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

Janne Pott[1], Marco Palma[1,2], Yi Liu[3], Jasmine A. Mack[4], Ulla Sovio[4], Gordon C. S. Smith[4], Jessica Barrett[1], Stephen Burgess[1]

[1] MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
[2] Population, Policy and Practice Research and Teaching Department, UCL Great Ormond Street Institute of Child Health, UK
[3] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[4] Department of Obstetrics and Gynaecology, University of Cambridge, NIHR Cambridge Biomedical Research Centre, Cambridge, UK

Corresponding authors:
- Janne Pott (janne.pott@mrc-bsu.cam.ac.uk)
- Stephen Burgess (stephen.burgess@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 twelve scenarios for power and type I error, depending on shared instruments between the mean, slope and variability, and regression model specifications. 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 shared SNPs 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 independent effect of the variability. However, we only had weak instruments in both data sets.

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, accounting for known biology and the trajectory of the exposure.

Introduction

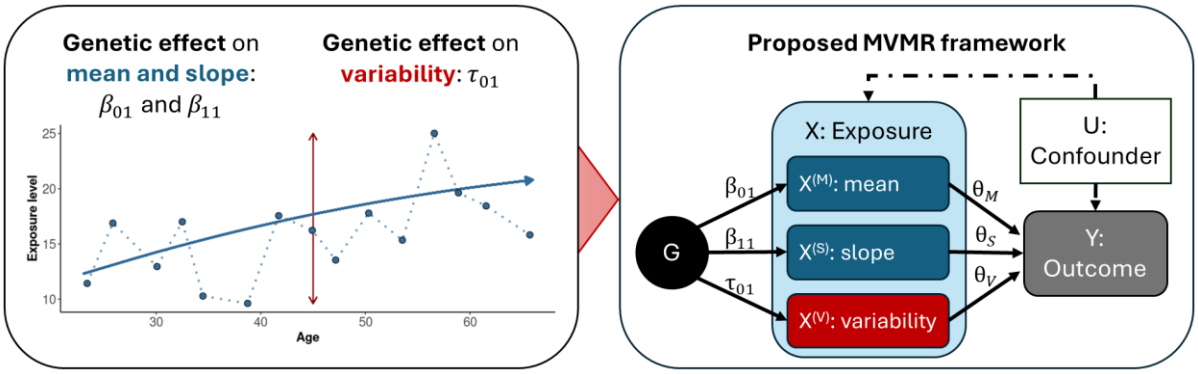
Mendelian Randomization (MR) allows to estimate the causal effect of an exposure X on an outcome Y using genetic variants (SNPs) as instrumental variables (IVs), as long as certain assumptions are met. [1] The SNPs must be strong predictors of X, while being independent of observed and unobserved confounders 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 effects on the mean level of the trait using linear mixed models (LMMs) 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 shortcomings. 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 LMM 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 [11,12] and/or the variability of the exposure (within-subject variability). [13] Wiegrebe at al. analysed genetic effects on trajectories of kidney decline, [14] and compared seven statistical approaches. Their comparison included different models, e.g. difference between two time points vs. LMM, and different time settings, e.g. time since baseline or age at each time point. They found that a LMM with random intercept and random slope using age was the most powerful approach with unbiased effect estimates. They also tested for effects on variability and concluded that decline and variability were two independent aspects of their longitudinal analyses.

In this work, we bring together methods to estimate SNP associations in longitudinal data in the presence of growth, decline, or variability over time using LMMs, 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**.

Figure 1: Overview of the proposed framework. First, we will use longitudinal exposure data to estimate the SNP effects on the mean, slope and variability of the exposure X (β_{01}, β_{11}, and τ_{01}, respectively). In the second step, these effects are used in a multivariable Mendelian Randomization (MVMR) approach, to test the genetically proxied mean levels of exposure X, denoted $X^{(M)}$, the time-dependent growth of X, denoted $X^{(S)}$, and the variability of X, denoted $X^{(V)}$, for independent effects θ_i on an outcome Y.



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.

Framework for longitudinal data in MVMR

Linear mixed models (LMMs)

Genome-wide association studies (GWAS) test millions of genetic variants for association with a specific trait or disease. [15] 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 linear 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. We choose age over time since baseline here, as others have shown that this choice works better in a GWAS setting. [14] β is the p-dimensional vector of regression coefficients for the predictors (fixed effects

common to 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:

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

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", [16] GMMAT (Generalized Linear Mixed Model Association Tests), [4,12] 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 a 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 model 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 (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. [17] 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". [18] 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. [13] developed a GWAS implementation, TrajGWAS, to estimate both the mean and within-subject variability using longitudinal data. It can be re-written as

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

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.

Throughout this work, we will focus on linear growth (slope) and the interaction between the linear age effect and SNPs (β_{11}). To allow potential non-linear growth, 2nd degree polynomials for the relationships of age can be added, e.g. adding age squared as covariable in the μ- and σ-function. However, we refrained from including SNP interactions with age squared to retain interpretability.

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] 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 [19]:

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

Here, $do(X = x_0)$ means forcing the continuous exposure X on level x_0. It indicates a causal expression, as in general $\mathbb{E}\left(Y\middle|do(X = x_0)\right) \neq \mathbb{E}(Y|X = x_0)$ ('correlation is not causation'). This approach has been extended to include multiple instruments, [20] 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, [1,21] 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 [21]:

(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 relevance assumption 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

$$\begin{aligned}\hat{\beta}_{GY} &= \theta\hat{\beta}_{GX} + \varepsilon \\ &= \theta_M\hat{\beta}_{GX^{(M)}} + \theta_S\hat{\beta}_{GX^{(S)}} \times T + \theta_V\hat{\beta}_{GX^{(V)}} + \varepsilon \\ &= \theta_M\hat{\beta}_{01} \quad + \theta_S\hat{\beta}_{11} \quad \times T + \theta_V\tau_{01} \quad + \varepsilon,\end{aligned}$$

with $X^{(M)}$, $X^{(S)}$, and $X^{(V)}$ denoting the mean, slope and variability of the exposure X, respectively (see also **Figure 1**), and T representing the average age in the cross-sectional outcome population, e.g. age at time of recruitment. In other words, the exposure X is partitioned into three distinct parts, and each part is tested for an independent effect on the outcome. While standard MVMR regression models such as MVMR-IVW should be able to directly detect unbiased causal estimates for $X^{(M)}$ and $X^{(V)}$, the effect of $X^{(S)}$ needs correction: either the SNP effects $\hat{\beta}_{11}$ or the estimated effect $\hat{\theta}_S$ must be multiplied with or divided by the mean age T, respectively. Throughout this work, we present the latter, with $\hat{\theta}_S$ representing the causal effect per time unit, e.g. per year.

One challenge of MVMR is that if the genetic variants are only weakly associated with the exposures, then the IVW approach will result in biased estimates for the exposure effect. To address the relevance assumption in MVMR, Sanderson et al. developed a conditional F-statistic framework to test if the SNPs strongly predict each exposure conditional on the other exposures included in the model. [22] We will use these conditional F-statistics in our work, but note that the setting differs from the traditional MVMR analysis: Sanderson et al. assumed multiple exposures, obtained from separate GWASs. [22] Hence, the summary statistics of one exposure are per default not conditional on the other exposures, unless they were included as covariates. In our longitudinal GWAS, the summary statistics are already conditional on each other, as they were obtained from the same model. Given the expected correlation between the mean and slope effects, we will check the conditional F-statistics in our simulation and real data analyses but not consider a hard threshold for reporting.

Simulation study

Overview

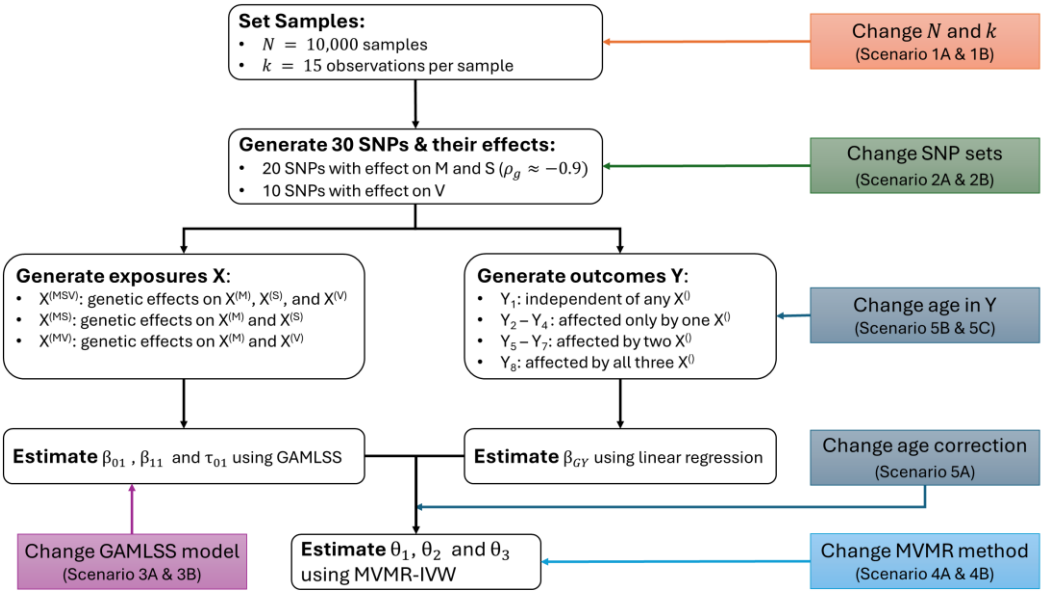
We performed a simulation study to evaluate the proposed method combining longitudinal SNP effect estimates in a MVMR approach. The workflow steps were as follows (see **Figure 2**):

1) Set $N = 10{,}000$ samples with $k = 15$ observations per sample, with one observation every 4 years (average age between 0 and 60)
2) Set $M = 30$ SNPs and simulate for 20 SNPs an effect on mean and slope with $\rho_g = -0.9$, and simulate for 10 SNPs an effect on the variability
3) Generate three exposures:

a. $X^{(MSV)}$: genetic effects on $X^{(M)}$, $X^{(S)}$, and $X^{(V)}$
b. $X^{(MS)}$: genetic effects on $X^{(M)}$ and $X^{(S)}$
c. $X^{(MV)}$: genetic effects on $X^{(M)}$ and $X^{(V)}$

4) Generate eight outcomes (all possible combinations of $X^{()}$s affecting Y_i), with $\theta = (\theta_M \quad \theta_S \quad \theta_V)^T = (1.2 \quad 0.3 \quad 1)^T$
5) Estimate SNP effects on the exposure's mean, slope, and variability using GAMLSS
6) Estimate SNP effects on outcome using linear regression
7) Estimate causal effects of mean, slope and variability of each $X^{()}$ on Y_i using MVMR-IVW
8) Repeat simulation $S = 500$ times

Figure 2: Overview of the main simulation study. Coloured boxes indicate where the additional scenarios differed from the main scenario. For all scenarios, we repeated the process $S = 500$ times. EAF: effect allele frequency; ρ_g: genetic correlation; M: mean; S: slope; V: variability



Given this workflow, we considered $3 \times 8 = 24$ sets of parameters, with three possibilities for the genetic effect on the exposure and eight possibilities for the effects of the exposure components on the outcome. All details for the simulation setup can be found in the **Supplemental Material**, following the ADEMP framework established in Morris et al. [23] In **Supplemental Table S1a and S1b** we report all used parameter settings. In addition to the main scenario (denoted scenario 0), we tested eleven additional scenarios, split into five categories:

1) <u>Change N and k</u>: to mimic the real data sets used later in the application, we reduced the sample size and number of observations (scenarios 1A and 1B, respectively).
2) <u>Change SNP sets</u>: In the main scenario, we used two distinct SNP sets to ensure that the variability was well powered and distinct from the mean effect. We compared that to one scenario in which mean, slope and variability had distinct SNP sets of each 10 SNPs (scenario 2A), and to one in which there was only one set for all (shared genetic effect of all 30 SNPs on mean, slope and variability, scenario 2B).
3) <u>Change GAMLSS model</u>: we compared the model specification by removing either the SNP x age interaction in the μ-function or the SNP effect in the σ-function (scenarios 3A and 3B, respectively).

4) Change MVMR method: We changed the MVMR method from the IVW to the more robust MVMR Generalized Method of Moments (GMM) approach for both the main scenario and 2B (scenarios 4A and 4B, respectively).
5) Change age correction and age in Y: To check the effect of the age correction on the slope estimates, we used wrong age correction and different outcome ages (scenarios 5A, 5B, and 5C).

Supplemental Figure S1 shows example trajectories of the main, 1A, 1B, 2A, and 2B scenarios. The other scenarios have the same exposure data as the main one.

Simulation results

Main scenario

In the main scenario, we simulated three different exposures, $X^{(MSV)}$, $X^{(MS)}$, and $X^{(MV)}$ and eight outcomes, Y_1 – Y_8, and tested if we can detect the independent effects of the mean, slope and variability on the outcome, depending on the exposure-outcome combination. All results including power, coverage, and bias are given in **Supplemental Table S2**. In this scenario, we only simulated a genetic correlation between mean and slope, but not with the variability. Indeed, when checking the correlation of the estimated SNP effects, we observed this correlation structure, when restricting to the two SNP sets. For example, the genetic correlation between the mean and slope was on average -0.89 when using the mean-specific SNP set. However, when using all SNPs, we also observed a significant genetic correlation between effects on mean and variability, and between effects on slope and variability (on average 0.68, and -0.43, respectively).

On average, there were 19 SNPs associated with the mean across all exposures with $p < 1 \times 10^{-6}$. For the slope, we also detected 19 SNPs on average for the exposure $X^{(MSV)}$ and $X^{(MS)}$, while no associations were found for $X^{(MV)}$. This was expected, as $X^{(MV)}$ was simulated without SNP effect on the slope. Similarly, we detected 10 SNPs on average for the variability in exposures $X^{(MSV)}$ and $X^{(MV)}$, while no association was found here for $X^{(MS)}$. This was directly reflected in the conditional F-statistics, which were above 10 for all exposure types of $X^{(MSV)}$, but below 10 for the variability in $X^{(MS)}$ and the slope in $X^{(MV)}$ (see **Supplemental Figure S2**). In addition, the median of the conditional F-statistics for the variability in $X^{(MV)}$ was also only 7.6, with an interquartile range from 2.3 to 51.7.

We observed a high power to detect all causal effects in the $X^{(MSV)}$ exposure, in combination with low type I error rates (see **Figure 3**). Here, high power means the effect was detected in all replicates (power = 100%). For the exposure $X^{(MS)}$, we detected high power for the mean and slope effects, but none for the variability. This was due to the lack of strong instruments in this setting. However, the situation for the exposure $X^{(MV)}$ was more complicated: due to the correlation structure, we had false positives for the mean whenever a slope effect was included, e.g. in outcomes Y_3 or Y_7. A false positive slope effect was detected for Y_2 and Y_6, probably also due to the correlation between mean and slope. For the variability, there was high power to detect the effect in case of variability only (Y_4) or mean and variability (Y_6), but only moderate power when combining it with the slope (e.g. Y_8 with a power of 56% to detect the variability effect).

Figure 3: Power/Type I error in main scenario. The columns are ordered by exposure type (mean, slope, and variability) and exposure composition ($X^{(MSV)}$, $X^{(MS)}$, and $X^{(MV)}$, while the rows are ordered by outcomes. We added for each outcome which true θ_i should be detectable, with θ_M, θ_S, and θ_V indicating a simulated mean, slope or variability effect, respectively. A value of 1 indicates that in all 500 replicates for this exposure, exposure type and outcome a significant θ_i was detected (with p(θ_i)<0.05).

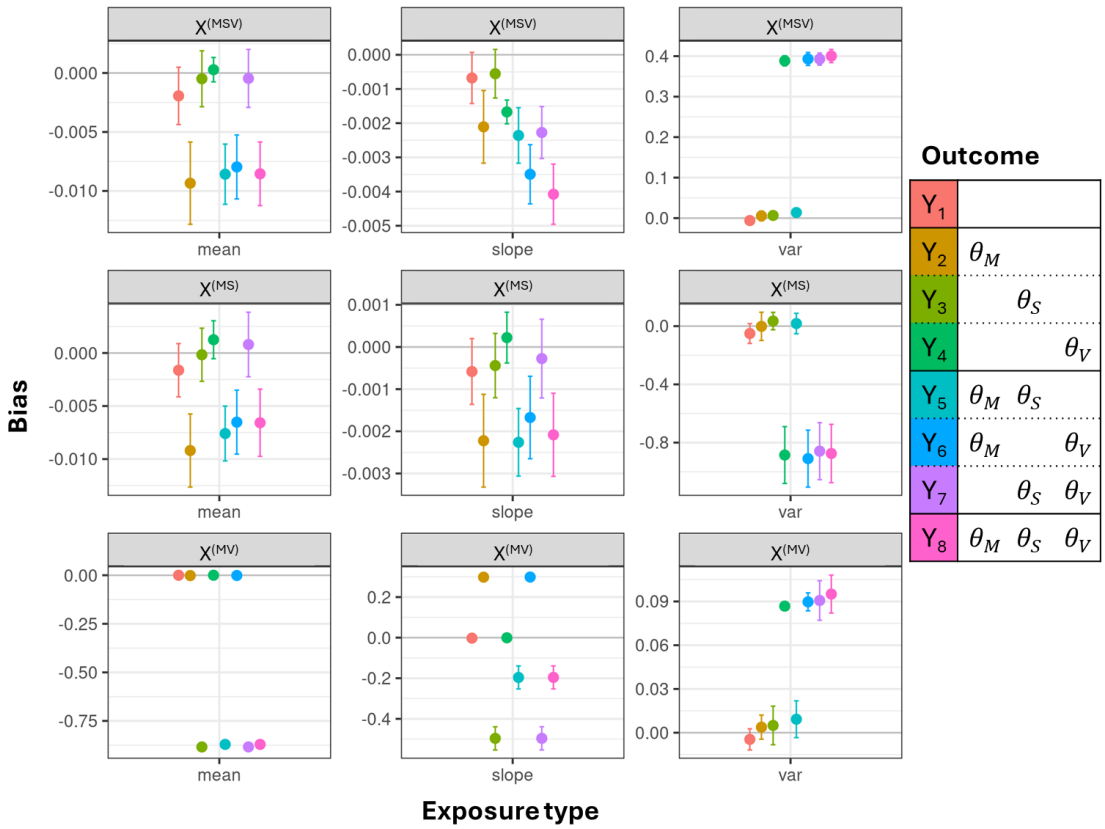
Main Scenario		$X^{(MSV)}$			$X^{(MS)}$			$X^{(MV)}$		
		Mean	Slope	Var.	Mean	Slope	Var.	Mean	Slope	Var.
	Y_1	0.04	0.05	0.05	0.04	0.04	0.05	0.04	0.05	0.06
θ_M	Y_2	1	0.03	0.04	1	0.03	0.03	1	0.57	0.01
θ_S	Y_3	0.01	1	0.01	0.02	1	0.01	1	0.06	0
θ_V	Y_4	0.05	0.09	1	0	0	0.06	0.07	0.05	1
θ_M θ_S	Y_5	1	1	0.06	1	1	0.05	0.87	0.06	0
θ_M θ_V	Y_6	1	0.06	1	1	0	0.07	1	0.8	1
θ_S θ_V	Y_7	0.02	1	1	0	1	0.08	1	0.06	0.54
θ_M θ_S θ_V	Y_8	1	1	1	1	1	0.07	0.87	0.06	0.56

Finally, we checked the bias of the estimates per exposure and outcome. For the well-powered mean and slope of $X^{(MSV)}$ and $X^{(MS)}$ the bias was marginal (average of -0.004 for the mean and -0.001 for the slope, see also **Supplemental Table S2** and **Figure 4**). For the variability we observed significant bias for $X^{(MSV)}$, where the effect was overestimated (average bias of 0.39 for Y_4, Y_6, Y_7 and Y_8). For $X^{(MS)}$, the effect of the variability was underestimated, as it was not detectable without strong instruments (average bias of -0.88 for Y_4, Y_6, Y_7 and Y_8). For the exposure $X^{(MV)}$, the bias was more outcome specific: the mean was underestimated in case of outcomes affected by the slope (Y_3, Y_5, Y_7, and Y_8), with an average bias of -0.88. This was due to the genetic correlation between the mean and slope: given the input genetic effects ($\beta_{01} = -0.2$ and $\beta_{11} = 0.01$) and the high negative correlation of -0.9, the genetic effect on the slope can be approximated with $\beta_{11} \approx -0.05 \times \beta_{01}$. Now, as the slope is not well-powered (cond. F = 1.8 on average), the slope effects are attributed to the mean. However, the mean was not corrected for the outcome age, which was on average 60. Taken together, when $\theta_S = 0.3$, we have an additional effect of the mean which is

$$-0.05 \times 60 \times \theta_S = -0.05 \times 60 \times 0.3 = -0.9$$

The effect of the slope of $X^{(MV)}$ was overestimated in case of Y_2 and Y_6 (false positives) and underestimated for all slope-related outcomes (Y_3, Y_5, Y_7, and Y_8, false negatives). The variability was overestimated, but less than in the $X^{(MSV)}$ case (average bias of 0.09).

Figure 4: Bias in main scenario. For each exposure type - exposure – outcome combination we estimated the bias and its error. The exposure types, mean, slope, and variability, are given on the x-axis, while the rows are displaying an exposure each. The colours indicate the outcome. The solid grey lines indicate no bias.



Other scenarios

In the following, we describe the results of the additional scenarios in relation to the main scenario. **Table 1** illustrates this comparison by providing power and bias of exposure $X^{(MSV)}$ on outcome Y_8 for all scenarios. We focus here on outcome Y_8, as it is the one with simulated effects of all three exposure types. The conditional F-statistics are the same for all outcomes, and in most settings the biases are similar when there is a non-null effect, with the exception being $X^{(MV)}$ and the slope effect. We restricted **Table 1** to $X^{(MSV)}$ for clarity; similar Tables for $X^{(MS)}$ and $X^{(MV)}$ can be found in **Supplemental Table S3 and S4**, respectively.

Changing the sample size of the data

We reduced the sample size and number of time points to approximate the real data in our later application approach. In scenario 1A, we reduced to $N = 3000$ individuals and $k = 3$ time points, similar to the POPS data. Here, we observed a strong reduction in the conditional F-statistics, which were no longer above 10 in any exposure for any exposure type. The only exception of that rule was the variability in $X^{(MSV)}$ with a median conditional F-statistic of 12.0 (see **Supplemental Figure S3 and S4** for all conditional F-statistics across the scenarios). Although the conditional F-statistics were low, we still detected high power similar to the main scenario for $X^{(MSV)}$ and $X^{(MS)}$ (see **Supplemental Figure S5** for power heat maps across all scenarios). The type I error increased, however, e.g. from 2% to 23% for outcome Y_7 and $X^{(MSV)}$. For $X^{(MV)}$, we found similar false

positives as before, but also detected higher power for the variability, e.g. from 56% to 84% for outcome Y_8. The bias for the mean increased due to the weaker instruments, but interestingly it stayed similar for the variability (see **Supplemental Figure S6**).

We observed a similar case for the second data reduction to approximate the UKB data (scenario 1B): more samples ($N \approx 17{,}000$), but randomly missing time points, with an average of $k = 7$ time points per individual. Although the conditional F-statistics were reduced for the mean and slope, we still detected overall similar power and bias as in the main scenario.

Table 1: Overview of simulation results for exposure $X^{(MSV)}$ and outcome Y_8. We expected for this exposure strong instruments for all three exposure types (mean, slope, and variability). The outcome was simulated with an effect of all three exposure types. Here we report the power to detect the respective effect, the bias, and the median conditional F-statistic over all 500 replicates per scenario. We highlighted power<0.8, |bias|>0.1 and cond. F-statistic<10. Scenarios 1: change in sample size; scenarios 2: change in genetic correlation structure; scenarios 3: change in GAMLSS regression; scenarios 4: change in MVMR approach; scenarios 5: change in outcome age and correction.

	Mean			**Slope**			**Variability**		
scenario	**Power**	**Bias**	**Cond. F**	**Power**	**Bias**	**Cond. F**	**Power**	**Bias**	**Cond. F**
main	1.000	-0.009	51.1	1.000	-0.004	52.1	1.000	**0.400**	115.4
1A	0.998	**-0.179**	**0.9**	0.942	-0.047	**0.9**	**0.752**	**-0.123**	12.0
1B	1.000	-0.024	19.6	1.000	-0.007	20.2	1.000	**0.205**	119.4
2A	1.000	-0.001	469.4	1.000	-0.001	498.6	1.000	**0.237**	149.9
2B	1.000	**-0.138**	**3.5**	1.000	-0.020	**4.5**	**0.068**	**-0.743**	**5.1**
3A	0.904	**-0.708**	502.6				0.936	**0.955**	116.5
3B	1.000	-0.009	49.6	1.000	-0.002	50.5			
4A	1.000	-0.009	234.4	1.000	-0.004	231.4	1.000	**0.400**	113.6
4B	1.000	**-0.139**	**6.5**	1.000	-0.020	10.7	**0.062**	**-0.743**	**7.0**
5A	1.000	-0.009	51.1	1.000	**0.218**	52.0	1.000	**0.400**	115.4
5B	1.000	-0.008	51.1	1.000	-0.007	52.0	1.000	**0.401**	115.4
5C	1.000	-0.008	51.1	**0.350**	-0.046	52.1	1.000	**0.402**	115.7

Changing the SNP sets

To test for an effect with distinct or shared SNP sets, we simulated data with distinct SNP sets for each exposure type (2A) and with one shared set for all (2B). As expected, using distinct instruments for the exposure types as in 2A increased the conditional F-statistics (see **Supplemental Figure S3**), while maintaining the high power as observed in the main scenario. Of note, this was the only scenario with no false positives for the mean of $X^{(MV)}$ (see **Supplemental Figure S5**). The bias further decreased for all exposure types.

When assuming shared genetics for all types (2B), the conditional F-statistics decreased, and were below 10 for all exposures and all types. For the mean of $X^{(MS)}$, we observed false positive estimates for Y_4 (100%) and Y_7 (42%). For the mean of $X^{(MV)}$, we found lower power for Y_5 and Y_8 (both 51%). For the slope, the false positive rate increased in both $X^{(MSV)}$ and $X^{(MS)}$ for outcome Y_4 and Y_6 (61% and 38%). For the variability in $X^{(MSV)}$, we were only able to detect the effects on Y_4 and Y_7 (100% and 87%) but had no power to detect those on Y_6 and Y_8 (8% and 7%, respectively). For the variability of exposure $X^{(MV)}$, we found a higher type I error for outcomes affected by the slope

(Y_3 and Y_5, 72% and 78%, respectively), compared to the main scenario. The bias increased in most cases. For exposure $X^{(MV)}$ we found an average bias of -5.3 for the variability on outcomes with an expected slope effect (Y_3, Y_5, Y_7, and Y_8, see **Supplemental Figure S7**), suggesting that in case of shared SNP effects and weak instruments for the slope, the variability incorporates the slope effect to some degree. The high bias is most likely a result from the missing age-adjustment in the variability SNP estimates.

Changing the GAMLSS regression model

Next, we used the same simulated data as in the main scenario but fit a mis-specified GAMLSS by either removing the SNP x age interaction term in the μ-function (3A) or the SNP effect in the σ-function (3B). Removing the interaction term resulted in drastic increase in the F-statistics. For example, for $X^{(MV)}$ the conditional F-statistic of the mean increased from 53.1 to 3065.7, and that of the variability from 7.6 to 181. For $X^{(MS)}$, which was simulated without a variability effect, we also observed an increase from 1.9 to 5.7. Here we now observed high power for all but Y_1 and Y_4. In other words, in lack of the interaction term, the variability incorporated the slope effect, similar to the situation in scenario 2B. In line with this, the bias of the variability was 11.6 and 10.7 for outcomes with both mean and slope effect (Y_5 and Y_8), -16.3 and -17.3 for outcomes with a slope effect (Y_3 and Y_7), and 28.1 and 27.3 for outcomes with a mean effect (Y_2 and Y_6). The different direction of the bias is due to the different correlation structure and combination of exposure types. For $X^{(MSV)}$ and $X^{(MV)}$, the slope effect was incorporated in the mean effect, with a strong increase in bias (see **Supplemental Figure S8**).

When using the GAMLSS regression without a SNP effect in the sigma function (scenario 3B), we obtain similar results as in the main scenario in terms of conditional F-statistics, power and bias, just without the additional estimates for the variability.

Changing the MVMR method

We then changed the MVMR method from IVW to GMM, which is more robust in case of weak instruments. Otherwise, we used the same settings as in the main scenario and scenario 2B (scenarios 4A and 4B, respectively). For scenario 4A, we observed an impact on the conditional F-statistics of the mean and slope of $X^{(MSV)}$ and $X^{(MS)}$, but not on the variability or $X^{(MV)}$. We detected a slightly better power to detect the variability effect in $X^{(MV)}$ (from 56% to 84%). The bias was similar to the main scenario throughout (see **Supplemental Figure S9**). We compared scenario 4B to 2B, as the same correlation structure was present here. Overall, the results were similar (see **Supplemental Figure S5**).

Changing the age correction or average age of outcome data

Finally, we ran scenarios to check the impact of age at outcome on the power and bias. We first ran a scenario with same average outcome age as in the main scenario, but with a wrong age correction of the GAMLSS regression coefficients $\hat{\beta}_{11}$: instead of correcting for the age of 70 we only used age 40 (scenario 5A). This did not change the power of any MVMR approach but introduced bias of the slope for both $X^{(MS)}$ and $X^{(MSV)}$ (e.g. from -0.002 to 0.221 in Y_8, see **Supplemental Figure S10** and **Supplemental Table S2**). As expected, the bias was positive, as the correction was less than necessary.

We then used an average age of 40 and 5 years for the generation of the outcome data and corrected the slope then for this age (scenarios 5B and 5C, respectively). Here, the findings for the mean did not change. In scenario 5B, we observed similar power for the slope as before in the main scenario. In scenario 5C we lost some power for the slope of $X^{(MSV)}$ and $X^{(MS)}$, e.g. for Y_8 we

had 1% for exposure $X^{(MS)}$ and 35% for $X^{(MSV)}$. However, for $X^{(MV)}$ we observed higher power to detect the variability (e.g. for Y_8 from 56.2% in the main scenario to 99.4% in 5B and 100% in 5C), while the bias was similar to the main scenario (0.095 and 0.096, respectively, see also **Supplemental Figure S10**). The power for the mean effect did not change, but the mean effect was less biased than before (from -0.871 to -0.500 and -0.064 in 5B and 5C for Y_8, respectively, see **Supplemental Table S4**).

Summary of simulation study findings

We performed a simulation study with one main and eleven sensitivity scenarios, comparing the effects of sample size, genetic correlation, GAMLSS regression, and MVMR method on the simulation 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. Mis-specifying the GAMLSS model without the SNP x time interaction had the greatest impact on power. In this scenario, we found that the variability then includes the missing slope effect, while in other cases it is attributable to the mean. Wrong age correction does introduce bias to the slope effect, while outcome observation during the observation of longitudinal exposures did not lead to significant changes.

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, [24–26] 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), [24] 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 similar to scenario 1A. We used two phenotype definitions here: the log-transformation of EFW to obtain nearly linear growth and to reduce heteroscedasticity, and the gestational age adjusted Z-scores per scan (trajectories are given in **Supplemental Figure S11A** and **S11B**, and study description in **Supplemental Table S5**).

In a second test, we used total cholesterol (TC) data from the UKB [26] in combination with electronic health data from GPs. Here, we had a sample size of 68,467 individuals, but unstructured data in terms of both number of time points per sample, and time difference between the time points (see **Supplemental Figures S11C** and **S12B** for histograms in POPS and UKB data, respectively). The trajectories of TC in 15 randomly selected samples of the UKB are shown in **Supplemental Figure S12A**, and a study description is given in **Supplemental Table**

S6. 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. [27]

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 S13**. We followed the same procedure for both applications.

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 (www.egg-consortium.org, downloaded 06/02/2024, European-only meta-analysis of up to 143,677 individuals and 16,245,523 SNPs). [28] For the UKB, we used data from the Global Lipids Genetics Consortium (GLGC) (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). [29] Please note that we focused on the mean effects for candidate selection due to limited information on vQTLs of foetal weight, and high sample overlap for the TC vQTLs publications and the data used here (UK Biobank).

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 (Bioconductor annotation package "SNPlocs.Hsapiens.dbSNP150.GRCh38" [30] 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. SNP summary statistics are given in **Supplemental Tables S7** and **S8** for EFW and TC, 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 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:

$$\mu(EFW) \sim \boldsymbol{SNP}, GA_f, \boldsymbol{SNP} \times \boldsymbol{GA_f}, GA_f^2, height_m, smoking_m, sex_f, PC_f\ 1-5$$
$$ln(\sigma(EFW)) \sim \boldsymbol{SNP}, GA_f, height_m, smoking_m, sex_f$$

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:

$$\mu(TC) \sim \boldsymbol{SNP}, age, \boldsymbol{SNP} \times \boldsymbol{age}, sex, lipidMed, \text{PCs } 1-10$$
$$ln(\sigma(TC)) \sim \boldsymbol{SNP}, age, sex, lipidMed$$

For both EFW and TC, the main model was the full model, estimating the SNP effect on the mean, slope and variability simultaneously. We included a random intercept in the μ-function. For computational purposes, we refrained from including a random intercept in the σ-function.

Sensitivity analyses

For sensitivity purposes, we also tested four modifications:

1A) "No-slope": we removed the SNP x time interaction (similar to scenario 3A)

$$\mu(EFW) \sim \boldsymbol{SNP}, GA_f, GA_f^2, height_m, smoking_m, sex_f, PC_f\ 1-5$$
$$ln(\sigma(EFW)) \sim \boldsymbol{SNP}, GA_f, height_m, smoking_m, sex_f$$

$$\mu(TC) \sim \boldsymbol{SNP}, age, sex, lipidMe\text{d}, \text{PCs } 1-10$$
$$ln(\sigma(TC)) \sim \boldsymbol{SNP}, age, sex, lipidMed$$

1B) "No-variability": we removed the SNP effect on the variability (similar to scenario 3B)

$$\mu(EFW) \sim \boldsymbol{SNP}, GA_f, \boldsymbol{SNP} \times \boldsymbol{GA_f}, GA_f^2, height_m, smoking_m, sex_f, PC_f\ 1-5$$
$$ln(\sigma(EFW)) \sim GA_f, height_m, smoking_m, sex_f$$

$$\mu(TC) \sim \boldsymbol{SNP}, age, \boldsymbol{SNP} \times \boldsymbol{age}, sex, lipidMed, PCs\ 1-10$$
$$ln(\sigma(TC)) \sim age, sex, lipidMed$$

2) "Sample set": we restricted the POPS data to British ancestry and EFW availability at 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, $n = 46{,}890$). Hence, the model for EFW was the same as in the main model, while for TC the lipid medication variable was removed:

$$\mu(TC) \sim \boldsymbol{SNP}, age, \boldsymbol{SNP} \times \boldsymbol{age}, sex, PCs\ 1-10$$
$$ln(\sigma(TC)) \sim \boldsymbol{SNP}, age, sex$$

SNP effects on the outcomes

For the birth weight (BW) outcome 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 (raw values, not adjusted for gestational age). [2]

For the 1-sample approach of the TC analysis, we used the phenotype definition for coronary atherosclerosis ("I9_CORATHER") as defined in FinnGen (31,198 cases and 382,052 controls). [3] For the 2-sample approach, we downloaded the latest publicly available summary statistics from Aragam et al. (meta-GWAS of CAD with up to 181,522 cases and 984,168 controls). [27]

MVMR and instrument selection

We used the same MVMR-IVW method as described in the simulation study. In addition, we performed univariable MR-IVW to compare the estimates and (conditional) F-statistics. We included all SNPs, as they were already identified as relevant instruments in the consortia data ("3-sample MR approach"). For sensitivity purposes, we also restricted to 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. In the UKB TC data, we found a sufficient number of SNPs to also test other SNP selections:

- 3A) using SNPs with an effect on either mean or variability (effectively removing the genetic correlation between mean and variability);
- 3B) the top 20 associated SNPs per exposure type, allowing overlaps;
- 3C) SNPs with pleiotropic effects: known association with alcohol consumption, obesity/high BMI, diabetes, smoking, blood pressure or physical (in)activity;
- 3D) SNPs at genes with known biological effect (*PCSK9*, *HMGCR*, *NPC1L1*, *CETP*, *APOB*, *LPA*, *FADS2*, *APOA5*, *LDLR*, *TM6SF2*, and *APOE*);
- 3E) SNPs with significant sex-interaction using the sex-stratified GLGC data. [31]

Results

EFW on BW

Genetic associations with the exposure

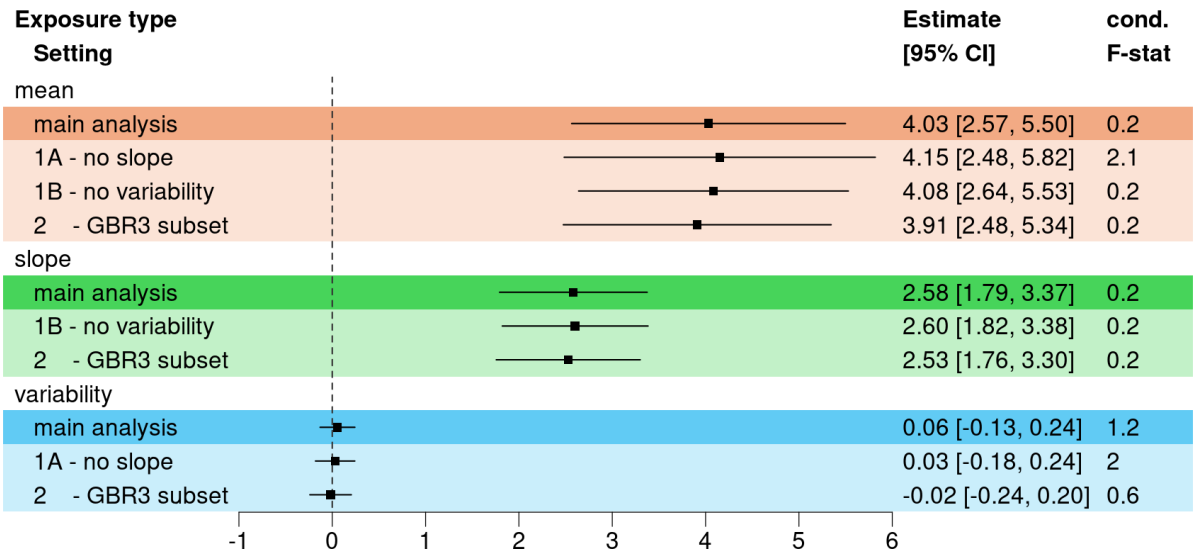
For EFW in POPS, we mainly got weak instruments using GAMLSS, due to the sample size and number of time points, and only one genome-wide significant association with the mean EFW (rs35261542, see **Supplemental Table S7** for all summary statistics). Of the 52 SNPs, 24 reached nominal significance (p<0.05) for at least one of the three exposure types using the log-transformed EFW. The genetic correlation 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 ($r_g = 0.15, p = 0.289$). The effect estimates were highly similar in most models with the notable exception being the "no-slope" model (sensitivity check 1A, see **Supplemental Figure S14**). Removing the SNP x time effect changed the estimates for the mean, and the correlation was significant with the slope estimates from the main model ($r_g = 0.30, p = 0.028$). This was to be expected, as the SNP effect in the "no-slope" model combines the effects on the intercept, β_{01}, and the slope of the trajectory, β_{11}. As a result, there were 17 genome-wide significant SNPs for the mean, but still none for the variability. Using the EFW Z-scores as trait, we found similar results for the genetic correlation (see **Supplemental Figure S15**, still high negative

correlation between mean and slope effects, and a weak positive correlation between slope and variability), but only 17 associated SNPs.

MVMR results

In the main analysis, we detected for EFW significant positive estimates of the mean and the slope on BW (estimates [95% confidence interval (CI)]: $\hat{\theta}_M = 4.08\ [2.62; 5.53]$; $\hat{\theta}_S = 2.63\ [1.84; 3.41]$), while no effect of the variability was observed ($\hat{\theta}_V = 0.04\ [-0.14; 0.22]$). **Figure 5** displays a Forest Plot of 2-sample MVMR results of log-transformed EFW and **Supplemental Figure S16** shows scatter plot of SNP effects on the exposure types and outcome. Similar results were observed when using the 1-sample approach (**Supplemental Figure S17**) or when using nominally significant SNPs only (**Supplemental Figure S18**).

Figure 5: Forest Plot of the causal estimates of estimated foetal weight on birth weight using the 2-sample approach. We used 52 instruments, and the genetic effect estimates on the slope were corrected using the average gestational age in POPS. The last column of the plot indicates the conditional F-statistics for the MVMR approaches. 1A - no slope: GAMLSS with no SNP x time interaction; 1B - no variability: GAMLSS with no SNP effect in the σ-function; 2 - GBR3: using British samples with data at all time points.



Finally, we compared the MVMR results to the MR results per exposure type. In the main model with three exposure types, the estimated mean effect in the MR was negative ($\hat{\theta}_M = -0.50\ [-1.09; 0.09]$). 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. When using the "no slope" model (sensitivity test 1A), we detected a significant effect of the mean of EFW on BW, and the effect size was similar to the detected MVMR effects ($\hat{\theta}_M = 4.19\ [2.55; 5.82]$, see **Supplemental Figure S19**). The variability of EFW was still not significant.

The results for the EFW Z-scores were overall very similar (see **Supplemental Table S9** and **S10**). Of note, the MR-IVW detected a significant variability effect here, which was lost in the MVMR ($\hat{\theta}_V = 0.37\ [0.15; 0.59]$ in the MR, and $\hat{\theta}_V = 0.14\ [-0.02; 0.31]$ in the MVMR).

Power

The power in the multivariable approach was low, with conditional F-statistics of 0.2 for mean and slope, and 1.4 for the variability. This was to be expected, given the relatively small sample size, instrument association p-values, and correlation structure between mean and slope effects. In the simulation study with similar sample size (scenario 1A) we had similar conditional F-statistics (e.g. 0.9 for mean and slope and 1.5 for the variability of exposure $X^{(MS)}$, see **Supplemental Table S3**). In the sensitivity analysis using GAMLSS without SNP x age interaction we observed slightly higher conditional F-statistics (2.2 for the mean, and 2.1 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: 108 for the mean, 66 for the slope and 75 for the variability. There was a high overlap between the associated SNP (see **Supplemental Figure S20** for Venn Diagram), with 26 SNPs shared between all three exposure types (e.g., genome-wide significant SNP effects on all three), and 24 and 17 specific for the mean and variability, respectively (using $p_M < 5 \times 10^{-8} \wedge p_V > 0.05$ and $p_V < 5 \times 10^{-8} \wedge p_M > 0.05$ as cut-offs). Similarly to the POPS data, we observed a high negative correlation between the mean and slope SNP effects ($r_g = -0.92, p = 7.99 \times 10^{-141}$). However, as expected with the observed high overlap, there was also a significant correlation between the mean and variability ($r_g = 0.65, p = 1.71 \times 10^{-42}$) and between the slope and variability ($r_g = -0.52, p = 4.95 \times 10^{-25}$). All SNP summary statistics are given in **Supplemental Table S8**.

MVMR results

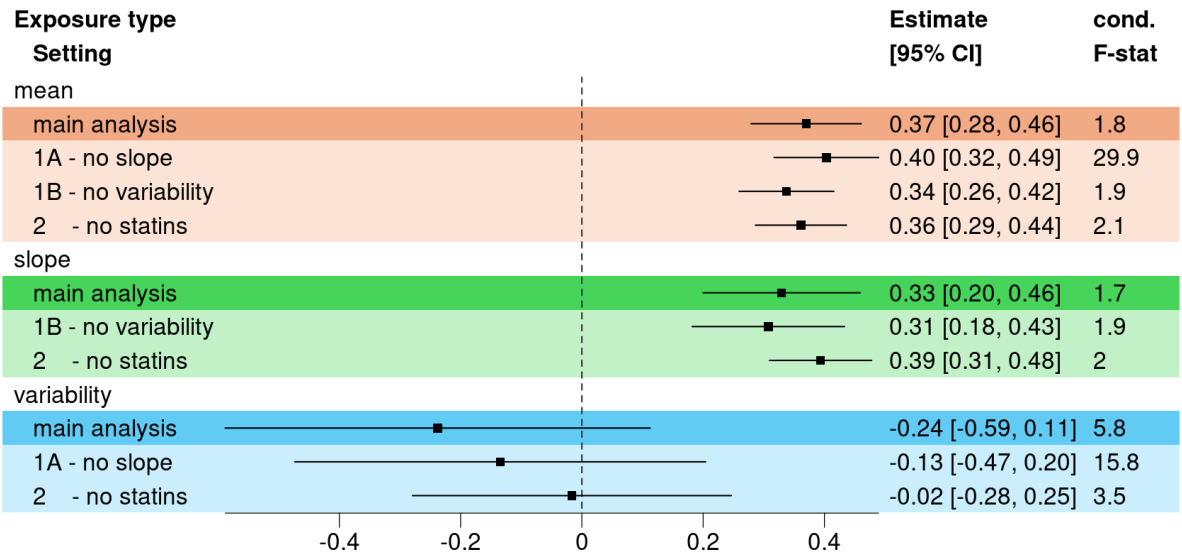
In the main analysis, we detected again significant positive effects of the mean and the slope on the risk for CAD ($\hat{\theta}_M = 0.37\ [0.28; 0.46]$, $\hat{\theta}_S = 0.33\ [0.20; 0.46]$). All results are summarized in **Supplemental Table S11 and S12**. In **Figure 6** the Forest Plot for the main analyses is given, and **Supplemental Figure S21** shows the respective scatter plots. However, even with the increased sample size and instrument strength, the conditional F-statistics were low (1.75 for mean, 1.68 for slope). The variability was better powered (5.84), but we did not detect any significant results here ($\hat{\theta}_V = -0.24\ [-0.59;\ 0.11]$). In the sensitivity analysis using a GAMLSS regression model with no SNP x age interaction, we observed high conditional F-statistics (29.9 for the mean, and 15.8 for the variability), and the effect estimates were of the same magnitude as in the main analysis. Using the statin-free subset, we observed similar estimates for the mean and slope, and the variability effect was null. We observed similar results in all sensitivity checks using 1-sample MVMR (see **Supplemental Figure 22**).

We also tested various SNP selections: using only genome-wide significant SNPs or the top 20 associated SNPs we found similar results as before (see **Supplemental Figure S23**). The same was true for SNPs with sex-interactions or known biological meaning but given the small number of instruments the error increased and the slope effect was no longer significant. When using only pleiotropic hits, we detected no significant effect anymore: the mean effect was estimated $\hat{\theta}_M = 0\ [-0.21;\ 0.22]$, and the slope was $\hat{\theta}_S = -0.12\ [-0.38;\ 0.14]$. For the variability, the effect was still not significant: $\hat{\theta}_V = 0.40\ [-0.30;\ 1.10]$.

Finally, we compared the MVMR and MR results (see **Supplemental Figure S24**). As in the POPS study, the "no slope" MR estimate for the mean was of similar size as the MVMR estimates ($\hat{\theta}_M =$

$0.38\ [0.32;\ 0.44]$). Of note, the variability was associated with the CAD risk in the univariate model: $\hat{\theta}_V = 0.95\ [0.68;\ 1.23]$.

Figure 6: Forest Plot of the causal estimates of total cholesterol levels on coronary artery disease risk using the 2-sample approach. We used 335 instruments, and the genetic effect estimates on the slope were corrected using the average age in Aragam et al.[27] The last column of the plot indicates the conditional F-statistics for the MVMR approaches. 1A - no slope: GAMLSS with no SNP x time interaction ; 1B - no variability: GAMLSS with no SNP effect in the σ-function; 2A – no statins: UKB data set restricted to individuals with no lipid-lowering medication.



Discussion

Overview

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. We tested our approach in a simulation study and observed an overall high power to detect unbiased 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.

General comments on the framework

The increased availability of electronic health records (EHR) linked to biobank data allows the analysis of genetics of the within-individual variability and the temporal change of a trait of interest or disease progression, [14,32–34] using tools such as trajGWAS or GAMLSS. [13,18] Compared to clinical trial data or longitudinal studies, EHR data points are spread seemingly random over time, with each individual having different number of observations at different ages. However, each observation represents a time point a patient visited their GP. In our simulation study, we assumed for the main scenario that each individual had observations at similar time points in life. In one sensitivity scenario, we then randomly removed 53% of all observations. Still, power and bias stayed similar to the main scenario. In another scenario, we used only 3 time points, to mimic typical longitudinal studies with two follow-ups after some years. Here, we still had acceptable power to detect an effect, with an increased bias. Taken together, our simulations

suggest that in this case the approach performs better in the unbalanced EHR data than in the balanced cohort data, as the number of observations appear to be more important than having similar age ranges in the dataset. However, the potential for informative observation bias with EHR data is a concern.

We included age squared as a covariate to allow non-linear growth of the trajectories but focused only on the linear SNP x age interactions. Before using these estimates in MVMR, they had to be corrected for the age at outcome, otherwise the MVMR estimate of the slope would have been on a different scale. Using the wrong age led to increased bias in our simulation study, while correct outcome age adjustment even allows usage of outcome events throughout the exposure observation. However, if the outcome occurs before the exposure data, the temporality criteria of causality is violated. [35] In this scenario, we observed no power to detect a slope effect but still identified unbiased effects of the mean and variability, as they are both time-independent. In other words, including the SNP x age interaction term in the longitudinal GWAS allows to estimate the true time-independent SNP effect on the mean and variability, which results in unbiased causal estimates regardless of temporality. This might be of use if limited exposure data is available, e.g. for children. In that case, one might borrow information of better powered adult studies.

In our simulation, we primarily focused on power as a performance metric, as the initial aim was to test if this framework can detect slope and variability effects. Alternatively, one could have concentrated on the conditional F-statistics as a measure of power and risk of weak instrument bias in MVMR. [22] Indeed, in the scenarios with smaller sample size or less frequent observations we had weaker instruments and larger bias. However, it is unclear if the conditional F-statistics as introduced by Sanderson and colleagues [22] are the correct measures here: they assumed that the exposures were taken from separate GWASs and hence requiring conditioning onto each other, we used SNP estimates that are already corrected for each other as they were estimated in one model. We would also like to point out that in MR the effect estimates might not be of high interest, as the life-long effect does not reflect a relevant intervention on the exposure. [36] Instead, the significance and effect direction is of more relevance. In our simulation, the power was good. Only in scenarios with extreme changes such as one shared SNP set or missing SNP x age interaction, we found drastic bias that changed the effect directions. In the following, we will discuss these two scenarios in more detail.

SNP selection

In our main simulation scenario, we stratified our SNP set and used 20 SNPs to simulate the genetic effect on the mean and slope of the exposures X, and 10 different SNPs to simulate the genetic effect on the variability of X. The main motivation for this was to ensure a scenario in which the variability was well-powered by having its own instruments. Although we achieved this in terms of conditional F-statistics and power, the variability estimate $\hat{\theta}_V$ was biased. In scenario 2B, we then used one shared SNP set for all exposure types. Although the simulated SNP effects of the mean and variability were still uncorrelated, the generated allele scores were not: the shared SNP set acts as confounder and introduces a correlation between the mean and variability, and between the slope and variability. Given this high genetic correlation structure, we expected the MVMR to be less powered, more biased, and less able to distinguish between the slope and variability effect. Indeed, for $X^{(MV)}$ we observed bias indicating that the slope effect was now included in the variability effect, given the lack of slope instruments and the correlation between genetic effects on slope and variability. Accordingly, we had high type I error rates for the variability, and low power for the mean, although we had good instruments with conditional F-

statistics >10. For the other two exposures, we found false positive results for the mean and slope on outcomes affected by the variability alone (outcome Y_4), and no power for the variability. As noted by others before, using only shared SNPs results in weak instruments. [22] In our case, using variants with shared effects on mean and slope resulted in high power, low type I error rate, and low bias (given strong instruments), while using variants with shared effects on all three exposure components led to lack of power and false positive results.

GAMLSS regression without a SNP x age interaction

When the SNP x age interaction term was missing as in scenario 3A, the true effect of the slope was attributed to either or both mean and variability. This attribution depended on the simulated exposure. For the exposure $X^{(MV)}$, in which we did not simulate a genetic effect on the slope, the power, type I error, and bias were similar to the main analysis. This suggests that when there is no genetic effect on the slope, e.g. no known or anticipated SNP x age interaction from other GWASs, the GAMLSS regression model can be changed without any further consideration. The advantage of the "no-slope" approach is that the MVMR has to estimate only two correlated exposures, leading to better power to distinguish between them. This also results in higher conditional F-statistics.

However, if there is a relevant SNP x age effect as simulated for exposure $X^{(MSV)}$ and $X^{(MS)}$, and the term is not considered in the GAMLSS regression model, the "missing" effect of the slope is attributed to either the mean or the variability, respectively. For $X^{(MSV)}$, where good genetic predictors of the variability are present, the main amount of the slope effect is attributed to the mean, resulting in high power, but also high type I error rate for the mean, as every slope effect is detected as mean effect as well. In turn, the observed estimate is always biased when there is a true effect of the mean or the slope. The direction of the bias depends on the true causal effects: if there should only be an effect of the mean, the bias is negative as we underestimate the causal effect; the average SNP effect on the exposure increases as it is the sum of the generated mean and slope, and the ratio of outcome to exposure effect then decreases. If there should only be an effect of the slope, the bias is positive as we overestimate the effect; again, the SNP effect is on average larger than before, and we compare it to a null effect. If there is a mix of mean and slope effect, the bias depends on the true causal effects and if one of them dominates the other. For $X^{(MS)}$ we found a similar pattern, but not as strong as the main part of the slope effect is attributed to the variability. This is the only scenario in which we actually have high power to detect a variability effect for $X^{(MS)}$, which was simulated without such effect. This can be recognized by the bias of the variability effect, which is in the range of the age-uncorrected slope effect.

Variability, SNP x environment interactions, and pleiotropy

This observation is in line with variance quantitative trait loci (vQTL) studies, which have suggested that SNP effects on the sample variance of a biomarker are candidates for unobserved SNP x environment interaction, [6–8] in our case SNP x age interaction. Others have already explored the use of these interactions in Mendelian Randomization approaches, [37,38] with main focus on correcting for pleiotropic effects which violate the MR assumptions. In more detail, the exclusion restriction assumption (assumption 3) is often also referred to as the "no horizontal pleiotropy" assumption, which states that the genetic variant does not influence the outcome independently of the hypothesised exposure. Although pleiotropy was not the focus of this work, we observed this effect in the modified GAMLSS regression scenario 3A: when a true SNP x age interaction effect was present but not estimated and no strong variability effects simulated, the variability captured some of this SNP x age interaction. While the MR-GxE [37] and MR-GENIUS [38] aim at estimating an unbiased causal effect estimates of the mean, our approach tries to estimate the

783 causal effects of the mean, slope and variability. While our mean and slope estimates are
784 unbiased in the main scenario of exposures $X^{(MSV)}$ and $X^{(MS)}$, the variability estimate is biased,
785 suggesting that it still captures some unobserved SNP x environment interaction. Of note, while
786 the variability captures the SNP x age interaction, the reverse is not true: in the GAMLSS
787 regression model without a SNP term in the σ-function (scenario 3B) we did not observe an
788 increase in bias in the slope estimates. In our real data application, we observed no relevant
789 differences between the main and the "no-slope" estimates. This could either be due to no true
790 variability effect, as we assume for the log-transformed EFW in the POPS data, or due to the weak
791 conditional instruments in our longitudinal data, as irrespective of the GAMLSS regression model
792 the variability is underpowered.

793 *Total cholesterol variability and coronary artery disease*

794 While the findings for the mean and slope of TC on CAD risk are plausible and in line with
795 literature, [25,39–41] the effect of the variability was not. There are some studies on lipid variability and
796 CAD related outcomes: in a randomized controlled trial the visit-to-visit lipid variability was
797 identified as independent predictor of cardiovascular events. [42,43] Other studies used data from
798 the Korean National Health Insurance System, and enrolees are recommended to undergo a
799 standardized medical examination at least every 2 years. They tested the variability of lipids and
800 other biomarkers and found high variability in lipid levels is associated with adverse health-
801 related outcomes. [44–46] Others used data from participants in the Kailuan Study without history of
802 myocardial infarction, stroke, and cancer, who underwent comprehensive biennial health
803 examinations, and observed also an association between higher lipid variability and increased
804 incidence of ischemic and haemorrhagic stroke. [47] Testing the univariate effect of TC variability
805 on CAD, we found a significant positive effect, but in the MVMR it was no longer significant. Again,
806 our analysis had low conditional F-statistics, and the results have to be interpreted with care.
807 Using the "no slope" model did increase the conditional F-statistics but did not change the null-
808 result for the variability. One major difference between our study and the reported literature is
809 that in the UKB the GP data of patients was used, as there are no regular health examinations
810 offered. Hence, we might have a biased exposure data set to begin with, as healthy people are
811 excluded due to missing data. In turn, patients who seek their GP more might have a better
812 surveillance, which in turn will reduce their risk for CAD as they benefit from prevention
813 strategies.

814 *Limitations*

815 Our study has several limitations. First, although we simulated the exposure data with a random
816 slope, we did not include that term in the GAMLSS regression, only the random intercepts. This
817 was due to computational reasons: the computation time increases further, and the covariance
818 matrix of the random effects is no longer invertible in all scenarios. Given that we still observed
819 associated SNPs and that the power / type I error rate was our primary performance metric, we
820 refrained from implementing a model with three random effects in the simulation, even if that
821 introduced bias. Another limitation was that we used only candidate SNPs in the real data
822 application. They were reported for mean level of the exposure, which gives the association with
823 the variability a disadvantage from the start. Also, we only used a simplistic position pruning to
824 ensure SNP-wise independence. However, a full longitudinal GWAS with extensive fine-mapping
825 would have been too onerous for illustrative examples. Thirdly, the conditional F-statistics in the
826 real data analyses were quite low, suggesting weak instruments. Although this might be the case,
827 we note that we obtained summary statistics per exposure type already conditional on each other
828 as they were estimated in the same model. Hence the conditional F-statistics might be to

conservative in this setting. Additionally, we did not explore other MVMR methods to address pleiotropic effects. Pleiotropy might be covered by including the variability term, but given the weak instruments, especially in the first application in the POPS data, this could have been incomplete. It was, however, not the main focus of this work. 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.

Conclusion

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 (https://github.com/pottj/LongitudinalMR_POPS_UKBB). We used R v4.3.1 and the R-packages "lme4" v1.1-35.1[16], "gamlss" v5.4-12[18], and "MendelianRandomization" v0.9.0[48] 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 S7** and **S8**. Supplemental Tables are accessible on zenodo [49] and github (https://github.com/pottj/LongitudinalMR_POPS_UKBB/tree/main/paper).

No Artificial Intelligence Generated Content (AIGC) tools such as ChatGPT and others based on large language models (LLMs) were used in developing any portion of this manuscript.

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) and his research is currently supported by the Ulverscroft Vision Research Group (UCL). 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. 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

13. 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

14. Wiegrebe S, Gorski M, Herold JM, et al. Analyzing longitudinal trait trajectories using GWAS identifies genetic variants for kidney function decline. *Nat Commun*. 2024;15(1):10061. doi:10.1038/s41467-024-54483-9

15. 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

16. 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

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

18. 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

19. 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

20. 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

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

22. Sanderson E, Spiller W, Bowden J. Testing and correcting for weak and pleiotropic instruments in two-sample multivariable Mendelian randomization. *Stat Med*. 2021;40(25):5434-5452. doi:10.1002/sim.9133

23. 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

24. 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

25. 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

26. 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

27. 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

28. 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

29. 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

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

31. Kanoni S, Graham SE, Wang Y, et al. Implicating genes, pleiotropy, and sexual dimorphism at blood lipid loci through multi-ancestry meta-analysis. *Genome Biol*. 2022;23(1):268. doi:10.1186/s13059-022-02837-1

32. Robinson-Cohen C, Triozzi JL, Rowan B, et al. Genome-Wide Association Study of CKD Progression. *J Am Soc Nephrol JASN*. 2023;34(9):1547-1559. doi:10.1681/ASN.0000000000000170

33. Tang W, Kowgier M, Loth DW, et al. Large-scale genome-wide association studies and meta-analyses of longitudinal change in adult lung function. *PloS One*. 2014;9(7):e100776. doi:10.1371/journal.pone.0100776

34. Couto Alves A, De Silva NMG, Karhunen V, et al. GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI. *Sci Adv*. 2019;5(9):eaaw3095. doi:10.1126/sciadv.aaw3095

35. Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med*. 1965;58(5):295-300. doi:10.1177/003591576505800503

36. Woolf B, Burgess S. The role of estimation in Mendelian randomization: should Mendelian randomization investigations provide estimates? *AJE Adv Res Epidemiol*. 2025;1(1):uuaf003. doi:10.1093/ajeadv/uuaf003

37. 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

38. 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

39. 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

40. 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

41. 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

42. Bangalore S, Breazna A, DeMicco DA, Wun CC, Messerli FH. Visit-to-Visit Low-Density Lipoprotein Cholesterol Variability and Risk of Cardiovascular Outcomes. *J Am Coll Cardiol*. 2015;65(15):1539-1548. doi:10.1016/j.jacc.2015.02.017

43. Waters DD, Bangalore S, Fayyad R, et al. Visit-to-visit variability of lipid measurements as predictors of cardiovascular events. *J Clin Lipidol*. 2018;12(2):356-366. doi:10.1016/j.jacl.2017.12.003

44. Kim MK, Han K, Kim HS, et al. Cholesterol variability and the risk of mortality, myocardial infarction, and stroke: a nationwide population-based study. *Eur Heart J*. 2017;38(48):3560-3566. doi:10.1093/eurheartj/ehx585

45. Kim MK, Han K, Park YM, et al. Associations of Variability in Blood Pressure, Glucose and Cholesterol Concentrations, and Body Mass Index With Mortality and Cardiovascular Outcomes in the General Population. *Circulation*. 2018;138(23):2627-2637. doi:10.1161/CIRCULATIONAHA.118.034978

46. Han B, Han K, Yoon K, Kim MK, Lee S. Impact of Mean and Variability of High-Density Lipoprotein-Cholesterol on the Risk of Myocardial Infarction, Stroke, and Mortality in the General Population. *J Am Heart Assoc*. 2020;9(7):e015493. doi:10.1161/JAHA.119.015493

47. Wang A, Li H, Yuan J, et al. Visit-to-Visit Variability of Lipids Measurements and the Risk of Stroke and Stroke Types: A Prospective Cohort Study. *J Stroke*. 2020;22(1):119-129. doi:10.5853/jos.2019.02075

48. 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

49. Pott J. pottj/LongitudinalMR_POPS_UKBB: Revised project. Published online November 17, 2025. doi:10.5281/ZENODO.17634559