Causal mediation analysis for time-varying heritable risk factors with Mendelian Randomization

Zixuan Wu[a], Ethan Lewis[a], Qingyuan Zhao[b], Jingshu Wang[a1]

[a]Department of Statistics, University of Chicago, Chicago, IL 60637, USA

[b]Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, CB2 0WB, UK

[1]Corresponding author

Abstract

Understanding the causal mechanisms of diseases is crucial in clinical research. When randomized experiments are unavailable, Mendelian Randomization (MR) leverages genetic mutations to mitigate confounding. However, most MR analyses assume static risk factors, oversimplifying dynamic risk factor effects. The framework of life-course MR addresses this but struggles with limited GWAS cohort sizes and correlations across time points. We propose FLOW-MR, a computational approach estimating causal structural equations for temporally ordered traits using only GWAS summary statistics. FLOW-MR enables inference on direct, indirect, and path-wise causal effects, demonstrating superior efficiency and reliability, especially with noisy data. By incorporating a spike-and-slab prior, it mitigates challenges from extreme polygenicity and weak instruments. Applying FLOW-MR, we uncovered a childhood-specific protective effect of BMI on breast cancer and analyzed the evolving impacts of BMI, systolic blood pressure, and cholesterol on stroke risk, revealing their causal relationships.

Keywords— life-course studies; Genetic Epidemiology; Structural Equation Modeling; Polygenicity; Disease Etiology

Introduction

Understanding the causal mechanisms behind diseases is a core challenge in clinical research. While randomized controlled trials are considered the gold standard for establishing causality, they can be impractical in certain situations. As a result, Mendelian Randomization (MR) has become increasingly popular as an alternative. MR leverages genetic variations as a form of "natural experiment", effectively reducing the influence of unmeasured environmental confounders in epidemiological studies [1, 2].

Traditional MR analyses often rely on cross-sectional designs, treating risk factors as constant over time. However, many heritable risk factors vary across the lifespan and may have time-specific effects on disease outcomes [3, 4]. Ignoring this temporal aspect can lead to simplistic and misleading conclusions [5]. For example, while ecological studies in epidemiology show that vitamin D levels during childhood are associated with multiple sclerosis risk, standard MR approaches, which focus on adult vitamin D levels, have instead pointed to adult vitamin D as the key factor in disease etiology [6].

In response to these limitations, life-course MR has emerged as a framework that considers how risk factors measured across an individual's lifetime influence later-life outcomes [7]. Though the idea of incorporating life-long information is appealing, performing life-course MR can be challenging due to the small cohort size in Genome-wide Association Studies (GWAS) of earlier life traits and high auto-correlations of risk factors over time. One commonly used approach is multivariable MR (MVMR), such as MVMR-IVW [8], which estimates the *direct* causal effects of the risk factor at each time point. However, MVMR will have limited efficiency and reliability on noisy GWAS data with a small cohort size. Additionally, current MVMR techniques struggle to evaluate indirect causal effects of early-life risk factors that influence outcomes through intermediate factors. Other methods, such as g-estimations of structural mean models [4, 9] or functional principal component analysis that aggregates effects across time points [10], offer some alternative solutions but require access to individual-level GWAS data that are often not available on a large scale.

To address these challenges, we propose FLOW-MR (Framework for Life-cOurse pathWay analysis using Mendelian Randomization), a computational method that estimates a full system of structural equations for temporally ordered heritable traits using only GWAS summary statistics. FLOW-MR enables comprehensive causal mediation analysis, allowing for the decomposition of total causal effects into direct, indirect, and path-specific components. As a longitudinal extension of GRAPPLE [11], FLOW-MR introduces a novel sufficient condition for causal identifiability that generalizes the well-known InSIDE assumption [12] to longitudinal settings, thereby accommodating both pervasive horizontal pleiotropy and unmeasured mediator-outcome confounding. FLOW-MR can be viewed as a Bayesian analogue of the full-information maximum likelihood (FIML) estimator commonly used in structural equation modeling in economics and social sciences [13], with the advantage of requiring only summary data. Its hierarchical Bayesian framework allows for flexible modeling of weak and invalid instruments and captures complex interactions between genetic variants and environmental factors.

When benchmarked against existing MVMR methods, FLOW-MR substantially improves the efficiency and robustness of the statistical inference, especially when the GWAS data are noisy and from smaller cohorts. By incorporating a spike-and-slab prior for genetic effects, FLOW-MR can tackle the different polygenic

patterns of complex traits and mitigate weak instrument bias. Applying FLOW-MR to publicly available GWAS summary statistics, we identify a protective effect of Body Mass Index (BMI) during childhood on breast cancer risk, confined to a specific developmental period. Our analysis further reveals a complex interplay between BMI, systolic blood pressure (SBP), and low-density cholesterol levels across different life stages and their effects on stroke risk. Our findings suggest that adulthood SBP may be the only factor with a direct causal effect on stroke, while the previously identified causal effect of BMI on SBP [14, 15] in adulthood may be attributable to confounding by childhood SBP.

Results

Method overview Consider a sequence of risk factors $(X_1, X_2, \ldots, X_{K-1})$ in temporal order, so that X_i precedes X_j when $i < j$. Our goal is to understand the causal relationship between these traits and their effects on a later outcome Y (also denoted as X_K for convenience), typically some disease status measured in adulthood. Causal effects among the traits must adhere to the temporal sequence (i.e., only an earlier trait can causally influence a later trait) (Figure 1a). We allow the genetic variants $\boldsymbol{Z}$ and the unmeasured environmental confounders $\boldsymbol{U}$ to have direct causal effects on any trait at any time point. Associated with this directed acyclic graph (DAG), we assume a series of structural equations for each individual (see Methods).

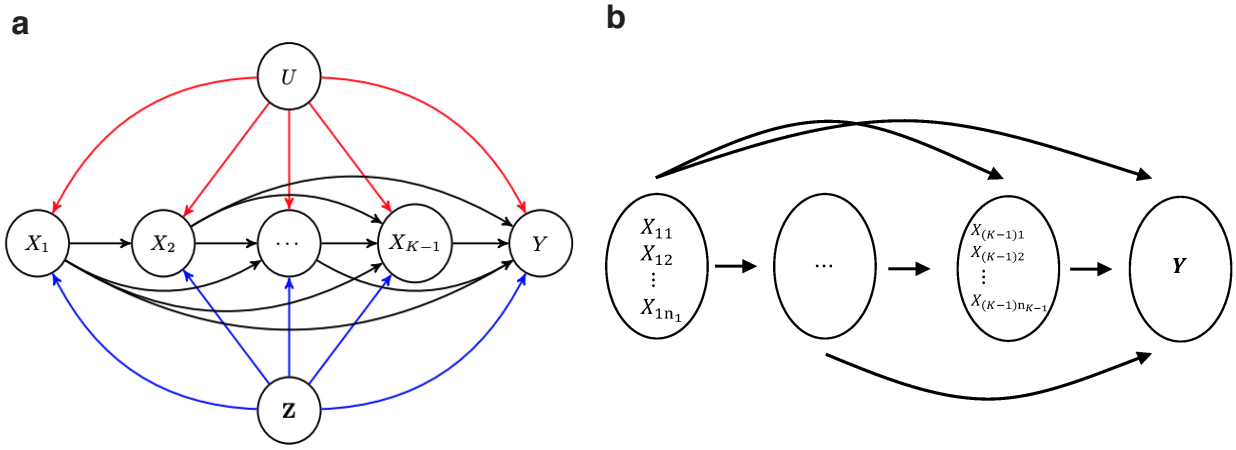

Figure 1: Model overview. a) $X_1, ..., X_{K-1}$ are exposure traits in increasing temporal order and Y is the outcome trait. The blue arrows indicate the causal effects of the genetic variants $\boldsymbol{Z}$ on the traits. The red arrows represent the effects of unmeasured non-heritable confounders $\boldsymbol{U}$. b) Illustration of causal relationships across traits, allowing for multiple traits at each time point.

FLOW-MR requires only GWAS summary statistics for each trait, achieved by projecting the individual-level structural equations onto each single SNP Z_j. Specifically, let γ_{kj} be the marginal associations between

a SNP j and trait X_k, we then have the following linear models:

$$\gamma_{kj} = \sum_{l=1}^{k-1} \beta_{kl}\gamma_{lj} + \alpha_{kj}, \quad k = 1, \ldots, K. \tag{1}$$

Here, β_{kl} is our parameter of interest, representing the direct causal effect of an earlier trait X_l on a later trait X_k. The term α_{kj} denotes the "direct" genetic effect of SNP j on trait k. If SNP Z_j is used as an instrument for trait X_k, then α_{kj} corresponds to the SNP-trait link used for instrumental variable analysis. In contrast, for $k' \neq k$, $\alpha_{k'j}$ represents pleiotropic effects of SNP Z_j on trait $X_{k'}$.

GWAS summary statistics provide estimates of the marginal associations γ_{kj} with their standard errors. They can be obtained either from linear regressions or, for binary traits, from logistic regressions, following [16] under the liability threshold model (Figure S1, SI text). For each SNP Z_j and trait X_k, we assume $\hat{\gamma}_{kj} \sim \mathcal{N}(\gamma_{kj}, \delta_{kj}^2)$, where $\hat{\gamma}_{kj}$ is the observed estimate and δ_{kj}^2 is the estimated variance. Our approach accommodates sample overlap across traits, thus allowing for correlations among $\hat{\gamma}_{kj}$ across different traits k. Following previous MR methods [11, 17], we can adaptively estimate this correlation matrix from the GWAS summary data.

Notice that, unlike marginal associations γ_{kj}, the "direct" genetic effects α_{kj} are not observed from the data. Existing literature shows that most SNPs may have non-zero but weak effects on complex traits, while a small subset may be responsible for the core biological process and strongly impact the traits [11]. Therefore, FLOW-MR assumes the following spike-and-slab distributional assumptions on α_{kj}:

$$\boldsymbol{\alpha}_{kj} \mid p_k, \sigma_{k0}^2, \sigma_{k1}^2 \overset{\text{ind}}{\sim} (1-p_k)\mathcal{N}(0, \sigma_{k0}^2) + p_k\mathcal{N}(0, \sigma_{k1}^2), \quad k = 1, 2, \ldots, K,$$

which allows for trait-specific effect sizes. To estimate the model, we use a hierarchical Bayesian framework and Gibbs sampler. The Gibbs sampler enables the convenient use of posterior samples to construct credible intervals for any function of the direct effects $\{\beta_{kl}\}_{k \geq l}$, including both direct and indirect effects among the traits, as well as the proportions of these effects relative to the total effects.

To address known confounders, we further extend our model to accommodate multiple traits at any given time point (Figure 1b). Specifically, at each time point k, we assume that there are $n_k \geq 1$ exposures, denoted as $X_{k1}, \ldots, X_{kn_k}$. To facilitate the MR analysis, we assume that no causal relationship exists between traits measured at the same time point. If a known causal direction exists between two traits measured at the same time point k, our method can be still be applied by introducing a "pseudo-time point" (or stage) $k+1$ following k, where the outcome trait of the two traits at time k is moved to stage $k+1$. More details on the model, estimation, inference procedures, and key assumptions are provided in Methods and SI Text.

Identifiability of direct and indirect causal effects Mediation analysis, as conducted in this study, is inherently vulnerable to unmeasured mediator-outcome confounding even when the treatment is fully randomized [18]. As illustrated in Figure 2, with two exposures and an outcome ($K = 3$), if we rely solely on a set of valid genetic instruments for the first risk factor X_1 (Figure 2a), it is not possible to separate the direct and indirect effects of X_1 due to unmeasured mediator-outcome confounding between X_2 and Y (see

Supplementary Text for a counter-example). In contrast, if each risk factor has its own set of valid instruments (Figure 2b), it becomes feasible to identify all direct and indirect causal effects. However, since the sequence of risk factors often represents the same factor measured at different time points, finding SNPs that exert their effects exclusively at a specific time point can be challenging.

FLOW-MR requires a more realistic assumption to account for pleiotropic effects of the SNPs (Figure 2c). Instead of assigning each risk factor its own set of valid instruments, we allow SNPs to have nonzero direct effects on all risk factors—thereby permitting any given SNP to be an invalid instrument for a particular trait. However, we require that the "direct" genetic effects $\boldsymbol{\alpha}_1 = (\alpha_{1j})_{j=1}^P, \ldots, \boldsymbol{\alpha}_K = (\alpha_{1j})_{j=1}^P$ be orthogonal across traits (see Methods). More generally, if the number of SNPs P is large, it suffices for these direct effects to be independent rather than exactly orthogonal, extending the well-known InSIDE assumption [12] to multiple traits. Notice that we only assume independence among α_{kj}, allowing the marginal associations γ_{kj} to remain possibly correlated across traits. This assumption enables FLOW-MR to address both pleiotropic effects and unmeasured mediator-outcome confounding. A formal mathematical statement of the identification conditions and its proof are provided in the SI Text.

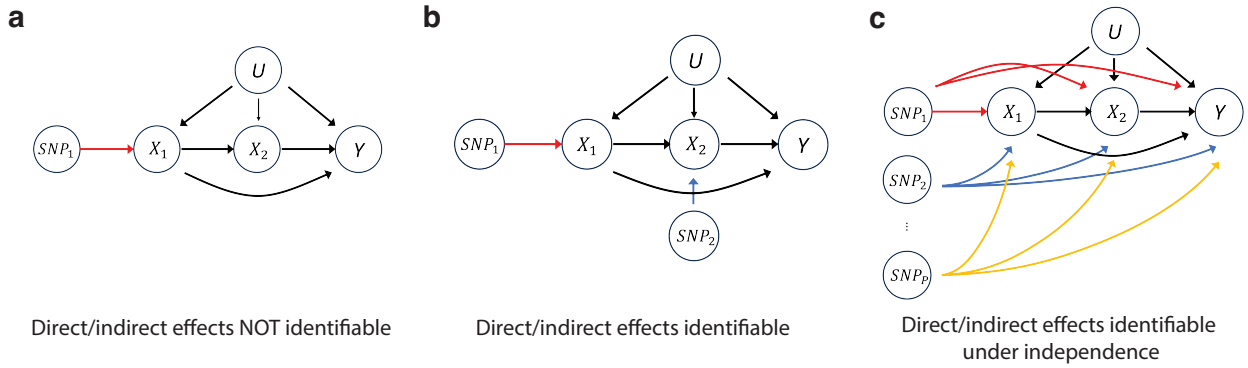

Figure 2: Model identifiability under three scenarios

Systematic benchmarking with synthetic GWAS summary statistics. We compare FLOW-MR with alternative MVMR approaches to perform life-course MR. Specifically, we benchmark our method against six methods, including MVMR-IVW [8], MVMR-Egger [19], MVMR-Robust [20], GRAPPLE [11], MVMR-Horse [21] and MVMR-cML [22]. Among these methods, MVMR-IVW is the most widely used in practice, and other methods are designed to deal with pleiotropic effects under different assumptions and weak instruments. To apply an MVMR method, we use a multiple-step approach where at each step k, we estimate the direct causal effects of $X_1, \cdots, X_{k-1}$ on X_k. One limitation of this approach is the challenge of providing reliable statistical inference on indirect and pathwise effects, as estimates across different steps are correlated.

To assess performance, we generate synthetic GWAS summary datasets by soft-thresholding real GWAS summary statistics, allowing the "true" direct genetic effects to be exactly 0 for some SNPs. Specifically, if we observe $\hat{\gamma}_{kj}^{\text{real}}$ with variance $\hat{\delta}_{kj}^2$ for a particular SNP j in a real GWAS dataset, we set $\alpha_{kj} = \text{sign}(\hat{\gamma}_{kj}^{\text{real}})(|\hat{\gamma}_{kj}^{\text{real}}| - 0.01)_+$ and randomly shuffle α_{kj} across j within each trait k to ensure independence across traits.

To specify the causal relationship across the temporally ordered traits, we simulate three scenarios: $K = 3$, $K = 4$, and a multivariate case with multiple exposures at each time point (Figure 3a). We generate synthetic data for earlier traits using childhood GWAS data to mimic real MR mediation analysis, where earlier childhood traits typically have smaller sample sizes than adulthood traits. Specifically, we use childhood BMI data from the Norwegian Mother, Father, and Child Cohort Study (MoBa) [23] and childhood lipid traits from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort [24]. Adult GWAS datasets are used to generate summary statistics for the remaining traits. Further details of the simulation design, along with additional sensitivity analyses and evaluation of computation costs, can be found in the SI Text. Tables S1-S2 present the number of SNPs and conditional F-statistics [25] computed at each p-value threshold as evaluations of instrument strength.

We evaluate the coverage and average lengths of the confidence and credible intervals of the direct and indirect causal effects across different methods (Figure 3bc, Figures S2-S3). For the direct causal effects, most MVMR methods suffer from severe under-coverage, largely because of limited capabilities to handle pervasive pleiotropic effects and weak-instrument bias. Specifically, MVMR-IVW does not allow any pleiotropic effects, whereas MVMR-Robust and MVMR-cML assume a plurality of valid genetic instruments for each step—a condition that is violated in our simulations and is also unlikely to hold in real life-course MR analyses. Although MVMR-Horse and MVMR-Egger can accommodate independent and pervasive pleiotropy, they remain under-covered in most scenarios, likely due to biases introduced by (conditionally) weak instruments (Table S2) and correlation of SNP-trait associations across risk factors. By contrast, both FLOW-MR and GRAPPLE demonstrate good coverage regardless of SNP strength. Compared to GRAPPLE, FLOW-MR has shorter intervals and offers more efficient analyses. Furthermore, for the indirect causal effects of X_1, only FLOW-MR can provide reliable inference, with credible intervals showing good coverage and reasonable power.

Effect of early life body size on breast cancer Recent studies suggest that early-life body size may protect against breast cancer [26–28]. Using MVMR-IVW, [3] found that this protective effect is not mediated by adult body size, which itself does not causally influence breast cancer. However, their study relied on a recalled early-life body score from the UK Biobank as a proxy for early-life body size. As noted by the original authors, such coarse measurements can raise concerns about the accuracy of their findings.

To address this, we revisited the analysis using FLOW-MR and childhood BMI GWAS summary statistics. First, we replicate the original study [3] using the same dataset: adult BMI from the UK Biobank for adult body size, recalled early-life body score from UK Biobank for early-life body size, and breast cancer data from Michailidou et al. [29] for the outcome. FLOW-MR, alongside all MVMR methods, successfully replicates the original findings (Figure 4a, Figure S4a), confirming no causal effect of adult BMI and a direct protective effect of childhood body size on breast cancer.

However, when we replaced the early-life body size trait with GWAS data on 8-year-old BMI, a direct measurement of childhood body size, MVMR-IVW and MVMR-Egger unexpectedly suggested a protective effect of adult BMI on breast cancer, while childhood BMI showed no direct effect (Figure 4b, Figure S4b). In contrast, GRAPPLE, MVMR-Robust and MVMR-Horse lost the power to detect any causal effects. Only

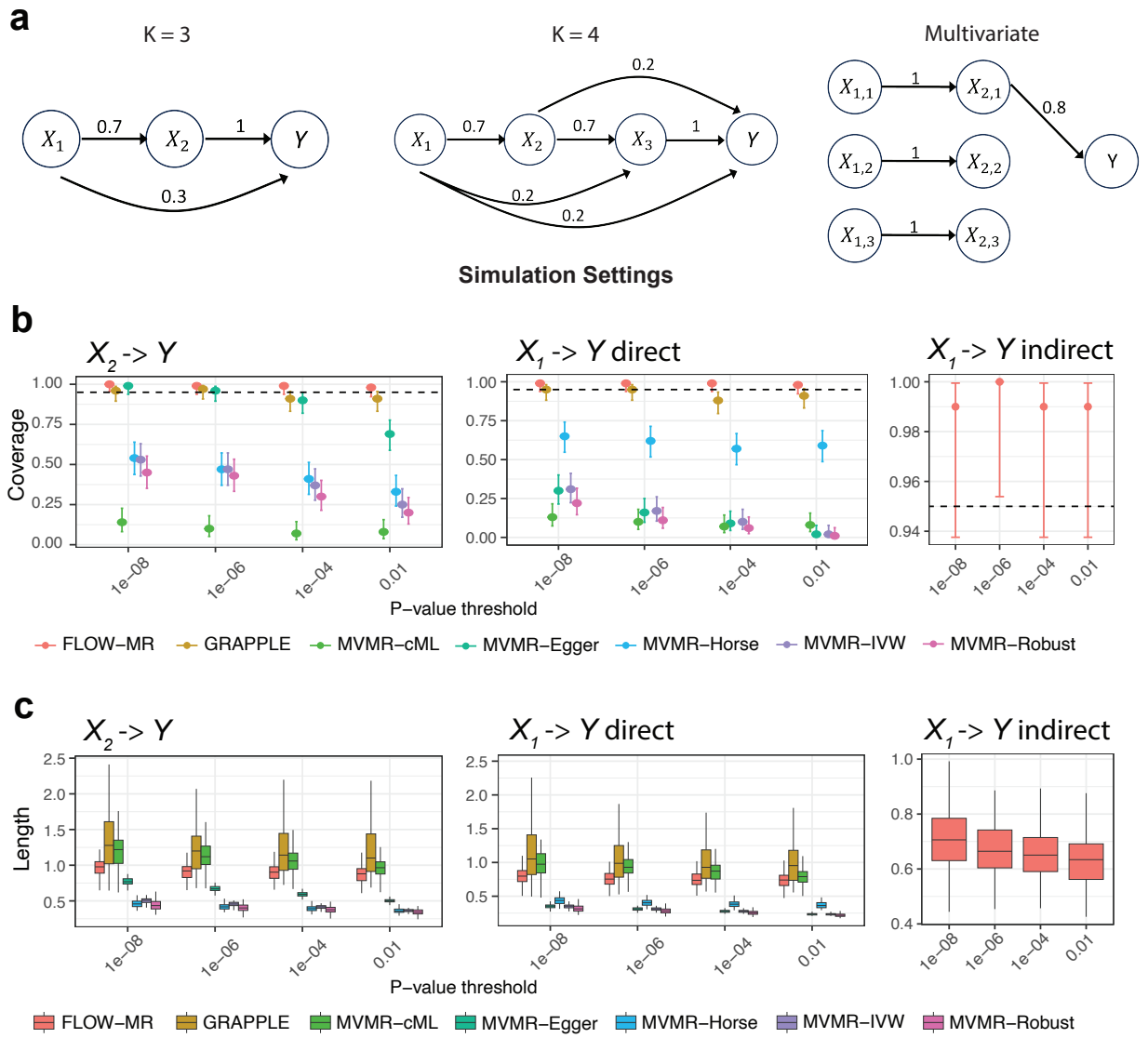

Figure 3: Systematic benchmarking. a) Three simulation scenarios for $K = 3$, $K = 4$ and the multivariate case. b) Empirical coverage of 95% credible intervals on direct effects of X_1 and X_2 on Y and indirect effects of X_1 over 100 repeated simulations when $K = 3$. The error bars are the 95% confidence intervals of the coverage. c) Boxplots of lengths of credible intervals over repeated simulations when $K = 3$. Each box represents the lower and upper quartiles, with the line inside indicating the median. The whiskers extend to 1.5 times the interquartile range beyond the box. Outliers have been removed for clarity.

FLOW-MR replicates the original conclusion that childhood BMI has a protective effect. Although 8-year-old BMI provides a more precise measurement, its use in MR is challenging due to its small sample size (3K samples in the MoBa cohort compared to half a million in the UK Biobank). This is reflected in the substantial decrease in the conditional F-statistics in the childhood dataset (Table S2), indicating a loss of instrument strength due to the limited cohort size of the MoBa study. FLOW-MR demonstrates both efficiency and robustness in this scenario.

We further explored the impact of including 1-year-old BMI in our analysis. Given that the GWAS

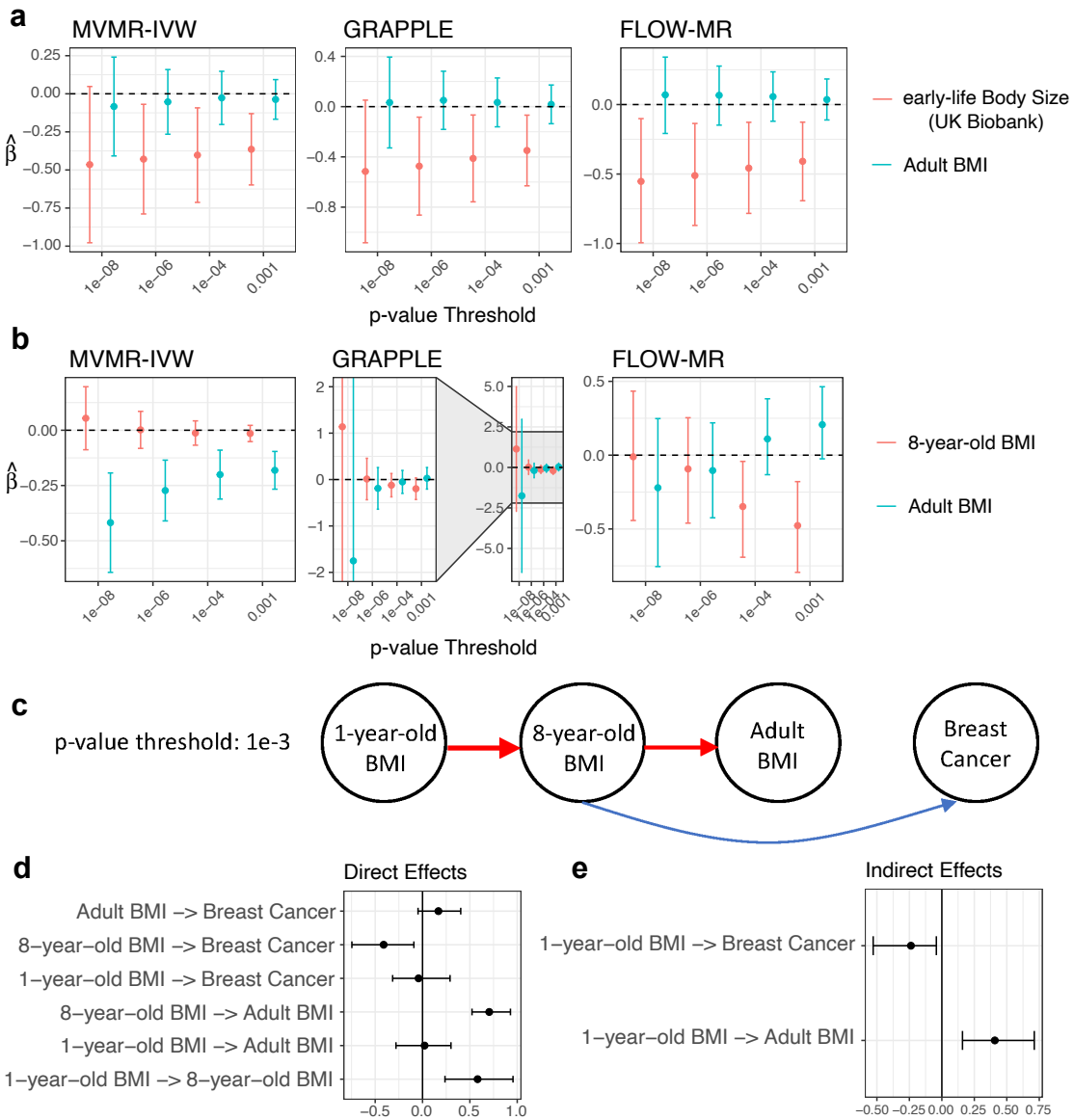

Figure 4: Evaluation of the effect of body size on breast cancer at different ages. a) 95% Confidence / credible intervals of childhood body size (from UK Biobank) and adult BMI on breast cancer risk estimated by MVMR-IVW, GRAPPLE, and FLOW-MR. b) 95% Confidence / credible intervals of childhood BMI and adult BMI on breast cancer risk estimated by these three methods. c) Estimated causal DAG using FLOW-MR at p-value threshold 0.001. The red (blue) arrows indicate significant positive (negative) direct effects at significance level $\alpha = 0.05$. 95% Credible intervals estimated by FLOW-MR are also provided for d) all the direct effects and e) indirect effects of 1-year-old BMI.

data of 1-year-old BMI and 8-year-old BMI traits are from the same cohort, we incorporate the estimated noise correlation matrix (Figure S4c). Despite significant genetic correlations between the 1-year-old and 8-year-old BMI traits (Figure S4d), FLOW-MR still detected the protective effect of 8-year-old BMI on breast cancer at the p-value threshold 0.001 (Figure 4cd). As the direct effect of 1-year-old BMI on breast cancer is insignificant, the analysis revealed that the protective effect of body size on preventing breast cancer is likely

confined to a specific period in childhood.

Additionally, we found that 1-year-old BMI directly influences 8-year-old BMI but does not have any additional direct effect on adult BMI. Interestingly, the estimated causal effects of 1-year-old BMI on 8-year-old BMI and 8-year-old BMI on adult BMI appear to be of similar magnitude (Figure 4d, Figure S4e). Furthermore, FLOW-MR has the power to detect path-wise indirect causal effects, indicating a significant protective effect of 1-year-old BMI against breast cancer, as well as a positive effect on adult BMI, both of which are mediated through 8-year-old BMI (Figure 4e).

Temporal causal relationships between high blood pressure, BMI and stroke. High blood pressure is a well-established modifiable risk factor for stroke [30]. Recent univariate MR studies have also suggested that adult BMI may be a potential causal risk factor for stroke and has a significant positive causal effect on high blood pressure in adulthood [14, 15, 31]. Our study aims to disentangle the complex relationships between high blood pressure, BMI, and stroke, focusing on how these causal connections evolve throughout life. We also consider potential confounders, such as lipid traits.

We analyze the causal effects of systolic blood pressure (SBP), BMI, and low-density lipoprotein cholesterol (LDL-C) in both childhood and adulthood on stroke using FLOW-MR. To capture potential causal effects of BMI on SBP, even at the same time point, we introduce additional "time points" (stages) for SBP traits both in childhood and adulthood (Figure 5a). We utilize GWAS summary statistics from the ALSPAC cohort [24] for childhood traits, UK Biobank for adult BMI and SBP, and the Global Lipids Genetics Consortium for adult LDL-C [32]. Stroke data comes from Malik et al. [33], and SNP selection is based on the GERA datasets [34, 35], the GIANT adult BMI dataset, and EGG childhood BMI dataset.

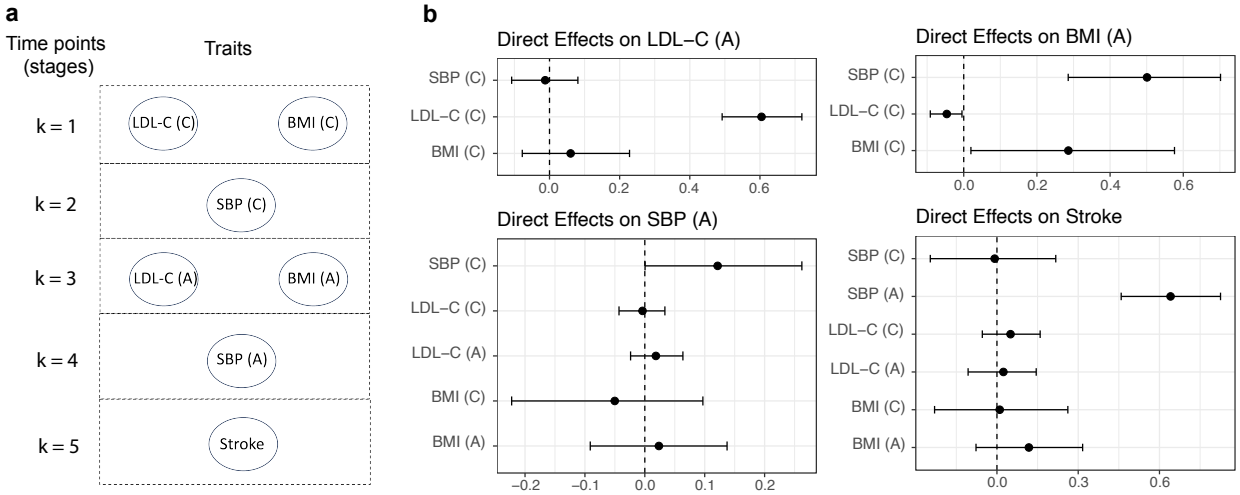

Figure 5: Multivariate mediation analysis on LDL-C, BMI, SBP and stroke. a) Design of the stages. Childhood traits are marked with label (C) and adulthood traits are marked with label (A). b) 95% Credible intervals estimated by FLOW-MR for direct effects on adult traits LDL-C, BMI, SBP and Stroke.

Our analysis reveals a clear positive causal effect of adulthood SBP on stroke, with no significant direct effects from other traits, including childhood traits and adult BMI (Figure 5b, Figure S5a). Interestingly, our findings challenge previous univariate MR results ([14, 15], Figure S5b), showing no evidence of a positive

causal effect of adult BMI on adult SBP after accounting for childhood SBP. Further MR analyses using only these three continuous traits also suggest that the previously observed association between adult BMI and SBP may be confounded by childhood SBP (Figure S5c). Additionally, we observe a stronger genetic correlation between childhood SBP and adult BMI than between adult SBP and BMI (Figure S5d). These results imply that the impact of adulthood BMI on SBP in earlier studies may have been overstated due to unaccounted confounding factors.

Discussion

We propose FLOW-MR using GWAS summary data for life-course Mendelian Randomization, based on a full structural model for all traits. This method allows users to assess time-varying causal effects of heritable risk factors and distinguish between direct and indirect causal effects. By addressing key challenges in life-course MR—the high genetic correlation of the same trait at different ages and the limited cohort size of age-specific GWAS, FLOW-MR demonstrates superior performance in simulation studies. Unlike other methods, FLOW-MR is not prone to biases arising from using weakly associated SNPs as instruments and can efficiently integrate information across all traits.

For the identification of causal effects in life-course MR, a similar argument has been discussed in [7]. They show that to identify the causal direct effects of traits, genetic variants must exert different effects on each exposure in the model, and these effects must be linearly independent, meaning the true SNP-trait association matrix must have full rank. This condition is necessary but not sufficient for identifying all direct and indirect effects. For instance, in univariable MR, pleiotropic effects must adhere to specific assumptions (such as the InSIDE assumption) to ensure the identification of a risk factor's casual effect. Merely assuming that the pleiotropic effects are not perfectly correlated with the marginal SNP-exposure associations is insufficient for identifying the causal effects. While we allow pervasive pleiotropy in our work, we assume that the direct genetic effects of SNPs are independent across traits, so there is no correlated pleiotropy between the traits. Our sensitivity analysis indicates that, although FLOW-MR models the life-course traits as a whole system, the statistical inference for a particular outcome demonstrates robustness to correlated pleiotropy between other traits (Figure S6). In practice, if confounding pathways are known to exist between traits with correlated pleiotropy, we recommend collecting additional GWAS data for these confounding factors and explicitly adjusting for them using FLOW-MR.

To assess the stability of estimates across different instrument selection criteria, in the paper we employed a range of p-value thresholds for selecting genetic instruments in both simulations and case studies, following the recommendations of [11]. In our simulations, the results remain broadly consistent across the various p-value thresholds. In contrast, larger discrepancies are observed in the real data case studies. This difference may be partly due to the stronger causal signals and the enforced independence of direct genetic effects across traits in the simulation design. Additionally, the number of selected SNPs varies much more substantially in the real data (Table S1), which contributes to differences in statistical power across thresholds.

Finally, as noted by previous studies [6, 36], life-course MR for mediation analysis also faces several

potential limitations. One such limitation is the absence of measurements at critical time points when the exposure exerts a substantial causal effect. However, if a missing time point X_k is not a hidden confounder for any pair of subsequent traits X_{k_1} and X_{k_2}, our assumption of independent direct genetic associations remains valid, and the omission of that time point does not bias the estimation of causal effects at observed time points. Another concern is the possibility of reverse causation, where the outcome may causally influence the exposure measured at the latest time point. Although life-course MR typically benefits from the temporal ordering of traits, a challenge arises when the final exposure and outcome are measured concurrently—such as when both are assessed in adulthood. In such cases, if the selected SNPs influence the outcome primarily through their effects on the exposures, MR estimates are generally robust to reverse causality [6, 11, 37]. Our proposed framework inherits this desirable property of MR.

Methods

Structural equations on individual-level data Denote $\boldsymbol{Z}$ as the vector containing all available genetic variants. Based on the causal DAG (Figure 1), we assume the following additive structural equations for each individual:

$$\begin{aligned} X_K := Y &= \sum_{l=1}^{K-1} \beta_{Kl} X_l + f_K(\boldsymbol{U}, \boldsymbol{Z}, \boldsymbol{E}_Y), \\ X_k &= \sum_{l=1}^{k-1} \beta_{kl} X_l + f_k(\boldsymbol{U}, \boldsymbol{Z}, \boldsymbol{E}_{X_k}), \quad k = 2, \ldots, K-1 \\ X_1 &= f_1(\boldsymbol{U}, \boldsymbol{Z}, \boldsymbol{E}_{X_1}). \end{aligned} \tag{2}$$

In these equations, the functions $f_1(\cdot), \cdots, f_K(\cdot)$ capture the direct genetic and environmental effects on the traits, which may be nonlinear and involve arbitrary interactions. Our primary assumption in (2) is that the causal effect β_{kl} for any earlier trait X_l $(1 \leq l \leq K-1)$ on any later trait X_k $(l+1 \leq k \leq K)$ is linear and homogeneous. When the outcome or risk factor traits are binary, we adopt a liability threshold model [16], treating each binary trait X_k as a discretized version of an underlying continuous liability $X_k^\star$, that follows the additive structural equations above (Figure S1). In this context, our method estimates the causal relationships among the latent liabilities. We also further generalize (2) to allow for nonlinear causal effects and interactions among earlier traits, so that our linear model for summary statistics can accommodate more complex causal relationships measured at different time points. Additional details are provided in the SI Text.

Given these structural equations representing the causal relationships among the traits, each β_{kl} represents the direct causal effect of X_l on X_k whenever $l < k$. A path-wise causal effect from X_l to X_k along a path $\mathcal{L} = (X_{r_0}, X_{r_1}, \cdots, X_{r_{t-1}}, X_{r_t})$ (where $X_{r_0} = X_l$ and $X_{r_t} = X_k$) of length t is defined as

$$\beta_{kl}^{\mathcal{L}} = \prod_{i=1}^{t} \beta_{r_i r_{i-1}}.$$

The total effect sums up the effect from X_l to X_k through all directed pathways, and the indirect effect is the difference between the total effect and the direct effects [38, 39]. In other words, the indirect effect β_{kl}^{ind} sums all path-wise causal effects of length $t \geq 2$:

$$\beta_{kl}^{\text{ind}} = \sum_{t=2}^{k-l} \sum_{\mathcal{L}=\{X_{r_0}=X_l,\cdots,X_{r_t}=X_k\}} \beta_{kl}^{\mathcal{L}}.$$

Model on summary statistics Denote $\gamma_{kj} \equiv \text{argmin}_\gamma \text{Var}[X_k - \gamma Z_j]$ as the marginal association between continous trait X_k and Z_j, and let $\alpha_{kj} \equiv \text{argmin}_\alpha \text{Var}[f_k(U, \boldsymbol{Z}, \boldsymbol{E}_{X_k}) - \alpha Z_j]$ as the projection of the genetic and environmental direct effects of trait X_k onto SNP Z_j, which can be heuristically understood as the "direct" genetic effect of SNP j on X_k. Then by projecting the structural equations (2) onto a single SNP Z_j, we can obtain a linear model (1) with measurement error on the GWAS summary statistics (for mathematical details, see SI Text). These equations can be more conveniently expressed in matrix form as:

$$\boldsymbol{\Gamma} = \tilde{\boldsymbol{B}} \cdot \boldsymbol{\Gamma} + \boldsymbol{A}, \tag{3}$$

where P is the number of SNPs and the matrices are defined as

$$\boldsymbol{\Gamma} \equiv \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1P} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2P} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{K1} & \gamma_{K2} & \cdots & \gamma_{KP} \end{bmatrix}, \ \boldsymbol{A} \equiv \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1P} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2P} \\ \vdots & \vdots & \cdots & \vdots \\ \alpha_{K1} & \alpha_{K2} & \cdots & \alpha_{KP} \end{bmatrix}, \ \tilde{\boldsymbol{B}} \equiv \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ \beta_{21} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \beta_{K1} & \beta_{K2} & \cdots & \beta_{K(K-1)} & 0 \end{bmatrix}.$$

Here, the matrix $\boldsymbol{\Gamma}$ represents the marginal SNP-trait associations, $\boldsymbol{A}$ denotes the direct effects of the SNPs on each trait and $\tilde{\boldsymbol{B}}$ is the matrix of direct causal relationships of earlier traits on later traits that we would like to infer. Equation (3) can be alternatively represented as

$$\boldsymbol{\Gamma} = (\boldsymbol{I} + \boldsymbol{B})\boldsymbol{A}, \tag{4}$$

where $\boldsymbol{I}$ is the $K \times K$ identity matrix, and $\boldsymbol{B}$ is the lower-triangular matrix that solves $(\boldsymbol{I} - \tilde{\boldsymbol{B}})^{-1} = \boldsymbol{I} + \boldsymbol{B}$. Compared to (3), the parametrization in (4) avoids matrix inversion, making it more amenable for statistical estimation. It is easy to show that $\boldsymbol{B}$ can be written as a Neumann series: $\boldsymbol{B} = \tilde{\boldsymbol{B}} + \tilde{\boldsymbol{B}}^2 + \ldots$. Intuitively, the (k, l) entry of the matrix $\boldsymbol{B}$ denotes the total causal effects of X_l on X_k through all directed pathways [40].

Since the GWAS cohort size is typically large, for each SNP Z_j, the summary statistics $\hat{\boldsymbol{\Gamma}}_j$ approximately follow a normal distribution $\hat{\boldsymbol{\Gamma}}_j \sim \mathcal{N}(\boldsymbol{\Gamma}_j, \boldsymbol{\Sigma}_j)$, where $\boldsymbol{\Gamma}_j := (\gamma_{1j}, \gamma_{2j}, \ldots, \gamma_{Kj})^T$ is a vector of marginal associations and $\boldsymbol{\Sigma}_j$ is a covariance matrix obtained from the GWAS standard errors and a correlation matrix that depends on the extent of sample overlap and the correlation between the traits. This correlation matrix is shared across the SNPs and can be estimated from the non-statistically significant GWAS summary statistics using the method described in [11] (SI Text).

Identification conditions for direct and indirect effects We provide identification conditions for the matrix $\tilde{\boldsymbol{B}}$ of all direct causal effects from earlier traits to later ones. Once these direct effects are identified, the indirect effects—arising from sums and products of direct effects under our additive structural equations (2)—become identifiable as well.

Under the relationship $\boldsymbol{B} = \left(\boldsymbol{I} - \tilde{\boldsymbol{B}}\right)^{-1} - \boldsymbol{I}$, it suffices to identify the lower-triangular matrix $\boldsymbol{B}$. Assuming an infinitely large cohort, the marginal association matrix $\boldsymbol{\Gamma}$ can be consistently estimated from GWAS data. We then provide conditions on the direct genetic effects $\boldsymbol{A}$ that make $\boldsymbol{B}$ uniquely solvable in equation (4).

Specifically, we prove that $\boldsymbol{B}$ is unique if matrix $\boldsymbol{A}$ lies in the set

$$\mathcal{A} := \{\boldsymbol{A} \in \mathbb{R}^{K\times P} : \boldsymbol{A}\boldsymbol{A}^\top = \text{diag}(d_1, \ldots, d_K) \text{ with } \min\{d_1, \ldots, d_{K-1}\} > 0\}.$$

Intuitively, rather than assuming that each trait k has its own set of valid SNPs (i.e., there is at least a SNP j where $\alpha_{jk} \neq 0$ and $\alpha_{jk'} = 0$ for any $k' \neq k$), we require only that the direct genetic effects across traits are mutually orthogonal. Mathematically, this is because $(\boldsymbol{I} + \boldsymbol{B})(\boldsymbol{A}\boldsymbol{A}^\top)(\boldsymbol{I} + \boldsymbol{B})^T$ provides a unique LDL decomposition of the identifiable matrix $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top$. To make this assumption more realistic, we further can show that when the number of SNPs $P \to \infty$, it is sufficient for these direct effects to be merely independent across traits rather than strictly orthogonal—an extension of the well-known InSIDE assumption to multiple traits.

Selection of genetic instruments We assume that we can use separate GWAS datasets to obtain p-values for SNP selection, thereby mitigating potential selection biases [11, 41]. Upon selecting the SNPs, another set of GWAS datasets is used to obtain the summary statistics for the exposure and outcome traits. The selection p-value for each SNP is defined as the Bonferroni-corrected p-value, computed from $K - 1$ GWAS summary datasets corresponding to traits X_1 to X_{K-1}. SNPs are subsequently chosen based on the selection p-values using LD clumping [42] (with r^2 cutoff at 0.001), ensuring that selected SNPs are approximately independent. Compared to a stringent p-value threshold (such as 10^{-8}) for SNP selection, our method allows for a higher threshold (such as 10^{-4} or 0.01) and uses SNPs that are weakly associated with the exposure traits. This can generally increase the power of MR analysis.

Model estimation and inference We use a hierarchical Bayesian framework and Gibbs sampler to infer the direct and indirect causal effects, which can be expressed as functions of the matrix $\boldsymbol{B}$. In addition to the spike-and-slab prior on the direct genetic effects, we put Gaussian priors on elements of the matrix $\boldsymbol{B}$ and conjugate priors on the hyperparameters:

$$\begin{aligned}
\beta_{kl} \mid \sigma^2 &\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \quad 1 \le l < k \le K \\
\sigma^2 &\sim \text{InvGamma}(\alpha, \beta) \\
p_k &\overset{\text{ind}}{\sim} \text{Beta}(a_k, b_k), \quad k = 1, 2, \ldots, K \\
\sigma^2_{k0} &\overset{\text{ind}}{\sim} \text{InvGamma}(\alpha^0_k, \beta^0_k) \quad k = 1, 2, \ldots, K \\
\sigma^2_{k1} &\overset{\text{ind}}{\sim} \text{InvGamma}(\alpha^1_k, \beta^1_k) \quad k = 1, 2, \ldots, K
\end{aligned}$$

To estimate the above Bayesian hierarchical model, we employ a Gibbs sampler algorithm which can efficiently generate posterior samples when K is small. Posterior samples of $\boldsymbol{B}$ directly give us posterior samples of $\tilde{\boldsymbol{B}}$, which can then be used to construct credible intervals for any direct and indirect causal effects, or their functions. Additionally, we use a simplified empirical Bayes approach to choose the hyper-parameters (α_k^0, β_k^0), (α_k^1, β_k^1) and (a_k, b_k), recognizing their significant impact on the posterior distributions of $\boldsymbol{B}$. For further computational and mathematical details, see SI Text.

Extension to multiple traits at each time point We also extend our model to accommodate multiple traits at any given time point, incorporating two key assumptions to facilitate the MR analysis. First, we assume that there is no causal effect between traits at the same time point. This assumption avoids the need to identify causal directions between any two traits that coexist temporally. If two sets of traits $\{X_{k1}, \cdots, X_{km}\}$ and $\{X_{k(m+1)}, \cdots, X_{kM}\}$ at the same time point k potentially causally related—with the first set influencing the second—our method can still be applied by introducing an additional "time point" (or stage) $k+1$ immediately following k. The second set of traits is then moved to this added stage, and all subsequent stages are shifted accordingly.

The second assumption is a generalization of the independence assumption on direct genetic effects. Specifically, we require that the direct genetic effects α_{kj} of any SNP j must be independent across all traits, including those measured at the same and at different time points. With these assumptions, we can adapt our model and Gibbs sampler to infer the direct and indirect causal effects.

Benchmarking methods We benchmarked the competing methods using publicly available software packages. Specifically, GRAPPLE (https://github.com/jingshuw/GRAPPLE), MVMR-IVW (MVMR; https://github.com/WSpiller/MVMR), MVMR-Egger (MendelianRandomization; https://CRAN.R-project.org/package=MendelianRandomization), MVMR-Robust (https://github.com/aj-grant/robust-mvmr), MVMR-cML (https://github.com/ZhaotongL/MVMR-cML), and MVMR-Horse (https://github.com/aj-grant/mrhorse) were all implemented with their standard procedures recommended in the respective software documentation. We used the same set of selected SNPs for each method.

Author contributions

J.W. supervised the work. J.W. and Q.Z. formulated the problem. J.W., Z.W. and E.L. developed the algorithm. Z.W. and E.L. implemented the method and performed simulations. Z.W. conducted real data analyses. J.W., Q.Z. and Z.W. wrote the paper.

Declaration of Interests

The authors have declared that no competing interests exist.

Acknowledgments

J.W. is partly supported by the National Science Foundation under grants DMS-2113646 and DMS-2238656. Q.Z. is partly supported by EPSRC grant EP/V049968/1. We acknowledge the University of Chicago's Research Computing Center for their support of this work.

Data availability

GWAS summary statistics are all downloaded from public sources. For adult BMI, data is obtained from the GIANT consortium website (https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files) and the UK Biobank Neale's lab (http://www.nealelab.is/uk-biobank/) using phenotype code 21001. Adult lipid traits (LDL-C, HDL-C, and triglycerides) summary statistics are acquired from the GLGC cohort via the GLGC website (https://csg.sph.umich.edu/willer/public/lipids2013/) and from the GERA cohort through GWAS Catalog with study accession numbers GCST007141, GCST007140, and GCST007142. Breast cancer GWAS summary statistics come from GWAS Catalog with study accession number GCST004988, while those for T2D are downloaded from the DIAGRAM website (https://diagram-consortium.org/downloads.html), and for stroke from GWAS Catalog with study accession number GCST006906.

For childhood traits, GWAS summary statistics on 1-year-old BMI and 8-year-old BMI are contributed by the Centre For Diabetes Research, University of Bergen, Norway, and the Norwegian Mother, Father and Child study, downloadable from their website (https://www.fhi.no/en/studies/moba/for-forskere-artikler/gwas-data-from-moba/). Childhood BMI GWAS data from the EGG Consortium can be found at http://egg-consortium.org/childhood-bmi.html. The ALSPAC datasets for childhood lipid traits, SBP, and BMI are obtained from GWAS Catalog with study accession numbers GCST90104679, GCST90104678, GCST90104680, GCST90104683, GCST90104677.

Code availability

The R package FLOW-MR for conducting our Bayesian mediation MR analysis is publicly available for installation at (https://github.com/ZixuanWu1/FLOW-MR). The code for all experiments, real data analyses, and the reproduction of all figures, along with source data, can be accessed at (https://github.com/ZixuanWu1/FLOW-MR-PAPER).

References

[1] Davey Smith, G. and Ebrahim, S. (2003). 'mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International journal of epidemiology *32*, 1–22.

[2] Davey Smith, G. and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Human molecular genetics *23*, R89–R98.

[3] Richardson, T.G., Sanderson, E., Elsworth, B., Tilling, K., and Smith, G.D. (2020). Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study. bmj *369*.

[4] Shi, J., Swanson, S.A., Kraft, P., Rosner, B., De Vivo, I., and Hernán, M.A. (2022). Mendelian randomization with repeated measures of a time-varying exposure: an application of structural mean models. Epidemiology *33*, 84–94.

[5] Labrecque, J.A. and Swanson, S.A. (2019). Interpretation and potential biases of mendelian randomization estimates with time-varying exposures. American journal of epidemiology *188*, 231–238.

[6] Sanderson, E., Richardson, T.G., Morris, T.T., Tilling, K., and Davey Smith, G. (2022). Estimation of causal effects of a time-varying exposure at multiple time points through multivariable mendelian randomization. PLoS Genetics *18*, e1010290.

[7] Power, G.M., Sanderson, E., Pagoni, P., Fraser, A., Morris, T., Prince, C., Frayling, T.M., Heron, J., Richardson, T.G., Richmond, R., et al. (2023). Methodological approaches, challenges, and opportunities in the application of mendelian randomisation to lifecourse epidemiology: A systematic literature review. European Journal of Epidemiology , 1–20.

[8] Burgess, S., Dudbridge, F., and Thompson, S.G. (2015). Re:"multivariable mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects". American journal of epidemiology *181*, 290–291.

[9] Shi, J., Swanson, S.A., Kraft, P., Rosner, B., De Vivo, I., and Hernán, M.A. (2021). Instrumental variable estimation for a time-varying treatment and a time-to-event outcome via structural nested cumulative failure time models. BMC medical research methodology *21*, 1–12.

[10] Cao, Y., Rajan, S.S., and Wei, P. (2016). Mendelian randomization analysis of a time-varying exposure for binary disease outcomes using functional data analysis methods. Genetic epidemiology *40*, 744–755.

[11] Wang, J., Zhao, Q., Bowden, J., Hemani, G., Davey Smith, G., Small, D.S., and Zhang, N.R. (2021). Causal inference for heritable phenotypic risk factors using heterogeneous genetic instruments. PLoS genetics *17*, e1009575.

[12] Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. International journal of epidemiology *44*, 512–525.

[13] Davidson, R., MacKinnon, J.G., et al. (1993). Estimation and inference in econometrics, volume 63. Oxford New York.

[14] Lee, M.R., Lim, Y.H., and Hong, Y.C. (2018). Causal association of body mass index with hypertension using a mendelian randomization design. Medicine *97*.

[15] Zhao, Q., Wang, J., Bowden, J., and Small, D. (2018). Statistical inference in two-sample summary-data mendelian randomization using robust adjusted profile score. Annals of Statistics *48*.

[16] Wu, Z. and Wang, J. (2024). Interpretation of two-sample mendelian randomization for binary exposures and outcome. bioRxiv , 2024–06.

[17] Morrison, J., Knoblauch, N., Marcus, J.H., Stephens, M., and He, X. (2020). Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. Nature genetics *52*, 740–747.

[18] VanderWeele, T.J., Vansteelandt, S., and Robins, J.M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology (Cambridge, Mass.) *25*, 300.

[19] Rees, J.M., Wood, A.M., and Burgess, S. (2017). Extending the mr-egger method for multivariable mendelian randomization to correct for both measured and unmeasured pleiotropy. Statistics in medicine *36*, 4705–4718.

[20] Grant, A.J. and Burgess, S. (2021). Pleiotropy robust methods for multivariable mendelian randomization. Statistics in medicine *40*, 5813–5830.

[21] Grant, A.J. and Burgess, S. (2024). A bayesian approach to mendelian randomization using summary statistics in the univariable and multivariable settings with correlated pleiotropy. The American Journal of Human Genetics *111*, 165–180.

[22] Lin, Z., Xue, H., and Pan, W. (2023). Robust multivariable mendelian randomization based on constrained maximum likelihood. The American Journal of Human Genetics *110*, 592–605.

[23] Helgeland, Ø., Vaudel, M., Juliusson, P.B., Lingaas Holmen, O., Juodakis, J., Bacelis, J., Jacobsson, B., Lindekleiv, H., Hveem, K., Lie, R.T., et al. (2019). Genome-wide association study reveals dynamic role of genetic variation in infant and early childhood growth. Nature communications *10*, 4448.

[24] O'Nunain, K., Sanderson, E., Holmes, M.V., Smith, G.D., and Richardson, T.G. (2023). A genome-wide association study of childhood adiposity and blood lipids. Wellcome Open Research *6*, 303.

[25] Sanderson, E., Spiller, W., and Bowden, J. (2021). Testing and correcting for weak and pleiotropic instruments in two-sample multivariable mendelian randomization. Statistics in medicine *40*, 5434–5452.

[26] Ooi, B.N.S., Loh, H., Ho, P.J., Milne, R.L., Giles, G., Gao, C., Kraft, P., John, E.M., Swerdlow, A., Brenner, H., et al. (2019). The genetic interplay between body mass index, breast size and breast cancer risk: a mendelian randomization analysis. International journal of epidemiology *48*, 781–794.

[27] Vatten, L.J. and Kvinnsland, S. (1990). Body mass index and risk of breast cancer. a prospective study of 23,826 norwegian women. International journal of cancer *45*, 440–444.

[28] Guo, Y., Warren Andersen, S., Shu, X.O., Michailidou, K., Bolla, M.K., Wang, Q., Garcia-Closas, M., Milne, R.L., Schmidt, M.K., Chang-Claude, J., et al. (2016). Genetically predicted body mass index and breast cancer risk: Mendelian randomization analyses of data from 145,000 women of european descent. PLoS medicine *13*, e1002105.

[29] Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., et al. (2017). Association analysis identifies 65 new breast cancer risk loci. Nature *551*, 92–94.

[30] Willmot, M., Leonardi-Bee, J., and Bath, P.M. (2004). High blood pressure in acute stroke and subsequent outcome: a systematic review. Hypertension *43*, 18–24.

[31] Chen, L., Peters, J.E., Prins, B., Persyn, E., Traylor, M., Surendran, P., Karthikeyan, S., Yonova-Doing, E., Di Angelantonio, E., Roberts, D.J., et al. (2022). Systematic mendelian randomization using the human plasma proteome to discover potential therapeutic targets for stroke. Nature communications *13*, 6143.

[32] Willer, C., Schmidt, E., Sengupta, S., Peloso, G., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M., Mora, S., et al. (2013). Discovery and refinement of loci associated with lipid levels. Nature genetics *45*, 1274–1283.

[33] Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., Rutten-Jacobs, L., Giese, A.K., Van Der Laan, S.W., Gretarsdottir, S., et al. (2018). Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nature genetics *50*, 524–537.

[34] Hoffmann, T.J., Ehret, G.B., Nandakumar, P., Ranatunga, D., Schaefer, C., Kwok, P.Y., Iribarren, C., Chakravarti, A., and Risch, N. (2017). Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. Nature genetics *49*, 54–64.

[35] Hoffmann, T.J., Theusch, E., Haldar, T., Ranatunga, D.K., Jorgenson, E., Medina, M.W., Kvale, M.N., Kwok, P.Y., Schaefer, C., Krauss, R.M., et al. (2018). A large electronic-health-record-based genome-wide study of serum lipids. Nature genetics *50*, 401–413.

[36] Tian, H. and Burgess, S. (2023). Estimation of time-varying causal effects with multivariable mendelian randomization: some cautionary notes. International journal of epidemiology *52*, 846–857.

[37] Burgess, S., Swanson, S.A., and Labrecque, J.A. (2021). Are mendelian randomization investigations immune from bias due to reverse causation? European Journal of Epidemiology *36*, 253–257.

[38] VanderWeele, T. (2015). Explanation in causal inference: methods for mediation and interaction. Oxford University Press.

[39] Baron, R.M. and Kenny, D.A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology *51*, 1173.

[40] Sullivant, S., Talaska, K., and Draisma, J. (2008). Trek separation for gaussian graphical models. Annals of Statistics *38*.

[41] Zhao, Q., Chen, Y., Wang, J., and Small, D.S. (2019). Powerful three-sample genome-wide design and robust statistical inference in summary-data mendelian randomization. International journal of epidemiology *48*, 1478–1492.

[42] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J., et al. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics *81*, 559–575.