

Bayesian causal graphical model for joint Mendelian randomization analysis of multiple exposures and outcomes

Authors

Verena Zuber, Toinét Cronjé, Na Cai,
Dipender Gill, Leonardo Bottolo

Correspondence

v.zuber@imperial.ac.uk (V.Z.),
lb664@cam.ac.uk (L.B.)

In real-life applications, exposures and outcomes are interconnected in complex relationships that cannot be modeled using available Mendelian randomization methods. We introduce MrDAG, a Bayesian causal graphical model to detect dependency relations between exposures and between outcomes to better estimate causal effects and shed light on complex pathways toward disease.



Zuber et al., 2025, The American Journal of Human Genetics 112, 1173–1198
May 1, 2025 Crown Copyright © 2025 Published by Elsevier Inc. on behalf of
American Society of Human Genetics.
<https://doi.org/10.1016/j.ajhg.2025.03.005>

Bayesian causal graphical model for joint Mendelian randomization analysis of multiple exposures and outcomes

Verena Zuber,^{1,2,3,*} Toinét Cronjé,⁴ Na Cai,^{5,6,7} Dipender Gill,¹ and Leonardo Bottolo^{8,9,*}

Summary

Current Mendelian randomization (MR) methods do not reflect complex relationships among multiple exposures and outcomes as is typical for real-life applications. We introduce MrDAG, a Bayesian causal graphical model for summary-level MR analysis to detect dependency relations within the exposures, the outcomes, and between them to improve causal effects estimation. MrDAG combines three causal inference strategies. It uses genetic variation as instrumental variables to account for unobserved confounders. It performs structure learning to detect and orientate the direction of the dependencies within the exposures and the outcomes. Finally, interventional calculus is employed to derive principled causal effect estimates. In MrDAG the directionality of the causal effects between the exposures and the outcomes is assumed known, i.e., the exposures can only be potential causes of the outcomes, and no reverse causation is allowed. In the simulation study, MrDAG outperforms recently proposed one-outcome-at-a-time and multi-response multi-variable Bayesian MR methods as well as causal graphical models under the constraint on edges' orientation from the exposures to the outcomes. MrDAG was motivated to unravel how lifestyle and behavioral exposures impact mental health. It highlights first, education and second, smoking as effective points of intervention given their important downstream effects on mental health. It also enables the identification of a novel path between smoking and the genetic liability to schizophrenia and cognition, demonstrating the complex pathways toward mental health. These insights would have been impossible to delineate without modeling the paths between multiple exposures and outcomes at once.

Introduction

Genetic evidence is increasingly used to infer causal relationships between human traits in Mendelian randomization (MR) analysis. The standard MR paradigm, one exposure and one outcome, can be biased by unmeasured pleiotropy. It occurs when the genetic variants used as instruments in the MR analysis act via separate pathways to the exposure under investigation. Extensions to consider multiple exposures¹ along with multi-response² of standard MR allow to model pleiotropy, acting via any of the exposures or any of the outcomes or both, respectively.

These and similar methods suffer an important limitation, since they are not designed to account for the dependency relations within the exposures and the outcomes to enhance the detection of causal effects between them and improve their accuracy. As we show in our motivating data application on mental health phenotypes, it is a common problem in practical applications that the effect of an exposure on an outcome can be confounded or (partially or completely) mediated by another exposure³ or mediated by another outcome, or both. However, this structure is latent and not known and consequently needs to be learned from the data.

The first attempt to provide a solution to this problem is a Bayesian network algorithm presented by Howey et al.⁴ Based on individual-level data, they apply a score-based method to determine the dependency structure among the variables (genetic variants and traits) under the constraint of directionality between the genetic variants used as instrumental variables (IVs) (called genetic anchors) and the traits. This is the only assumption regarding the directionality, so their method can be used in a “bidirectional” or “reciprocal” fashion to determine the direction of causation between two traits. However, unobserved confounding that operates between traits that are not directly linked with the genetic anchors might bias the results. Moreover, the directed acyclic graph (DAG) that they identify might not be unique, since other DAGs can hold the same conditional independencies and, thus, the same score. A similar approach for individual-level data is proposed in Badsha and Fu⁵ with a constraint-based method to detect the dependency relations among the variables. Besides the problem of unobserved confounders, since the data could be of mixed type (discrete and continuous variables), the specification of a unique type of conditional independence test for the entire dataset is also problematic.

¹Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK; ²MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK; ³UK Dementia Research Institute, Imperial College London, London, UK; ⁴Department of Public Health, University of Copenhagen, Copenhagen, Denmark; ⁵Helmholtz Pioneer Campus, Helmholtz Munich, Neuherberg, Germany; ⁶Computational Health Centre, Helmholtz Munich, Neuherberg, Germany; ⁷School of Medicine and Health, Technical University of Munich, Munich, Germany; ⁸Department of Genomic Medicine, School of Clinical Medicine, University of Cambridge, Cambridge, UK; ⁹MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, Cambridge, UK

*Correspondence: v.zuber@imperial.ac.uk (V.Z.), lb664@cam.ac.uk (L.B.)

<https://doi.org/10.1016/j.ajhg.2025.03.005>.

Crown Copyright © 2025 Published by Elsevier Inc. on behalf of American Society of Human Genetics.
This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



Other solutions to this problem have been proposed recently from the same lab.^{6–10} They are bidirectional MR models that consider the problem of invalid and weak IVs with inference performed by using a frequentist approach. Some of these methods have been developed for individual-level data,^{8,10} while others work with summary-level statistics^{6,7,9} Lin et al.^{6,7,9} present a two-step approach that first utilizes bidirectional MR on every pair of traits to construct a total effect causal graph and then applies network deconvolution to the estimated (total) causal network to estimate the direct causal effect graph between two traits conditionally on the mediating effects of other traits in the graph. In the first step, the MRcML method¹¹ for invalid IV screening is used to infer both causal directions. In the second step, a graph deconvolution algorithm¹² is employed on perturbed datasets¹³ to perform accurate inference in finite samples and mitigate the effect of weak instruments. However, to construct the network of direct effects, Lin et al.⁶ rely on a critical assumption regarding the spectral radius for network deconvolution that might be violated in practice. Moreover, the total effects are decomposed into the direct effects of all possible trait pairs, including the non-significant ones. This might lead to the “dilution” of the causal effect size of significant trait pairs. Chen et al.⁷ present one- and two-sample summary-level approaches for causal network inference based on structure equation modeling that accounts for the possible presence of some invalid IVs and consider the possibility of bidirectional relationships between traits. In the two-sample approach, the methodology requires complete sample overlap of the exposures (for instance, when molecular traits from the same study are considered), but limits its applicability to more general cases of exposures derived from cohorts of different sizes. In addition, it does not incorporate underlying graph uncertainty, and causal graphical model selection is performed by thresholding adjusted *p* values. Li et al.⁸ solve the problem of identifiability of Gaussian DAGs for individual-level and Zilinskas et al.⁹ for summary-level data, respectively. They distinguish between the primary variables (traits) and intervention variables (genetic variants). To make their model identifiable, they assume (among others) that any intervention variables, called instruments, cannot intervene on multiple primary variables (exclusion restriction) and that each primary variable is intervened by at least one instrument. However, their exclusion-restriction definition does not include unobserved confounders differently from a similar condition for valid IVs in MR. This limitation has been removed in Chen et al.¹⁰ for Gaussian individual-level data.

Here, we contribute to the solution of this problem by taking a different “unidirectional” approach. We propose the MrDAG model, an MR method with essential graphs (EGs) learning and causal effects estimation. MrDAG uses summary-level genetic associations from genome-wide association studies (GWASs) to learn how inter-related expo-

sure affect multiple outcomes which, in turn, are interconnected in a complex fashion. The estimated relationships within the exposures and the outcomes are then used for the estimation of the direct causal effects. MrDAG combines three causal inference strategies, the first of which is the MR paradigm, which uses genetic variation as IVs^{14,15} to ensure unconfoundedness. As such, the directionality of the causal effects between the exposures and the outcomes is assumed known, in the sense that exposures can only be potential causes of the outcomes and no reverse causation from the outcomes to the exposures is allowed. This reflects a hypothesis-driven research question that aims at understanding how certain exposures affect a set of related outcomes, and it is a plausible assumption when only the designed exposures are modifiable, i.e., an intervention can be carried out. The second strategy is structure learning,¹⁶ i.e., graphical models selection to define the graphs that best describe the dependency structure in a given dataset, thus accounting for graph uncertainty, under the constraint on edges’ orientation from the exposures to the outcomes. The third strategy is interventional calculus to derive principled causal effects estimates¹⁷ given the identified graphical models, thus shrinking to zero the causal effects of unimportant dependency relations.

Our motivating real data application considers the impact of six common modifiable lifestyle and behavioral exposures on seven mental health phenotypes. Mental health describes patterns of cognitive, emotional, and behavioral dysregulations that limit daily functioning and cause distress. One in eight individuals suffers from one or more mental health phenotypes worldwide, most commonly anxiety-, attention-deficit hyperactivity-, autism-spectrum-, bipolar-, eating-, personality-, or schizophrenia-related diseases.¹⁸ Collectively, they contribute to more than 15% of total years lived with disability.¹⁹ Clinically, mental health phenotypes are notoriously difficult to disentangle and diagnose due to the lack of objective biological biomarkers and distinct disease impressions.²⁰ No symptom can be uniquely ascribed to one disease, and each disease comprises experiencing a group of inter-related traits. In research, this complexity is reinforced by the multi-faceted mechanisms that cause and sustain mental health.^{20,21} In addition to genetic liability, numerous behavioral and lifestyle factors such as alcohol consumption, smoking, sleep hygiene, physical activity, and education contribute to the risk of developing a mental health trait.^{21,22} Notably, these factors are also affected by existing disease and treatment.²³ It is essential to appreciate these complexities when attempting to identify underlying mechanisms of mental health. While MR studies have been effective in circumventing some of the limitations of traditional epidemiology such as environmental confounding and reverse causation, MR remains largely unable to fully disentangle the interplay between traits that cause or result from mental health.²⁴

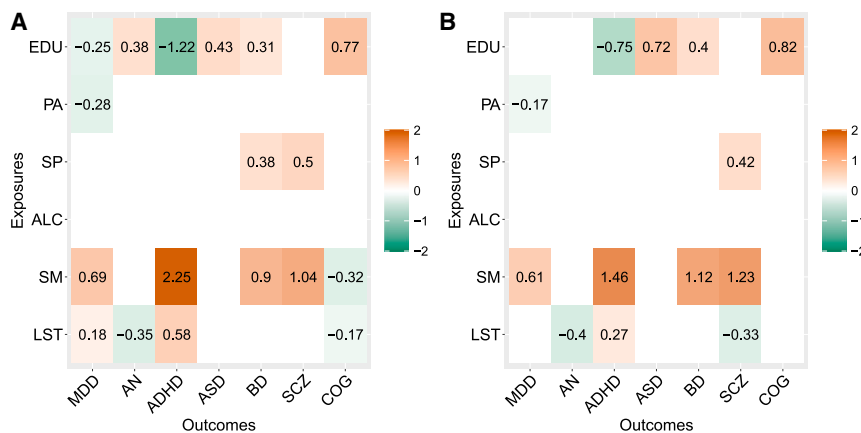


Figure 1. Results of standard and multi-variable MR (MVMR) methods regarding how lifestyle and behavioral exposures impact mental health outcomes

(A) Total causal effects estimated by standard MR²⁵ of one exposure and one outcome at a time at 5% Benjamini-Hochberg (BH) false discovery rate (FDR)²⁹ across all exposures and outcomes. (B) Direct causal effects estimated by a Bayesian MVMR method (MR-BMA)¹ after adjusting for multiple testing³⁰ at 5% BH FDR across all exposures and outcomes.

To illustrate these ideas, [Figure 1](#) shows the estimated total causal effects obtained by standard MR of one exposure and one outcome at a time²⁵ and the direct effects from a multi-variable Bayesian MR model (MVMR)¹ that accounts for the other exposures considered in the real data application. Irrespective of the method, three lifestyle and behavioral exposures seem important when considering the number of associated mental health phenotypes and the size of the causal effects: lifetime smoking index (SM), education (in years) (EDU), and leisure screen time (LST). There are a few differences between the two MR methods, in particular the role of the genetically predicted level of SM and LST on cognition (COG). However, these methods do not consider the dependency relations that might exist between lifestyle and behavioral exposures and have been already reported in MR literature, for instance, between EDU and SM and EDU and LST (EDU has a positive effect on reducing SM and LST^{26,27}). Similarly, they do not model the relationships that might be present and have been detected among mental health phenotypes. For instance, bipolar disorder (BD) and anorexia nervosa (AN) might be consequences of genetic liability to major depressive disorder (MDD).²⁸

In this study we show that, if these dependencies within multiple related exposures and multiple related outcomes are not considered, the results are severely biased by falsely detected causal effects (despite false discovery rate [FDR] control) and inflated effect sizes (see [Tables S2](#) and [S3](#)). In contrast, by estimating the relationships within modifiable lifestyle and behavioral exposures and within mental health phenotypes, MrDAG provides more interpretable results regarding the direct and indirect, i.e., partially or completely mediated, effects of each exposure on the outcomes with fewer false positives and false negatives, and thus informs precise strategies for the prevention and therapeutic intervention of mental health (see the results of real data application obtained by the proposed MrDAG model in [Figure 7](#)). A detailed discussion of these results is provided in “[real data application: the impact of lifestyle and behavioral traits on mental health](#).”

Methods

Causal inferential strategies in MrDAG

MrDAG combines three causal inference strategies.

First, MR has pioneered the ability to use genetic data as IVs to derive causal statements from observational data despite the presence of unobserved confounders.^{31,32}

Second, in its standard formulation of one exposure and one outcome, the conditional dependencies between the outcome Y , the exposure X , the IV G , and the unobserved confounder U are all given as well as their graphical representation.¹⁵ When multiple exposures $\mathbf{X}^1$ and multiple outcomes $\mathbf{Y}^2$ are considered along with multiple IVs $\mathbf{G}$, (partial) correlation between $\mathbf{X}$ and conditional dependencies between $\mathbf{Y}$ are included in the models to perform the selection of important exposures whose causal effects can be shared or are distinct across the responses. However, no dependency relations within the exposures and the outcomes are estimated by these methods, although, in practical applications, the effect of an exposure on an outcome can be confounded or (partially or completely) mediated by another exposure or mediated by another outcome, or both (see [Figures 2A](#) and [2B](#) for an illustration).

In real data applications, complex dependency relations between the traits are generally not known in advance and they need to be learned from the data. To detect them, we rely on EGs and structure learning. Graphical models are multi-variate distributions associated with a graph and are very effective for encoding conditional dependencies³³ between random variables. They are represented in a graph as nodes (vertices), while edges denote conditional dependence relationships between the corresponding random variables. A DAG is a directed graph, where each edge has an orientation with no directed cycles. Structure learning is a model selection problem¹⁶ to estimate the graph (or competing graphs) that best describes the dependency structure in a given dataset. However, without identifiability conditions,^{8–10,34} it is not possible to estimate uniquely the underlying DAG, since its conditional independencies can be associated with several alternative DAGs. The set of DAGs that hold the same conditional independencies is known as Markov equivalent class (MEC), and the best that can be done from observational data is to estimate this class (or competing classes). Moreover, all DAGs with the same conditional independencies can be represented by an EG.³⁵ Thus, this study aims to illustrate how to perform EG learning (and thus the exploration of distinct DAGs in the identified EGs whose

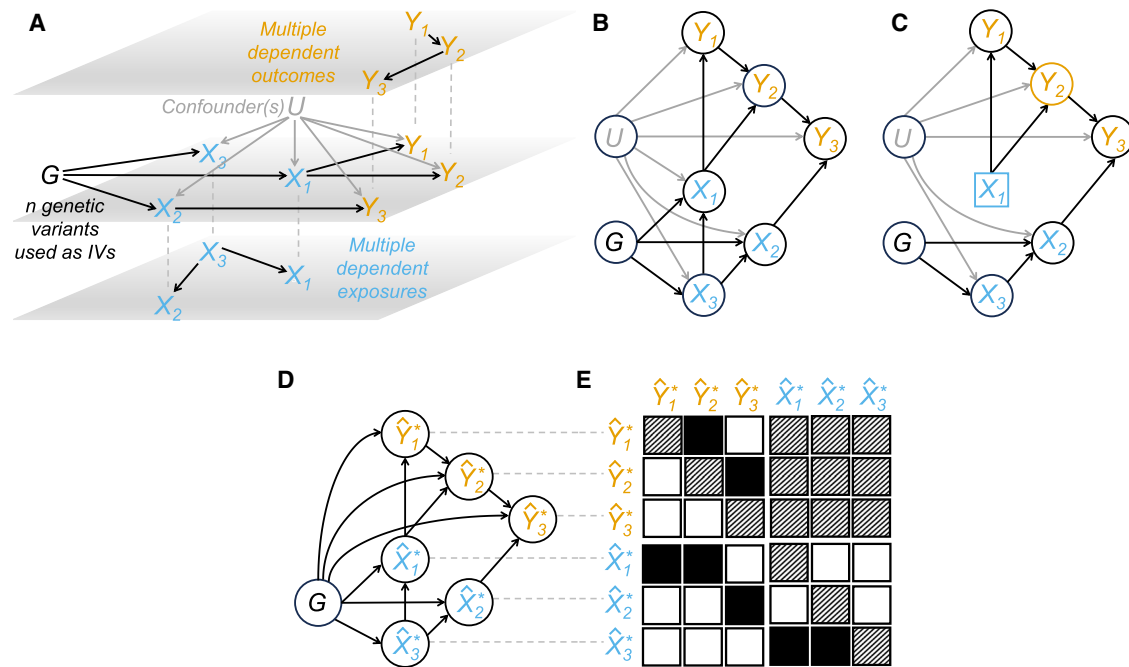


Figure 2. Representation of the proposed multiple exposures and multiple outcomes Mendelian randomization model and causal effects estimation where reverse causation is not allowed

(A) (Middle) Multi-variable Mendelian randomization for multiple responses with $G = (G_1, \dots, G_n)^T$: genetic variants (black) or instrumental variables (IVs); $X = (X_1, X_2, X_3)^T$: exposures (blue); $Y = (Y_1, Y_2, Y_3)^T$: responses (orange); U : unobserved confounder(s) (gray). True (unconfounded by U) exposure-outcome dependency relations are depicted in the middle panel. (Bottom) True fork structure within the exposures with X_3 regarded as the common cause of X_1 and X_2 . (Top) True chain structure within the outcomes, where Y_1 affects Y_3 through Y_2 .

(B) Directed acyclic graph (DAG) obtained by combining the panels in (A).

(C) Estimation of the causal effect under intervention in X_1 on Y_2 , highlighted in blue and orange, respectively. The representation of X_1 has changed to emphasize that, under intervention, it is no longer a random variable. Intervention affects only the conditional distribution of X_1 , i.e., $X_1|(X_3, G, U)$, and it leaves unaltered all the others. After removing the effect of Y_1 by marginalization (see Appendix A), it would be sufficient to condition on X_3, G , and U (graphically, the directed edges to X_1 from X_3, G , and U are removed) to guarantee that the association between X_1 and Y_2 is purely causative (see Figure S1). However, since U is unobserved, the estimation of the causal effects cannot be obtained only by conditioning.

(D) Genetically predicted exposures $\hat{X}^* = (\hat{X}_1^*, \hat{X}_2^*, \hat{X}_3^*)^T$ and outcomes $\hat{Y}^* = (\hat{Y}_1^*, \hat{Y}_2^*, \hat{Y}_3^*)^T$ depend only on G , which are chosen to be associated with X and not with Y . Graphically, no directed edges to $\hat{X}^*$ and $\hat{Y}^*$ from U are pictured. True (unconfounded by U) dependency relations between the traits in the original (individual-level) data shown in (B) are obtained by using $\hat{X}^*$ and $\hat{Y}^*$.

(E) Adjacency matrix describing the Markov properties of the DAG involving the genetically predicted exposures and outcomes (the variables in the x axis are dependent on the variables in the y axis) that are function of the IVs and the inverse-variance weighting (IVW)

(depicted with an asterisk) summary-level statistics $\hat{B}_X^* = (\hat{\beta}_{X_1}^*, \hat{\beta}_{X_2}^*, \hat{\beta}_{X_3}^*)^T$ and $\hat{B}_Y^* = (\hat{\beta}_{Y_1}^*, \hat{\beta}_{Y_2}^*, \hat{\beta}_{Y_3}^*)^T$. Neither reverse causation (top-right submatrix) nor phenotypic traits feedback loops (main diagonal) are allowed. Color code: black, directed edge between variables; white, no causal relationship between variables; black-white strips, directed edge not allowed (feedback loop and reverse causation between exposures and outcomes).

importance will be apparent in the next paragraph) that best fit the data under the constraint on the orientation of the edges, known as partial ordering,³⁶ from the exposures to the outcomes implied by the MR paradigm.

Third, along with the identification of the exposure-outcome relations as well as the dependency patterns within the exposures and the outcomes, we are also interested in causal effects estimation under intervention.¹⁷ Intervention has to be interpreted as a manipulation of an exposure to be forced to take a particular value ("doing") in contrast to the natural value that can be observed ("seeing").³⁷ This objective is possible, since graphical models based on DAGs are suited for causal reasoning based on the notion of interventional distribution¹⁷ (see Appendix A for details). An intervention on the exposures can be made explicit by a suitable modification of the multi-variate distribution associated with the DAG, under the assumption that the intervention does

not affect any other variable in the joint distribution besides the conditional distribution of the exposure under intervention.³⁸ DAGs in which it is possible to perform an intervention on any arbitrary node are called causal DAGs.³⁷ After intervention, it is possible to use graphical rules to convert the conditioning on "doing" (intervention) into conditioning on "seeing" (observation), derive the interventional distribution, and finally estimate causal effects.¹⁷ Figure 2C presents an example of the intervention on an exposure and the estimation of the causal effect on an outcome.

In the formulation described above, all confounders should be measurable to perform structure learning and causal effects estimation (causal sufficiency assumption³⁹). This condition is (explicitly or implicitly) assumed^{4,5,8,9} and is usually not met in real data applications where, instead, unobserved confounders are ubiquitous and affect exposures and responses at the same

time. To solve this problem, we demonstrate (see [Appendix A](#)) and show in an extensive simulation study (see [Results](#)) that, under partial ordering, we can estimate the dependency structure that exists between the traits in the original (individual-level) data unconfounded by U by using their genetically predicted values. Since the genetically predicted traits depend only on the selected IVs, the confounders do not mask the true dependency relations required in causal effects estimation. See [Figure 2D](#), where the graphical model estimated by using genetically predicted exposures and outcomes approximates the corresponding graph in the individual-level data not affected by U . Our proposed approach shares similarities with methods based on the genetic correlation and developed to analyze the joint genetic architecture of complex traits,⁴⁰ where the genetically predicted exposures and outcomes can be seen as the estimated genetic components of the traits. We use this analogy to show the unconfoundedness of the estimated dependency structure that exists between the traits. Finally, for a given DAG in the identified EGs, the genetically predicted values of the exposures and the outcomes are used to derive the causal effect estimator that complies with Pearl's back-door criterion,¹⁷ which indicates the variables that must be added to the regression equation to eliminate what is known as "omitted variable bias."

The MrDAG model can be summarized as follows:

$$[\mathbf{g}^\top \hat{\mathbf{B}}_Y^\top \mathbf{g}^\top \hat{\mathbf{B}}_X^\top]^\top \sim N_{q+p}([\mathbf{g}^\top \mathbf{B}_Y^\top \mathbf{g}^\top \mathbf{B}_X^\top]^\top, \Sigma^*), \quad (\text{Equation 1})$$

where $\mathbf{g}$ are the IVs after pruning or clumping, $\hat{\mathbf{B}}_Y^\top$ and $\hat{\mathbf{B}}_X^\top$ are the inverse-variance weighted (IVW)⁴¹ estimated genetic associations with the outcomes and the exposures, $\hat{\mathbf{Y}}^* = \mathbf{g}^\top \hat{\mathbf{B}}_Y^\top$ and $\hat{\mathbf{X}}^* = \mathbf{g}^\top \hat{\mathbf{B}}_X^\top$ are the genetically predicted values of the outcomes and exposures based on the IVs (see [Figure 2D](#)), which are normally distributed for large sample sizes, and Σ^* is the covariance matrix that can be partitioned into Σ_{YY}^* , Σ_{XX}^* , and Σ_{XY}^* , the genetic covariances within the outcomes, the exposures, and between them. Note that the genetically predicted values of the outcomes and exposures do not need to be available/calculated, since the proposed causal graphical model uses as input data the sufficient statistic for Σ^* , which is a function of the summary-level data and the linkage disequilibrium (LD) structure between the genetic variants selected as IVs. Thus, the only additional information that is required from individual-level data is the LD matrix $\mathbf{V}$. However, this information is not necessary when independent genetic variants are considered after pruning or clumping, as we have done in the simulation study and the real data application, and, thus, $\mathbf{V} = \mathbf{I}_n$ (see [Appendix A](#)).

By using a summary-level MR design, the MrDAG model allows us to find a solution to the two problems highlighted before. First, we perform structure learning under partial ordering by using $\Omega^* = \Sigma^{*-1}$ to learn the unconfounded dependency relations within the exposures, the outcomes, and between them and to understand the genetic paths that link exposures and outcomes. Second, we estimate the causal effects of the intervention on the exposures as a function of trait-specific elements of the genetic associations $\hat{\mathbf{B}}_Y^\top$ and $\hat{\mathbf{B}}_X^\top$ informed by distinct DAGs in the identified EGs, unconfounded by any measured and unmeasured pleiotropic effects⁴² within the exposures and the outcomes, respectively, and any unobserved confounder.

Finally, the uncertainty regarding which EGs best describe the data is fully accounted for in the Bayesian implementation of the proposed model (see [Appendix A](#)). Posterior inference allows us to rank the identified graphical models according to their importance and to obtain causal effects by Bayesian model averaging,

which shows advantages compared to frequentist approaches.⁴³ Sparsity to detect important causal effects is obtained by specifying the *a priori* number of edges (or its probability) in the graphical model, which is easier to elicit than Lasso-type penalization on the space of causal effects used in frequentist approaches.⁴⁴

Selection of instrumental variables

MrDAG uses the same instrument selection procedure employed in MVMR.¹ A genetic variant is considered a valid instrument for MVMR when three core conditions hold.³ (IV1) Independence: the variant is independent of all confounders of each of the exposure-outcome associations. (IV2) Relevance: the variant must not be conditional independent of each exposure given the other exposures. (IV3) Exclusion restriction: The variant is independent of the outcome conditional on the exposures and confounders. We revise these core conditions for multiple exposures and extend them for multiple outcomes in the [Appendix A](#). In practice, only IV2 can be computationally evaluated from the available data. Tests for weak IV bias that arises because some genetic variants are weakly associated, with some exposures conditional on the other exposures are available.⁴⁵ A recent solution to mitigate the effects of weak IVs in MVMR is presented in Wu et al.⁴⁶

There is an important distinction between IV selection in MVMR, as used by MrDAG, and bidirectional MR. Let us consider two traits A and B. In bidirectional MR, two MR analyses are conducted, one for trait A on trait B and vice versa. First, specific IVs are selected for trait A and the first MR model is fit. Another set of specific IVs is then selected for trait B, and the second MR model tests the opposite effects direction. In contrast, in MVMR, IVs are chosen to be the union of genome-wide significant genetic variants for any exposure. By combining MVMR IV selection approach with EG learning, MrDAG can infer the bidirectionality of the relationships within exposures based on $\Omega_{XX}^* = \Sigma_{XX}^{*-1}$ without repeated IV selection and subsequent analyses. A similar comment can be made for the estimation of the bidirectionality of the relationships within the outcomes based on $\Omega_{YY}^* = (\Sigma_{YY}^* - \Sigma_{YX}^* \Sigma_{XX}^{*-1} \Sigma_{XY}^*)^{-1}$ (see [Appendix A](#)). The dependencies within the outcomes can be interpreted as an indication of a violation of condition IV3, i.e., pleiotropy not explained by the estimated causal effects from the exposures to the outcomes.² The detected relationships within the exposures also suggest the existence of pleiotropy, which, in the proposed framework, comprises confounding, mediation and independent pleiotropic pathways.³ In this study, measured pleiotropy is accounted for via any of the exposures included in the model.³ Unmeasured pleiotropy, in contrast, is accounted for when an unmeasured pathway impacts more than one outcome jointly and introduces unidirectional and/or bidirectional effects among the genetic associations of the outcomes.²

While MrDAG can account for the impact of these pleiotropic effects, other direct effects of some IVs on the outcomes might exist and bias the results. Indeed, another unmeasured pleiotropy might be present in which some genetic variants are directly associated with a single outcome at a time and not via the exposures. To deal with this scenario, a possible extension of MrDAG model in [Equation 1](#) is in analogy with MR-Egger for MVMR,⁴² where an intercept is added. However, the InSIDE assumption must be imposed, and its violation can cause further bias.⁴⁷ Instead of extending the MrDAG model in this direction, an alternative strategy is to check whether any of the genetic variants used as IVs do not follow the proposed model. This is equivalent to detecting outliers as proposed for

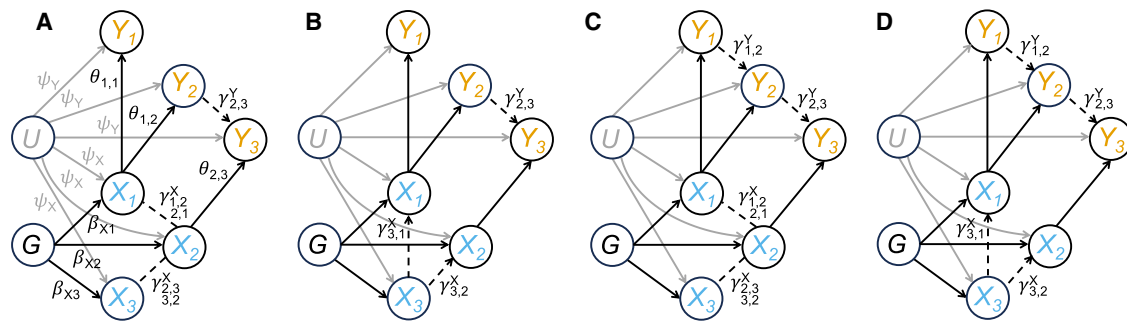


Figure 3. Schematic illustration of different dependency structures simulated between the traits at the individual-level data and the parameters employed in the simulation study

Directed edges indicate dependency relations, while undirected edges denote partial correlations. Dashed lines depict the true (unconfounded by U) dependency structure within the exposures and the outcomes, while solid lines indicate true causal effects between them. Parameters ψ_Y and ψ_X indicate the simulated effects of the unobserved confounder U on the exposures and the outcomes, respectively, and $\mathbf{B}_X = (\beta_{X_1}, \beta_{X_2}, \beta_{X_3})$ are the simulated genetic effects on the exposures. For simplicity, they are shown only on the left panel. $\Theta = (\theta_{1,1}, \theta_{1,2}, \theta_{2,3})$ are the simulated causal effects from the exposures to the outcomes, while $\Gamma_X = (\gamma_{3,1}^X, \gamma_{3,2}^X)$ and $\Gamma_Y = (\gamma_{1,2}^Y, \gamma_{2,3}^Y)$ are the mediation parameters within the exposures and the outcomes, respectively, where the subscripts denote their directionality. When partial correlations are simulated within the exposures, bidirectional effects are depicted with double subscripts, i.e., $\Gamma_X = (\gamma_{1,2}^X, \gamma_{2,1}^X, \gamma_{2,3}^X, \gamma_{3,2}^X)$.

(A) Simulated scenario “UndG_X-Med_Y,” where an undirected graph (“UndG_X”) encodes the dependency pattern within $\mathbf{X}$ and, within the responses, an outcome (Y_3) is completely mediated (“Med_Y”) by another response (Y_2), which, in turn, is affected by a different exposure (X_1). Although there is another partial mediation between X_1 and Y_3 through X_2 , this mediation happens within $\mathbf{X}$, so it does not affect the definition of complete mediation within $\mathbf{Y}$.

(B) Simulated scenario “DAG_X-Med_Y,” where a topologically ordered DAG within the exposures (“DAG_X”) is simulated. Specifically, in the example depicted, a fork structure is simulated, i.e., X_3 affects both X_1 and X_2 . A complete mediation is still considered within the responses.

(C) Simulated scenario “UndG_X-DAG_Y.” Here, the dependency structure between the individual-level responses is obtained by simulating a topologically ordered DAG (“DAG_Y”). Specifically, a chain structure is considered, i.e., Y_1 affects Y_2 , which, in turn, affects Y_3 , whereas an undirected graph encodes the dependency pattern within $\mathbf{X}$.

(D) Simulated scenario “DAG_X-DAG_Y,” where two topologically ordered DAGs are simulated within the exposures (fork structure) and outcomes (chain structure), respectively.

univariable⁴⁸ and multi-variable MR models³ to identify specific IVs that might be invalid due to a direct unmeasured pleiotropic pathway impacting one outcome at a time. This approach, called conditional predictive ordinate (CPO), has already been pursued in a Bayesian multi-response MR model² and is also included in MrDAG. Details are presented in the [Appendix A](#).

Overall, only the direction from exposures to outcomes is fixed in MrDAG, and no reverse causation is allowed, reflecting the standard MR paradigm. Thus, one of the key design decisions for MrDAG is which variables are considered exposures and, consequently, which instruments are selected for these exposures.

Results

Simulation study

We compare MrDAG in a comprehensive simulation study where four different *in silico* scenarios have been generated on individual-level data for $N = 100,000$ individuals with $N_Y = N_X = 50,000$. The simulated datasets include $n = 100$ independent genetic variants $\mathbf{G}$, an unobserved confounder U , 15 exposures $\mathbf{X}$, and 5 outcomes $\mathbf{Y}$. All exposures $\mathbf{X}$ were measured on the same individuals in the first sample and have complete overlap, and all outcomes $\mathbf{Y}$ were measured on the same individuals in the second sample independent of the first sample. In all simulations, the unconfounded dependency relations between the traits are simulated at the individual level, while the algo-

rithms use as input data the corresponding IVW summary-level statistics.

The four simulation scenarios are built by combining two different strategies we used to simulate the dependency patterns within the exposures and the responses.

- (1) “UndG_X-Med_Y.” A sparse undirected graphical model (“UndG_X”) encodes the dependency pattern within the exposures $\mathbf{X} = (X_1, \dots, X_{15})$. Regarding the responses $\mathbf{Y} = (Y_1, \dots, Y_5)$, one outcome is completely mediated by another one (“Med_Y”). This simulated scenario aims to assess the ability of MrDAG to detect the most common type of relationship within the exposures assumed in MVMR methods.^{1,3,49} For a visual representation of this scenario, see [Figure 3A](#).
- (2) “DAG_X-Med_Y.” The dependency relations within the exposures are more complex than in scenario (1), since a topologically ordered DAG within the exposures (“DAG_X”) is simulated.⁵⁰ A complete mediation is still considered within the responses. This second scenario is illustrated in [Figure 3B](#).
- (3) “UndG_X-DAG_Y.” Here, a more complex dependency structure within the individual-level responses (“DAG_Y”) is simulated. This scenario is represented in [Figure 3C](#). An example of the complex dependency patterns generated in the simulation

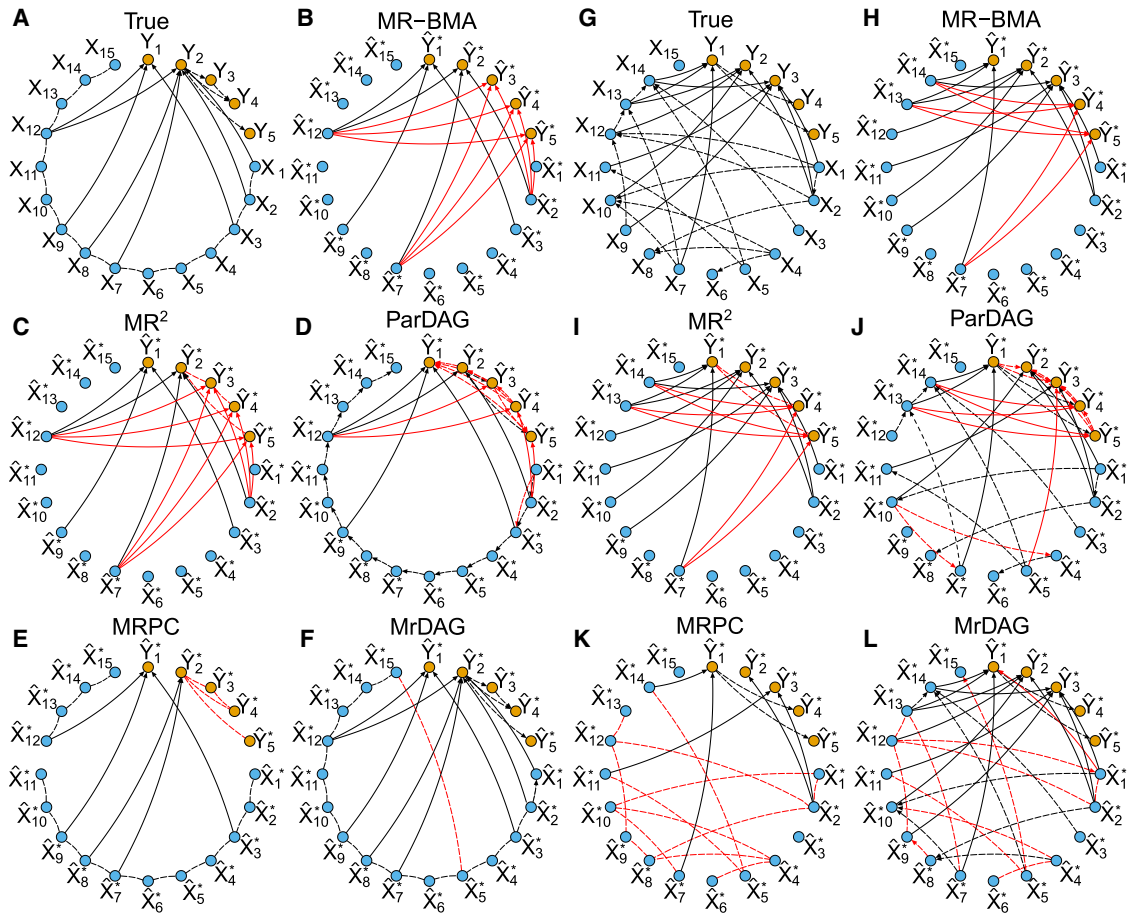


Figure 4. Examples of unconfounded dependency structure simulated at the individual-level data and estimated by using summary-level statistics within the exposures, the outcomes, and between them in two different scenarios

In each panel, individual-level outcomes $\mathbf{Y} = (Y_1, \dots, Y_5)$ and exposures $\mathbf{X} = (X_1, \dots, X_{15})$ as well as genetically predicted outcomes $\hat{\mathbf{Y}}^* = (\hat{Y}_1^*, \dots, \hat{Y}_5^*)$ and exposures $\hat{\mathbf{X}}^* = (\hat{X}_1^*, \dots, \hat{X}_{15}^*)$ are represented with orange and blue nodes, respectively. Directed edges indicate dependency relations, while undirected edges denote partial correlation. Dashed lines depict the true (unconfounded by U) and estimated dependency structure within the exposures and the outcomes, while solid lines indicate true and estimated causal effects between them. Red color denotes false positives, either falsely detected effects (regardless of the directionality) or wrong directionality of the edges. Besides the proposed model, alternative methods considered Mendelian randomization with Bayesian model averaging (MR-BMA),¹ multi-response Mendelian randomization (MR²),² Mendelian randomization with PC algorithm (MRPC),⁵¹ and partition-DAG (ParDAG).⁴⁴ We report the results of MR-BMA and MR² obtained by thresholding the marginal posterior probability of inclusion (mPPI) > 0.5 , which correspond to the median models.⁵² No threshold is applied to MrDAG posterior probability of edge inclusion (PPEI). MRPC partially directed acyclic graphs (PDAGs) are obtained by specifying the type I error rate for the conditional independence test at $\alpha = 0.01$. ParDAG results are the solutions of causal effects estimation with Lasso penalization set at $\lambda = 0.9$.

(A–F) Single replicate of the simulated scenario UndG_X-DAG_Y, where an undirected graph encodes the dependency pattern within $\mathbf{X}$ and a DAG represents the dependency relations within $\mathbf{Y}$ along with the simulated causal effects from the exposures to the outcomes, resulting in an overall partially oriented DAG. In this scenario, the strength of correlation between consecutive $\mathbf{X}$ is set at $r_X = 0.6$ and then decreases exponentially for non-consecutive exposures, and the average level of the mediation parameters within $\mathbf{Y}$ is set at $m_Y = 1$. (G–L) Single replicate of the simulated scenario DAG_X-DAG_Y, where two topologically ordered DAGs have been independently simulated within $\mathbf{X}$ and $\mathbf{Y}$ along with the simulated causal effects from the exposures to the responses, resulting in an overall fully oriented DAG. In this scenario, the average level of mediation parameters for $\mathbf{X}$ and $\mathbf{Y}$ are set at $r_X = 0.6$ and $m_Y = 1$, respectively.

study between the traits for one replicate of scenario UndG_X-DAG_Y is shown in Figure 4A.

- (4) “DAG_X-DAG_Y.” This is the most complex simulated scenario, where two independent topologically ordered DAGs have been simulated within the exposures and outcomes. Figure 3D presents a schematic illustration of this scenario, while Figure 4G shows the intricate dependency structure simulated between the traits for one replicate of DAG_X-DAG_Y scenario.

Taken together, in scenarios (2) and (4), the overall individual-level DAGs obtained by combining two different simulation strategies for $\mathbf{X}$ and $\mathbf{Y}$ are fully oriented, while in scenarios (1) and (3) the overall DAGs are partially oriented. Details regarding the parameters ψ_X and ψ_Y , the simulated levels of the effects of the unobserved confounder U on the responses and the outcomes, $\mathbf{B}_X$, the simulated levels of the genetic effects on the exposures, and $\mathbf{I}_X$ and $\mathbf{I}_Y$, the simulated levels of the mediation parameters within the exposures and the outcomes and their

average value r_X and m_Y , are presented in the [Appendix A](#). Finally, all simulations are replicated 25 times and initialized with a different random seed.

We compare MrDAG with published MVMR methods and their software implementations, excluding the comparisons from naive one-exposure one-outcome MR models, since it has been shown that they are outperformed by MVMR methods when there is measured pleiotropy among exposures.³ Specifically, we consider MR with Bayesian model averaging (MR-BMA),¹ an MVMR algorithm that allows for many exposures to be included but does not model explicitly the dependency relations within the exposures.³ MR-BMA estimates the sparse direct causal effects between the exposures and one outcome, providing the marginal posterior probability of inclusion (mPPI) along with the (Bayesian model-averaged) direct causal effects. We treat MR-BMA as the baseline algorithm for the comparisons, since it analyzes one outcome at a time. Second, we present the results of a sparse multi-variable Bayesian summary-level MR model for the joint analysis of multiple responses (MR²).² MR² estimates mPPIs, the (posterior mean of) direct causal effects between the exposures and the outcomes as well as the residual covariation between the outcomes not explained by the exposures. Similarly to MR-BMA, it allows for correlation among the exposures, while unmeasured pleiotropy between responses is accounted for by a Gaussian decomposable graphical model. Third, we include an in-house modified version for summary-level data of the PC algorithm⁵³ with the principle of MR (MRPC),⁵¹ which uses the PC algorithm for the estimation of the causal graphical model among the variables (in the original implementation: genetic variants and traits; in our modified version: exposures and outcomes) under partial ordering. At a specified type I error rate for the Gaussian conditional independence test, MRPC returns the estimated partially directed acyclic graphs (PDAGs)³⁶ (see [Appendix A](#)) in which some undirected edges are present along with the directed ones (recall that an undirected edge $z - v$ is equivalent to $z \rightarrow v$ and $v \rightarrow z$) as well as the p values of all conditional independence tests. For a given PDAG detected by MRPC in each replicate and scenario, we utilize Kalisch et al.⁵⁰ to estimate the causal effects between the exposures and outcomes. Fourth, Partition-DAG (ParDAG)⁴⁴ provides a solution to the structure learning problem once the summary-level statistics have been partitioned into two groups and the orientation of the edges from the exposures to the outcomes has been enforced. ParDAG computes the causal effects estimates under Lasso regularization. It has not been combined with instrumental variable estimation and applied to genetic data to date. Finally, we consider Graph-MRcML,⁶ which is based on a bidirectional MR framework and does not distinguish between exposures and outcomes. Among the causal graphical models considered, ParDAG is the only one that returns the estimation of a fully oriented DAG, while MRPC, Graph-MRcML, and MrDAG return PDAGs according to the designed level of

type I error rate, Bonferroni-adjusted significance level, and posterior probability of edge inclusion (PPEI), respectively. However, since in ParDAG no identifiability conditions are assumed, the reported DAG is the sparsest DAG in the MEC. All methods use summary-level statistics as input data after IVW. For each method and algorithmic implementation, details of the parameter settings are provided in [supplemental text](#). Finally, all algorithms were run on the same Cambridge high-performance computer (HPC), taking for each replicate (and based on the designed parameters settings) on average 1 min for MR-BMA and MRPC, 2 min for MR², <10 s for ParDAG, >2 h for Graph-MRcML, and 10 min for MrDAG.

Regarding the evaluation criteria, we use a precision-recall curve (PRC) that shows the relationship between precision (i.e., positive predictive value, on the y axis) and recall (i.e., sensitivity, on the x axis) for every possible cut-off and is not impacted by the over-representation of null effects. In drawing PRCs, for MR-BMA, MR², and MrDAG, we rank the estimated mPPIs and PPEIs, respectively. For Graph-MRcML, we rank p values obtained by the perturbation scheme. This allows us to represent the PRC as a smooth step line, where each step corresponds to a different cutoff on mPPIs, PPEIs, and p values. As ParDAG is based on a Lasso-type penalization, unimportant causal effects are forced to zero and excluded from the model. Consequently, ParDAG does not provide a full ranking of important dependency relations but rather a single cutoff. Thus, it is presented as a single point along with its standard error for each specified value of the penalization parameter λ instead of a continuous line. Similarly, we report as a single cutoff the PDAG estimated by MRPC in two steps (graph-skeleton selection at a specified α followed by edges-orientation step with further conditional independence tests at the same type I error rate) for different values of α . See [supplemental text](#) for a detailed discussion regarding how we implemented a fair comparison between the methods considered.

Finally, to evaluate the quality of the causal effects estimation, we calculate the sum of squared errors (SSE), defined as the sum of the squared differences between the estimated and the simulated causal effect. In contrast to the evaluation of the recovery obtained by each method of the simulated dependencies within the exposures, the outcomes, and between them, we do not report the SSE of the mediation parameters Γ_X and Γ_Y , since they are considered nuisance parameters in the proposed model (see [supplemental text](#)).

MrDAG more accurately detects unconfounded dependency relations within the exposures and the outcomes and between them

[Figure 4](#) presents the results of MrDAG and alternative methods for one replicate of the simulated scenario UndG_X-DAG_Y ([Figures 4A–4F](#)) and DAG_X-DAG_Y ([Figures 4G–4L](#)) for a particular choice of the parameters $r_X = 0.6$ and $m_Y = 1$ used in the simulation study to

control the average value of the mediation parameters I_X within the exposures and I_Y within the outcomes, and $\psi_X = 2$ and $\psi_Y = 1$ for the level of confounding on the exposures and the outcomes, respectively (see Appendix A).

For its applicability, Graph-MRcML requires the assumption that the spectral radius of the direct causal graph is less than 1,^{6,7,9} which is violated not only in these two replicates but in most of the simulated scenarios' replicates. Therefore, although originally considered, we omit Graph-MRcML from the simulation study. The general performance of the other competing algorithms is already apparent from it. In scenario UndG_X-DAG_Y, if a causal effect is simulated from an exposure to an outcome and there are dependency relations from this outcome to other responses (Figure 4A), MR-BMA adds erroneously causal effects to all linked responses with severe false positive (FP) inflation (Figure 4B, FPs between $\hat{X}_{12}^*$ and $\hat{Y}_3^*, \hat{Y}_4^*, \hat{Y}_5^*$ depicted in red). On the other hand, MR-BMA estimates neither the dependency pattern within **X**, since the (partial) correlation between summary-level exposures is assumed in the model³ but not estimated, nor the dependencies within **Y**, since it considers one response at a time. MR² detects bidirectional (as assumed by the model) relationships between the outcomes, although this is not sufficient to prevent FPs, similarly to MR-BMA (Figure 4C). ParDAG results regarding the causal effects simulated from the exposures to the outcomes are extremely sparse but very dense within the responses. Only oriented dependencies are estimated within the responses, as assumed by the model (Figure 4D). MRPC infers correctly most of the dependencies within **X**, but it does not have the power to detect all simulated causal effects Θ at the specified type I error rate for the conditional independence test ($\alpha = 0.01$) with a few false negatives (FNs) (Figure 4E, FNs between $\hat{X}_1^*, \hat{X}_2^*$ and $\hat{Y}_2^*$) and well as FPs within **Y** (FPs between $\hat{Y}_2^*, \hat{Y}_3^*, \hat{Y}_4^*, \hat{Y}_5^*$, where bidirectionally is erroneously detected). MrDAG performs better than alternative methods to detect both directed and bidirected edges, with only one FP between $\hat{X}_5^*$ and $\hat{X}_{15}^*$ (Figure 4F).

Similar comments can be made for a particular replicate of scenario DAG_X-DAG_Y, although in this scenario the dependency patterns are more complex, since a topological ordered DAG is simulated also within the outcomes (Figure 4G). MR² does not detect any unmeasured pleiotropy within the outcomes, and the results coincide with MR-BMA, both with several FPs (Figures 4H and 4I). MrDAG confirms its good performance except for the directionality of the dependency relations within **X**, where bidirectional edges are found with a few FPs (Figure 4L, FPs between $\hat{X}_{12}^*$ and $\hat{X}_1^*$ and between $\hat{X}_8^*$ and $\hat{X}_9^*$) and an FP between $\hat{X}_1^*$ and $\hat{Y}_1^*$ (although no threshold has been applied to PPEIs) and no FNs.

Figure 5 generalizes the results depicted in Figure 4, averaging the results over 25 replicates of the simulated scenarios UndG_X-DAG_Y (Figures 5A–5C) and DAG_X-DAG_Y (Figures 5D–5F) with the same parameters setting used in Figure 4. The results are presented separately for the simu-

lated dependency structures from the exposures to the outcomes (Figures 5A and 5D), within the exposures (Figures 5B and 5E), and within the outcomes (Figures 5C and 5F), respectively.

On average, MRPC and MrDAG have good performance in both simulated scenarios (Figures 5A and 5D). MRPC's best results are obtained at a stringent type I error rate $\alpha = 0.01$ for the conditional independent tests (blue dots), although they are quite similar across different values of α and thus robust to this choice. However, it fails to detect the simulated dependency pattern within **X** in scenario DAG_X-DAG_Y (Figure 5E). The performance of MR-BMA can be only evaluated for the detection of the causal effects from the exposures to the outcomes (Figures 5A and 5D). As we noticed above, the large number of FPs degrades the results of this method, which was not developed to deal with multiple related responses. MR² is not able to detect complex dependency relations simulated within the outcomes (Figures 5C and 5F), although this is expected given the assumed bidirectionality within **Y**. It performs better than MR-BMA in scenario DAG_X-DAG_Y but not in scenario UndG_X-DAG_Y, where dependencies within **Y** are wrongly estimated, showing that its results crucially depends on the quality of the detected unmeasured pleiotropy. The performance of ParDAG is the worst among the methods considered for all types of designed relationships, slightly better within the exposures (Figures 5B and 5E) and between the exposures and outcomes (Figures 5A and 5D), and worse within the outcomes (Figures 5C and 5F), likely due to the very dense solutions within **Y** as already noted in Figures 4D and 4J. Since ParDAG detects only directed edges, in Figure 5B, where the partial correlation between exposures is simulated, the method has 50% recall rate. The results also seem quite different according to the penalty parameter λ .

MrDAG has a strong performance in both scenarios. In contrast to MR-BMA and MR², in scenario DAG_X-DAG_Y (Figures 5D–5F), there is only a small reduction of the precision in the estimation of the dependency relations between the exposures and the outcomes and within the latter, compared to the scenario UndG_X-DAG_Y (Figures 5A–5C).

The comments above can be extended to the scenarios where the relationships within outcomes are completely mediated (UndG_X-Med_Y depicted in Figures S2A–S2C and DAG_X-Med_Y shown in Figures S2D–S2F). In these scenarios, the mediation within the outcomes is easier to detect (Figures S2C and S2F) than a topologically ordered DAG simulated within **Y**.

Figure S4 shows the results of the area under the curve of precision recall (AUCPR) to detect the causal effects Θ and the sensitivity of the methods to different specifications of r_X and m_Y . MrDAG is confirmed to be the best method, with stable AUCPR for any combination of r_X and m_Y and with similar AUCPR when partial correlation or a topological ordered DAG is simulated within **X**. MR-BMA performs well, especially in the scenario UndG_X-Med_Y

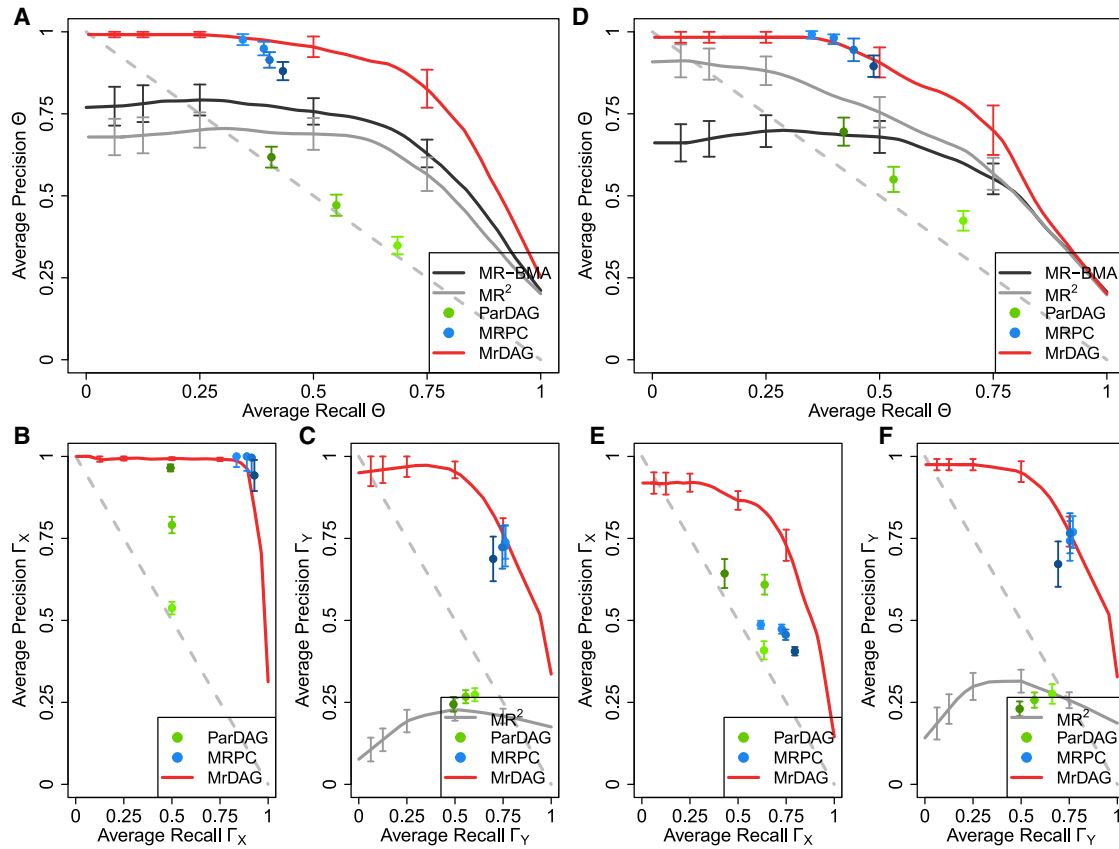


Figure 5. Precision-recall curves (PRCs) for all methods considered in the simulated scenarios

UndG_X-DAG_Y and DAG_X-DAG_Y show recall (= sensitivity = TP/(TP+FN)) in the *x*-axis and precision (= positive predictive value = TP/(TP+FP)) in the *y*-axis with TP = true positive, FN = false negative and FP = false positive averaged over 25 replicates in each scenario. In scenario UndG_X-DAG_Y (A–C), the strength of correlation between consecutive **X** is set at $r_X = 0.6$ and then decreases exponentially for non-consecutive exposures, and the average level of the mediation parameters within **Y** is set at $m_Y = 1$, while in scenario DAG_X-DAG_Y (D–F), the average level of the mediation parameters within **X** and **Y** is set at $r_X = 0.6$ and $m_Y = 1$, respectively. For details, see [Appendix A](#). In both scenarios, the results are presented separately for the simulated dependency structures from the exposures to the outcomes (A and D), within the exposures (B and E) and the outcomes (C and F), respectively. Vertical bars in each PRC, at specific recall levels 0.0625, 0.125, 0.25, 0.50, and 0.75, indicate standard error. For the MRPC algorithm, the type I error rate for the conditional independence test is set at $\alpha = \{0.01, 0.05, 0.10, 0.20\}$ (from light- to dark-blue dots), and for the ParDAG algorithm we specify three different values for the Lasso penalization $\lambda = \{0.5, 0.7, 0.9\}$ (from light- to dark-green dots). See [supplemental text](#) for details.

([Figure S4A](#)), which is the scenario that is most compatible for this method, as well as in scenario DAG_X-Med_Y ([Figure S4C](#)), where its performance slightly decreases. Despite the limitations highlighted above, MR² is overall the second-best method, although it shows a drop of power in scenario UndG_X-DAG_Y ([Figure S4B](#)). Both MRPC and ParDAG seem to be less precise at higher levels of r_X irrespective of the simulated scenario, with ParDAG also influenced by the value of m_Y . Similarly, [Figures S5](#) and [S6](#) show the sensitivity of the algorithms to detect the simulated patterns within **X** and within **Y** for different specifications of r_X and m_Y .

In summary, MrDAG outperforms competing methods in estimating the dependency relations, unconfounded by *U*, within the exposures, within the outcomes, and between them, simulated at the individual level and estimated by using summary-level data in a variety of scenarios and with different levels of the strength of the

dependencies (mediation parameters r_X and m_Y) within the exposures and the outcomes.

MrDAG improves the estimation of the causal effects over existing methods

[Figure 6A](#) shows the SSE of the causal effects Θ between the exposures and the outcomes for all methods considered in the simulated scenario UndG_X-DAG_Y and in [Figure 6B](#) for the simulated scenario DAG_X-DAG_Y across 25 replicates in each scenario with the same parameter settings and implementation of algorithms described above. For MRPC and ParDAG algorithms, we only show the results obtained at type I error rate for the conditional independence test $\alpha = 0.01$ and Lasso penalization $\lambda = 0.9$, respectively. These values provide the best results for the two algorithms as shown in [Figures 5](#) and [S2](#).

MrDAG has the lowest SSE mean and median (white dots and horizontal black line, respectively) in both

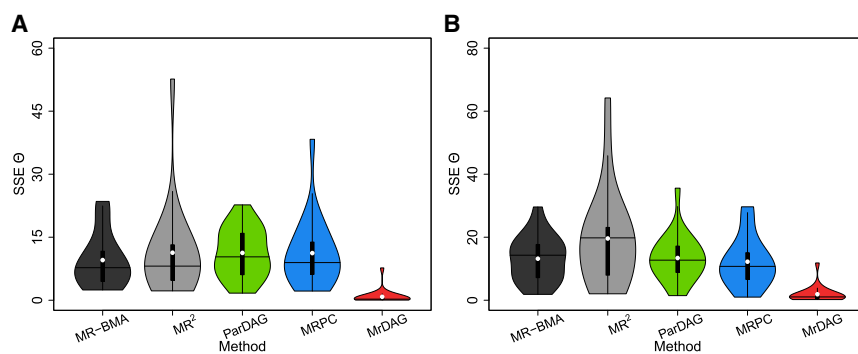


Figure 6. Violin plots of the sum of squares error (SSE) of the causal effects Θ between the exposures and the outcomes for all methods considered in the simulated scenarios UndG_X-DAG_Y and DAG_X-DAG_Y across 25 replicates in each scenario

(A) In scenario UndG_X-DAG_Y, the strength of correlation between consecutive **X** is set at $r_X = 0.6$ and then decreases exponentially for non-consecutive exposures, and the average level of the mediation parameters within **Y** is set at $m_Y = 1$.

(B) In scenario DAG_X-DAG_Y, the average level of the mediation parameters within **X** and **Y** is set at $r_X = 0.6$ and $m_Y = 1$,

respectively. For details, see [Appendix A](#). In each violin plot, the vertical black thick line displays the interquartile range, the black horizontal line denotes the median, and the white dot denotes the mean. For MRPC and ParDAG algorithms, we only show the results obtained at type I error rate for the conditional independence test $\alpha = 0.01$ and Lasso penalization $\lambda = 0.9$, respectively. These values provide the best results for the two algorithms as shown in [Figures 5](#) and [S2](#).

scenarios. As expected, when a topological ordered DAG is simulated within the exposures ([Figure 6B](#)), the violin plots have a wider range, showing more variable results, although the median is almost similar to the scenario with simulated partial correlation within **X** ([Figure 6A](#)). Alternative methods have larger SSEs.

Similar comments can be made for simulated scenarios UndG_X-Med_Y ([Figure S3A](#)) and DAG_X-Med_Y ([Figure S3B](#)), where a complete mediation is considered within the outcomes. MrDAG is confirmed as the best method.

We conclude this section by inspecting the sensitivity of the SSE of the causal effects between the exposures and the outcomes for different values of the average level of the mediation parameters r_X and m_Y . The estimation of the causal effects displayed in [Figure S7](#) shows that both MR-BMA and MRPC depend on the combination of r_X and m_Y , with almost similar performance when a complete mediation is simulated ([Figures S7A](#) and [S7C](#)). MR² has good performances across all scenarios compared to the other methods, but its behavior depends largely on the simulated level of m_Y , which, in turn, affects the estimated dependency relations within **Y** (see [Figure S6](#)). Compared to the other methods, MrDAG is not only the best, but it is rather insensitive to different levels of the mediation parameters within **X** and **Y**.

In summary, MrDAG has the lowest bias in the estimation of the causal effects in all simulation scenarios and for any combinations of the parameters r_X and m_Y that control the strengths of the pleiotropy within the exposure and the outcomes. The advantage of MrDAG is more pronounced when there are complex relationships within the responses or the outcomes, and, in particular, when both cases are simulated, which reflects more closely what happens in real-life applications.

Robustness to noisy genetic association estimates and mis-specification of the exposure-outcome groups' definition

We evaluate the robustness of the proposed MrDAG model by looking at the effect of noisy genetic association esti-

mates and the mis-specification of the exposure-outcome groups' definition.

Regarding the effect of imprecise genetic association estimates, we replicate the setup used in the simulated scenarios described above, but we decrease the number of individuals from $N = 100,000$ to $N = 20,000$, equally split between the exposures and the outcomes, i.e., $N_Y = N_X = 10,000$. Although this can be considered an extreme case because the sample size in modern GWASs is much larger (see [Table S1](#) for the number of individuals considered in the real data application), it reflects the presence of noisy genetic association estimates.

The results concerning the detection of the simulated sparse signals and the quality of the causal effect estimates are presented in [Figures S8–S11](#) for all simulated scenarios and a particular choice of the parameters $r_X = 0.6$ and $m_Y = 1$. While there is a less clear advantage of the proposed MrDAG model over MR-BMA and MR² in all scenarios considered and a similar behavior regarding the detection of the simulated causal effects with the MRPC algorithm (see [Figures S8](#) and [S9](#)), MrDAG outperforms the other methods in the quality of the causal effect estimates when dependency relations are simulated within the outcomes (see [Figure S11](#)). Only in one case do we record a worse performance of the proposed model than alternative methods, specifically against MR-BMA and MR² in their most favourable simulated scenario UndG_X-Med_Y (see [Figure S10A](#)).

Regarding the robustness to the mis-specification of the exposure-outcome groups' definition, we took the datasets originally simulated and incorrectly defined the two groups. Specifically, for all algorithms considered, 5 exposures are now mis-specified as outcomes, which reduced the group of exposures from 15 to 10 and increased the group of outcomes from 5 to 10. In doing so, reverse causation, originally not considered in the simulation study, is now present. For instance, this happens if, in a relation dependence between two exposures, the parent node is mis-specified and wrongly assigned to the outcomes group.

We evaluate the ability of the algorithms to detect the simulated sparse signals based not only on a subset of the causal effects Θ originally simulated from the exposures to the outcomes (i.e., selecting 10×5 exposure-outcome combinations from the original simulated 15×5) but also on a subset of the sparse signals T_X that originally was simulated within the exposures (i.e., including 10×5 exposure-exposure combinations from the original simulated 15×15). Overall, in all mis-specified scenarios, the number of exposure-outcome combinations where the sparse causal effects might be present increases from $15 \times 5 = 75$ to $10 \times 10 = 100$.

The results of this experiment regarding the detection of the simulated sparse signals and the quality of the causal effects estimation for all simulated scenarios, and a particular choice of the parameters $r_X = 0.6$ and $m_Y = 1$, are presented in Figures S12–S15. As expected, the detection of the simulated causal effects is more difficult, although MR-BMA depends less on the mis-specification given that it does not model dependency relations within either the exposures or the outcomes. MrDAG is less influenced than the other causal graphical algorithms and particularly MRPC, and it still has a clear edge on MR-BMA when dependency relations are simulated within the outcomes (see Figures S13A–S13D). Regarding the quality of the causal effects estimation, MrDAG is confirmed to be the overall best method (see Figures S14 and S15). MR-BMA is the second-best method when no dependency structure is simulated within the outcomes.

Real data application: The impact of lifestyle and behavioral traits on mental health

We apply MrDAG to investigate its ability to detect the effects of lifestyle and behavioral exposures on the risk of mental health phenotypes as well as potential forms of interventions for their prevention. As exposures, we chose seven lifestyle and behavioral traits that have previously been investigated for their effects on mental health, including EDU, physical activity (PA), sleep duration (SP), alcohol consumption (ALC), SM, and LST. As outcomes, we selected seven mental health phenotypes, including MDD, AN, attention-deficit hyperactivity disorder (ADHD), BD, autism spectrum disorder (ASD), schizophrenia (SCZ), and COG. See Table S1 for the description of the summary-level statistics, the data sources, and the number of IVs for each trait, and the Appendix A for the pre-processing steps. In a separate analysis, we also investigate the reverse direction, i.e., whether the same mental health phenotypes have an impact on the group of lifestyle and behavioral traits selecting IVs for the mental health phenotypes (see Appendix A for the respective pre-processing steps).

Figure 7 presents the results of MrDAG. In particular, Figures 7A and 7C show the estimated PPEI (Equation A14) after structure learning, and Figures 7B and 7D show the (Bayesian-model-averaged) causal effects (95% credible intervals [CI]) between the exposures and the outcomes. We ran the MrDAG algorithm for 10^6 Markov chain

Monte Carlo (MCMC) sweeps, 10^5 of which as burn-in (for details see supplemental text). The computational time is 1 h 40 min on Cambridge HPC. Figure S16A shows no sign of aberrant behavior of the MCMC or the Markov chain being trapped in local maxima. The number of sweeps seems sufficient for the convergence of the MCMC algorithm to draw samples from the posterior distribution. Results on PPEI and the causal effects are not thresholded, and sparsity is enforced by assigning *a priori* on the number of expected edges. We set it at $\pi^{\text{edge}} = 0.16$, i.e., we expect *a priori* one edge for each of the 13 traits (see Appendix A and supplemental text). Post-processing of the MrDAG output and corresponding outliers detection (CPO estimation) show no invalid IVs due to unmeasured pleiotropy that acts on a single outcome at a time (see Appendix A and Figure S17). The time required to estimate CPO is 3 h 30 min, higher than the MrDAG algorithm, since the marginal likelihood has to be calculated for each observation across all the graphical models visited during the MCMC. However, it is less than the computation time of the MrDAG algorithm $\times$ the number of observations, since no structure learning needs to be performed and, thus, there is no need to evaluate the time-consuming Metropolis-Hastings ratio in Equation A11.

As shown in Figures 7C and 7D, there are two key shared exposures with important downstream effects on mental health phenotypes, which are EDU and SM, on which we focus our discussion. For each of them, we describe how MrDAG can disentangle complex dependency relations within the exposures and the outcomes and detect (partial or complete) mediation, which prevents spurious findings.

As could be expected due to its centrality in the global health agenda⁵⁴ and the high level of confounding of this phenotype with other genetically associated biological, behavioral, and socioeconomic traits, genetically predicted EDU shows the most inter-exposure and exposure-outcome dependency relations (Figure 7C, bottom part). Previous work has supported the broad mental health implications of education.⁵⁵ First, in keeping with previous findings,^{26,56–58} our results show that genetically predicted EDU increases COG, and liability to ASD and BD, as well as decreasing liability to ADHD. In contrast, genetically predicted EDU has no effects on SP, the amount of ALC, or the liability to MDD,²⁶ AN,⁵⁹ or SCZ⁵⁸ (Figure 7D). Second, we investigate the detected dependency relations of EDU with other exposures that contribute to the reported associations. We find bidirectional relationships between genetically predicted EDU, PA, and LST consistent with a large body of literature.^{26,60} Dependency relations have been also identified between EDU and SM.^{26,27} Supported by the existing literature, these results confirm the ability of MrDAG to disentangle complex relationships that exist between inter-related exposures.

We find that SM is second only to EDU in its association with several outcomes. Specifically, genetically predicted levels of SM are associated with an increased liability to MDD and ADHD, as previously reported.^{61,62} It is also

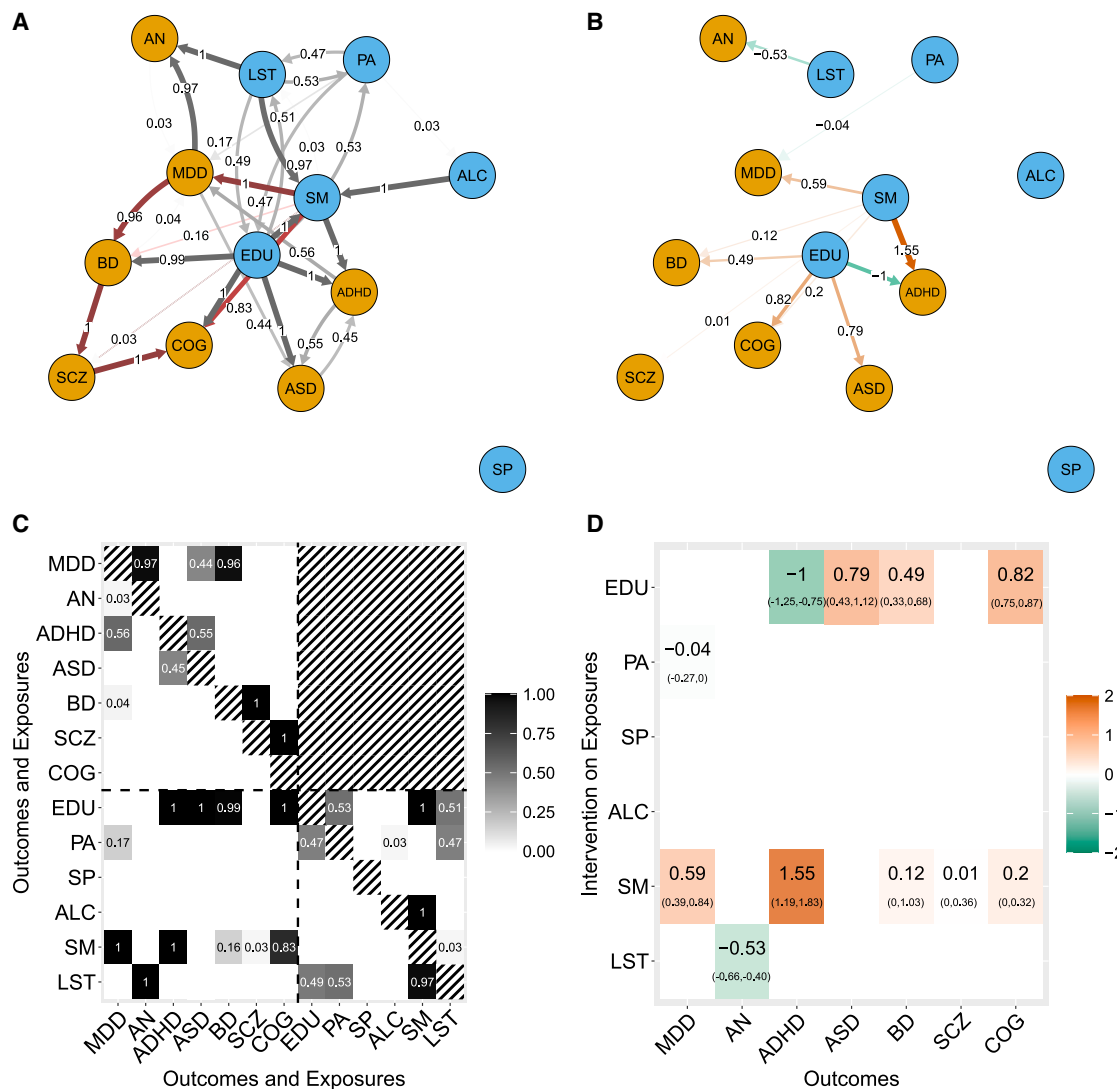


Figure 7. Results of MrDAG algorithm regarding how lifestyle and behavioral exposures impact mental health outcomes

(A) PDAG of the posterior probability of edge inclusion (PPEI) within the exposures (lifestyle and behavioral traits, blue nodes), the outcomes (mental health phenotypes, orange nodes), and between them. Undirected edges are represented as bidirectional edges; see, for instance, edges between PA (physical activity) and LST (leisure screen time) or ASD (autism spectrum disorder) and ADHD (attention-deficit hyperactivity disorder). Red edges indicate the estimated direct and indirect path from SM (smoking) to COG (cognition), including the path from SM to SCZ (schizophrenia). Neither reverse causation from the outcomes to the exposures nor phenotypic traits feedback loops are allowed.

(B) (Bayesian model-averaged) Causal effects on the outcomes (orange nodes) under intervention on the exposures (blue nodes). Red and green edges indicate positive and negative (Bayesian model-averaged) causal effects, respectively.

(C) Posterior probability of edge inclusion (PPEI) for each combination of outcomes (mental health phenotypes) and exposures (lifestyle and behavioral traits). Horizontal and vertical dotted lines separate the exposures (bottom-right submatrix) from the outcomes (top-left submatrix). PPEIs between exposures and outcomes are depicted in the bottom-left submatrix) nor phenotypic traits feedback loops (main diagonal) are allowed (black-white strips).

(D) (Bayesian model-averaged) Causal effects (95% credible intervals) on the outcomes (y axis) under intervention on the exposures (x axis).

associated with BD and SCZ (although these effects are small) and COG. Notably, the association between genetically predicted levels of SM and COG is also detected by standard MR but with a negative effect (see Figure 1A), while MR-BMA¹ does not declare it significant after FDR control across all exposures and outcomes, although it estimates a positive effect (see Table S3). As discussed above, we also check the detected dependency relations of SM with other exposures.

MrDAG appropriately identifies the relationship between ALC and SM, but not vice versa. In a recent MR publication,⁶³ the opposite association is observed. However, in contrast to Reed et al.,⁶³ who conceptualize SM with smoking initiation, we use a lifetime smoking index,⁶¹ which captures smoking duration, heaviness, and cessation.

As important as the discussion of existing associations between the exposures and the outcomes is, it is similarly

insightful to discuss the absence of causal effects, especially those relationships that are reported in the literature or found by standard (one exposure and one outcome) MR models. For example, we do not replicate all previous evidence for genetically predicted levels of SM being associated with mental health phenotypes. Although we find a strong effect of genetically predicted levels of SM on MDD,⁶¹ we do not find the same strong effect of SM on SCZ⁶¹ as observed in observational studies.^{64,65} By looking at Figure 7C, this might be due to pleiotropic effects that have been identified by MrDAG within the mental health phenotypes. In line with prior findings, evidence from MrDAG supports dependency relations between genetic liability to MDD and AN, ASD, and BD²⁸ as well as between genetic liability to BD and SCZ.⁶⁶ Lastly, in keeping with prior findings of possible bidirectional ASD-ADHD relationships,⁶⁷ we observed genetic dependency relations between ASD and ADHD, and vice versa. These results suggest that the genetic effects of SM on SCZ can be mediated by pleiotropic effects within the responses. By considering the results above, we hypothesize that the SM-SCZ relationship is partly mediated first by MDD and then by BD. Moreover, there is another path that goes from the genetically predicted level of SM to SCZ through a positive weak association identified by MrDAG between SM and BD.⁶⁸ Both genetic paths are illustrated in Figure 7A and highlighted in red. Conditionally on these relationships that are not considered in standard MR or MVMR, MrDAG does not detect a strong causal effect between SM and SCZ.

We further note that the effect of SM on ADHD is both direct and indirect, the latter mediated first by MDD and then by ASD. Thus, our analysis pinpoints the important role of MDD, which partly or entirely accounts for many paths within mental health phenotypes and their causal exposures. This might be due to the potentially high levels of confounding and non-specific genetic associations present in the original MDD GWAS^{69,70} as well as the high levels of symptom-level and, therefore, diagnostic overlap between MDD and all other psychiatric disorders.⁷¹ Nonetheless, the implications of our results, assuming the validity of all GWAS findings, are that prevention and/or therapeutic intervention on MDD⁷² can have a cascade of important effects for the prevention of several mental health phenotypes.

To investigate this hypothesis, Figures S18A and S18B show the results of MrDAG when MDD is removed from the list of outcomes. Regarding the association between genetically predicted levels of SM and ADHD, it is still present with the same strength and similar CI depicted in Figure 7D, suggesting that the indirect effect mediated first by MDD and then by ASD is negligible. Figure S18B also shows that, after removing MDD, genetically predicted levels of SM are positively associated with SCZ, as reported in the literature, as well as negatively associated with LST, as shown in Figure 1. Combined with our main findings, this result indicates that the absence of a link between

SM and SCZ (and the link between LST and SCZ) in the MrDAG model is likely due to the mediation of MDD. Similar results are obtained if BD, which appears in the path highlighted in red in Figure 7A, is removed from the list of outcomes (see Figure S19).

To check whether the hypothesized edges' orientation between the exposures and the responses might not be supported by the data, we have also tested reverse causation, whereby we assess the impact of mental health phenotypes on lifestyle and behavioral traits by selecting genetic variants to be associated with the mental health phenotypes. As before, we used 10^6 MCMC sweeps of which 10^5 were burn-in, and the computational time is 2 h on the Cambridge HPC. Post-processing of MrDAG output to check for the presence of invalid IVs due to unmeasured pleiotropy that acts on a single outcome at a time identifies 84 IVs (18%) as invalid (see Figure S26). CPO estimation took 3 h to run. We removed these IVs and rerun MrDAG on the remaining 386 IVs. Figures 8 and S27 show the results of this analysis where, besides the positive effect of genetically predicted COG on EDU,⁷³ genetic liability to MDD⁶¹ and ADHD is associated with SM, the latter well documented in epidemiological studies⁷⁴ and confirmed in a randomized clinical trial of smoking cessation,⁷⁵ although the respective effect size is small.

We conclude the analysis of the real data application by assessing the validity of the results obtained by MrDAG and adding the comparison with other methods. We divide this internal check into sensitivity to hyper-prior specification and robustness of structure learning. Regarding the first point, Figure S20 shows that the (Bayesian model-averaged) causal effects as well as the 95% CIs for different values of the *a priori* probability of edge inclusion are not influenced by this choice. For the second internal check, we bootstrap MrDAG repeatedly on the data⁷⁶ (see supplemental text). In Figure S21 we present the bootstrap frequency of edge inclusion for each permitted combination of exposures and outcomes and the scatterplot of the PPEI against the bootstrap frequency of edge inclusion. The results show that there is a satisfactory agreement between a single run of the algorithm and the bootstrap results for the reported causal associations. Extended results are presented in supplemental text.

Figures S22–S25 show results of MR², MRPC, ParDAG, and Graph-MRcML applied to the same real data. We discuss in detail the results of MR² and Graph-MRcML.

As in the simulation study, MR² shows some difficulties in teasing out complex relationships within the responses, which are mostly directed, e.g., MDD → BD, or mediated, e.g., ADH → MDD → AN, and detected by MR² as MDD-BD and ADHD-AN, once the causal effects have been detected, and vice versa.

Graph-MRcML is the only method considered that treats all traits as equal and does not distinguish between exposures and outcomes. To perform this task, it requires a different format for the data input (see supplemental text). The results show generally a good agreement with

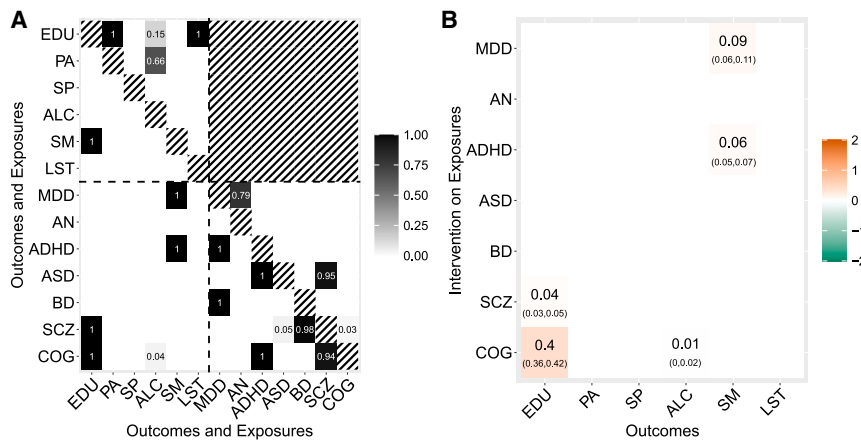


Figure 8. Results of MrDAG regarding how liability to mental health phenotypes affects lifestyle and behavioral traits after removing invalid IVs

(A) Posterior probability of edge inclusion (PPEI) for each combination of outcomes (lifestyle and behavioral traits) and exposures (mental health phenotypes). Horizontal and vertical dotted lines separate the exposures (bottom-right submatrix) from the outcomes (top-left submatrix). PPEIs between exposures and outcomes are depicted in the bottom-left submatrix. Neither reverse causation (top-right submatrix) nor phenotypic traits feedback loops (main diagonal) are allowed (black-white strips). (B) (Bayesian model-averaged) Causal effects (95% credible intervals) on the outcomes (y axis) under intervention on the exposures (x axis).

MrDAG regarding EDU and SM as effective points of intervention, and the absence of any direct effect of SM on SCZ. They also largely agree on the lack of reverse causation, with the effect of COG on EDU detected by both. However, two important differences are apparent. They are related to MrDAG shrinkage of unimportant causal effects and sparse structure learning. In Graph-MRcML, the total effects are decomposed into the direct effects of all possible trait pairs (see Figure S25B), including the non-significant ones (see Figure S25C), leading to the “dilution” of the causal effect size of significant trait pairs. An example of this phenomenon is the effect of SM on SCZ. Albeit not selected at 5% Bonferroni-adjusted significance level ($p_{\text{adj}} = 1$), the direct causal effect of SM on SCZ is one of the largest estimated by Graph-MRcML, weakening the estimated causal effects of the other trait pairs; for example, the size of the causal effect of SM on ADHD obtained by Graph-MRcML (1.03) is 50% lower compared to MrDAG (1.55). In contrast, MrDAG estimates the direct causal effects given the sparse visited EGs, on average with ≈ 21 edges out of 114 possible ones (Figure S16), thus shrinking to zero the causal effects of unimportant dependency relations. The second difference is linked to the selection of important relationships. In MrDAG this is accomplished by a score-based (marginal likelihood) structure learning of sparse EGs visited during the MCMC, which separates important dependency relations from less important ones. In Figure S20 we show that the results do not depend on the *a priori* probability of edge inclusion. Contrarily, in Graph-MRcML small differences in the adjusted significance level lead to different models. An example is the effect of SM on ADHD (Figure S25A), which is not significant at 5% Bonferroni-adjusted significance level ($p_{\text{adj}} = 0.051$), although this relation has been detected by MrDAG and the majority of alternative approaches. At higher Bonferroni-adjusted significance level it becomes significant. Another example is the effect of EDU on ASD, which has an adjusted p value that is only slightly higher than 20% ($p_{\text{adj}} = 0.208$). It is not clear whether or not it should be included in the causal graphical model.

Discussion

In this study, we have introduced MrDAG, the first Bayesian causal graphical MR model for joint analysis of multiple exposures and outcomes. The proposed method can detect dependency patterns within the exposures as well as within the outcomes, thus allowing for a more precise estimation of the causal effects from the exposures to the outcomes.

In a comprehensive simulation study, MrDAG outperforms recently proposed one-outcome-at-a-time and multi-response multi-variable MR methods and causal graphical models under the constraint on edges’ orientation from the exposures to the outcomes. We showcased the advantage of MrDAG also in a real data application to disentangle how lifestyle and behavioral traits interact to cause mental health phenotypes and, separately, the opposite. We highlighted how MrDAG can recover information on the genetic paths that link exposures to outcomes compared to existing MR methods that ignore these dependency relations. Specifically, we highlighted primarily, education and secondarily, smoking as solely effective points of intervention given their distinct downstream effects on multiple mental health phenotypes. In contrast to widely used uni- and multi-variable MR methods, and a recently proposed multi-response model, neither leisure screen time (LST) nor sleep duration (SP) have been identified as key exposures of intervention. Compared to other causal graphical models considered in this study under the assumption of known directionality between the exposures and the outcomes, more significant causal effects are detected by MrDAG, especially those linked with education. Finally, by enforcing sparsity, MrDAG better separates important dependency relations from less important ones and better estimates direct causal effects than an alternative bidirectional causal graphical model where no regularization of the estimated causal effects is implemented.

These insights are possible because three methodological advances are considered in MrDAG. First, in structure learning, the hypothesis of no unobserved confounding

is a fundamental underlying assumption. This assumption, known as causal sufficiency, is difficult to justify in real data applications, and its violation produces biased results. By using IVs within the MR paradigm, we bypass the need to remove the effects of the unobserved confounder from the individual-level data.^{39,77,78} Specifically, we avoid the assumption of causal sufficiency by employing genetically predicted exposures and outcomes that depend only on the genetic variants chosen as IVs. Genetically predicted exposures are key in the derivation of the two-stage least-square causal effect estimator,⁴¹ but in MrDAG we have extended it to include genetically predicted outcomes. On both predicted groups of traits, we perform EG exploration to learn the unconfounded dependency relations that exist within and between the exposures and the outcomes. Our second contribution is the estimation of causal effects under intervention on the exposures conditionally on a given DAG. We showed that they can be estimated based on Pearl's interventional calculus.¹⁷ Moreover, differently from Kalisch et al.⁵⁰ and its application in the MRPC algorithm,^{5,51} in the proposed Bayesian implementation the estimation of the causal effects is averaged over the visited graphical models,⁴³ thus taking into account the uncertainty regarding the EGs that best describe the dependency structure in a given dataset. Third, MrDAG allows the possibility of including domain-knowledge relations between the traits. In the designed MrDAG model, constraints between the exposures and the outcomes descend directly from the MR paradigm. Our Bayesian implementation of structure learning under restrictions offers clear advantages over alternative methods.⁴⁴ In particular, adding an acceptance/rejection step to check whether the proposed EG satisfies the edge-orientation constraints is simple and effective. Although not presented in this study, other restrictions can be straightforwardly included—for instance, known relations regarding disease progression or time-dependent outcomes, e.g., smoking initiation and cessation.⁷⁹

MrDAG is an MR approach that is best suited in the context of biologically informed relationships, since exposures and outcomes need to be specified before the analysis. This insight should be used to inform the design of the study as well as the data sources selected to test specific research questions and hypotheses. The lack of sufficient appreciation of underlying biology is one of the reasons for the current crisis faced in the application of MR more widely.⁸⁰ In the case of no prior domain knowledge, agnostic causal network models should be preferred.⁶

There are some limitations in the proposed method. A drop in power to detect causal effects is apparent in the simulation study when the summary-level data are derived from noisy genetic association estimates. This is expected, since the detection of dependency relations requires informative data. As we have shown in the real data application, modern GWASs are performed on large cohorts, so noisy genetic association estimates are less of a concern, although this might still be an issue (shared

with most causal graphical models) when molecular traits are considered, since their sample size could be small.⁸¹ Mis-specification of the exposure-outcome groups' definition has the same effect on the power, since the hypothesized edges' orientation between the exposures and the responses might not be supported by the data, e.g., if reverse causation is present. This can be investigated by flipping the comparison and treating the outcomes as exposures and vice versa. This entails selecting a new set of genetic variants as IVs to instrument the outcome traits and rerunning MrDAG to check for reverse causation, as illustrated in the real data application where we investigated the reverse direction, i.e., the impact of liability to mental health phenotypes on lifestyle and behavioral traits. We also suggest monitoring the rate of the rejection step. High levels might be an indicator of reverse causation and, thus, mis-specification of the exposure/outcome groups' definition.

In the real data application, while the use of existing summary-level statistics of GWASs facilitates the integration of diverse phenotypes measured in different cohorts, we are also limited by the biases suffered by the initial GWASs. Specifically, studies of mental health rely on the presence of a clinical diagnosis. Consequently, it is not truly the genetic liability of the disease itself as much as the probability of having access to diagnoses or treatment. Our findings on the relationship between higher genetically predicted EDU and increased risk of ASD and BD, but decreased ADHD risk, provide an example of such bias. In these analyses, the predicted number of school years completed is unlikely to be causally implicated in the development of ASD traits. While the typical age of onset of ASD precedes the start of formal education (and is therefore unlikely to be caused by it), ASD-related traits are more likely to be recognized and referred, particularly in those who are undiagnosed or untreated, when individuals are within a schooling system where standardized testing and progress reports by peer comparison are performed. Moreover, current GWASs consider one trait or disease at a time and do not consider to what extent cases are comorbid with other diseases. Future GWASs on comorbidity⁸² might provide more fine-grained genetic associations, allowing disentanglement of some of these relationships.

In conclusion, MrDAG, with its unidirectional approach and Bayesian implementation, represents an alternative contribution to how we can learn complex relationships among phenotypic traits. It provides analysts with the opportunity to derive a more comprehensive picture of causal mechanisms between complex phenotypes. The real data application is an example of the proposed holistic approach, whereby we leverage MrDAG and large-scale genome-wide association data to offer novel mechanistic insight into the causal behavioral determinants of mental health phenotypes to delineate between their overlapping pathophysiology and phenotypic presentation toward translational progress in the field of mental health.

Data and code availability

Data sources are presented in [supplemental text](#) with associated URL links. Social Science Genetic Association Consortium (SSGAC) summary-level statistics are available through a standard registration procedure (<https://thessgac.com/register/>).

The MrDAG learning R package is freely available on <https://github.com/lb664/MrDAG/>. It includes the data of the real data applications and how to run the algorithm. Post-processing routines to estimate the (Bayesian model-averaged) causal effects presented in this article are also included along with PPEI.

Acknowledgments

The authors are thankful to Federico Castelletti and Guido Consonni for their insightful comments and suggestions and to the participants of the 1st Danish International Conference on Personalized Medicine, Aarhus, Denmark, whose remarks regarding preliminary results of the real data application led to its substantial improvement. The authors are also grateful to the editor and two anonymous referees for their valuable comments that greatly improved the presentation of the article and for pointing out useful references on causal graphical models.

The authors gratefully acknowledge United Kingdom Research and Innovation Medical Research Council grants MR/W029790/1 (V.Z. and L.B.), the Marmaduke Sheild Fund (L.B.), and the British Heart Foundation Centre of Research Excellence at Imperial College London grant RE/18/4/34215 (D.G.). This research was supported by the UK Dementia Research Institute (V.Z.), which receives its funding from UK DRI, funded by the UK MRC, Alzheimer's Society, and Alzheimer's Research UK, and by the NIHR Cambridge Biomedical Research Centre (NIHR203312) (L.B.). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Author contributions

Conceptualization, L.B., V.Z., and D.G.; methodology: L.B. and V.Z.; formal analysis, V.Z. and L.B.; resources, all collaborators; data curation, V.Z. and D.G.; writing – original draft, V.Z., L.B., T.C., D.G., and N.C.; writing – review & editing, L.B., V.Z., and D.G.; visualization, V.Z. and L.B.; funding acquisition, V.Z.; supervision, L.B. and D.G.

Declaration of interests

D.G. is the Chief Executive Officer of Sequoia Genetics. D.G. has financial interests in several biotechnology companies. T.C. is employed by Sequoia Genetics.

Appendix A

In the following, we denote with capital letters the random variables Y, X, G , and U for the observed outcome, exposure, instrumental variable, and unobserved confounder, respectively, and with small letters y, x, g , and u their corresponding observations. Multivariate random variables and corresponding observations are presented in bold. A marginal element of a vector of random variables is specified by a suitable subscript index, e.g., $Y_k, k \in K = \{1, \dots, q\}, X_j, j \in J = \{1, \dots, p\}$, and $G_i, i \in I = \{1, \dots, n\}$. $\mathbf{Y}_{\setminus k}$

and $\mathbf{X}_{\setminus j}$ consist of all the outcomes and exposures except those that are related to the k th response and j th exposure, respectively. Finally, vectors understood as column vectors and matrices are indicated in bold, the latter also in capital letters.

We indicate with β_{ij}^X and β_{ik}^Y the effect of the genetic variant $i \in I$ on the exposure $j \in J$ and outcome $k \in K$, respectively, with β_{X_j} and β_{Y_j} the n -dimensional vector of genetic effects on the j th exposure and k th outcome, respectively, and, finally, with $\mathbf{B}_X$ and $\mathbf{B}_Y$ the $(n \times p)$ - and $(n \times q)$ -dimensional matrices of the genetic effects on all exposures and outcomes. θ_{jk} denotes the causal parameter of interest, i.e., the direct causal effect of X_j on Y_k , and γ_{hj}^X and γ_{hk}^Y the mediation effect of X_h on $X_j, h \neq j$ and Y_j on $Y_k, h \neq k$, respectively. Θ, Γ_X , and Γ_Y indicate the corresponding $(p \times q)$ -, $(p \times p)$ -, and $(q \times q)$ -dimensional matrices of the causal parameters of interest (Θ) and the mediation parameters Γ_X and Γ_Y whose average value is r_X and m_Y , respectively. The symbol “ $\wedge$ ” denotes the estimator of a parameter or its estimated value and “ $*$ ” an IVW parameter.

Let $\mathcal{D} = (V, E)$ be a directed acyclic graph (DAG), where V denotes a set of vertices (nodes) and $E = V \times V$ a set of directed edges, i.e., if $(z, v) \in E$, then $(z, v) \notin E$. For a given DAG $\mathcal{D}$, if $z \rightarrow v$, then z is a parent of v and, conversely, v is a child of z . Moreover, if $z \rightarrow \dots \rightarrow v$, then z is an ancestor of v and v is a descendant of z . We denote the parent set of v in $\mathcal{D}$ as $\text{pa}_{\mathcal{D}}(v)$ and $v \text{Upa}_{\mathcal{D}}(v) = \text{fa}_{\mathcal{D}}(v)$ the family of v . Unless otherwise stated, for ease of notation, we remove the subscript $\mathcal{D}$. We indicate with $\mathbf{Y}_{\text{pa}(k)}$ and $\mathbf{X}_{\text{pa}(k)}$ the outcome and exposure parents of the outcome $k \in K$, with $\mathbf{X}_{\text{pa}(j)}$ the exposure parents of the exposure $j \in J$ and with $\mathbf{X}_{\text{fa}(j)}$ its family, i.e., $\mathbf{X}_{\text{fa}(j)} = X_j \cup \mathbf{X}_{\text{pa}(j)}$.

In research by Didelez and co-workers,^{15,37,38} key results regarding standard Mendelian randomization (MR) (single exposure with single instrumental variable and single outcome) are presented. Here, we use them to show that MrDAG is an extension of standard MR when multiple exposures and outcomes are considered. Technical details are provided in [supplemental text](#).

These results are conditioned on a given DAG. Since the underlying dependency relations within and between exposures and outcomes are not known (latent) and need to be estimated, structure learning is performed on the genetically predicted values of the traits. We draw the connection between them and the genetic components of the phenotypic covariance, which are orthogonal to any confounders U . Thus, by using the genetically predicted values, the causal sufficiency condition (usually assumed explicitly or implicitly in structure learning) is not required. We conclude this section with a description of the Bayesian implementation of the proposed model and details regarding the data-generation steps used in the simulation study.

Multi-exposure and multi-outcome core conditions for instrumental variables

Let $\mathbf{Y}, \mathbf{X}$, and $\mathbf{G}$ be the q -, p - and n -dimensional vector of the outcomes, exposures and instruments (genetic variants) random variables, respectively.

Let us assume the following “multi-variate core conditions” (MCC) for valid instrumental variables (IVs), which are the extensions of the core conditions that $\mathbf{G}$ has to satisfy in standard MR¹⁵:

- (IV1) $G_i \perp\!\!\!\perp U, \forall i \in I$, i.e., G_i must be independent of U ;
- (IV2) $G_i \not\perp\!\!\!\perp X_j \mid \mathbf{X}_{\setminus j}, \forall i \in I$ and $\forall j \in J$, i.e., G_i must not be independent of X_j conditionally on $\mathbf{X}_{\setminus j}$; and
- (IV3) $G_i \perp\!\!\!\perp Y_k \mid (\mathbf{X}, U), \forall i \in I$, and $\forall k \in K$, i.e., G_i must be independent of Y_k conditionally on $\mathbf{X}$ and U .

The first multi-exposure and multi-outcome core condition (MCC) for instrumental variables is similar to the first CC in standard MR.¹⁵ The second MCC imposes that G_i should be associated with X_j conditionally on the other exposures. The third MCC establishes that the instrumental variables and outcomes are conditionally independent given the exposures and the unobserved confounder.

From the DAG $\mathcal{D}$ involving $\mathbf{Y}, \mathbf{X}, \mathbf{G}$, and U that satisfies the MCC, the corresponding Markov properties say that $G_i \perp\!\!\!\perp U, \forall i \in I$, since G_i is not a descendant of U and vice versa and $G_i \not\perp\!\!\!\perp X_j \mid \mathbf{X}_{\text{pa}(j)}, \forall i \in I$ and $\forall j \in J$, because X_j is a descendant of G_i . The Markov property for the third MCC is $G_i \perp\!\!\!\perp Y_k \mid (\mathbf{Y}_{\text{pa}(k)}, \mathbf{X}_{\text{pa}(k)}, U), \forall i \in I$ and $\forall k \in K$, since G_i is a non-descendant of Y_k and $(\mathbf{Y}_{\text{pa}(k)}, \mathbf{X}_{\text{pa}(k)}, U)$ are the parents of Y_k .

Interventional distributions and causal effects estimation

The conditional dependencies associated with the multi-exposure and multi-outcome DAG $\mathcal{D}$ lead to the following factorization of the joint density of all random variables considered:

$$f(\mathbf{y}, \mathbf{x}, \mathbf{g}, u) = \prod_{k \in K} f(y_k \mid \mathbf{y}_{\text{pa}(k)}, \mathbf{x}_{\text{pa}(k)}, u) \prod_{j \in J} f(x_j \mid \mathbf{x}_{\text{pa}(j)}, \mathbf{g}, u) f(\mathbf{g}) f(u),$$

which is known as pre-intervention distribution and assumed to be faithful to the DAG,⁵³ i.e., there are no conditional dependence relationships between the variables in the model that do not follow directly from the Markov properties.

The post-intervention distribution under intervention on the h th exposure sets to take the value $\tilde{x}_h$ is obtained by the truncated factorization¹⁷

$$f(\mathbf{y}, \mathbf{x}_{\setminus h}, \mathbf{g}, u \mid \text{do}(X_h = \tilde{x}_h)) = \prod_{k \in K} f(y_k \mid \tilde{x}_h, \mathbf{y}_{\text{pa}(k)}, \mathbf{x}_{\text{pa}(k)}, u) \prod_{j \in J \setminus \{h\}} f(x_j \mid \mathbf{x}_{\text{pa}(j)}, \mathbf{g}, u) \mathbb{I}_{x_h}(\tilde{x}_h) f(\mathbf{g}) f(u), \quad (\text{Equation A2})$$

where $\mathbb{I}_x(\tilde{x})$ is the indicator function which is equal to one if $x_h = \tilde{x}_h$ and zero otherwise. Graphically, the directed edges to X_h from its parents in $\mathbf{X}, \mathbf{G}$ and U are removed.

A post-intervention distribution under intervention on the h th exposure is obtained from Equation A2 by marginalizing all variables but the selected outcome and the exposure on which an intervention is carried out:

$$f(y_k \mid \text{do}(X_h = \tilde{x}_h)) = \int f(\mathbf{y}, \mathbf{x}_{\setminus h}, \mathbf{g}, u \mid \text{do}(X_h = \tilde{x}_h)) d\mathbf{y}_{\setminus k} d\mathbf{x}_{\setminus h} d\mathbf{g} du = \int f(y_k \mid \tilde{x}_h, \mathbf{x}_{\text{pa}(h)}, u) f(\mathbf{x}_{\text{pa}(h)}, u) \mathbb{I}_{x_h}(\tilde{x}_h) d\mathbf{x}_{\text{pa}(h)} du \quad (\text{Equation A3})$$

This result is derived from Pearl¹⁷ and follows directly from the Markov properties of the DAG. It establishes that the parents of the variable on which an intervention is carried out are the only variables that need to be measured to estimate the causal effect on an outcome.⁸³

The post-intervention distribution (Equation A3) can be summarized by taking the expectation and defining the causal effect of an intervention⁸⁴ as

$$\theta_{h,k} = \frac{\partial}{\partial x_h} \mathbb{E}(Y_k \mid \text{do}(X_h = x_h)) \Big|_{x_h = \tilde{x}_h}, h \in J, k \in K.$$

In [supplemental text](#), we show the identifiability of the causal effect ([proposition S2](#)) and the derivation of its estimand in multiple

exposures and multiple outcomes MR framework ([proposition S3](#)). We also show the consistency of the effects of the regressions of each outcome and exposure on $\mathbf{G}$ ([proposition S1](#)), i.e., the estimated genetic effects on the outcomes and exposures contain all information regarding the causal parameters of interest and the mediation parameters within the exposures and the outcomes unconfounded by U .

Here, for a given DAG $\mathcal{D}$, we report the IVW estimator of the causal effect of the intervention in X_h on Y_k :

$$\hat{\theta}_{h,k} = \left[\left(\hat{\mathbf{B}}_{X_{\text{fa}(h)}}^{*\top} \hat{\mathbf{B}}_{X_{\text{fa}(h)}}^* \right)^{-1} \hat{\mathbf{B}}_{X_{\text{fa}(h)}}^{*\top} \hat{\beta}_{Y_k}^* \right]_1, \quad (\text{Equation A4})$$

where the subscript indicates the first element of the solution of the linear least squares (LLS) regression, since $\text{fa}(v) = v \cup \text{pa}(v)$, $\mathbf{X}_{\text{fa}(h)}$ denotes the exposures that are the family of the exposure X_h under intervention, $\hat{\mathbf{B}}_{X_{\text{fa}(h)}}^*$ are the IVW estimated coefficients of the regressions of each exposure in $\mathbf{X}_{\text{fa}(h)}$ on $\mathbf{G}$, and $\hat{\beta}_{Y_k}^*$ is the IVW estimated coefficient of a regression of Y_k on $\mathbf{G}$. Equation A4 resembles the standard IVW estimator of the causal effect that approximates the estimate that would have been obtained if individual-level data were available.³ However, in contrast to general proposed solutions in MVMR, in Equation A4 the set of regressors is with regard to the family of the exposure under intervention.

Dependency structure under the effect of unobserved confounders

To estimate Equation A4, structure learning of the graphical models needs to be performed to detect the parents $\mathbf{X}_{\text{pa}(h)}$ of the exposure X_h under intervention. However, structure learning assumes causal sufficiency,³⁹ i.e., it requires that there are no hidden (or latent) variables that are common causes of two or more traits. Instead, in this study we explicitly assume that an unobserved confounder U acts on both outcomes and exposures.

Links between the genetic correlation and MR causal effect estimate have been already discussed in Bulik-Sullivan et al.⁸⁵ Here we provide further connections with genetic covariance,⁸⁶ which is key to showing that, by working with summary-level statistics, it is possible to recover the dependency structure between the corresponding traits in the original (individual-level) data unconfounded by U .

Let us assume that the genetic effect on a phenotypic trait is linear and consider two traits,

$$\begin{aligned} Y_k &= \mathbf{G}^\top \beta_{Y_k} + \psi_Y U + \epsilon_{Y_k}, k \in K, \\ X_j &= \mathbf{G}^\top \beta_{X_j} + \psi_X U + \epsilon_{X_j}, j \in J, \end{aligned}$$

where $\mathbf{G}$ is a set of genetic variants of dimension n , either spanning the whole genome or region(s)-specific or selected to be associated with a trait, β_{Y_k} and β_{X_j} are the genetic effects, U is an unobserved confounder that affects both traits with ψ_Y and ψ_X the effects sizes, and ϵ_{Y_k} and ϵ_{X_j} are white noises that can be interpreted as environmental effects. We assume that $\mathbf{G} \perp\!\!\!\perp U$ and, similarly, $\mathbf{G} \perp\!\!\!\perp \epsilon_{Y_k}$ and $\mathbf{G} \perp\!\!\!\perp \epsilon_{X_j}$. Finally, we assume that $U \perp\!\!\!\perp \epsilon_{Y_k}$ and $U \perp\!\!\!\perp \epsilon_{X_j}$, i.e., the unobserved confounder U exerts its effect on both traits and it is distinct from other environmental factors. Under this model, the phenotypic covariance is

$$\begin{aligned} \text{Cov}(Y_k, X_j) &= \text{Cov}(\mathbf{G}^\top \beta_{Y_k} + \psi_Y U + \epsilon_{Y_k}, \mathbf{G}^\top \beta_{X_j} + \psi_X U + \epsilon_{X_j}) \\ &= \beta_{Y_k}^\top \mathbb{V}(\mathbf{G}) \beta_{X_j} + \psi_Y \psi_X \mathbb{V}(U) + \text{Cov}(\epsilon_{Y_k}, \epsilon_{X_j}) \end{aligned} \quad (\text{Equation A5})$$

The phenotypic covariance can be decomposed into $c_g(Y_k, X_k) = \text{Cov}(\mathbf{G}^\top \beta_{Y_k}, \mathbf{G}^\top \beta_{X_k}) = \beta_{Y_k}^\top \mathbb{V}(\mathbf{G}) \beta_{X_k}$, the genetic covariance between the two traits, i.e., the covariance between the genetic components of the two traits, $\mathbf{G}^\top \beta_{Y_k}$ and $\mathbf{G}^\top \beta_{X_k}$, and the environmental covariance, i.e., the covariance between the environmental effects of two traits that we have split into the effect of the unobserved confounder, $c_u(Y_k, X_k) = \psi_Y \psi_X \mathbb{V}(U)$, and other environmental factors, $c_e(Y_k, X_k) = \text{Cov}(\epsilon_{Y_k}, \epsilon_{X_k})$. If the environmental factors are trait specific, since U includes all common confounding factors, $c_e(Y_k, X_k) = 0$ and Equation A5 shows that an estimand of the covariance between two traits unconfounded by U is c_g .

Assuming that the individuals for the two phenotypic traits are drawn from the same population with LD matrix between the genetic variants $\mathbf{V} = \mathbf{G}^\top \mathbf{G}$, the sampling distribution of the genetic effects are $N_{Y_k}^{1/2}(\hat{\beta}_{Y_k} - \beta_{Y_k}) \xrightarrow{d} N_n(\mathbf{0}, \sigma_{Y_k}^2 \mathbf{V}^{-1})$ and $N_{X_k}^{1/2}(\hat{\beta}_{X_k} - \beta_{X_k}) \xrightarrow{d} N_n(\mathbf{0}, \sigma_{X_k}^2 \mathbf{V}^{-1})$, where “ d ” denotes convergence in distribution. Under infinite sample sizes, $\hat{\beta}_{Y_k} \xrightarrow{p} \beta_{Y_k}$ and $\hat{\beta}_{X_k} \xrightarrow{p} \beta_{X_k}$, where “ p ” denotes convergence in probability, and an estimator of the genetic covariance between the two traits is

$$\hat{c}_g(Y_k, X_k) = \hat{\beta}_{Y_k}^\top \mathbf{V} \hat{\beta}_{X_k}.$$

In the finite sample sizes case, the estimates of β_{Y_k} and β_{X_k} are noised and $\hat{c}_g(Y_k, X_k)$ is biased⁸⁶:

$$\mathbb{E}(\hat{c}_g(Y_k, X_k)) = \beta_{Y_k}^\top \mathbf{V} \beta_{X_k} + \frac{N_o \sum_{i=1}^n \mathbf{V}_{ii}}{N_{Y_k} N_{X_k}} c_u(Y_k, X_k), \quad (\text{Equation A6})$$

where N_o is the sample size overlap between the two traits and n is the number of genetic variants considered. However, even in the scenario of complete overlap, the bias in Equation A6 is negligible if the sample sizes of the two traits are large, as usually happens in modern GWAS, and n is small compared to the sample sizes.

The same considerations can be made for all phenotypic traits under investigation to reconstruct their joint genetic covariance unconfounded by U :

$$\Sigma = \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_Y^\top \mathbb{V}(\mathbf{G}) \mathbf{B}_Y & \mathbf{B}_Y^\top \mathbb{V}(\mathbf{G}) \mathbf{B}_X \\ \mathbf{B}_X^\top \mathbb{V}(\mathbf{G}) \mathbf{B}_Y & \mathbf{B}_X^\top \mathbb{V}(\mathbf{G}) \mathbf{B}_X \end{bmatrix} \quad (\text{Equation A7})$$

where Σ_{YY} , Σ_{XX} , and Σ_{XY} are the genetic covariances within the outcomes, the exposures, and between them, and $\mathbf{B}_Y$ and $\mathbf{B}_X$ are the genetic effects on the outcomes and the exposures, respectively.

MrDAG model

Assuming that the individuals for two phenotypic traits Y_k and X_j are drawn from the same population with LD matrix $\mathbf{V}$, we have $N_{Y_k}^{1/2}(\mathbf{g}^\top \hat{\beta}_{Y_k} - \mathbf{g}^\top \beta_{Y_k}) \xrightarrow{d} N(0, \sigma_{Y_k}^2)$ and $N_{X_j}^{1/2}(\mathbf{g}^\top \hat{\beta}_{X_j} - \mathbf{g}^\top \beta_{X_j}) \xrightarrow{d} N(0, \sigma_{X_j}^2)$, where $\mathbf{g}$ are the selected IVs, $\mathbf{g}^\top \hat{\beta}_{Y_k}$ and $\mathbf{g}^\top \hat{\beta}_{X_j}$ are the k th and the j th genetically predicted values of the outcome and exposure, i.e., $\hat{Y}_k$ and $\hat{X}_j$, respectively.

The joint distribution of all genetically predicted values of the outcomes and exposures based on the IVs is

$$[\mathbf{g}^\top \hat{\mathbf{B}}_Y \mathbf{g}^\top \hat{\mathbf{B}}_X]^\top \sim N_{q+p}([\mathbf{g}^\top \mathbf{B}_Y \mathbf{g}^\top \mathbf{B}_X]^\top, \Sigma),$$

i.e., for large sample sizes they are normally distributed with mean $[\mathbf{g}^\top \mathbf{B}_Y \mathbf{g}^\top \mathbf{B}_X]^\top$ and covariance matrix $\Sigma \in \mathcal{C}_D$, the space of the symmetric positive definite covariance matrices Markov with respect to the DAG $\mathcal{D}$.

If we assume that IVW is performed on the estimated genetic effects and IVs are independent after pruning or clumping, i.e., $\mathbf{V} = \mathbf{I}_n$, the MrDAG model becomes

$$[\mathbf{g}^\top \hat{\mathbf{B}}_Y \mathbf{g}^\top \hat{\mathbf{B}}_X]^\top \sim N_{q+p}([\mathbf{g}^\top \mathbf{B}_Y \mathbf{g}^\top \mathbf{B}_X]^\top, \Sigma^*), \quad (\text{Equation A8})$$

where $[\mathbf{g}^\top \hat{\mathbf{B}}_Y \mathbf{g}^\top \hat{\mathbf{B}}_X]^\top = [\mathbf{g}^\top \bar{\sigma}_Y^{-1} \hat{\mathbf{B}}_Y \mathbf{g}^\top \bar{\sigma}_X^{-1} \hat{\mathbf{B}}_X]^\top$ with $\bar{\sigma}_Y^2 \mathbf{I}_n = q^{-1} \sum_{k \in K} \mathbb{V}(\hat{\beta}_{Y_k})^2$ and similarly for $[\mathbf{g}^\top \hat{\mathbf{B}}_X \mathbf{g}^\top \hat{\mathbf{B}}_Y]^\top$. The covariance matrix can be partitioned into

$$\Sigma^* = \begin{bmatrix} \Sigma_{YY}^* & \Sigma_{YX}^* \\ \Sigma_{XY}^* & \Sigma_{XX}^* \end{bmatrix},$$

where Σ_{YY}^* , Σ_{XX}^* , and Σ_{XY}^* are the genetic covariances within the outcomes, the exposures, and between them and its inverse (Harville,⁸⁷ theorem 8.5.11) into

$$\Omega^* = \Sigma^{*-1} = \begin{bmatrix} \Omega_{YY}^* & -\Omega_{YY}^* \Sigma_{YX}^* \Sigma_{XX}^{*-1} \\ -\Sigma_{XX}^{*-1} \Sigma_{XY}^* \Omega_{YY}^* & \Sigma_{XX}^{*-1} + \Sigma_{XX}^{*-1} \Sigma_{XY}^* \Omega_{YY}^* \Sigma_{YX}^* \Sigma_{XX}^{*-1} \end{bmatrix}, \quad (\text{Equation A9})$$

with $\Omega^* \in \mathcal{P}_D$, the space of the precision matrices Markov with respect to the DAG $\mathcal{D}$ and $\Omega_{YY}^* = (\Sigma_{YY}^* - \Sigma_{YX}^* \Sigma_{XX}^{*-1} \Sigma_{XY}^*)^{-1}$. However, since by partial ordering $\Omega_{YX}^* = \Omega_{YX}^* \Sigma_{YX}^* \Sigma_{XX}^{*-1} = \mathbf{0}$, Equation A9 becomes

$$\Omega^* = \begin{bmatrix} \Omega_{YY}^* & \mathbf{0} \\ -\Sigma_{XX}^{*-1} \Sigma_{XY}^* \Omega_{YY}^* & \Sigma_{XX}^{*-1} \end{bmatrix}.$$

By using Ω^* , Gaussian graphical models³³ can be used to estimate the conditional dependence relationships between the traits in the original (individual-level) data unconfounded by U , since genetically predicted outcomes and exposures depend only on the selected IVs.

Finally, for a given DAG $\mathcal{D}$, the estimand of the causal effect under intervention⁸⁸ is

$$\theta_{h,k} = \left[\Sigma_{\text{fa}(h), \text{fa}(h)}^{*-1} \Sigma_{\text{fa}(h), k}^* \right]_1, \quad (\text{Equation A10})$$

where $\Sigma_{\text{fa}(h), \text{fa}(h)}^*$ indicates the submatrix of Σ^* whose rows and columns are $\text{fa}(h)$, $\Sigma_{\text{fa}(h), k}^*$ indicates the subvector of Σ^* whose rows are $\text{fa}(h)$ and the columns correspond to the k th outcome, and where the subscript indicates the first element of the vector. By using Equation A7 and since $\mathbb{V}(\mathbf{G}) = \mathbf{I}_n$, Equation A10 becomes

$$\theta_{h,k} = \left[\left(\mathbf{B}_{X_{\text{fa}(h)}}^\top \mathbf{B}_{X_{\text{fa}(h)}}^* \right)^{-1} \mathbf{B}_{X_{\text{fa}(h)}}^\top \beta_{Y_k}^* \right]_1,$$

where $\mathbf{B}_{X_{\text{fa}(h)}}^*$ are the IVW coefficients of the regressions of each exposure in $\mathbf{X}_{\text{fa}(h)}$ on $\mathbf{G}$, and $\beta_{Y_k}^*$ is the IVW coefficient of a regression of Y_k on $\mathbf{G}$. The corresponding estimator coincides with Equation A4. As noted above, in contrast to general proposed solutions in MVMR, by conditioning on a given DAG, the set of regressors used in the estimator of the causal effect is with regard to the family of the exposure under intervention.

MrDAG algorithm

Markov equivalent class, completed partial DAGs, essential graphs, and partial DAGs

The estimation of a DAG from observational data suffers the known problem of identifiability, i.e., it is not possible to estimate uniquely the underlying true DAG, since its conditional independencies can be encoded in several alternative DAGs. This set of DAGs that hold the same conditional independencies is known as Markov equivalent class (MEC), and the best that can be done from observational data is to estimate this class. All DAGs with the same conditional independencies can be represented by a completed partial DAG (CPDAG)⁸⁹ or essential graph (EG).³⁵ EGs

are chain graphs (CGs) whose chain components are decomposable undirected graphs.³³ A CPDAG or EG is a partially directed graph that might contain both directed and undirected edges without directed cycles. Finally, Partially DAGs (PDAGs) contain both directed and undirected edges, and directed cycles might be present.

Posterior probability of edge inclusion

In the following, we refer to $\hat{\mathbf{Y}}^* = \mathbf{g}^\top \hat{\mathbf{B}}_Y^*$ and $\hat{\mathbf{X}}^* = \mathbf{g}^\top \hat{\mathbf{B}}_X^*$, the genetically predicted values of the outcomes and exposures based on the genetic variants selected as IVs, as the “data.” Note that these quantities are not directly observable, but they are readily available as the product of the IVW summary-level statistics and the genotypes.

Technical details of the algorithm for graphical model exploration that we used to develop the MrDAG algorithm are presented in Castelletti et al.⁹⁰ In brief, it is based on a MCMC algorithm devised to explore the space of EGs whose enumeration is infeasible, since their number grows super-exponentially with the number of nodes. The EG $\mathcal{G}$ is sampled from a proposal distribution described in He et al.⁹¹ Specifically, given EG $\mathcal{G}$, six types of operators in the proposal distribution are considered: (1) inserting an undirected edge, (2) deleting an undirected edge, (3) inserting a directed edge, (4) deleting a directed edge, (5) converting two adjacent undirected edges in a v structure, and (6) converting a v structure in two adjacent undirected edges. Each operator $o_{\mathcal{G}}$ determines the transition of $\mathcal{G}$ into another EG $\mathcal{G}'$ with the probability of transition from $\mathcal{G}$ to $\mathcal{G}'$ given by $p_{\mathcal{G},\mathcal{G}'} = 1/|O_{\mathcal{G}}|$, where $O_{\mathcal{G}}$ is the set of operators that are allowed in the current EG $\mathcal{G}$. The proposed EG $\mathcal{G}'$ is accepted with a probability given by a Metropolis-Hastings (M-H) ratio defined to guarantee the convergence of the algorithm to the correct posterior distribution

$$\alpha = \min \left\{ 1, \frac{m_{\mathcal{G}'}(\text{data})p(\mathcal{G}')p_{\mathcal{G},\mathcal{G}'}}{m_{\mathcal{G}}(\text{data})p(\mathcal{G})p_{\mathcal{G},\mathcal{G}'}} \right\}, \quad (\text{Equation A11})$$

where $p(\mathcal{G})$ is the prior on $\mathcal{G}$ (Equation A12) and $p_{\mathcal{G},\mathcal{G}'}$ is the transition probability defined above. The key ingredient of the M-H ratio is a closed-form expression for the marginal likelihood $m_{\mathcal{G}}(\text{data})$. For its computation, along with some hyper-parameters based on the data input (including the number of IVs), it requires the sufficient statistic for Σ^* which is function of $\hat{\mathbf{Y}}^*$ and $\hat{\mathbf{X}}^*$. Since the IVs are chosen independently after pruning or clumping, the sufficient statistic becomes a function only of the summary-level data. We refer interested readers to Castelletti et al.⁹⁰ for the derivation of the marginal likelihood.

In the MrDAG algorithm, we added an acceptance/rejection step to guarantee that in the proposed graph the partial ordering that corresponds to the orientation of the edges from the exposures to the outcomes is satisfied (see Figure 2E). Specifically, let $\{\mathcal{D}_l(\mathcal{G}); l = 1, \dots, L_{\mathcal{G}}\}$ denote the set of distinct DAGs in the MEC represented by $\mathcal{G}$. The acceptance/rejection step checks that the partial ordering is fulfilled in the set of distinct DAGs. To check the efficiency of this step, we also monitor its acceptance rate. We also included a tempering scheme⁹² by considering an annealing parameter T in the M-H ratio to facilitate the convergence of the MCMC algorithm to the posterior distribution and the exploration of regions of high posterior mass. The temperature $1/T$ exponentiates the M-H ratio, and its value increases linearly during the burn-in until $T = 1$ at the end of the burn-in.

Sparsity is enforced by assigning a prior to $\mathcal{G}$ and specifically on $\mathcal{G}_l^U$, the skeleton of $\mathcal{G}$, which contains the same edges of $\mathcal{G}$ but without orientation:

$$\mathcal{G}_{(l)}^U | \pi^{\text{edge}, \text{i.i.d.}} \sim \text{Ber}(\pi^{\text{edge}}), l = 1, \dots, (q+p)(q+p-1)/2, \quad (\text{Equation A12})$$

where $\mathcal{G}_{(l)}^U$ is the l th element of the vectorized lower triangular part of the adjacency matrix of $\mathcal{G}^U$ and $(q+p)(q+p-1)/2$ is the maximum number of edges in an EG on $q+p$ nodes.

The posterior distribution of $\mathcal{G}$ is

$$\mathbb{P}(\mathcal{G}|\text{data}) = \frac{m_{\mathcal{G}}(\text{data})\mathbb{P}(\mathcal{G})}{\sum_{\mathcal{G} \in S} m_{\mathcal{G}}(\text{data})\mathbb{P}(\mathcal{G})},$$

with S the set of all EGs with $q+p$ nodes.

The posterior probability of edge inclusion (PPEI) is defined as

$$\mathbb{P}_{z \rightarrow v}(\text{data}) = \sum_{\mathcal{G} \in S_{z \rightarrow v}} \mathbb{P}(\mathcal{G}|\text{data}),$$

where $S_{z \rightarrow v}$ is the set of EGs with $q+p$ nodes containing the directed edge $z \rightarrow v$.

Let $\{\mathcal{G}^s, s = 1, \dots, S\}$ the set of visited EGs by MrDAG. The posterior probability of $\mathcal{G}^s$ can be approximated by

$$\mathbb{P}(\mathcal{G}^s|\text{data}) \approx \frac{m_{\mathcal{G}^s}(\text{data})\mathbb{P}(\mathcal{G}^s)}{\sum_{s=1}^S m_{\mathcal{G}^s}(\text{data})\mathbb{P}(\mathcal{G}^s)} \quad (\text{Equation A13})$$

and, similarly, the PPEI by

$$\mathbb{P}_{z \rightarrow v}(\text{data}) \approx \sum_{s=1}^S \mathbb{I}_{\mathcal{G} \in S_{z \rightarrow v}}(\mathcal{G}^s \in S_{z \rightarrow v}) \mathbb{P}(\mathcal{G}^s|\text{data}), \quad (\text{Equation A14})$$

where $\mathbb{I}_{\mathcal{G} \in S_{z \rightarrow v}}(\mathcal{G}^s \in S_{z \rightarrow v})$ is the indicator function that is equal to one if the visited EG belongs to the set of EGs with $q+p$ nodes containing the directed edge $z \rightarrow v$. Note that, although MrDAG explores the space of EGs, the graphs obtained by thresholding the PPEIs might give rise to a PDAG.⁹⁰

Bayesian causal effects estimation

Here, we summarize the results reported in Castelletti and Consonni⁸⁸ that we employed to derive the Bayesian estimation of the causal effects under unobserved confounders.

Let us rewrite Equation A8 as

$$[\mathbf{g}^\top \hat{\mathbf{B}}_Y^* \mathbf{g}^\top \hat{\mathbf{B}}_X^*]^\top | \Sigma_D^* \sim N_{q+p}([\mathbf{g}^\top \hat{\mathbf{B}}_Y^* \mathbf{g}^\top \hat{\mathbf{B}}_X^*]^\top, \Sigma_D^*),$$

where $\Sigma_D^* \in \mathcal{C}_D$, the space of s.p.d. (symmetric positive definite) covariance matrices Markov with respect to D . For ease of notation, in the following, we drop the subscript D .

Let $\Omega^* = \Sigma^{*-1} = \mathbf{L}^* \mathbf{D}^{*-1} \mathbf{L}^{*\top}$ be the modified Cholesky decomposition of the precision Ω^* . The DAG Cholesky parameterization of Ω^* is given by the node-parameters $\omega_l^* = (\mathbf{D}_{ll}^*, \mathbf{L}_{fa(l)}^*), l = 1, \dots, q+p$, with

$$\mathbf{D}_{ll}^* = \Sigma_{ll|pa(l)}^* - \Sigma_{pa(l)}^{*-1} \Sigma_{fa(l)}^*,$$

where $\Sigma_{pa(l)}^*$ indicates the submatrix of Σ^* whose rows and columns are $pa(l)$.

For a given DAG D ,⁸⁸ derive the posterior distribution of $\omega_l^*, l = 1, \dots, q+p$, in an objective Bayes framework, which has the advantage of not depending on subjective priors hyper-parameters. In turn, the posterior draws of the Cholesky parameters ω_l^* provide posterior draws from $(\Omega^*|\text{data}) = (\mathbf{L}^* \mathbf{D}^{*-1} \mathbf{L}^{*\top} | \text{data})$ and, finally, by using Equation S9 for a given DAG, posterior samples of the causal effects between the exposures and the outcomes.

In contrast to frequentist approaches^{50,84} where, for an estimated EG $\mathcal{G}$ by the PC algorithm⁵³ (or any other score-based method), the causal effects are calculated over all DAGs within

the MEC $\mathcal{G}$ or, more efficiently, only “the unique possible” causal effects within a given EG $\mathcal{G}$ are provided, in this study we also consider the uncertainty related to the estimation of the EGs. Averaging over the visited EGs, the (Bayesian-model-averaged) causal effect under intervention in the h th exposures on the k th outcomes is

$$\theta_{h,k}|\text{data} = \sum_{s=1}^S \mathbb{E}(\theta_{h,k}(\mathcal{G}^s)|\text{data}, \mathcal{G}^s) \mathbb{P}(\mathcal{G}^s|\text{data}), h \in J, k \in K, \quad (\text{Equation A15})$$

where $\mathcal{G}^s$ is an unique EGs visited during the MCMC, $\mathbb{E}(\theta_{h,k}(\mathcal{G}^s)|\text{data}, \mathcal{G}^s)$ is the posterior expectation of the causal effect given $\mathcal{G}^s$, $\mathbb{P}(\mathcal{G}^s|\text{data})$ is defined in Equation A13. Note that $\mathbb{E}(\theta_{h,k}(\mathcal{G}^s)|\text{data}, \mathcal{G}^s)$ is implicitly defined over the distinct DAGs in the EG $\mathcal{G}$. Specifically, let $\{\mathcal{D}_l(\mathcal{G}); l = 1, \dots, L_{\mathcal{G}}\}$ denote the set of distinct DAGs in the MEC represented by $\mathcal{G}$. Then,

$$\mathbb{E}(\theta_{h,k}(\mathcal{G}^s)|\text{data}, \mathcal{G}^s) = \frac{1}{L_{\mathcal{G}^s}} \sum_{l=1}^{L_{\mathcal{G}^s}} \mathbb{E}(\theta_{h,k}(\mathcal{D}_l(\mathcal{G}^s))|\text{data}, \mathcal{G}^s),$$

where $\theta_{h,k}(\mathcal{D}_l(\mathcal{G}^s))|\text{data}, \mathcal{G}^s$ are posterior draws obtained during the MCMC. Finally, by a suitable modification of Equation A15, credible intervals of the causal effects between the exposures and outcomes can be derived.

Detection of invalid IVs due to unmeasured pleiotropy

Detection of invalid IVs due to unmeasured pleiotropy in which some genetic variants are directly associated with a single outcome at a time and not via the exposures is performed by using the conditional predictive ordinate (CPO).⁹³ The CPO is also known as the leave-one-out cross-validation predictive density,⁹⁴ which is defined for the MrDAG model as

$$\begin{aligned} \text{CPO}_i &= m(\text{data}_i|\text{data}_{-i}) \\ &= \int_S m_{\mathcal{G}}(\text{data}_i) \mathbb{P}(\mathcal{G}|\text{data}_{-i}) d\mathcal{G}, \\ &= \mathbb{E}_{\mathcal{G}|\text{data}}^{-1}(m_{\mathcal{G}}^{-1}(\text{data}_i)) \end{aligned}$$

where $m_{\mathcal{G}}(\text{data}_i)$ is the marginal likelihood evaluated at the i th IV conditionally on the EG $\mathcal{G}$ and $\mathbb{P}(\mathcal{G}|\text{data}_{-i})$ is the posterior probability of the EG $\mathcal{G}$ conditionally on all remaining IVs, with S the set of all EGs with $q + p$ nodes. The CPO is based on the idea that is behind leave-one-out cross-validation: if the likelihood of observing an IV, having observed all the remaining IVs, is “small,” then the IV is considered an outlier and therefore an invalid IV.

An estimate of the CPO is based on the visited EGs during the MCMC. Specifically,

$$\widehat{\text{CPO}}_i = \frac{1}{\frac{1}{S} \sum_{s=1}^S m_{\mathcal{G}^s}^{-1}(\text{data}_i)},$$

with $\{\mathcal{G}^s, s = 1, \dots, S\}$ the set of EGs visited during the MCMC. Thus, the Monte Carlo estimate of CPO_i is obtained without actually omitting the i th IV from the estimation of the posterior distribution of $\mathcal{G}$ and is provided by the harmonic mean of the marginal likelihood across the visited graphical models.

Recommendations are available regarding the threshold for the detection of outliers. Log-inverse-CPOs larger than 40 can be considered as possible outliers and higher than 70 as extreme values. Ntzoufras^{95,96} and Congdon^{95,96} recommend scaling CPOs by dividing each one by its maximum recorded across the MCMC sweeps and considering observations with scaled CPOs under 0.01 to be outliers. If few CPOs are less than 0.01, the model

is considered to fit adequately. For the detection of outliers, we employ scaled CPO definition.

Simulation study

We share several aspects of the simulation study with Zuber et al.² It is formulated in a two-sample summary-level MR design, where $N = 100,000$ independent individuals are simulated, of which $N_Y = 50,000$ are used to compute the genetic associations with the exposures and $N_X = 50,000$ to compute the genetic associations with the outcomes. Thus, we assume that the quantitative exposures $X_j, j \in J = \{1, \dots, p\}$, and the quantitative responses $Y_k, k \in K = \{1, \dots, q\}$, are measured on the same individuals N_X and N_Y , respectively, with 100% sample overlap within the exposures and 100% sample overlap within the outcomes, but the samples are independent between the two groups, i.e., no sample overlap between N_X and N_Y .

In all simulated scenarios, we consider $p = 15$ exposures, $q = 5$ outcomes, and $n = 100$ independent genetic variants as IVs. Genetic variants for the i th genetic variant and each individual ℓ are simulated independently according to a binomial distribution with minor allele frequency (MAF) equal to 0.05, i.e., $g_{\ell,i} \stackrel{\text{i.i.d.}}{\sim} \text{Bin}(2, 0.05), \ell \in L = \{1, \dots, N\}, i \in I = \{1, \dots, n\}$. The resulting matrix of genetic variants $\mathbf{G}$ is split into two equally sized groups, $\mathbf{G}_X$ and $\mathbf{G}_Y$, of dimension $N_X \times n$ and $N_Y \times n$, respectively. Thus, no IVW is needed in the simulation study given that the same MAF at 5% is used to simulate the genetic variants.

Overall, the data-generation process consists of two stages. In the first stage, the raw data for the exposures $\mathbf{X}$ and the outcomes $\mathbf{Y}$ are simulated. Then, in the second stage, summary-level statistics are obtained as the linear regression coefficients $\hat{\beta}_{ij}^X$ from a univariable linear regression in which the j th exposure is regressed on the i th genetic variant in sample 1 and the linear regression coefficients $\hat{\beta}_{ik}^Y$ from a univariable linear regression in which the k th outcome is regressed on the i th genetic variant in sample 2.

In the following, we detail each stage and how we simulate the quantities involved. We start with the first stage which is divided into two steps.

- (1) In the first step, the exposures are generated as follows:

$$\mathbf{x}_j = \mathbf{G}_X \boldsymbol{\beta}_{Xj} + \psi_X \mathbf{u}_X + \boldsymbol{\epsilon}_{Xj}, j \in J, \quad (\text{Equation A16})$$

where $\mathbf{G}_X$ and $\mathbf{u}_X$ are the genetic variants of the n IVs and the values of the confounder U measured on the same N_X individuals, respectively, and where $\boldsymbol{\beta}_{Xj}$ and ψ_X are the corresponding genetic and confounding effects. $\boldsymbol{\epsilon}_{Xj} \sim N_{N_X}(\mathbf{0}, h_{Xj} \mathbf{I}_{N_X})$, with h_{Xj} the j th diagonal element of the $(p \times p)$ -dimensional matrix

$$\mathbf{H}_X = \frac{1 - v_X}{v_X} \left(\mathbf{G}_X \mathbf{B}_X + \psi_X \mathbf{u}_X \mathbf{1}_p^\top \right)^\top \left(\mathbf{G}_X \mathbf{B}_X + \psi_X \mathbf{u}_X \mathbf{1}_p^\top \right), \quad (\text{Equation A17})$$

where v_X is the desired level of heritability, or how much variation $\mathbf{G}$ can explain of X_j , fixed at 10% for all exposures and in all simulated scenarios. In Equation A17, $\mathbf{B}_X = \{\boldsymbol{\beta}_{Xj}\}_{j \in J}$ is an $(n \times p)$ -dimensional matrix of the effects of the genetic variants on the exposures.

The confounder U is drawn from a multi-variate standard Gaussian distribution, i.e., $\mathbf{u} \sim N_N(\mathbf{0}, \mathbf{I}_N)$ and then split into two equally sized vectors $\mathbf{u}_X$ and $\mathbf{u}_Y$, with effect ψ_X impacting all exposures and ψ_Y effecting all outcomes.

The effects β_{X_j} of the n genetic variants on the j th exposure are drawn following Castelletti et al.⁹⁰ We randomly generate a topologically ordered DAG among the p exposures with a probability of edge inclusion $p_X^{\text{edge}} = 2/(p-1)$ using the function `randomDAG()` in the R package `pcalg`.⁵⁰ Thus, the resulting DAG implies the following system of equations³⁴:

$$\beta_{X_j} = \sum_{h \in \text{pa}(j)} \gamma_{h,j}^X \beta_{X_h} + \epsilon_{X_j} \quad (\text{Equation A18})$$

with $\epsilon_{X_j} \sim N(\mathbf{0}, \mathbf{I}_n)$. For each $j \in J$, the effect within the exposures $\gamma_{h,j}^X$ are uniformly chosen in the interval $[-1.1r_X, -0.9r_X] \cup [0.9r_X, 1.1r_X]$. This construction procedure for β_{X_j} corresponds to the simulated scenario that we call “DAG_X,” i.e., DAG within $\mathbf{X}$, which, in turn, is paired with two different simulated scenarios for the effects β_{Y_k} described in the second step (first stage) of the simulation study.

We also simulate the effects β_{X_j} following Zuber et al.² Specifically, we simulate $\beta_{X_j} \sim N_n(\mathbf{0}, \mathbf{R}_X)$, where $\mathbf{R}_X$ is the $(p \times p)$ -dimensional Toeplitz matrix with $r_X^{|j-j'|}$ for $j, j' \in J$. The matrix $\mathbf{R}_X$ implies a tridiagonal sparse inverse correlation matrix $\mathbf{\Omega}_X = \mathbf{R}_X^{-1}$. The interpretation of non-zero elements of $\mathbf{\Omega}_X$ coincides with the effects simulated in Equation A18. We call this second scenario for the effects of the genetic variants on the exposures “UndG_X,” i.e., undirected graph within $\mathbf{X}$.

In both simulated scenarios for $\mathbf{X}$ we use different levels of r_X , ranging from independence to a strong dependence, i.e., $r_X = \{0, 0.2, 0.4, 0.6, 0.8\}$, where $r_X = 0.6$ represents a medium dependence between the genetic associations with the exposures. We use this value in the figures presented in the “simulation study” section.

- (2) In the second step (first stage) of the simulation study, the outcomes are generated on another independent set of N_Y individuals based on the following set of equations:

$$\mathbf{y}_k = \mathbf{X}\boldsymbol{\theta}_k + \sum_{h \in \text{pa}(k)} \gamma_{h,k}^Y \mathbf{y}_h + \psi_Y \mathbf{u} + \epsilon_{Y_k}, k \in K, \quad (\text{Equation A19})$$

where $\mathbf{X}$ is the $(N_X \times p)$ -dimensional matrix of exposures simulated using Equation A16, $\boldsymbol{\theta}_k = (\theta_{1k}, \dots, \theta_{pk})^T$ is p -dimensional (sparse) vector the causal effects from the exposures to the k th outcome and where ψ_Y is the effect of the confounder U on the outcomes. $\epsilon_{Y_k} \sim N_{N_Y}(\mathbf{0}, h_{Y_k} \mathbf{I}_{N_Y})$, with h_{Y_k} the k th diagonal element of the $(q \times q)$ -dimensional matrix

$$\mathbf{H}_Y = \frac{1 - v_Y}{v_Y} \left(\mathbf{X}\boldsymbol{\theta}_k + \sum_{h \in \text{pa}(k)} \gamma_{h,k}^Y \mathbf{y}_h + \psi_Y \mathbf{u} + \epsilon_{Y_k} \right)^T, \\ \left(\mathbf{X}\boldsymbol{\theta}_k + \sum_{h \in \text{pa}(k)} \gamma_{h,k}^Y \mathbf{y}_h + \psi_Y \mathbf{u} + \epsilon_{Y_k} \right)$$

where v_Y is the desired level of the proportion of variance explained, fixed at 25% for all outcomes and in all simulated scenarios.

In Equation A19, the term $\sum_{h \in \text{pa}(k)} \gamma_{h,k}^Y \mathbf{y}_h$ depends on a randomly generated topologically ordered DAG among the q outcomes with probability of edge inclusion $p_Y^{\text{edge}} = 1/(q-1)$. For each $k \in K$, the effects within the outcomes $\gamma_{h,k}^Y$ are uniformly drawn in the interval $[0.9m_Y, 1.1m_Y]$. In analogy with the first step, we call this scenario “DAG_Y,” i.e., DAG within $\mathbf{Y}$.

We also simulate a simplified scenario where

$$\mathbf{y}_k = \gamma_{h,k}^Y \mathbf{y}_h + \psi_Y \mathbf{u} + \epsilon_{Y_k}, \quad (\text{Equation A20})$$

i.e., a randomly selected outcome k is completed mediated by another randomly selected response chosen between the remaining ones. We call this scenario “Med_Y,” i.e., complete mediation of an outcome, since in the previous scenario “DAG_Y” partial mediations⁹⁷ are likely simulated, while here we exclude this case. In this second simulated scenario for the outcomes, the matrix $\mathbf{H}_Y$ is calculated according to Equation A20. Moreover, we use different levels of m_Y , ranging from small to a strong level of (partial or complete) mediation, i.e., $m_Y = \{0.25, 0.50, 0.75, 1, 1.5, 2\}$, where $m_Y = 1$ represents a medium (partial or complete) mediation effect. We use this value in the figures presented in the section “simulation study.”

Finally, the causal effects $\boldsymbol{\theta}_k$ are drawn independently from a multi-variate Gaussian distribution, i.e., $\boldsymbol{\theta}_k \sim N_p(\mathbf{0}, \mathbf{I}_p)$.

In both simulated scenarios for $\mathbf{Y}$, we consider a $(q \times p)$ -dimensional sparse matrix of causal effects $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_k\}_{k \in K}$, where 30 cells of the matrix are non-zero and where several exposures are either shared or distinct for the outcomes. Specifically, we select at random the same proportion of cells in the matrix $\boldsymbol{\Theta}$ and assign them the simulated values, while the other cells are set to zero.

After the first stage, four scenarios are created by combining the simulations for $\mathbf{X}$ and $\mathbf{Y}$: (1) “UndG_X-Med_Y,” i.e., undirected graph within $\mathbf{X}$ and complete mediation of an outcome in $\mathbf{Y}$; (2) “DAG_X-Med_Y,” i.e., topologically ordered DAG within $\mathbf{X}$ and complete mediation of a response within $\mathbf{Y}$; (3) “UndG_X-DAG_Y,” i.e., undirected graph within $\mathbf{X}$ and topologically ordered DAG within $\mathbf{Y}$; and (4) “DAG_X-DAG_Y,” topologically ordered DAGs within $\mathbf{X}$ and $\mathbf{Y}$. In (2) and (4) the overall DAGs, obtained by combining different simulation patterns for $\mathbf{X}$ and $\mathbf{Y}$, are fully oriented, while in (1) and (3) they are partially oriented.

After creating the data at the individual level, in the second stage, we compute the summary-level statistics from the two independent groups of individuals. The input data for the simulation study are the summary-level statistics $\hat{\mathbf{B}}_X = \{\hat{\beta}_{ij}^X\}_{i \in I, j \in J}$, an $(n \times p)$ -dimensional matrix, and $\hat{\mathbf{B}}_Y = \{\hat{\beta}_{ik}^Y\}_{i \in I, k \in K}$, an $(n \times q)$ -dimensional matrix, derived from a univariable linear regression model, where each genetic variant G_i is regressed against each exposure X_j and each outcome Y_k , one at a time.

Real data application: Pre-processing and data preparation

The first step of the data processing merges the summary-level data (beta regression coefficients, their standard errors, and associated p values) of all exposures by their unique “rs” identifier and aligns the effect direction of the genetic associations with each exposure according to the same effect allele. As IVs, we select the genetic variants that are associated with any of the exposures at genome-wide significance (minimum p value $< 5 \times 10^{-8}$ across all exposures). Next, we merge the genetic variants selected as IVs with the outcome data by their unique “rs” identifier and align the effect direction of the genetic associations with each outcome according to the same effect allele. Finally, we clump the genetic variants to be independent at $r^2 < 0.01$ using a European reference panel.³⁰ This results in $n = 708$ independent genetic variants selected as IVs. See Table S1 for the description of the summary-level statistics, the data sources, and the number of non-unique IVs that were genome-wide significant for each exposure, along with the contribution (%) of each exposure on the selected IVs.

Finally, we perform reverse causation using the same traits with mental health phenotypes as exposures and lifestyle and behavioral traits as outcomes. We apply the same procedure described above, resulting in 470 IVs for mental health phenotypes. See [Table S1](#) for details regarding the number of non-unique IVs that were genome-wide significant for each exposure along with the contribution (%) of each exposure on the selected IVs.

Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2025.03.005>.

Received: July 25, 2024

Accepted: March 7, 2025

Published: April 2, 2025

References

- Zuber, V., Colijn, J.M., Klaver, C., and Burgess, S. (2020). Selecting likely causal risk factors from high-throughput experiments using multivariable Mendelian randomization. *Nat. Commun.* 11, 29. <https://doi.org/10.1038/s41467-019-13870-3>.
- Zuber, V., Lewin, A., Levin, M.G., Haglund, A., Ben-Aicha, S., Emanuel, C., Damrauer, S., Burgess, S., Gill, D., and Bottolo, L. (2023). Multi-response Mendelian randomization: Identification of shared and distinct exposures for multimorbidity and multiple related disease outcomes. *Am. J. Hum. Genet.* 110, 1177–1199. <https://doi.org/10.1016/j.ajhg.2023.06.005>.
- Sanderson, E., Davey Smith, G., Windmeijer, F., and Bowden, J. (2019). An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int. J. Epidemiol.* 48, 713–727. <https://doi.org/10.1093/ije/dyy262>.
- Howey, R., Shin, S.-Y., Relton, C., Davey Smith, G., and Cordell, H.J. (2020). Bayesian network analysis incorporating genetic anchors complements conventional Mendelian randomization approaches for exploratory analysis of causal relationships in complex data. *PLoS Genet.* 16, e1008198. <https://doi.org/10.1371/journal.pgen.1008198>.
- Badsha, M.B., and Fu, A.Q. (2019). Learning causal biological networks with the principle of Mendelian randomization. *Front. Genet.* 10, 460. <https://doi.org/10.3389/fgene.2019.00460>.
- Lin, Z., Xue, H., and Pan, W. (2023). Combining Mendelian randomization and network deconvolution for inference of causal networks with GWAS summary data. *PLoS Genet.* 19, e1010762. <https://doi.org/10.1371/journal.pgen.1010762>.
- Chen, S., Lin, Z., Shen, X., Li, L., and Pan, W. (2023). Inference of causal metabolite networks in the presence of invalid instrumental variables with GWAS summary data. *Genet. Epidemiol.* 47, 585–599. <https://doi.org/10.1002/gepi.22535>.
- Li, C., Shen, X., and Pan, W. (2023). Inference for a large Directed Acyclic Graph with unspecified interventions. *J. Mach. Learn. Res.* 24, 73–3308. <https://jmlr.org/papers/v24/21-0855.html>.
- Zilinskas, R., Li, C., Shen, X., Pan, W., and Yang, T. (2024). Inferring a Directed Acyclic Graph of phenotypes from GWAS summary statistics. *Biometrics* 80, ujad039. <https://doi.org/10.1093/biometc/ujad039>.
- Chen, L., Li, C., Shen, X., and Pan, W. (2024). Discovery and inference of a causal network with hidden confounding. *J. Am. Stat. Assoc.* 119, 2572–2584. <https://doi.org/10.1080/01621459.2023.2261658>.
- Xue, H., Shen, X., and Pan, W. (2021). Constrained maximum likelihood-based Mendelian randomization robust to both correlated and uncorrelated pleiotropic effects. *Am. J. Hum. Genet.* 108, 1251–1269. <https://doi.org/10.1016/j.ajhg.2021.05.014>.
- Feizi, S., Marbach, D., Médard, M., and Kellis, M. (2013). Network deconvolution as a general method to distinguish direct dependencies in networks. *Nat. Biotechnol.* 31, 726–733. <https://doi.org/10.1038/nbt.2635>.
- Shen, X., and Ye, J. (2002). Adaptive model selection. *J. Am. Stat. Assoc.* 97, 210–221. <https://doi.org/10.1198/016214502753479356>.
- Hernán, M.A., and Robins, J.M. (2006). Instruments for causal inference. An epidemiologist's dream? *Epidemiology* 17, 360–372. <https://doi.org/10.1097/01.ede.0000222409.00878.37>.
- Didelez, V., and Sheehan, N. (2007). Mendelian randomization as an instrumental variable approach to causal inference. *Stat. Methods Med. Res.* 16, 309–330. <https://doi.org/10.1177/0962280206077743>.
- Drton, M., and Maathuis, M.H. (2017). Structure learning in graphical modeling. *Annu. Rev. Stat. Appl.* 4, 365–393. <https://doi.org/10.1146/annurev-statistics-060116-053803>.
- Pearl, J. (2009). *Causality: Models, Reasoning and Inference*, 2nd edn (Cambridge: Cambridge University Press).
- GBD 2019 Mental Disorders Collaborators (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry* 9, 137–150. [https://doi.org/10.1016/s2215-0366\(21\)00395-3](https://doi.org/10.1016/s2215-0366(21)00395-3).
- Arias, D., Saxena, S., and Verguet, S. (2022). Quantifying the global burden of mental disorders and their economic value. *eClinicalMedicine* 54, 101675. <https://doi.org/10.1016/j.eclinm.2022.101675>.
- Tyrer, P. (2014). A comparison of DSM and ICD classifications of mental disorder. *Adv. Psychiatr. Treat.* 20, 280–285. <https://doi.org/10.1192/apt.bp.113.011296>.
- Saxe, G.N., Bickman, L., Ma, S., and Aliferis, C. (2022). Mental health progress requires causal diagnostic nosology and scalable causal discovery. *Front. Psychiatry* 13, 898789. <https://doi.org/10.3389/fpsy.2022.898789>.
- Arango, C., Dragioti, E., Solmi, M., Cortese, S., Domschke, K., Murray, R.M., Jones, P.B., Uher, R., Carvalho, A.F., Reichenberg, A., et al. (2021). Risk and protective factors for mental disorders beyond genetics: An evidence-based atlas. *World Psychiatry* 20, 417–436. <https://doi.org/10.1002/wps.20894>.
- Leichsenring, F., Steinert, C., Rabung, S., and Ioannidis, J.P.A. (2022). The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: An umbrella review and meta-analytic evaluation of recent meta-analyses. *World Psychiatry* 21, 133–145. <https://doi.org/10.1002/wps.20941>.
- Saccaro, L.F., Gasparini, S., and Rutigliano, G. (2022). Applications of Mendelian randomization in psychiatry: A comprehensive systematic review. *Psychiatr. Genet.* 32, 199–213. <https://doi.org/10.1097/ypg.0000000000000327>.
- Burgess, S., Foley, C.N., and Zuber, V. (2018). Inferring causal relationships between risk factors and outcomes from genome-wide association study data. *Annu. Rev. Genomics*

- Hum. Genet. 19, 303–327. <https://doi.org/10.1146/annurev-genom-083117-021731>.
26. Davies, N.M., Dickson, M., Davey Smith, G., van den Berg, G.J., and Windmeijer, F. (2018). The causal effects of education on health outcomes in the UK Biobank. *Nat. Hum. Behav.* 2, 117–125. <https://doi.org/10.1038/s41562-017-0279-y>.
 27. Gage, S.H., Bowden, J., Davey Smith, G., and Munafò, M.R. (2018). Investigating causality in associations between education and smoking: A two-sample Mendelian randomization study. *Int. J. Epidemiol.* 47, 1131–1140. <https://doi.org/10.1101/184218>.
 28. Kendler, K.S., Ohlsson, H., Sundquist, J., and Sundquist, K. (2022). Risk for mood, anxiety, and psychotic disorders in individuals at high and low genetic liability for bipolar disorder and major depression. *JAMA Psychiatry* 79, 1102–1109. <https://doi.org/10.1001/jamapsychiatry.2022.2873>.
 29. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
 30. Zuber, V., Gill, D., Ala-Korpela, M., Langenberg, C., Butterworth, A., Bottolo, L., and Burgess, S. (2021). High-throughput multivariable Mendelian randomization analysis prioritizes apolipoprotein B as key lipid risk factor for coronary artery disease. *Int. J. Epidemiol.* 50, 893–901. <https://doi.org/10.1093/ije/dyaa216>.
 31. Smith, G.D., and Ebrahim, S. (2003). Mendelian randomization: Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* 32, 1–22. <https://doi.org/10.1093/ije/dyg070>.
 32. Thomas, D.C., and Conti, D.V. (2004). Commentary: The concept of 'Mendelian Randomization'. *Int. J. Epidemiol.* 33, 21–25. <https://doi.org/10.1093/ije/dyh048>.
 33. Lauritzen, S.L. (2004). *Graphical Models. Repr. With Corrections* (Oxford, New York: Clarendon Press).
 34. Peters, J., and Bühlmann, P. (2013). Identifiability of Gaussian structural equation models with equal error variances. *Biometrika* 101, 219–228. <https://doi.org/10.1093/biomet/ast043>.
 35. Andersson, S.A., Madigan, D., and Perlman, M.D. (1997). A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Statist.* 25, 505–541. <https://doi.org/10.1214/aos/1031833662>.
 36. Perković, E., Kalisch, M., and Maathuis, M.H. (2017). In Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence (UAI). G. Elidan, K. Kersting, and A.T. Ihler, eds., (AUAI Press). <http://auai.org/uai2017/proceedings/papers/120.pdf>.
 37. Didelez, V. (2018). Causal concepts and graphical models. In *Handbook of Graphical Models*, M. Maathuis, M. Drton, S. Lauritzen, and M. Wainwright, eds. (Boca Raton, FL: Chapman and Hall/CRC), pp. 353–376. <https://doi.org/10.1201/9780429463976-15>.
 38. Didelez, V., Meng, S., and Sheehan, N.A. (2010). Assumptions of IV methods for observational epidemiology. *Stat. Sci.* 25, 22–40. <https://doi.org/10.1214/09-sts316>.
 39. Frot, B., Nandy, P., and Maathuis, M.H. (2019). Robust causal structure learning with some hidden variables. *J. Roy. Stat. Soc. B* 81, 459–487. <https://doi.org/10.1111/rssb.12315>.
 40. Grotzinger, A.D., Rhemtulla, M., de Vlaming, R., Ritchie, S.J., Mallard, T.T., Hill, W.D., Ip, H.F., Marioni, R.E., McIntosh, A.M., Deary, I.J., et al. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* 3, 513–525. <https://doi.org/10.1038/s41562-019-0566-x>.
 41. Burgess, S., and Bowden, J. (2015). Integrating summarized data from multiple genetic variants in Mendelian randomization: Bias and coverage properties of inverse-variance weighted methods. Preprint at arXiv. Unpublished manuscript. <https://doi.org/10.48550/ARXIV.1512.04486>.
 42. Rees, J.M.B., Wood, A.M., and Burgess, S. (2017). Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat. Med.* 36, 4705–4718. <https://doi.org/10.1002/sim.7492>.
 43. Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. (1999). Bayesian model averaging: A tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors). *Stat. Sci.* 14, 382–417. <https://doi.org/10.1214/ss/1009212519>.
 44. Rahman, S., Khare, K., Michailidis, G., Martínez, C., and Carulla, J. (2023). Estimation of Gaussian directed Acyclic Graphs using partial ordering information with applications to DREAM3 networks and dairy cattle data. *Ann. Appl. Stat.* 17, 929–960. <https://doi.org/10.1214/22-aos1636>.
 45. Sanderson, E., Spiller, W., and Bowden, J. (2021). Testing and correcting for weak and pleiotropic instruments in two-sample multivariable Mendelian randomization. *Stat. Med.* 40, 5434–5452. <https://doi.org/10.1002/sim.9133>.
 46. Wu, Y., Kang, H., and Ye, T. (2024). A more credible approach to multivariable Mendelian randomization. Preprint at arXiv. <https://doi.org/10.48550/ARXIV.2402.00307>.
 47. Burgess, S., and Thompson, S.G. (2017). Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* 32, 377–389. <https://doi.org/10.1007/s10654-017-0255-x>.
 48. Verbanck, M., Chen, C.Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* 50, 693–698. <https://doi.org/10.1038/s41588-018-0099-7>.
 49. Burgess, S., and Thompson, S.G. (2015). Multivariable Mendelian randomization: The use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* 181, 251–260. <https://doi.org/10.1093/aje/kwu283>.
 50. Kalisch, M., Mächler, M., Colombo, D., Maathuis, M.H., and Bühlmann, P. (2012). Causal inference using graphical models with the R package pcalg. *J. Stat. Software* 47, 1–26. <https://doi.org/10.18637/jss.v047.i11>.
 51. Badsha, M.B., Martin, E.A., and Fu, A.Q. (2021). MRPC: An R package for inference of causal graphs. *Front. Genet.* 12, 651812. <https://doi.org/10.3389/fgene.2021.651812>.
 52. Barbieri, M.M., and Berger, J.O. (2004). Optimal predictive model selection. *Ann. Statist.* 32, 870–897. <https://doi.org/10.1214/009053604000000238>.
 53. Spirtes, P., Glymour, C., and Scheines, R.C. (2000). *Prediction and Search, 2nd edn* (Cambridge, MA: MIT Press).
 54. Raghupathi, V., and Raghupathi, W. (2020). The influence of education on health: An empirical assessment of OECD countries for the period 1995–2015. *Arch. Public Health* 78, 20. <https://doi.org/10.1186/s13690-020-00402-5>.
 55. Amin, V., Fletcher, J.M., Lu, Q., and Song, J. (2023). Re-examining the relationship between education and adult mental health in the UK: A research note. *Econ. Educ. Rev.* 93, 102354. <https://doi.org/10.1016/j.econeduc.2023.102354>.

56. Dardani, C., Riglin, L., Leppert, B., Sanderson, E., Rai, D., Howe, L.D., Davey Smith, G., Tilling, K., Thapar, A., Davies, N.M., et al. (2022). Is genetic liability to ADHD and ASD causally linked to educational attainment? *Int. J. Epidemiol.* 50, 2011–2023. <https://doi.org/10.1093/ije/dyab107>.
57. Verhoef, E., Grove, J., Shapland, C.Y., Demontis, D., Burgess, S., Rai, D., Børglum, A.D., and St Pourcain, B. (2021). Discordant associations of educational attainment with ASD and ADHD implicate a polygenic form of pleiotropy. *Nat. Commun.* 12, 6534. <https://doi.org/10.1038/s41467-021-26755-1>.
58. Cai, J., Wei, Z., Chen, M., He, L., Wang, H., Li, M., and Peng, Y. (2022). Socioeconomic status, individual behaviors and risk for mental disorders: A Mendelian randomization study. *Eur. Psychiatry* 65, e28. <https://doi.org/10.1192/j.eurpsy.2022.18>.
59. Lloyd, E.C., Reed, Z.E., and Wootton, R.E. (2023). The absence of association between anorexia nervosa and smoking: Converging evidence across two studies. *Eur. Child Adolesc. Psychiatry* 32, 1229–1240. <https://doi.org/10.1007/s00787-021-01918-z>.
60. Kari, J.T., Viinikainen, J., Böckerman, P., Tammelin, T.H., Pitkänen, N., Lehtimäki, T., Pahlkala, K., Hirvensalo, M., Raitakari, O.T., and Pehkonen, J. (2020). Education leads to a more physically active lifestyle: Evidence based on Mendelian randomization. *Scand. J. Med. Sci. Sports* 30, 1194–1204. <https://doi.org/10.1111/sms.13653>.
61. Wootton, R.E., Richmond, R.C., Stuijzand, B.G., Lawn, R.B., Sallis, H.M., Taylor, G.M.J., Hemani, G., Jones, H.J., Zammit, S., Davey Smith, G., and Munafò, M.R. (2020). Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: A Mendelian randomisation study. *Psychol. Med.* 50, 2435–2443. <https://doi.org/10.1017/S0033291719002678>.
62. Treur, J.L., Demontis, D., Smith, G.D., Sallis, H., Richardson, T.G., Wiers, R.W., Børglum, A.D., Verweij, K.J.H., and Munafò, M.R. (2021). Investigating causality between liability to ADHD and substance use, and liability to substance use and ADHD risk, using Mendelian randomization. *Addict. Biol.* 26, e12849. <https://doi.org/10.1111/adb.12849>.
63. Reed, Z.E., Wootton, R.E., and Munafò, M.R. (2022). Using Mendelian randomisation to explore the gateway hypothesis: Possible causal effects of smoking initiation and alcohol consumption on substance use outcomes. *Addiction* 117, 741–750. <https://doi.org/10.1111/add.15673>.
64. Lohr, J.B., and Flynn, K. (1992). Smoking and schizophrenia. *Schizophr. Res.* 8, 93–102. [https://doi.org/10.1016/0920-9964\(92\)90024-Y](https://doi.org/10.1016/0920-9964(92)90024-Y).
65. Kendler, K.S., Lönn, S.L., Sundquist, J., and Sundquist, K. (2015). Smoking and schizophrenia in population cohorts of Swedish women and men: A prospective co-relative control study. *Am. J. Psychiatr.* 172, 1092–1100. <https://doi.org/10.1176/appi.ajp.2015.15010126>.
66. Richards, A.L., Cardno, A., Harold, G., Craddock, N.J., Di Florio, A., Jones, L., Gordon-Smith, K., Jones, I., Sellers, R., Walters, J.T.R., et al. (2022). Genetic liabilities differentiating bipolar disorder, schizophrenia, and major depressive disorder, and phenotypic heterogeneity in bipolar disorder. *JAMA Psychiatry* 79, 1032–1039. <https://doi.org/10.1371/j10.1001/jamapsychiatry.2022.2594>.
67. Peyre, H., Schoeler, T., Liu, C., Williams, C.M., Hoertel, N., Havdahl, A., and Pingault, J.B. (2021). Combining multivariate genomic approaches to elucidate the comorbidity between autism spectrum disorder and attention deficit hyperactivity disorder. *J. Child Psychol. Psychiatry* 62, 1285–1296. <https://doi.org/10.1111/jcpp.13479>.
68. Vermeulen, J.M., Wootton, R.E., Treur, J.L., Sallis, H.M., Jones, H.J., Zammit, S., van den Brink, W., Goodwin, G.M., de Haan, L., and Munafò, M.R. (2021). Smoking and the risk for bipolar disorder: Evidence from a bidirectional Mendelian randomisation study. *Br. J. Psychiatry* 218, 88–94. <https://doi.org/10.1192/bjp.2019.202>.
69. Cai, N., Revez, J.A., Adams, M.J., Andlauer, T.F.M., Breen, G., Byrne, E.M., Clarke, T.K., Forstner, A.J., Grabe, H.J., Hamilton, S.P., et al. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nat. Genet.* 52, 437–447. <https://doi.org/10.1038/s41588-020-0594-5>.
70. Cai, N., Choi, K.W., and Fried, E.I. (2020). Reviewing the genetics of heterogeneity in depression: Operationalizations, manifestations and etiologies. *Hum. Mol. Genet.* 29, R10–R18. <https://doi.org/10.1093/hmg/ddaa115>.
71. Forbes, M.K., Neo, B., Nezami, O.M., Fried, E.I., Faure, K., Michelsen, B., Twose, M., and Dras, M. (2024). Elemental psychopathology: Distilling constituent symptoms and patterns of repetition in the diagnostic criteria of the DSM-5. *Psychol. Med.* 54, 886–894. <https://doi.org/10.1017/S0033291723002544>.
72. Kendrick, T., Pilling, S., Mavranouzouli, I., Megnin-Viggars, O., Ruane, C., Eadon, H., Kapur, N., and Guideline Committee (2022). Management of depression in adults: summary of updated NICE guidance. *Br. Med. J.* 378, o1557.
73. Anderson, E.L., Howe, L.D., Wade, K.H., Ben-Shlomo, Y., Hill, W.D., Deary, I.J., Sanderson, E.C., Zheng, J., Korologou-Linden, R., Stergiakouli, E., et al. (2020). Education, intelligence and Alzheimer's disease: Evidence from a multivariable two-sample Mendelian randomization study. *Int. J. Epidemiol.* 49, 1163–1172. <https://doi.org/10.1093/ije/dydz280>.
74. van Amsterdam, J., van der Velde, B., Schulte, M., and van den Brink, W. (2018). Causal factors of increased smoking in ADHD: A systematic review. *Subst. Use Misuse* 53, 432–445. <https://doi.org/10.1080/10826084.2017.1334066>.
75. Green, R., Baker, N.L., Ferguson, P.L., Hashemi, D., and Gray, K.M. (2023). ADHD symptoms and smoking outcomes in a randomized controlled trial of varenicline for adolescent and young adult tobacco cessation. *Drug Alcohol Depend.* 244, 109798. <https://doi.org/10.1016/j.drugalcdep.2023.109798>.
76. Glymour, C., Zhang, K., and Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Front. Genet.* 10, 524. <https://doi.org/10.3389/fgene.2019.00524>.
77. Wang, Y., and Blei, D.M. (2019). The blessings of multiple causes. *J. Am. Stat. Assoc.* 114, 1574–1596. <https://doi.org/10.1080/01621459.2019.1686987>.
78. Shah, R.D., Frot, B., Thanei, G.-A., and Meinshausen, N. (2020). Right singular vector projection graphs: Fast high dimensional covariance matrix estimation under latent confounding. *J. Roy. Stat. Soc. B Stat. Methodol.* 82, 361–389. <https://doi.org/10.1111/rssb.12359>.
79. Sanderson, E., Davey Smith, G., Bowden, J., and Munafò, M.R. (2019). Mendelian randomisation analysis of the effect of educational attainment and cognitive ability on smoking behaviour. *Nat. Commun.* 10, 2949. <https://doi.org/10.1038/s41467-019-10679-y>.
80. Burgess, S., Woolf, B., Mason, A.M., Ala-Korpela, M., and Gill, D. (2024). Addressing the credibility crisis in Mendelian

- p randomization.
- BMC Med.*
- 22, 374.
- <https://doi.org/10.1186/s12916-024-03607-5>
- .
81. Haglund, A., Zuber, V., Abouzeid, M., Yang, Y., Ko, J.H., Wiemann, L., Otero-Jimenez, M., Muhammed, L., Feleke, R., Nott, A., et al. (2025). Cell state-dependent allelic effects and contextual Mendelian randomization analysis for human brain phenotypes. *Nat. Genet.* 57, 358–368. <https://doi.org/10.1038/s41588-024-02050-9>.
 82. LaBianca, S., Brikell, I., Helenius, D., Loughnan, R., Mefford, J., Palmer, C.E., Walker, R., Gådin, J.R., Krebs, M., Appadurai, V., et al. (2024). Polygenic profiles define aspects of clinical heterogeneity in attention deficit hyperactivity disorder. *Nat. Genet.* 56, 234–244. <https://doi.org/10.1038/s41588-023-01593-7>.
 83. Pearl, J. (2010). An introduction to causal inference. *Int. J. Biostat.* 6, 7. <https://doi.org/10.2202/1557-4679.1203>.
 84. Maathuis, M.H., Kalisch, M., and Bühlmann, P. (2009). Estimating high-dimensional intervention effects from observational data. *Ann. Stat.* 37, 3133–3164. <https://doi.org/10.1214/09-aos685>.
 85. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, and Duncan, L., et al. (2015). An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* 47, 1236–1241. <https://doi.org/10.1038/ng.3406>.
 86. Shi, H., Mancuso, N., Spendlove, S., and Pasaniuc, B. (2017). Local genetic correlation gives insights into the shared genetic architecture of complex traits. *Am. J. Hum. Genet.* 101, 737–751. <https://doi.org/10.1016/j.ajhg.2017.09.022>.
 87. Harville, D. (1997). *Matrix Algebra from a Statistician's Perspective* (New York: Springer-Verlag). Repr. with corrections.
 88. Castelletti, F., and Consonni, G. (2021). Bayesian inference of causal effects from observational data in Gaussian graphical models. *Biometrics* 77, 136–149. <https://doi.org/10.1111/biom.13281>.
 89. Chickering, D.M. (2002). Learning equivalence classes of Bayesian-network structures. *J. Mach. Learn. Res.* 2, 445–498. <https://jmlr.csail.mit.edu/papers/v2/chickering02a.html>.
 90. Castelletti, F., Consonni, G., Vedova, M.L.D., and Peluso, S. (2018). Learning Markov equivalence classes of Directed Acyclic Graphs: An objective Bayes approach. *Bayesian Analysis* 13, 1235–1260. <https://doi.org/10.1214/18-ba1101>.
 91. He, Y., Jia, J., and Yu, B. (2013). Reversible MCMC on Markov equivalence classes of sparse Directed Acyclic Graphs. *Ann. Statist.* 41, 1742–1779. <https://doi.org/10.1214/13-aos1125>.
 92. Bottolo, L., and Richardson, S. (2010). Evolutionary stochastic search for Bayesian model exploration. *Bayesian Anal.* 5, 583–618. <https://doi.org/10.1214/10-ba523>.
 93. Geisser, S., and Eddy, W.F. (1979). A predictive approach to model selection. *J. Am. Stat. Assoc.* 74, 153–160. <https://doi.org/10.1080/01621459.1979.10481632>.
 94. Gelfand, A.E. (1996). Model determination using sampling-based methods. In *Markov Chain Monte Carlo in Practice*, W. Gilks, S. Richardson, and D. Spiegelhalter, eds. (Boca Raton, FL: Chapman & Hall), pp. 145–161.
 95. Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS* (Hoboken, N.J.: Wiley). <https://doi.org/10.1002/9780470434567>.
 96. Congdon, P. (2005). *Bayesian Models for Categorical Data* (Chichester: John Wiley & Sons).
 97. Zeng, P., Shao, Z., and Zhou, X. (2021). Statistical methods for mediation analysis in the era of high-throughput genomics: Current successes and future challenges. *Comput. Struct. Biotechnol. J.* 19, 3209–3224. <https://doi.org/10.1016/j.csbj.2021.05.042>.