit your research proposal for consideration by the ALSPAC Executive Committee. You will receive a response within 10 working days to advise you whether your proposal has been approved. If you have any questions about accessing data, please email alspac-data@bristol.ac.uk.

References

Advisory committee of NICE. n.d. "NICE Recommendations: Clopidogrel Genotype Testing After Ischaemic Stroke or Transient Ischaemic Attack." Accessed April 26, 2024. <https://www.nice.org.uk/consultations/1710/2/recommendations>.

Bowden, J., and M. V. Holmes. 2019. "Meta-Analysis and Mendelian Randomization: A Review." *Research Synthesis Methods* 10, no. 4: 486–496. <https://doi.org/10.1002/jrsm.v10.4>.

Bowden, J., L. Pilling, D. Türkmen, C. L. Kuo, and D. Melzer. 2021. "The Triangulation Within a STudy (TWIST) Framework for Causal Inference Within Pharmacogenetic Research." *PLoS Genetics* 17, no. 9: e1009783. <https://doi.org/10.1371/journal.pgen.1009783>.

Boyd, A., J. Golding, J. Macleod, et al. 2013. "Cohort Profile: The 'Children of the 90s'—The Index Offspring of the Avon Longitudinal Study of Parents and Children." *International Journal of Epidemiology* 42, no. 1: 111. <https://doi.org/10.1093/ije/dys064>.

Cnattingius, S. 2004. "The Epidemiology of Smoking During Pregnancy: Smoking Prevalence, Maternal Characteristics, and Pregnancy Outcomes." *Nicotine and Tobacco Research* 6, no. Suppl. 2: S125–S140. <https://doi.org/10.1080/14622200410001669187>.

Davey Smith, G., and S. Ebrahim. 2003. "Mendelian Randomization: Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease?" *International Journal of Epidemiology* 32, no. 1: 1–22. <https://doi.org/10.1093/ije/dyg070>.

Doll, R., and A. B. Hill. 1950. "Smoking and Carcinoma of the Lung." *BMJ* 2: 4682. <https://doi.org/10.1136/bmj.2.4682.739>.

Fraser, A., C. Macdonald-wallis, K. Tilling, et al. 2013. "Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC Mothers Cohort." *International Journal of Epidemiology* 42, no. 1: 97–110. <https://doi.org/10.1093/ije/dys066>.

Freathy, R. M., S. M. Ring, B. Shields, et al. 2009. "A Common Genetic Variant in the 15q24 Nicotinic Acetylcholine Receptor Gene Cluster (CHRNA5-CHRNA3-CHRNA4) Is Associated With a Reduced Ability of Women to Quit Smoking in Pregnancy." *Human Molecular Genetics* 18, no. 15: 2922–2927. <https://doi.org/10.1093/hmg/ddp216>.

Hernán, M., and J. Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

Hernán, M. A., and J. M. Robins. 2006. "Instruments for Causal Inference: An Epidemiologist's Dream?" *Epidemiology* 17, no. 4: 360–372. <https://doi.org/10.1097/01.ede.0000222409.00878.37>.

Jaitner, A., M. Vaudel, K. Tsaneva-Atanasova, et al. 2024. "Smoking During Pregnancy and Its Effect on Placental Weight: A Mendelian Randomization Study." *BMC Pregnancy and Childbirth* 24, no. 1: 238.

Larsen, S., C. Haavaldsen, E. K. Bjelland, J. Dypvik, A. M. Jukic, and A. Eskild. 2018. "Placental Weight and Birthweight: The Relations With Number of Daily Cigarettes and Smoking Cessation in Pregnancy. A Population Study." *International Journal of Epidemiology* 47, no. 4: 1141–1150. <https://doi.org/10.1093/ije/dyy110>.

Lawlor, D. A., R. M. Harbord, J. A. Sterne, N. Timpson, and G. DaveySmith. 2008. "Mendelian Randomization: Using Genes As Instruments for Making Causal Inferences in Epidemiology." *Statistics in Medicine* 27, no. 8: 1133–1163. <https://doi.org/10.1002/sim.3034>.

Liu, M., Y. Jiang, R. Wedow, et al. 2019. "Association Studies of up to 1.2 Million Individuals Yield New Insights Into the Genetic Etiology of Tobacco and Alcohol Use." *Nature Genetics* 51, no. 2: 237–244. <https://doi.org/10.1038/s41588-018-0307-5>.

Moen, G. H., B. Brumpton, C. Willer, et al. 2020. "Mendelian Randomization Study of Maternal Influences on Birthweight and Future Cardiometabolic Risk in the HUNT Cohort." *Nature Communications* 11, no. 1: 5404. <https://doi.org/10.1038/s41467-020-19257-z>.

Morris, T. P., I. R. White, and M. J. Crowther. 2019. "Using Simulation Studies to Evaluate Statistical Methods." *Statistics in Medicine* 38, no. 11: 2074–2102. <https://doi.org/10.1002/SIM.8086>.

U.S. Department of Health and Human Services. 2014. *The Health Consequences of Smoking: 50 Years of Progress: A Report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Northstone, K., Y. BenShlomo, A. Teyhan, et al. 2023. "The Avon Longitudinal Study of Parents and Children Alspac G0 Partners: A Cohort Profile [version 2; peer review: 1 approved]." *Wellcome Open Research* 8: 37. <https://doi.org/10.12688/wellcomeopenres.18782.2>.

Pereira, R. D., C. A. Rietveld, and H. van Kippersluis. 2022. "The Interplay Between Maternal Smoking and Genes in Offspring Birth Weight." *Journal of Human Resources* 59, no. 4: 1020–11266R2. <https://doi.org/10.3368/jhr.1020-11266R2>.

Pilling, L. C., D. Türkmen, H. Fullalove, et al. 2021. "Analysis of Cyp2c19 Genetic Variants With Ischaemic Events in UK Patients Prescribed Clopidogrel in Primary Care: A Retrospective Cohort Study." *BMJ Open* 11, no. 12: e053905. <https://doi.org/10.1136/bmjopen-2021-053905>.

Population Health, Clinical Audit, Team., Specialist Care, and Lead Analyst: Walt Treloar. 2023. Statistics on Women's Smoking Status at Time of Delivery: England, Quarter 3 2022-23. <https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-womens-smoking-status-at-time-of-delivery-england/statistics-on-womens-smoking-status-at-time-of-delivery-england-quarter-3-2022-23>.

Rajagopal, V. M., K. Watanabe, J. Mbatchou, et al. 2023. "Rare Coding Variants in CHRN2 Reduce the Likelihood of Smoking." *Nature Genetics* 55, no. 7: 1138–1148. <https://doi.org/10.1038/s41588-023-01417-8>.

Sanderson, E., G. DaveySmith, F. Windmeijer, and J. Bowden. 2019. "An Examination of Multivariable Mendelian Randomization in the Single-Sample and Two-Sample Summary Data Settings." *International Journal of Epidemiology* 48, no. 3: 713–727. <https://doi.org/10.1093/ije/dyy262>.

Tyrrell, J., V. Huikari, J. T. Christie, et al. 2012. "Genetic Variation in the 15q25 Nicotinic Acetylcholine Receptor Gene Cluster (CHRNA5-CHRNA3-CHRN4) Interacts With Maternal Selfreported Smoking Status During Pregnancy to Influence Birth Weight." *Human Molecular Genetics* 21, no. 24: 5344–5358. <https://doi.org/10.1093/hmg/dd372>.

Warrington, N. M., R. N. Beaumont, M. Horikoshi, et al. 2019. "Maternal and Fetal Genetic Effects on Birth Weight and Their Relevance to Cardio-Metabolic Risk Factors." *Nature Genetics* 51, no. 5: 804–814. <https://doi.org/10.1038/s41588-019-0403-1>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.