Mendelian Randomization with longitudinal exposure data: simulation study and real data application

Janne Pott[1], Marco Palma[1], Yi Liu[2], Jasmine A. Mack[3], Ulla Sovio[3], Gordon C. S. Smith[3], Jessica Barrett[1], Stephen Burgess[1]

[1] MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
[2] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[3] Department of Obstetrics and Gynaecology, University of Cambridge, NIHR Cambridge Biomedical Research Centre, Cambridge, UK

Corresponding author: Janne Pott (janne.pott@mrc-bsu.cam.ac.uk)

Abstract

Background and aim: Mendelian Randomization (MR) is a widely used tool to estimate causal effects using genetic variants as instrumental variables. MR is limited to cross-sectional summary statistics of different samples and time points to analyse time-varying effects. We aimed at using longitudinal summary statistics for an exposure in a multivariable MR setting and validating the effect estimates for the mean, slope and within-individual variability.

Simulation study: We tested our approach in various scenarios for power and type I error, depending on the correlation structure between the mean, slope and variability, and regression model. We observed high power to detect causal effects of the mean and slope throughout the simulation, but the variability effect was low powered in case of correlation between the mean and variability. Mis-specified regression models led to lower power and increased the type I error.

Real data application: We applied our approach to two real data sets (POPS, UK Biobank). We detected significant causal estimates for both the mean and the slope in both cases, but no effect of the variability. However, even in the UK Biobank we only had weak instruments.

Conclusion: We used a new approach to test a time-varying exposure for causal effects of the exposure's mean, slope and variability. The simulation with strong instruments seems promising, but also highlights three crucial points: 1) the difficulty to define the correct exposure regression model, 2) the dependency on the genetic correlation, and 3) the lack of

strong instruments in real data. Taken together, this demands a cautious evaluation of the results, taking into account known biology and trajectory of the exposure.

Introduction

Mendelian Randomization (MR) allows to estimate the causal effect of an exposure X on an outcome Y using genetic variants (SNPs) as instrumental variables (IVs), as long as certain assumptions are met[1]. The SNPs must be strong predictors of X, while being independent of observed and unobserved confounders U of X and Y. In addition, the effect of the SNPs on the outcome must solely be mediated by X, e.g. no direct effect or other pathway from the SNP to Y. As this method only requires summary statistics from genome-wide association studies (GWAS)[1], it is widely used and applied to various genetically regulated exposures and disease outcomes.

So far, the majority of MR analyses have used data from cross-sectional studies, due to the availability of large-scale GWAS summary statistics from the UK Biobank (e.g. Neale Lab[2]), FinnGen[3], or meta-analyses of other cohort data. Most GWAS studies have tested for SNP effect on the mean level of the trait using linear mixed models as implemented in GMMAT[4] or REGENIE[5]. Others have tested the sample variance for SNP effects, using for example Levene's test[6–8]. These variance quantitative trait loci (vQTLs) can then be tested as candidates for gene-environment interactions. Age is one possible environmental factor for interactions, as age acts as a proxy for the different biological processes during different phases of life. As a result, causal effects might also change over time[9,10]. So far, GWAS summary statistics from different time points have been used in a multivariable setting to estimate the conditional effects of each time point on the outcome of interest[9,10]. Doing so has nonetheless some short-comings. On the one hand, the identification of the most critical time point with the strongest effect of exposure on outcome will not necessarily be detected in the multivariable MR (MVMR) approach[9]. On the other hand, finding GWAS summary statistics of the same exposure with similar size and ancestry is still challenging.

In the recent years, more follow-up data in large biobanks from re-invitations or electronic health records (EHR) have become available. Accordingly, novel GWAS methods to analyse large scale longitudinal data have been developed[11–13]. They allow estimation of the main SNP effect as well as for the time-dependent effect of the SNP on the slope of the trajectory [11,13] and/or the variability of the exposure (within-subject variability)[12].

In this work, we bring together methods to estimate SNP associations in longitudinal data in the presence of growth over time, and combine them in the multivariable Mendelian Randomization framework to estimate the causal effects of the mean of an exposure (intercept of the trajectory), of the change of the exposure over time (slope of the trajectory), and of the within-subject variability (standard deviation of the exposure per individual). A directed acyclic graph of this concept is given in Figure 1.

The manuscript is organized as follows: the regression methods used in this work are introduced in Section 2, followed by an evaluation of the proposed method in simulation studies in Section 3. We then apply the method to real data from the Pregnancy Outcome Prediction Study (POPS) and from the UK Biobank (UKB) in Section 4, and discuss the results, draw conclusions, and point out limitations in Section 5.

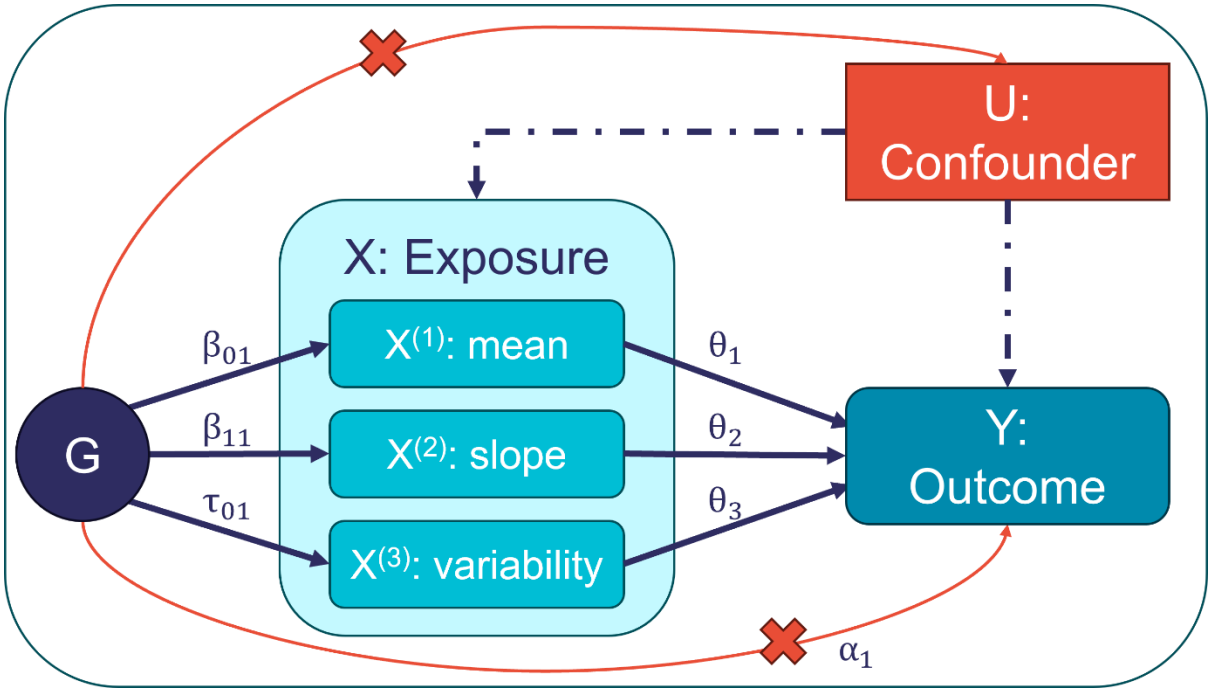

Figure 1: Directed acyclic graph of the presented MVMR analysis. The genetic variant G has an effect on the mean level of the exposure X, denoted as $X^{(1)}$, on the time-dependent growth, $X^{(2)}$, and on the within-subject variability, $X^{(3)}$. The SNP effects β_{01}, β_{11}, and τ_{01} can be obtained from a GAMLSS analysis. Each exposure type $X^{(i)}$ can have an independent causal effect θ_i on the outcome Y.

Framework for longitudinal data in MVMR

Linear mixed models

Genome-wide association studies (GWAS) test millions of genetic variants for association with a specific trait or disease[14]. For example, the UK Biobank includes the data of approximately 500,000 individuals and allows well-powered GWAS of hundreds of traits. However, the main focus of these GWAS has been on cross-sectional data gathered at time of recruitment. GWAS on longitudinal phenotypes have the advantage that they can identify genetic variants associated with the development of a trait over time using linear mixed models (LMM)[11]. In more detail, a mixed model for a longitudinally measured phenotype X including both random intercept and random slope effects has the following form:

$$X_{ij} = C_{ij}\beta + Z_{ij}b_i + \varepsilon_{ij}, \qquad i = 1, \dots, n, \;\; j = 1, \dots, k_i$$
$$b_i \sim N(0, D)$$
$$\varepsilon_{ij} \sim N(0, \sigma^2)$$
$$b_1, \dots, b_n, \varepsilon_{11}, \dots, \varepsilon_{nk_n} \text{ independent.}$$

Here, X_{ij} is the value of phenotype X for individual i at observation j (with j from $1, \dots, k_i$), C_{ij} is a p-dimensional vector with all p predictors at observation j, and Z_{ij} is a 2-dimensional

vector with 1 at the first entry and t_{ij}, the age of individual i at observation j, in the second entry. β is the p-dimensional vector of regression coefficients for the predictors (fixed effects for all individuals), b_i is a 2-dimensional vector of the random intercept, b_{0i}, and random slope, b_{1i}, per individual i, and ε_{ij} is the measurement error for individual i at observation j. The matrix D and σ^2 denote the variance-covariance matrix of the random effects and the variance of the measurement error, respectively. In a GWAS, the linear mixed model for one SNP will have the form:

$$\begin{aligned} X_{ij} &= C_{ij}\beta + Z_{ij}b_i + \varepsilon_{ij} \\ &= \beta_{00} + \beta_{01}G_i + \beta_{10}t_{ij} + \beta_{11}G_it_{ij} + C^*_{ij}\beta^{COV} + b_{0i} + b_{1i}t_{ij} + \varepsilon_{ij} \\ &= \beta_{00} + b_{0i} + \beta_{01}G_i + C^*_{ij}\beta^{COV} + (\beta_{10} + b_{1i} + \beta_{11}G_i)t_{ij} + \varepsilon_{ij} \\ &= \text{intercept} + \text{slope} \times t_{ij} + \varepsilon_{ij} \end{aligned}$$

in which G_i is the genotype of individual i of the tested SNP G (constant over time), and C^*_{ij} is the vector of additional covariates, either constant over time or time-varying, e.g. biological sex or weight, respectively. This model is implemented in common tools such as the R-package lme4[15], GMMAT (Generalized Linear Mixed Model Association Tests)[4,13] and the GALLOP algorithm (Genome-wide Analysis of Large-scale Longitudinal Outcome using Penalization)[11]. While lme4 performs well for a handful of SNPs, GALLOP allows fast genome-wide calculation of the fixed SNP effects by 1) estimating the mixed model without genetics, and 2) reusing the variance-covariance information for each SNP. As result, we obtain the main SNP effect, β_{01}, and the SNP x time interaction effect, β_{11}, from a single model. GMMAT additionally allows to model the random effects from a genetic relationship matrix (GRM).

Generalized additive models for location, scale, and shape

Testing for a SNP x time interaction term makes sense for longitudinal traits with a clear trend, e.g. growth from childhood to adulthood, or steady decline of a biomarker such as eGFR in patients with chronic kidney disease. However, traits can also be time-independent or normalized to adjust for time. For example, foetal weight is usually standardized to gestational age adjusted scores or percentiles, as it can be more clinically relevant to know if the infant's weight is in the lowest 10th percentile than the absolute value. In these traits without a slope, it is unlikely to observe a SNP x time interaction. However, as well as affecting the mean of the trait, the SNP might also affect the within-subject variability, which cannot be tested in a LMM as described above.

In this case, Generalized Additive Models for Location, Scale and Shape[16] (GAMLSS) offer an alternative, as they are distributional regression models, in which the explanatory variables C might affect all parameters of the distribution of X, e.g. the mean (location), variance (scale), skewness, kurtosis, or quantiles. This work will focus on estimating two parameters, representing the mean, μ, and the variance, σ^2, of the normally-distributed phenotype X, using the R package gamlss[17] . The corresponding mixed model can be written as

$$\begin{aligned} X_{ij} &\sim N(\mu_{ij}, \sigma^2_{ij}) \\ g^{\mu}(\mu_{ij}) &= C_{ij}\beta + Z_{ij}b_i \\ g^{\sigma}(\sigma_{ij}) &= W_{ij}\tau + v_i \end{aligned}$$

where C_{ij} and W_{ij} contain the explanatory variables associated with the mean and the standard deviation, v_i represents the random intercept on the variability, while the link functions $g^{\mu}(\cdot)$ and $g^{\sigma}(\cdot)$ are taken to be the identity and log function, respectively. This represents a more general version of the LMM, as σ is no longer constant for all individuals (allowing for heteroscedasticity).

Ko et al.[12] developed a GWAS implementation, TrajGWAS, to estimate both the mean and within-subject variability using longitudinal data. It can be re-written as

$$\begin{aligned} \mu_{ij} &= \beta_{00} + b_{0i} + \beta_{01}G_i + (\beta_{10} + b_{1i} + \beta_{11}G_i)t_{ij} + C_{ij}^{*}\beta^{COV} \\ ln(\sigma_{ij}) &= exp(W_{ij}\tau + v_i) \\ &= exp(\tau_{00} + v_{0i} + \tau_{01}G_i + W_{ij}^{*}\tau^{COV}) \end{aligned}$$

in which μ_{ij} is defined as before in the LMM, and for σ the fixed effects are denoted by τ, and the random intercept on the variability is given by $v_i = v_{0i}$. The vector W_{ij}^{*} contains additional covariates. As a result, we obtain the main SNP effect, β_{01}, the SNP x time interaction effect, β_{11}, and the SNP effect on the variability, τ_{01}. So far, TrajGWAS has only been developed for unrelated samples, e.g. no inclusion of the GRM has been possible.

Multivariable Mendelian Randomization (MVMR)

In Mendelian Randomization (MR), genetic variants G associated with an exposure X are used as instrumental variables (IVs) to make inferences about the causal effect θ of X on an outcome Y[1(p20)]. In more detail, assuming

(1) G strongly affects X (e.g., regression coefficient $\hat{\beta}_{GX} \neq 0$),
(2) G is independent of any confounders U ($G \perp\!\!\!\perp U$), and
(3) G affects Y only via its effect on X ($Y \perp\!\!\!\perp G|(X, U)$),

the ratio of the genetic regression coefficients on Y and X, β_{GY} and β_{GX}, is a robust and unbiased estimate of the true causal effect θ_1, as full mediation can be assumed[18]:

$$\begin{aligned} \mathbb{E}(X|G = g, U = u) &= \beta_0 + \beta_1 g + \beta_2 u \\ \mathbb{E}(Y|do(X = x_0), U = u) &= \theta_0 + \theta_1 x_0 + \theta_2 u \\ \mathbb{E}(\beta_1) = \hat{\beta}_{GX} \text{ and } \mathbb{E}(\beta_1\theta_1) = \hat{\beta}_{GY} &\Rightarrow \hat{\theta}_1 = \frac{\hat{\beta}_{GY}}{\hat{\beta}_{GX}} \end{aligned}$$

Here, $do(X = x_0)$ means forcing the continuous exposure X on level x_0. It indicates a causal expression, as in general $\mathbb{E}\big(Y\big|do(X = x_0)\big) \neq \mathbb{E}(Y|X = x_0)$ ('correlation is not causation'). This approach has been extended to include multiple instruments[19], e.g. using inverse-weighted variance (IVW) method to meta-analyse the single ratios, and into a multivariable framework to simultaneously estimate the independent effects of multiple exposures[1,20], namely the multivariable Mendelian Randomization (MVMR). In a 2-sample MVMR approach, the genetic estimates for the exposures and the outcome are obtained from independent GWAS summary statistics. The causal estimates per exposure can be retrieved from a linear regression of l exposures[1]:

$$\hat{\beta}_{GY} = \theta_1 \times \hat{\beta}_{GX^{(1)}} + \cdots + \theta_l \times \hat{\beta}_{GX^{(l)}} + \varepsilon$$

As in all MR analyses, three main assumptions must still be met for MVMR[20]:

(1) Relevance: the exposures $X^{(i)}$ must be strongly predicted by the SNPs given the other exposures included in the model
(2) Independence: the SNPs must be independent of all confounders U of any of the exposures $X^{(i)}$ and the outcome Y
(3) Exclusion restriction: the SNPs must be independent of the outcome Y given all of the exposures $X^{(i)}$ included in the model

In this work, we use the classic MVMR approach with the genetic effects obtained from the GAMLSS as described in the subsection above, which ensures that assumption 1 is met, as the effects are estimated together. The other two assumptions are met as long as they were met in the univariate case. Using this framework and estimates from the GAMLSS model we have

$$\begin{aligned}\hat{\beta}_{GY} &= \theta_1\hat{\beta}_{GX^{(1)}} + \theta_2\hat{\beta}_{GX^{(2)}} + \theta_3\hat{\beta}_{GX^{(3)}} + \varepsilon \\ &= \theta_1\hat{\beta}_{01} \quad + \theta_2\hat{\beta}_{11} \quad + \theta_3\tau_{01} \quad + \varepsilon,\end{aligned}$$

with $X^{(1)}, X^{(2)}$, and $X^{(3)}$ denoting the mean, slope and variability of the exposure X, respectively. While the MVMR regression model can detect unbiased causal estimates for the mean and the variability, the estimate for the slope will be on a different scale: the SNP effect on the slope of the exposure will be given by time units (e.g. SNP effect per year or week), while for the outcome, we only obtain the main SNP effect, which is not scaled for time. Hence, the regression model will estimate $\hat{\theta}_2$ as $\theta_2 \times T$, with T representing the mean (calendar) age in the outcome population. If T is known, it is possible to rescale the time to obtain a more interpretable estimate (e.g. causal effect per year).

Simulation study

Overview

We performed a simulation study with four scenarios to evaluate the proposed method combining longitudinal SNP effect estimates in a MVMR approach. These scenarios comprised:

A. Genetic correlation between mean and slope effects (“independent variability”)
    1. Full GAMLSS, estimating all three SNP effects (denoted as “main”)
    2. Reduced GAMLSS, estimating only the mean and variability SNP effect (denoted as "no-slope”)
B. Genetic correlation between mean, slope and variability effects (“full correlation”)
    1. Main
    2. No Slope

In the following, we describe our simulation approach in general within the ADEMP framework established in Morris et al.[21]. The scenario-specific parameters and their respective values can be found in Supplemental Table S1. Additional technical notes on bias correction can be found in the Supplemental Material.

Simulation setup

Aim

The aim of our simulation study was to investigate different settings for 2-sample MVMR where the exposure is a multilevel model of exposure progression. We wanted to validate that the MVMR correctly identifies the causal effects when using instrument effect estimates obtained from general additive models.

Data generating mechanisms

For all scenarios, we simulated $n = 10{,}000$ samples with $k = 15$ observations per sample and $m = 30$ SNPs.

Generating SNPs

All SNPs were simulated using random sampling, with genotype distributions in Hardy-Weinberg equilibrium (2 alleles, minor allele frequency $p_{MA} = 0.25$). In each simulation, we tested the SNPs for pairwise correlation, but detected no SNPs in linkage disequilibrium (LD $r^2 > 0.1$). SNP effects were drawn from a multivariate normal distribution with

$$\begin{bmatrix} \beta_{01,s} \\ \beta_{11,s} \\ \tau_{01,s} \end{bmatrix} \sim N\left(\begin{bmatrix} -0.2 \\ 0.01 \\ 0.2 \end{bmatrix}, \begin{bmatrix} 0.1 & c_1 & c_2 \\ c_1 & 0.01 & c_3 \\ c_2 & c_3 & 0.05 \end{bmatrix} \right)$$

for $j = 1, \ldots, m$ SNPs, with $c^T = (c_1 \quad c_2 \quad c_3)$ setting the genetic correlation between the SNP effects. Motivated by the observed genetic correlation on the real data approaches (see Section 4 and Supplemental Figures S11 and S16), we set $c_1 = -0.9$ and $c_2 = c_3 = 0$ for scenario A, and $c_1 = -0.9$, $c_2 = 0.5$, and $c_3 = -0.3$ for scenario B. The allele scores were then obtained as weighted sum of the simulated genotype matrix G:

$$AS_{mean} = \sum_{s \in S^{(1)}} \beta_{01,s} \times G_s$$

$$AS_{slope} = \exp\left(0.5 \sum_{s \in S^{(2)}} \beta_{11,s} \times G_s \right)$$

$$AS_{var} = \sum_{s \in S^{(3)}} \tau_{01,s} \times G_s$$

with S being the set of all m SNPs, and $S^{(1)}$, $S^{(2)}$, and $S^{(3)}$ being subsets of S used to simulate the mean, slope and variability SNP effects. In scenario A, $S^{(1)} = S^{(2)}$ included the first 20 SNPs, and $S^{(3)}$ the remaining 10 SNPs. In scenario B, $S^{(1)} = S^{(2)} = S^{(3)}$, e.g. all SNPs were used in all allele scores. The score for the mean effect, AS_{mean}, was then standardized to mean

of 0 and standard deviation (SD) of 1. We used the exponential score for the slope, AS_{slope}, here to ensure positive trend throughout all simulations. In addition, the allele scores were multiplied with scenario-specific factors to ensure strong instruments for the later MVMR (see Supplemental Table S1). The transformed scores where then used to generate both exposure and outcome.

Generating exposure X

In both scenarios A and B, we simulated three exposures for individuals i at time points k:

$$X_{123,i,k} = \beta_{00} + b_{0i} + AS_{mean,i} + (\beta_{10} + b_{1i} + AS_{slope,i})t_{i,k} + \beta_{20}t_{i,k}^2 + \epsilon_{i,k}$$
$$X_{12,i,k} = \beta_{00} + b_{0i} + AS_{mean,i} + (\beta_{10} + b_{1i} + AS_{slope,i})t_{i,k} + \beta_{20}t_{i,k}^2 + \epsilon_{i,k}$$
$$X_{13,i,k} = \beta_{00} + b_{0i} + AS_{mean,i} + (\beta_{10} + b_{1i})t_{i,k} + \beta_{20}t_{i,k}^2 + \epsilon_{i,k}$$

with $\beta_{00} = 0$, $\beta_{10} = 1$, and $\beta_{20} = 0.1$. The time points were independently sampled, using $t_{i,k} \sim N(4 \times k, 0.5)$, for all time points k. The random intercept and random slope were simulated using a multivariate normal distribution with

$$\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0.25 \end{bmatrix}\right).$$

The error terms were either dependent on AS_{var} with $\epsilon_{i,k} \sim N\left(0, exp\left(AS_{var,i}\right)\right)$ for X_{123} and X_{13}, or independently sampled with $\epsilon_{i,k} \sim N(0,1)$ for X_{12}.

Generating outcome Y

In each simulation, we simulated eight continuous outcomes, $Y_1, \dots, Y_8$, which were either independent of X, causally affected by a combination of the mean, slope, or variability of X:

$$Y_1 \sim N(0,1)$$
$$Y_2 = \theta_1 \times AS_{mean} + N\left(0, SD(AS_{mean})\right)$$
$$Y_3 = \theta_2 \times AS_{slope} \times T + N\left(0, SD\left(AS_{slope}\right)\right)$$
$$Y_4 = \theta_3 \times AS_{var} + N\left(0, SD(AS_{var})\right)$$
$$Y_5 = \theta_1 \times AS_{mean} + \theta_2 \times AS_{slope} \times T + N\left(0, SD(AS_{mean}) \times SD\left(AS_{slope}\right)\right)$$
$$Y_6 = \theta_1 \times AS_{mean} + \theta_3 \times AS_{var} + N\left(0, SD(AS_{mean}) \times SD(AS_{var})\right)$$
$$Y_7 = \theta_2 \times AS_{slope} \times T + \theta_3 \times AS_{var} + N\left(0, SD\left(AS_{slope}\right) \times SD(AS_{var})\right)$$
$$Y_8 = \theta_1 \times AS_{mean} + \theta_2 \times AS_{slope} \times T + \theta_3 \times AS_{var} + N\left(0, SD(AS_{mean}) \times SD\left(AS_{slope}\right) \times SD(AS_{var})\right)$$

In all simulations, we set $\theta^T = (1.2 \quad 0.3 \quad 1)$. The time parameter for the outcome simulation was independently sampled, using $T_i \sim N(70,1)$.

Estimands/targets of analysis

Our estimands of interest are the causal estimates for the mean, slope, and variability of X. For this, we used the multivariable inverse-variance weighted method *mr_mvivw* as implemented in the R-package MendelianRandomization[22]. For each exposure-outcome

combination, we extracted the number of strong instruments per exposure type, the estimated causal effects, $\hat{\theta}_i$, its standard error, $SE(\hat{\theta}_i)$, p-values associated with the estimates, conditional F-statistics, heterogeneity statistic (Cochran's Q statistic) and its associated p-value.

Methods

SNP association with the exposure X

To estimate the SNP effects on the exposure, we used GAMLSS with and without SNP x time interaction term for the main and no-slope model. We included a random intercept and estimated the SNP effects on the mean, (slope,) and variability of X (β_{01}, β_{11}, and τ_{01}), adjusting both the μ and the σ model on time and time squared.

SNP association with the outcome Y

We used linear regression models to estimate the SNP effects on each of the continuous outcomes without any further adjustments.

Multivariable Mendelian Randomization (MVMR)

The MVMR models were used as described in Section 2 above. MVMR estimates of the slope were rescaled for the mean age of the outcome, and estimates of the variability were rescaled depending on the allele score factor used in the respective scenario (see Supplemental Table S1 and Supplemental Material).

Performance metrics

In line with our aim to test the detection of causal effects, we focused on the power / type I error rate as the main performance metric. We used the number of significant estimates per outcome and scenario (MVMR association p-value $< \alpha = 0.05$ for any rescaled $\hat{\theta}_i$) divided by the number of simulations. We expected high detection rates for $\hat{\theta}_1$ in Y_2, Y_5, Y_6, and Y_8; for $\hat{\theta}_2$ in Y_3, Y_5, Y_7 and Y_8; and for $\hat{\theta}_3$ in Y_3, Y_6, Y_7 and Y_8, while the other estimates were expected to have a low power as expected by chance (around 5%, type I error rate):

$$\text{Power} = P = \Pr(p_i \leq \alpha), \qquad \hat{P} = \frac{1}{s}\sum_{i=1}^{s} 1(p_i \leq \alpha)$$

Other performance metrics such as bias and coverage can be found in the Supplemental Material.

Sensitivity analyses

In addition to the four main scenarios, we tested 11 further settings. This included further GAMLSS modifications (no SNP effect in the σ function, and random intercept in the σ function), reduced sample size ($n = 3000$ samples with $k = 5$ observations), different genetic correlation structures (no correlation, positive correlation, distinct SNP sets), additional $G \times E$ effect, and binary outcomes (see Supplemental Table S1 for all scenarios and used parameters).

Simulation results

X_{12}: exposure with instruments for mean and slope

The first exposure, X_{12}, was simulated with a genetic effect on both mean and slope, while the variability was not affected by the SNPs. In line with this, we had strong instruments in the main GAMLSS approach for both mean and slope (median conditional F-statistic: $F_m = 171.7$ and $F_s = 153.7$ for scenario A, and $F_m = 167.8$ and $F_s = 145.4$ for scenario B), but relative weak instruments for the variability ($F_v = 14.8$ in scenarios A and $F_v = 13.2$ in B, respectively). In the no-slope GAMLSS approach, we found that the conditional F-statistics increased for the variability to $F_v = 50.9$ and $F_v = 40.2$ in scenario A and B, respectively (see Supplemental Figure S1, left panel).

We observed 100% power for both mean and slope across all scenarios (see Figure 2 and Supplemental Figure S2). However, in case of high genetic correlation structures (B scenarios) we also detected high type 1 error rates (false positives) for Y_4 for the mean and the slope (99% and 63% in the main B scenario, respectively), and for Y_7 for the mean (46%). In the "B - no slope" scenario, we found additional false positives for the mean for Y_3 and Y_7. Of note, in the no slope scenarios, the variability estimates reached power of 69-95% in the A and 46-91% in the B scenario for all but Y_1 and Y_4, with causal effect estimates in the range of the true slope effect times age (see Supplemental Figure S3). The variability had no power in any other scenario. Changing the parameter θ_1 did not change the main findings (see Supplemental Table S2 for full simulation results).

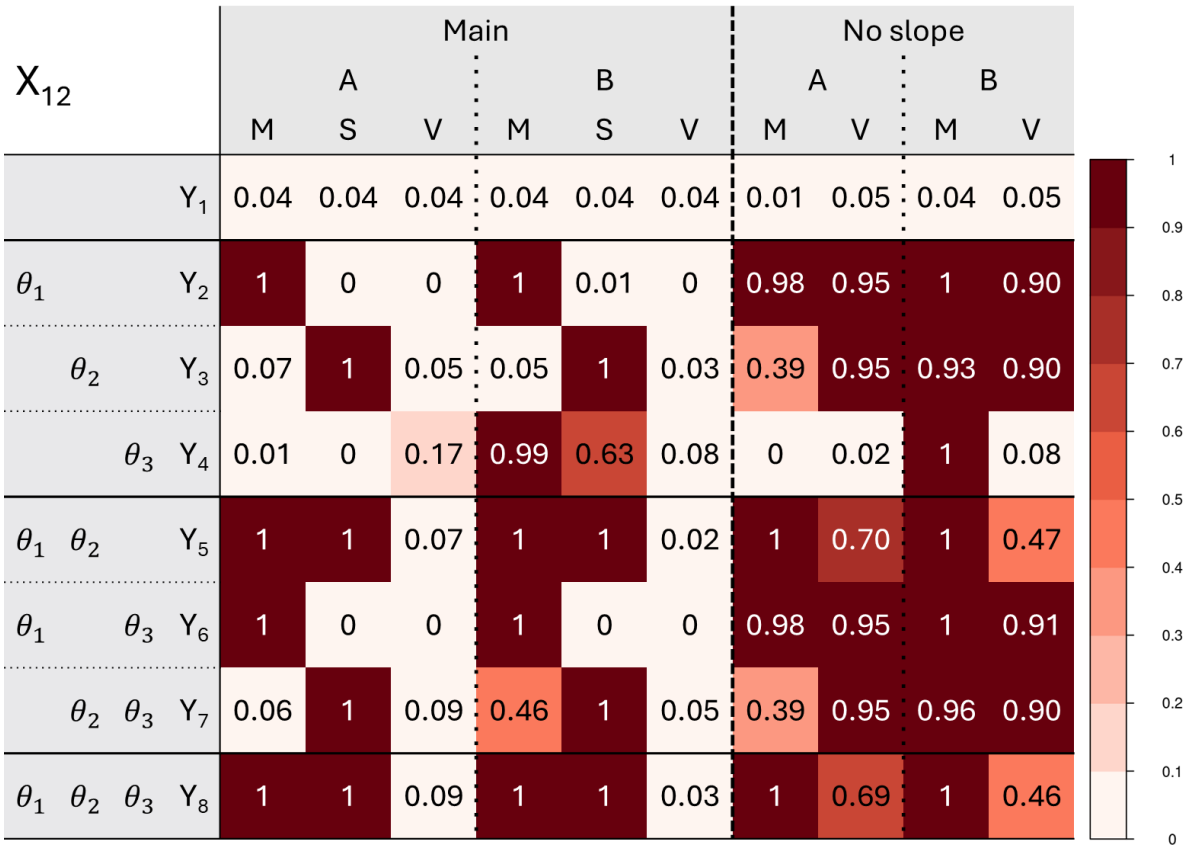

Figure 2: Power to detect the causal effects of X_{12} on all eight outcomes in the main and in the "no slope" scenarios. A and B denote no and positive genetic correlation between mean and variability, respectively. In the main model, all three SNP effects on the mean, slope and variability were estimated, while in the "no slope" model, the SNP x time interaction was left out. The power is order by exposure type (M: mean, S: slope and V: variability), genetic correlation scenario A or B, and per regression model, with true values $\theta_1 = 1.2$, $\theta_2 = 0.3$, and $\theta_3 = 1$.

X_{13}: exposure with instruments for mean and variability

The second exposure, X_{13}, was simulated with a genetic effect on both mean and variability, while the slope was simulated without a SNP effect. The median conditional F-statistics for mean and variability were well above the classical threshold of 10, only in the sample size reduced scenario we observed $F_m = 8.3$. Analogous to X_{12} and the variability, the conditional F-statistics for the slope were low here with $F_s = 7.7$ and $F_s = 2.6$ in the main A and B scenarios (see Supplemental Figure S1, right panel, and Supplemental Table S2 for all scenarios).

Across all scenarios, we observed high power ranging between 83-100% to detect an effect of the mean in outcomes Y_2, Y_3, and Y_5-Y_8, in other words, in all outcomes with either mean and/or slope effect (see Figure 3 and Supplemental Figure S4). The exceptions were the scenarios with three distinct SNP sets per exposure type and in case of positive correlation between mean and slope SNP effects (denoted as "A - dist. SNPs" and "A – pos. cor." in Supplemental Figure S4). The high type I error rate of the mean for Y_3 and Y_7 can be explained by the simulated genetic correlation structure and is also reflected in the bias of the estimates

(see Supplemental Figure S5 and Supplemental Material for technical notes on the bias of the estimate for the mean effect). The slope was low powered in terms of available instruments, and accordingly we found no power to detect the true causal effects of the slope (power 6-41%). We only detected high type I error rates for Y_2 and Y_6 in the B scenario, probably due to the correlation between mean and slope.

X_{13}				Main						No slope			
				A			B			A		B	
				M	S	V	M	S	V	M	V	M	V
			Y_1	0.04	0.04	0.06	0.06	0.04	0.06	0.06	0.06	0.07	0.07
θ_1			Y_2	1	0.04	0.02	1	0.51	0.02	1	0.05	1	0.05
	θ_2		Y_3	1	0.07	0	1	0.10	0.63	1	0	1	0.70
		θ_3	Y_4	0.04	0.06	1	0.05	0.05	1	0.04	1	0.04	1
θ_1	θ_2		Y_5	0.86	0.06	0	1	0.08	0.64	0.83	0	1	0.69
θ_1		θ_3	Y_6	1	0.02	0.97	1	0.70	0.09	1	0.57	1	0.09
	θ_2	θ_3	Y_7	1	0.08	0	1	0.11	0.51	1	0	1	0.58
θ_1	θ_2	θ_3	Y_8	0.87	0.07	0	1	0.08	0.53	0.84	0	1	0.60

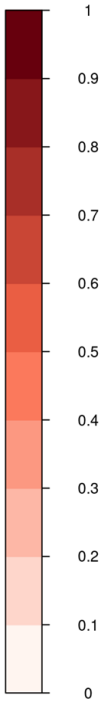

Figure 3: Power to detect the causal effects of X_{13} on all eight outcomes in the main and in the "no slope" scenarios. A and B denote no and positive genetic correlation between mean and variability, respectively. In the main model, all three SNP effects on the mean, slope and variability were estimated, while in the "no slope" model, the SNP x time interaction was left out. The power is order by exposure type (M: mean, S: slope and V: variability), genetic correlation scenario A or B, and per regression model, with true values $\theta_1 = 1.2$, $\theta_2 = 0.3$, and $\theta_3 = 1$.

For the variability, we observed high power for Y_4 (100%), the outcome only affected by the variability. For Y_6, which was expected to be affected by the mean and the variability, we detected high power in scenario A (97%), but no power in scenario B (9%, no vs. pos. genetic correlation between mean and variability). For Y_7 and Y_8, which were both simulated with a slope and variability effect, we detected no power in scenario A (0%), and weak power in scenario B (e.g. 53% and 60% in main and no slope for Y_8). Of note, in scenario A the effects were unbiased but with high uncertainty, while in B the effects were biased with reversed effect direction in all outcomes with a simulated slope effect (see Supplemental Figure S5). In addition to the low power, we found moderate type I error rates for the variability and Y_3 and Y_5 in the B scenarios (63% and 64% in B, 70% and 69% in B – no slope).

X_{123}: exposure with instruments for mean, slope and variability

Finally, the third simulated exposure, X_{123}, was simulated with genetic effects on all three of the exposure types. We found strong conditional F-statistics in our simulation approach for all three exposure types (see Supplemental Figure S1, middle panel, and Supplemental Table S2 for all scenarios).

Here, we found 100% power for both mean and slope across all relevant outcomes in almost all scenarios, with exceptions the "no slope" scenarios regarding the mean (23-29% to detect mean effect on Y_2, 22-31% to detect mean effect on Y_6) (see Figure 4 and Supplemental Figure S6). For the variability, Y_4 was well powered again (100%), but for $Y_6 - Y_8$ we only detected high power in the main and four other A scenarios (93-99%). For the other outcomes, we did not find high type I error rates. For the variability, the effect was slightly overestimated in scenario A, and underestimated in B. As before, the lack of power is likely due to the imprecision (high empirical SE, see Supplemental Figure S7).

X_{123}		Main						No slope			
		A			B			A		B	
		M	S	V	M	S	V	M	V	M	V
	Y_1	0.03	0.04	0.06	0.03	0.05	0.05	0.04	0.08	0.05	0.04
θ_1	Y_2	1	0.01	0.02	1	0.03	0.04	0.29	0.36	0.23	0.93
θ_2	Y_3	0.08	1	0.05	0.07	1	0.07	0.77	0.37	0.61	0.93
θ_3	Y_4	0.05	0.07	1	0.07	0.09	1	0.42	1	0.19	1
θ_1 θ_2	Y_5	1	1	0.05	1	1	0.06	0.97	0.33	0.83	0.94
θ_1 θ_3	Y_6	1	0.01	0.93	1	0.03	0.08	0.31	0.43	0.22	0.93
θ_2 θ_3	Y_7	0.07	1	0.78	0.08	1	0.32	0.77	0.33	0.61	0.94
θ_1 θ_2 θ_3	Y_8	1	1	0.78	1	1	0.18	0.98	0.57	0.84	0.91

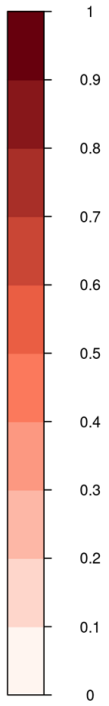

Figure 4: Power to detect the causal effects of X_{123} on all eight outcomes in the main and in the "no slope" scenarios. A and B denote no and positive genetic correlation between mean and variability, respectively. In the main model, all three SNP effects on the mean, slope and variability were estimated, while in the "no slope" model, the SNP x time interaction was left out. The power is order by exposure type (M: mean, S: slope and V: variability), genetic correlation scenario A or B, and per regression model, with true values $\theta_1 = 1.2$, $\theta_2 = 0.3$, and $\theta_3 = 1$.

The exception of the rule was again the no slope scenarios. In the A scenario setting (no correlation between mean and variability SNP effects), there was low power for the mean effect on Y_2 (29%) and Y_6 (31%), while the type I error rate for the mean on Y_3 and Y_7 was high

with 77%. It also affected the power of the variability, which decreased for $Y_6 - Y_8$. In contrast, in the B setting (positive correlation between mean and variability SNP effects), we observed high detection rates for all $Y_2 - Y_8$ (91-100%), e.g. high power for Y_4, $Y_6 - Y_8$, and high type I error rate for Y_2, Y_3, and Y_5. The estimated effect sizes of the variability on Y_2, Y_3, Y_6 and Y_7 reflected again the time-uncorrected slope effect (see Supplemental Figure S7, lower right panel). For the mean, the power/type I error de-/increased similar to the A – no slope scenario.

Sensitivity checks

Here, we briefly summarize the results of the sensitivity runs of the simulation scenarios. All parameter settings and results are summarized in Supplemental Tables S1 and S2.

Binary outcomes: The mean and slope results for the binary outcomes were very similar to the main scenario, regardless of the outcomes prevalence (20% and 50%, denoted as "binary20" and "binary50" in Table S2). Only the variability was less powered for X_{123} and X_{13}.

Different genetic correlation / SNP sets: The results for the scenarios with distinct SNP sets, no genetic correlation and positive genetic correlation were very similar to the main scenario (denoted as "distSNPs", "noCor", "posCor" in Table S2). Only for X_{13} the type I error rate was reduced in case of positive correlation or distinct SNP sets.

Different GAMLSS model: Excluding the variability SNP effect in the σ-function resulted in the same results for mean and slope as in the main scenario ("noVar" in Table S2). Including a random intercept term in the σ-function reduced the power of the variability of X_{13} on Y_6 and of X_{123} on $Y_6 - Y_8$ ("RIsigma" in Table S2).

Smaller sample size ("sampleSize" in Table S2): In scenario A, the power loss due to decreased sample size led to power of the variability of X_{13} on Y_6 and of X_{123} on $Y_6 - Y_8$, while increasing the type I error rate of the slope of X_{13} on Y_2 and on Y_6. In scenario B, we also observed high type I error rate of the slope of X_{13} on Y_2 and on Y_6, but also high power on Y_5 and Y_8. For X_{123}, the results are very similar, but with reduced power of the variability and slope.

Summary

We performed a simulation study with four main scenarios, comparing the effects of genetic correlation and GAMLSS specification on the MVMR results. In the presence of strong instruments, we were able to detect the true mean and slope effect with high power in all scenarios for almost all exposure-outcome combinations. Only for X_{123} in the no-slope approach we observed low power for the mean effect alone (Y_2) or for the combination of mean and variability effect (Y_6). The power for the variability depended highly on the underlying genetic correlation structure, with high power in case of independence between mean and variability, but low power in the fully correlated scenarios. False positives were mainly caused by GAMLSS misspecifications, e.g. in the no-slope approach.

Application in real data

Material and Methods

Study description and exposures

We tested our method of combining longitudinal GWAS summary statistics and MVMR in real data on two positive controls, i.e. well-established causal pairs. Please note, some of our analysis choices are made to illustrate the method or because of limitations in the available data, rather than to reflect best practice. Detailed study descriptions can be found elsewhere[23–25], and additional information on sample filtering and genotyping can be found in the Supplemental Material.

First, we applied the method using exposure data from the Pregnancy Outcome Prediction Study (POPS)[23], in which the estimated foetal weight (EFW) was measured at three distinct time points during pregnancy in 2,996 participants most similar to GBR ancestry with foetal genotype data. The outcome was birth weight (BW), both within POPS (1-sample MVMR) as well as in the UKB (2-sample MVMR)[2]. This study offered very structured longitudinal data of an age-dependent exposure, but was limited in power due to sample size. We used the log-transformation of EFW to obtain nearly linear growth and to reduce heteroscedasticity (trajectories are given in Supplemental Figure S8A, and study description in Supplemental Table S3).

In a second test, we used total cholesterol (TC) data from the UKB[25] in combination with electronic health data from GPs. Here, we had a sample size of 73,778 individuals, but unstructured data in terms of both number of time points per sample, and time difference between the time points (see Supplemental Figures S8B and S9B for histograms in POPS and UKB data, respectively). The trajectories of TC in 10 randomly selected samples of the UKB are shown in Supplemental Figure S9A, and a study description is given in Supplemental Table S4. Again, we used 1- and 2-sample MVMR approaches with outcome data for coronary artery disease (CAD) from the UKB and from a large meta-GWAS for CAD (Aragam et al.[26]).

Candidate SNPs

To avoid analysing the whole genome, we selected candidate SNPs associated with BW for the POPS analysis, and SNPs associated with TC for the UKB analysis. A flowchart for the SNP selection and processing is given in Supplemental Figure S10. We followed the same procedure for both approaches.

First, we selected suitable consortia with publicly available summary statistics. For the analysis in POPS, we obtained data from the Early Growth Genetics (EGG) Consortium[27] (www.egg-consortium.org, downloaded 06/02/2024, European-only meta-analysis of up to 143,677 individuals and 16,245,523 SNPs). For the UKB, we used data from the Global Lipids Genetics Consortium (GLGC)[28] (https://csg.sph.umich.edu/willer/public/glgc-lipids2021/, downloaded 16/09/2024, European-only meta-analysis of up to 1.3 million individuals and 46 million SNPs).

In the second step, we filtered the consortium data for strong, genome-wide significant instruments (association p-value of the respective mean trait $p < 5 \times 10^{-8}$), minor allele frequency ($MAF > 1\%$), and availability of rsID and position in both hg19 and hg38 (Bioconductors annotation package SNPlocs.Hsapiens.dbSNP150.GRCh38[29] to lift the positions to hg38 and dropping SNPs that could not be matched). To simplify our approach, we also excluded tri-allelic SNPs, which could complicate the harmonization, and excluded chromosome X SNPs, to avoid more elaborate model specification due to gene dosage differences between males and females.

Next, we harmonized the generated consortium SNP lists with the respective exposure study genetic data and the publicly available outcome summary statistic data for matching position and alleles. The exposure genetic data sets were restricted to SNPs with high imputation quality (info > 0.8). The effect alleles were harmonized to the exposure genetic data set, to simplify later combination of the *gamlss* regression estimates with the publicly available data. We tested SNPs for difference in allele frequency in the different data sets, and excluded those with a difference of more than 10%.

Finally, we performed clumping based on the consortia's p-values and genomic position to obtain independent SNPs. In more detail, we sorted the SNPs by increasing p-value as reported by EGG or GLGC, selected the best associated SNP and excluded all SNPs within a range of $\pm$ 1MB. This procedure was repeated with the next best SNP until there were no more lower ranking SNPs left. This resulted in 52 and 346 SNPs for the EFW and TC analyses, respectively. We used the exposure data sets to estimate the pairwise LD of SNPs on the same chromosome. We observed no strong correlation between SNPs for the EFW analysis (all pairwise LD r^2<0.05), while there were five correlated SNP clusters for the TC analysis. We performed the *gamlss* regression for all SNPs, and selected the best-associated SNP per cluster for the MVMR analysis.

SNP effects on the exposures

Main analyses

In the POPS analysis for EFW we adjusted for maternal height, maternal smoking status, foetal sex, foetal gestational age (GA) at the exact time of the scan, GA squared, and the first five foetal principal components (PCs). GA was used as time parameter in the SNP x time interaction, e.g. we estimated the SNP effect per week.

In the UKB analysis for TC we adjusted for sex, age, lipid lowering medication (statins), and the first 10 PCs. The participants' age was used as time parameter, e.g. we estimated the SNP effect per year.

For both EFW and TC, the main model was the full model, estimating the SNP effect on the mean, slope and variability simultaneously. For both the μ- and σ-function we used the same adjustment models, but the only random effect included was the random intercept for the mean.

Sensitivity analyses

For sensitivity purposes, we also tested three modifications of the *gamlss* regression model: (1) we excluded the SNP x time interaction, (2) we excluded the SNP effect on the variability, and (3) we included the random intercept in the σ-function. In addition, we restricted the POPS data to British ancestry and all three time points ($n = 2{,}509$). In the UKB, we restricted the samples to age within 40 to 70, no statin treatment, and first TC measurement per year (ceiling of age per person and time point, $n = 42{,}193$).

SNP effects on the outcomes

For the outcome birth weight (BW) in POPS (1-sample approach), we used linear regression models. The same samples and covariables as in the *gamlss* regression were used here. For the 2-sample approach, we downloaded the data on self-reported BW in the UK Biobank from the Neale lab[2] (raw values, not adjusted for gestational age).

For the 1-sample approach of the TC analysis, we used the phenotype definition for coronary atherosclerosis ("I9_CORATHER") as defined in FinnGen[3] (31,198 cases and 382,052 controls). For the 2-sample approach, we downloaded the latest publicly available summary statistics from Aragam et al.[26] (meta-GWAS of CAD with up to 181,522 cases and 984,168 controls).

MVMR and instrument selection

We used the same MVMR IVW method as described in the simulation study. To ensure strong instruments, we included SNPs associated with $p < 5 \times 10^{-8}$ with one of the exposure types of TC in UKBB. Given the lower sample size and power in POPS, this threshold was relaxed to $p < 0.05$ for the association with EFW exposure types. As sensitivity check, we also allowed all SNPs to be included, as they were already identified as valid instruments in the consortia data. For comparison, we also estimated the causal univariate effect, using MR-IVW on either the strong instruments per exposure type, or all possible SNPs.

Results

EFW on BW

Genetic associations with the exposure

For EFW in POPS, we mainly got weak instruments using *gamlss*, due to the sample size and number of time points, and only one genome-wide significant association with the mean EFW (rs35261542, see Supplemental Table S5 for all summary statistics). Of the 52 SNPs, 24 reached nominal significance (p<0.05) for at least one of the three exposure types. The genetic correlation between the 52 candidate SNPs was significant between mean and slope ($r_g = -0.93, p = 3.3 \times 10^{-115}$), but not between mean and variability ($r_g = -0.07, p = 0.644$) or slope and variability ($r_g = 0.15, p = 0.289$). While the effect estimates were highly similar when using most other regression specifications (see Supplemental Figure S11), removing the SNP x time effect changed the estimates for the mean, and the correlation between these effect estimates and the slope or variability were higher than with other mean effects. This was to be expected, as the main effect now incorporates both the mean and the interaction term. As a result, there were 17 genome-wide significant SNPs for the mean, but still none for the variability.

MVMR results

In the main analysis, we detected for EFW significant positive estimates of the mean and the slope on BW ($p_{mean} = 1.03 \times 10^{-4}$, $p_{slope} = 1.97 \times 10^{-6}$), while no effect of the variability was observed ($p_{var} = 0.360$). Figure 5 displays a Forest Plot of 2-sample MVMR and MR results and Supplemental Figure S12 shows scatter plot of SNP effects on the exposure types and outcome. Of note, the mean and slope effect in the MR were of opposite direction, while in the MVMR they were both positive. This was to be expected due to the opposite SNP effect directions for the mean and slope for all associated SNPs, the resulting high negative correlation and its correction in the MVMR. Similar results were observed when using the 1-sample approach or when using all SNPs instead of nominally significant SNPs only. Also, the results from all sensitivity analyses were in general in good accordance with the main analysis: mean and slope of EFW had positive effects on BW, regardless of the model specification (see Supplemental Figures S13 – S15), and the variability was never detected with any significance. All results are summarized in Supplemental Table S6.

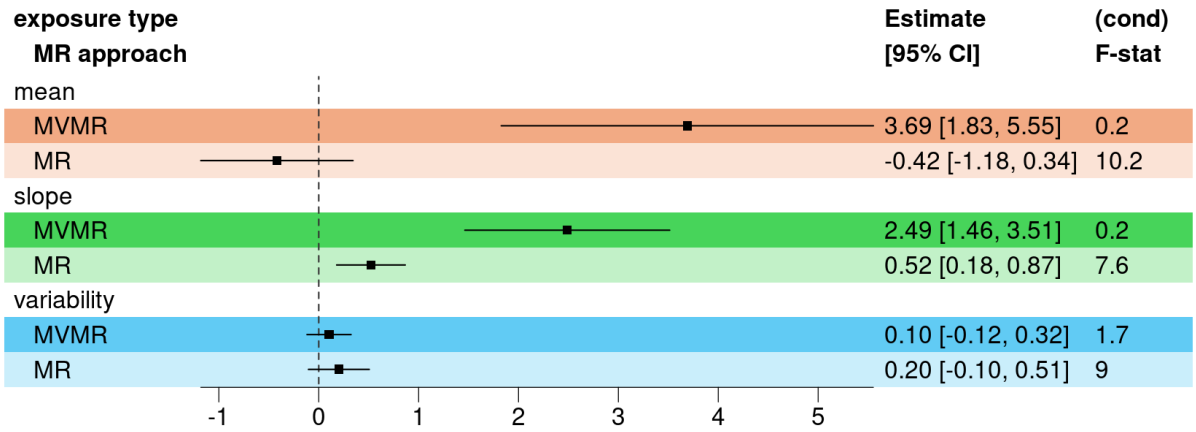

Figure 5: Forest Plot of the causal estimates of estimated foetal weight on birth weight using the 2-sample approach and nominally significant SNPs. In the MVMR approach, we used 24 SNP which were at least nominally associated (p<0.05) with one of the exposure types mean, slope or variability. In the MR approach, only the significant SNPs per type were used, resulting in 15, 19, and 6 instruments for mean, slope and variability, respectively. The slope estimates were scaled to effect per gestational week. The last column of the plot indicates the (conditional) F-statistics for the (MV)MR approaches.

Power

The power was moderate for the univariable MR approach, with F-statistics of 10.15, 7.59, and 8.95 for mean, slope, and variability, respectively (see Supplemental Table S7). However, in the multivariable approach, the conditional F-statistics were low, with 0.25 for mean and slope, and 1.71 for the variability. This was to be expected, given the relatively low sample size, instrument association p-values, and correlation structure between mean and slope effects. In the sensitivity analysis using no time-interaction we observed slightly higher conditional F-statistics (2.76 for the mean, and 2.57 for the variability).

TC on CAD

GX associations

For TC, we tested 346 candidate SNPs for association on mean, slope and variability of TC. We detected for all three exposure types genome-wide significant associations in the main analysis: 101 for the mean, 61 for the slope and 107 for the variability. There was a high overlap between the associated SNP (see Figure S16 for Venn Diagram), with 31 SNPs shared between all three exposure types (e.g., genome-wide significant SNP effects on all three), and only 23 and 21 specific for the mean and variability, respectively (using $p_{mean} < 5 \times 10^{-8} \wedge p_{var} > 0.05$ and $p_{var} < 5 \times 10^{-8} \wedge p_{mean} > 0.05$ as cut-offs). Similarly to the POPS data, we observed a high negative correlation between the mean and slope SNP effects ($r_g = -0.92, p = 4.64 \times 10^{-144}$). However, as expected with the observed high overlap, there was also a significant correlation between the mean and variability ($r_g = 0.68, p = 8.97 \times 10^{-48}$) and between the slope and variability ($r_g = -0.54, p = 2.66 \times 10^{-27}$). All SNP summary statistics are given in Supplemental Table S8.

MVMR results

In the main analysis, we detected again significant positive effects of the mean and the slope on the risk for CAD ($p_{mean} = 1.76 \times 10^{-14}$, $p_{slope} = 1.15 \times 10^{-7}$), but even with the increased sample size and instrument strength, the conditional F-statistics were low (2.24 for mean, 2.12 for slope). In contrast, the variability was well powered (8.52), but we did not detect any significant results here ($p_{var} = 0.239$, see Figure 6 for Forest Plot of 2-sample MVMR and MR results, and Supplemental Figure S17 for scatter plots).

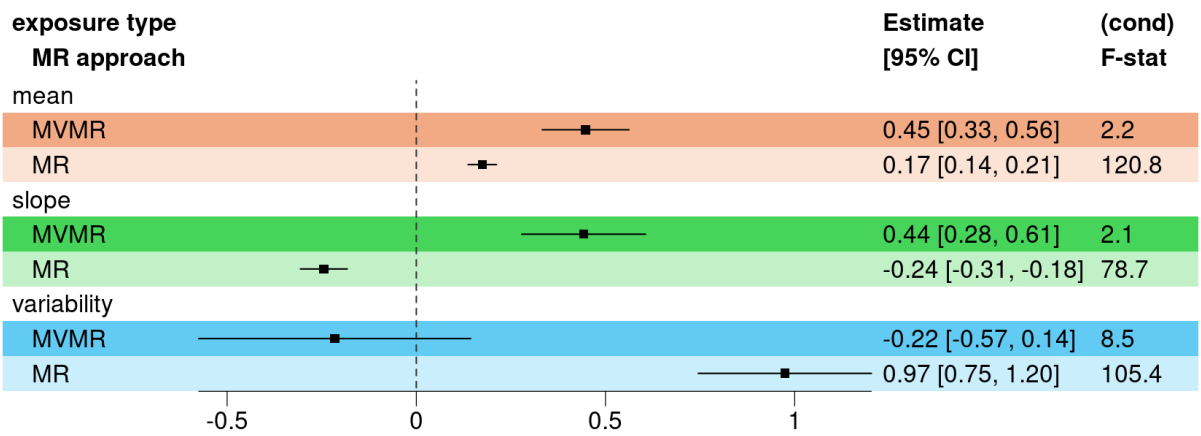

Figure 6: Forest Plot of the causal estimates of total cholesterol levels on coronary artery disease using the 2-sample approach and genome-wide significant SNPs. In the MVMR approach, we used 168 SNP which were at least genome-wide significantly associated ($p < 5 \times 10^{-8}$) with one of the exposure types mean, slope or variability. In the MR approach, only the significant SNPs per type were used, resulting in 100, 61, and 106 instruments for mean, slope and variability, respectively. The slope estimates were scaled to effect per year. The last column of the plot indicates the (conditional) F-statistics for the (MV)MR approaches.

We observed similar results in all sensitivity checks, using 1-sample MVMR, all SNPs, different *gamlss* regression specifications, or a different sample set. In the sensitivity analysis using no

time-interaction, we observed high conditional F-statistics (39.4 for the mean, and 22.46 for the variability), and the effect estimates were of the same magnitude as in the main analysis. All results are summarized in Supplemental Table S9, S10 and shown in Supplemental Figures S18-S20.

Cherry picking

Given the high number of strong and specific instruments for mean and variability, we repeated the analysis in the main setting, restricting it to the 23 SNPs only associated with the mean and the 21 SNPs only associated with the variability (total of 44 SNPs). The genetic correlation between the mean and slope stayed the same ($r_g = -0.93, p = 7.72 \times 10^{-20}$), while the correlation between mean and variability and slope and variability decreased ($r_g = 0.06, p = 0.677$ and $r_g = 0.13, p = 0.399$, respectively). In the MVMR, we still found significant positive effects of mean and slope ($p_{mean} = 1.30 \times 10^{-7}$, $p_{slope} = 1.73 \times 10^{-6}$), but detected in addition a significant negative effect of the variability ($p_{var} = 9.63 \times 10^{-3}$, see Figure 7).

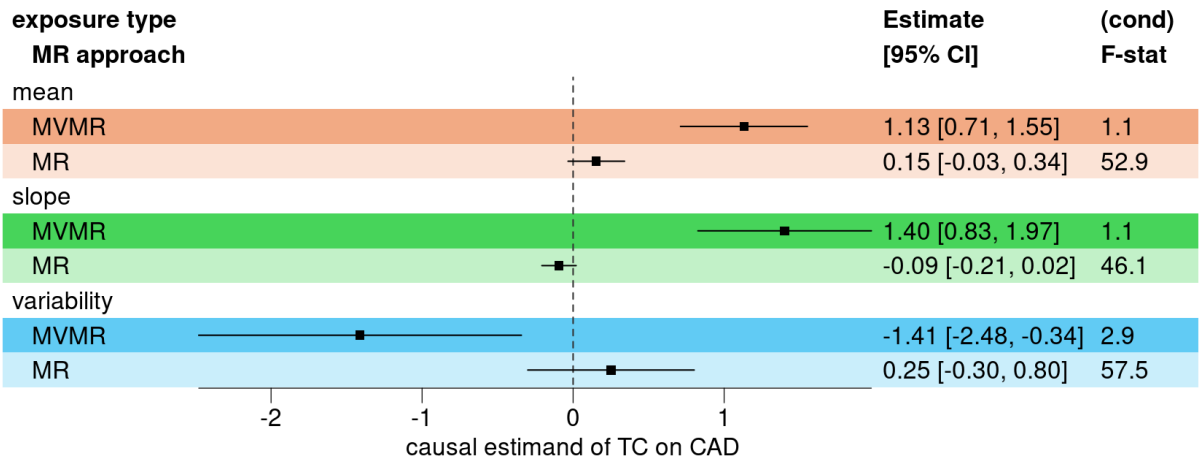

Figure 7: Forest Plot of the causal estimates of total cholesterol levels on coronary artery disease using the 2-sample approach and 44 genome-wide significant SNPs. These 44 SNPs comprise 23 SNPs associated with the mean but not the variability ($p_{mean} < 5 \times 10^{-8} \wedge p_{var} > 0.05$), and 21 SNP associated with the variability but not the mean ($p_{var} < 5 \times 10^{-8} \wedge p_{mean} > 0.05$). The slope estimates were scaled to effect per year. The last column of the plot indicates the (conditional) F-statistics for the (MV)MR approaches.

Discussion

In this study, we presented a framework combining longitudinal GWAS summary statistics and MVMR to estimate causal effects of the mean of the exposure as well as of the time-dependent slope and within-subject variability. In our simulation study, we observed an overall high power to detect the causal mean and slope effects, but mixed results for the variability. In our positive controls using real data, we could indeed detect significant causal effects of mean levels and the slope on the respective outcomes, but not of the exposures' variability.

640 We simulated three different exposures, denoted X_{12}, X_{13}, and X_{123}, in 15 different scenarios.
641 The simulation revealed high power and low rates of false positives, given the GWAS model
642 was correctly specified and included the SNP x time interaction term. In the simulation
643 without the interaction term ("no slope" scenarios), this unobserved SNP effect was attributed
644 to both the mean and variability in both X_{12} and X_{123}, leading to biased causal estimates. In X_{13}
645 this problem did not occur as it was simulated without genetic effects on the slope. This
646 observation is in line with variance quantitative trait loci (vQTL) studies, which have suggested
647 that SNP effects on the sample variance of a biomarker are candidates for unobserved gene-
648 environment interaction[6–8]. While the sample variance and within-individual variability differ
649 in their definition and interpretation, it is plausible to assume that the observed SNP effects
650 on the variability are caused by the unobserved SNP x time interaction in the "no slope"
651 model. Others have already explored the use of gene x environment (GxE) interactions in
652 Mendelian Randomization approaches[30,31], with main focus on correcting for pleiotropic
653 effects which violate the MR assumptions. Similarly, our approach was unbiased of the
654 pleiotropic effects introduced by the time interaction, given the GWAS model was correctly
655 specified. In addition, using the conditional significant effect of the GAMLSS regression in the
656 MVMR approach allowed to test for further unobserved interactions, which might still be
657 attributed to the within-individual variability. In contrast, the MR-GxE[30] and MR-GENIUS[31] only
658 aim at an unbiased causal effect estimates of the mean.

659 Another aspect influencing the power in the MVMR besides the model specification was the
660 genetic relationship between mean, slope, and variability effect. In the presence of significant
661 correlations between all three ("B" scenarios) and exposures with instruments for the
662 variability (X_{13} and X_{123}), we found that empirical standard error of the variability increased
663 from A ("independent variability scenarios") to B, due to the more complex correlation
664 structure and instrument strength needed to precisely distinguish between the three
665 exposure types in the MVMR. This led to low power to detect the variability effect of X_{123} on
666 outcomes affected by the variability and something else (Y_6, Y_7 and Y_8), while the power to
667 detect the variability effect only (Y_4) was high in both A and B. In exposure X_{13} we observed
668 better power, but the effect was also biased with reversed effect direction in all outcomes with
669 a slope effect. This might be due to the simulation setup, in which the SNP effects and their
670 correlations were simulated the same way for all three exposures, regardless of their later use
671 in the exposure simulation. This resulted in a different correlation structure in the GAMLSS
672 estimates when compared to the simulation input.

673 To better understand the genetic correlation between mean, slope, and variability, we used
674 two real data applications. First, we tested the estimated foetal weight for causal effects on
675 birth weight. Here, we only observed a correlation between mean and slope. However, our
676 power for the variability was low, as we only had three time points in the POPS. Using the GP
677 records of TC levels for UK biobank participants, we increased the number of observations
678 (mean of 7 observations per sample). Using known lipid SNPs and this increased sample size
679 and numbers of observations, we now detected significant SNP effects on the variability, and
680 high correlation between mean, slope, and variability. In both approaches, we only detected
681 causal effects of the mean and slope on the respective outcomes, birth weight and CAD.

While the findings are plausible and in line with literature[24,32–34], their interpretation has to be taken with caution, as even in the UK Biobank analyses we had weak power in terms of conditional F-statistics. We performed sensitivity checks with a “no slope” model, which was well powered in the MVMR, but still did not detect a variability effect. Hence, one potential explanation is that the slope effect is rather a false positive due to correlation structure and weak power, as otherwise the variability of the “no slope” model would have picked this effect up.

On the other hand, we were able to cherry pick 44 SNPs with either mean or variability effect on TC, which dissolved the correlation structure between mean and variability. Here, we still found the mean and slope effect, but additionally a negative effect of the variability on CAD risk, which is contrary to reported effects of lipid variability on CAD related outcomes[35]. Again, this analysis had low conditional F-statistics, and the results have to be interpreted with care. One explanation of the negative effect is that individuals with high variability are more likely to see their GP on regular basis, which in turn will reduce their risk for CAD as they benefit from prevention strategies. Another interpretation angle comes from the potential GxE interactions. Indeed, two of the 21 variability SNPs have already been reported for a SNP-interaction effect on LDL, including alcohol intake and fish oil supplementation[36,37]. Also, nine of the 21 SNPs have been reported to also affect sex-hormone binding globulin (SHBG)[38], which might indicate a hidden sex-interaction. However, although these explanations seem to be plausible, the result was only detected by cherry-picking, and selectively choosing instruments can introduce bias and a false positive result.

Our study has several limitations. First, although we simulated the exposure data with a random slope, we did not include that term in the GAMLSS regression, only the random intercepts. This was due to computational reasons: the computation time increases further, and the covariance matrix of the random effects is no longer invertible in all scenarios. Given that we still observed associated SNPs and that the power / type I error rate was our primary performance metric, we refrained from implementing a model with three random effects in the simulation, even if that introduced bias. Another limitation was that we used only candidate SNPs in the real data application. They were reported for mean level of the exposure, which gives the association with the variability a disadvantage from the start. However, a full longitudinal GWAS would have been too onerous for illustrative examples. Finally, as seen in the simulation, the approach is sensitive to the regression model specification. In our real data approach, we kept the analysis as simple as possible, as it was just to illustrate the method. However, the unknown true model might have resulted in different findings.

In conclusion, we presented a framework combining the estimates from longitudinal GWAS summary statistics with the MVMR approach. While the data availability for such summary data is yet scarce, it will be more standard in the near future. We demonstrate that the mean and slope effects can be estimated, while a positive variability finding needs cautious evaluation on both GWAS level regarding potential environmental interactions, and on MR level for pleiotropic effects.

Declarations

Availability of data and code

The code for both simulation study and real data application is available on github. We used R v4.2.2 and the R-packages lme4 v1.1-35.1[15], gamlss v5.4-12[17], and MendelianRandomization v0.9.0[22] throughout the analyses. Individual level data of POPS are not publicly available due to ethical restrictions, but they can be made available upon reasonable request to GS. Individual level UKB data can be accessed after an approved application. All generated genetic summary statistics will be made publicly available upon publication in the Supplemental Tables S5 and S8. Supplemental Tables are accessible on github and zenodo[39].

Competing interests

JKB has received research funding for unrelated work from F. Hoffmann-La Roche Ltd.

Funding

JP was supported by grants from the Wellcome Trust (225790/Z/22/Z) and the United Kingdom Research and Innovation Medical Research Council (MC_UU_00002/7) to SB. MP was supported by the MRC grant "Looking beyond the mean: what within-person variability can tell us about dementia, cardiovascular disease and cystic fibrosis" (MR/V020595/1). JKB was supported by MRC Unit Programme MC_UU_00002/5 and MRC Unit Theme MC_UU_00040/02 (Precision Medicine). POPS was supported by the Women's Health theme of the NIHR Cambridge Biomedical Research Centre.

Authors' contribution

Janne Pott: methodology, data curation, formal analysis, writing – original draft preparation
Marco Palma: methodology
Yi Liu: investigation
Jasmine Mack: data curation (POPS genotyping)
Ulla Sovio: data curation (POPS phenotypes), study management of POPS
Gordon Smith: PI of POPS
Jessica Barrett: methodology, funding
Stephen Burgess: study design, methodology, funding, supervision
All authors reviewed, edited and approved the final manuscript.

Acknowledgement

We thank all the women who participated in POPS, and all the staff in the Rosie Hospital (Cambridge, UK) and NIHR Cambridge Clinical Research Facility who provided direct or indirect assistance for the study. We thank Colin Starr for the IT support using the HPC of the Cambridge University. This research has been conducted using the UK Biobank Resource (http://www.ukbiobank.ac.uk/) under Application Number 98032 and 7439. Data on birth

weight trait has been contributed by the EGG Consortium using the UK Biobank Resource and has been downloaded from www.egg-consortium.org. Data on total cholesterol has been contributed by the Global Lipid Genetics Consortium and has been downloaded from https://csg.sph.umich.edu/willer/public/glgc-lipids2021/. Data on coronary artery disease has been contributed by Aragam et al. and has been downloaded from https://www.ebi.ac.uk/gwas/publications/36474045.

References

1. Burgess S, Thompson SG. Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects. *Am J Epidemiol*. 2015;181(4):251-260. doi:10.1093/aje/kwu283

2. Neale Lab. Neale Lab. http://www.nealelab.is/uk-biobank/

3. Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023;613(7944):508-518. doi:10.1038/s41586-022-05473-8

4. Chen H, Conomos M, Pham D. GMMAT: Generalized Linear Mixed Model Association Tests. Published online April 14, 2023:1.4.2. doi:10.32614/CRAN.package.GMMAT

5. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet*. 2021;53(7):1097-1103. doi:10.1038/s41588-021-00870-7

6. Paré G, Cook NR, Ridker PM, Chasman DI. On the Use of Variance per Genotype as a Tool to Identify Quantitative Trait Interaction Effects: A Report from the Women's Genome Health Study. Storey JD, ed. *PLoS Genet*. 2010;6(6):e1000981. doi:10.1371/journal.pgen.1000981

7. Wang H, Zhang F, Zeng J, et al. Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. *Sci Adv*. 2019;5(8):eaaw3538. doi:10.1126/sciadv.aaw3538

8. Hillary RF, Gadd DA, Kuncheva Z, et al. Systematic discovery of gene-environment interactions underlying the human plasma proteome in UK Biobank. *Nat Commun*. 2024;15(1):7346. doi:10.1038/s41467-024-51744-5

9. Sanderson E, Richardson TG, Morris TT, Tilling K, Davey Smith G. Estimation of causal effects of a time-varying exposure at multiple time points through multivariable mendelian randomization. Kutalik Z, ed. *PLOS Genet*. 2022;18(7):e1010290. doi:10.1371/journal.pgen.1010290

10. Hao Y, Xiao J, Liang Y, et al. Reassessing the causal role of obesity in breast cancer susceptibility: a comprehensive multivariable Mendelian randomization investigating the distribution and timing of exposure. *Int J Epidemiol*. 2023;52(1):58-70. doi:10.1093/ije/dyac143

11. Sikorska K, Lesaffre E, Groenen PJF, Rivadeneira F, Eilers PHC. Genome-wide Analysis of Large-scale Longitudinal Outcomes using Penalization —GALLOP algorithm. *Sci Rep*. 2018;8(1):6815. doi:10.1038/s41598-018-24578-7

12. Ko S, German CA, Jensen A, et al. GWAS of longitudinal trajectories at biobank scale. *Am J Hum Genet*. 2022;109(3):433-445. doi:10.1016/j.ajhg.2022.01.018

13. Chen H, Wang C, Conomos MP, et al. Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *Am J Hum Genet*. 2016;98(4):653-666. doi:10.1016/j.ajhg.2016.02.012

14. Uffelmann E, Huang QQ, Munung NS, et al. Genome-wide association studies. *Nat Rev Methods Primer*. 2021;1(1):59. doi:10.1038/s43586-021-00056-9

15. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using **lme4**. *J Stat Softw*. 2015;67(1). doi:10.18637/jss.v067.i01

16. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape,(with discussion). *Appl Stat*. 2005;54:507-554.

17. Stasinopoulos DM, Rigby RA. Generalized Additive Models for Location Scale and Shape (GAMLSS) in *R*. *J Stat Softw*. 2007;23(7). doi:10.18637/jss.v023.i07

18. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res*. 2007;16(4):309-330. doi:10.1177/0962280206077743

19. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med*. 2016;35(11):1880-1906. doi:10.1002/sim.6835

20. Sanderson E. Multivariable Mendelian Randomization and Mediation. *Cold Spring Harb Perspect Med*. 2021;11(2):a038984. doi:10.1101/cshperspect.a038984

21. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38(11):2074-2102. doi:10.1002/sim.8086

22. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol*. 2017;46(6):1734-1739. doi:10.1093/ije/dyx034

23. Pasupathy D, Dacey A, Cook E, Charnock-Jones DS, White IR, Smith GC. Study protocol. A prospective cohort study of unselected primiparous women: the pregnancy outcome prediction study. *BMC Pregnancy Childbirth*. 2008;8(1):51. doi:10.1186/1471-2393-8-51

24. Sovio U, White IR, Dacey A, Pasupathy D, Smith GCS. Screening for fetal growth restriction with universal third trimester ultrasonography in nulliparous women in the Pregnancy Outcome Prediction (POP) study: a prospective cohort study. *The Lancet*. 2015;386(10008):2089-2097. doi:10.1016/S0140-6736(15)00131-2

25. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z

26. Aragam KG, Jiang T, Goel A, et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat Genet*. 2022;54(12):1803-1815. doi:10.1038/s41588-022-01233-6

27. CHARGE Consortium Hematology Working Group, Early Growth Genetics (EGG) Consortium, Horikoshi M, et al. Genome-wide associations for birth weight and correlations with adult disease. *Nature*. 2016;538(7624):248-252. doi:10.1038/nature19806

28. Graham SE, Clarke SL, Wu KHH, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2021;600(7890):675-679. doi:10.1038/s41586-021-04064-3

29. Pagès H. SNPlocs.Hsapiens.dbSNP150.GRCh38. Published online 2017. doi:10.18129/B9.BIOC.SNPLOCS.HSAPIENS.DBSNP150.GRCH38

30. Spiller W, Slichter D, Bowden J, Davey Smith G. Detecting and correcting for bias in Mendelian randomization analyses using Gene-by-Environment interactions. *Int J Epidemiol*. Published online November 20, 2018. doi:10.1093/ije/dyy204

31. Tchetgen Tchetgen E, Sun B, Walter S. The GENIUS Approach to Robust Mendelian Randomization Inference. *Stat Sci*. 2021;36(3). doi:10.1214/20-STS802

32. Si S, Hou L, Chen X, et al. Exploring the Causal Roles of Circulating Remnant Lipid Profile on Cardiovascular and Cerebrovascular Diseases: Mendelian Randomization Study. *J Epidemiol*. 2022;32(5):205-214. doi:10.2188/jea.JE20200305

33. Wang K, Shi X, Zhu Z, et al. Mendelian randomization analysis of 37 clinical factors and coronary artery disease in East Asian and European populations. *Genome Med*. 2022;14(1):63. doi:10.1186/s13073-022-01067-1

34. Sovio U, Moraitis AA, Wong HS, Smith GCS. Universal *vs* selective ultrasonography to screen for large-for-gestational-age infants and associated morbidity. *Ultrasound Obstet Gynecol*. 2018;51(6):783-791. doi:10.1002/uog.17491

35. Smit RAJ, Jukema JW, Postmus I, et al. Visit-to-visit lipid variability: Clinical significance, effects of lipid-lowering treatment, and (pharmaco) genetics. *J Clin Lipidol*. 2018;12(2):266-276.e3. doi:10.1016/j.jacl.2018.01.001

863 36. de Vries PS, Brown MR, Bentley AR, et al. Multiancestry Genome-Wide Association Study
864 of Lipid Levels Incorporating Gene-Alcohol Interactions. *Am J Epidemiol*.
865 2019;188(6):1033-1054. doi:10.1093/aje/kwz005

866 37. Francis M, Li C, Sun Y, et al. Genome-wide association study of fish oil supplementation
867 on lipid traits in 81,246 individuals reveals new gene-diet interaction loci. *PLoS Genet*.
868 2021;17(3):e1009431. doi:10.1371/journal.pgen.1009431

869 38. Ruth KS, Day FR, Tyrrell J, et al. Using human genetics to understand the disease impacts
870 of testosterone in men and women. *Nat Med*. 2020;26(2):252-258. doi:10.1038/s41591-
871 020-0751-5

872 39. Pott J. pottj/LongitudinalMR_POPS_UKBB: Link to zenodo. Published online March 19,
873 2025. doi:10.5281/ZENODO.15052248

874