

MRS�: A phenome-wide causal discovery algorithm based on GWAS summary data

Lei Hou^{1,2}, Zhi Geng³, Xu Shi⁴, Chuan Wang⁵, Hongkai Li^{1,2*†}, Fuzhong Xue^{1,2*†}

† Fuzhong Xue and Hongkai Li contributed equally to this work.

Author affiliations:

1. Department of Epidemiology and Health Statistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, People's Republic of China, 250000
2. Institute for Medical Dataology, Cheeloo College of Medicine, Shandong University, Jinan, People's Republic of China, 250000
3. School of Mathematics and Statistics, Beijing Technology and Business University, Beijing, People's Republic of China, 100048
4. Department of Biostatistics, University of Michigan, Ann Arbor, USA
5. Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, People's Republic of China, 250000

Corresponding author:

1. Fuzhong Xue, E-mail: xuefzh@sdu.edu.cn, Telephone: 13906405997,
Address: School of public health, Shandong University, 44 Wenhua West Road, Jinan, Shandong province, China
2. Hongkai Li, E-mail: lihongkaiyouxiang@163.com, Telephone: 18310601363,
Address: School of public health, Shandong University, 44 Wenhua West Road, Jinan, Shandong province, China

Abstract

Causal discovery is a powerful tool to disclose underlying structures by analyzing purely observational data. Genetic variants can provide useful complementary information for structure learning. Here, we propose a novel algorithm MRSL (Mendelian Randomization (MR)-based Structure Learning algorithm), which combines the graph theory with univariable and multivariable MR to learn the true structure using only GWAS summary statistics. Specifically, MRSL also utilizes topological sorting to improve the precision of structure learning and provides three adjusting categories for multivariable MR. Results of simulation reveal that MRSL has up to two-fold higher F1 score than other eight competitive methods. Additionally, the computing time of MRSL is 100 times faster than other methods. Furthermore, we apply MRSL to 26 biomarkers and 44 ICD10-defined diseases from UK Biobank. The results cover most of expected causal links which have biological interpretations and several new links supported by clinical case reports or previous observational literatures.

Introduction

Causal discovery aims to infer causal structure by analyzing purely observational data [1-2]. It can be widely applied in the social and natural sciences, and it is a powerful tool for discovering biological networks [3-4] and disease diagnostic purposes [5-6], etc. Graphical models reveal the generating process of the observed data and they can be identified under three assumptions [1-2, 7-8]: (1) the causal Markov condition, (2) the causal faithfulness assumption and (3) the causal sufficiency assumption. The causal Markov condition means that all nodes are independent of their non-descendants when conditionally on their parents. The causal faithfulness assumption requires all conditional independences in true underlying distribution \mathbb{P} are represented in the graph and are invariant to changes in parameterization. The causal sufficiency assumption states that any pair of nodes in the graph has no common external cause, and it implies there is no unobserved confounding variable. Various algorithms have been developed and can learn causal structures from purely or mostly observational data [9]. Constraint-based methods start with a fully connected graph and carry out a series of marginal and conditional independence tests to decide which edges should be removed, such as PC [10] and Markov blanket detection algorithm (e.g. Grow-Shrink [11] and Incremental Association algorithm [12-13]), etc. The outputs of such algorithms are equivalence classes. Score-based methods find the most plausible Directed Acyclic Graph (DAG) by optimizing a properly defined score function, such as the hill climbing (HC) greedy search algorithm [14], etc. The hybrid algorithms have become widely used because they combine the advantages of the constraint-based and score-based algorithms [15]. One popular strategy is to use constraint-based algorithms to determine an initial network structure and then to use score-based algorithms to find the highest-scoring network structure. For example, Max-Min Hill-Climbing (MMHC) [16] and Restricted Maximization algorithm [17], etc. All of these methods require the causal sufficiency assumption and require that the input data sets must be individual data. Recently several algorithms have been developed to learn the structure with latent variables, that is, relax the causal sufficiency assumption, such as Fast Causal Inference (FCI) algorithm [18-19] and Greedy FCI [20], etc. However, they output Partial Ancestral Graphs (PAG), but not complete structure information. All of the above algorithms are for structural learning but not for parameter learning

Utilization of genetic variants which are robustly associated with a risk factor provides a directional causal anchor for causal discovery. Thus the combination of Mendelian Randomization (MR) and causal discovery becomes popular. MR methods use genetic

variants as instrumental variables (IVs) to infer the causal relationship from the exposure to the outcome [21]. These IVs can be used to remove confounding biases and to avoid reversed causal relationships. A valid IV must satisfy three assumptions: (1) Relevance – IV is robustly associated with the exposure; (2) Exchangeability – IV is not associated with any confounder of the exposure–outcome relationship; (3) Exclusion restriction – IV is independent of the outcome conditional on the exposure and all confounders of the exposure-outcome relationship. Richard et al. [22-23] concluded that Bayesian network (BN) incorporating genetic anchors is a useful complementary method to conventional MR for exploring causal relationships in complex omic data sets. A novel machine learning algorithm named MRPC incorporates the Principle of Mendelian randomization (PMR) in the PC algorithm, to learn causal graphs [24]. Nevertheless the algorithm also requires causal sufficiency assumption, i.e. no unobserved confounders among all the variables. Actually, this method only uses the information of genetic variants but not the idea of MR. Afterwards, David et al. [25] presented a pipeline, named causal Graphical Analysis Using Genetics (cGAUGE), using IV filters with provable properties to perform univariable MR (UVMR) [26-27] then to obtain a causal graph. This algorithm allows the unobserved confounders among all the variables and requires individual genetic and phenotypic data. A flexible two stage procedure called bidirectional mediated Mendelian randomization (BIMMER) can be used to infer sparse networks of direct causal effects (DCEs) from phenome-scale GWAS summary statistics [28]. However, this process is implemented by inverse sparse regression under the assumption that the DCE matrix is sparse. Furthermore, BIMMER is very time consuming.

Multivariable MR (MVMR) [27, 29-30] is able to compute DCEs when there are multiple potential exposures and a single outcome. In MVMR, each genetic variant must satisfy the following criteria: (1) the variant is associated with at least one of the exposures; (2) the variant is independent of all confounders between exposures and outcomes; and (3) the variant is independent of the outcome conditional on the exposures and confounders. In this paper, we propose a novel algorithm called MRSL based on UVMR and MVMR for structural learning using summarized genetic data without requiring individual data. MRSL starts with an empty graph. For the first step, a marginal causal graph can be obtained by using bi-directional MR [31-32] in pairs. This process is a UVMR analysis, which estimates all total causal effects for each pair of variables. For the second step, we find the topological sorting for the marginal causal graph. For the third step, based on the above topological sorting [33], MVMR is performed to estimate the direct effects for each pair of variables by adjusting for

the genetic associations with the phenotypes in a sufficient separating set. After an iteration process of step 2 and step 3, MRSL outputs a true causal graph. We apply MRSL to 26 biomarkers and 44 ICD10-defined diseases in 337,198 European from UK Biobank using summarized genetic data.

Results

[Please insert Figure 1 here]

Method overview. We present a novel algorithm MRSL for structural learning based on summarized genetic data. An illustration diagram of MRSL is displayed in Figure 1. Consider a DAG \mathcal{G} with d phenotypes $\{X_1, X_2, \dots, X_d\}$. U represents a set of unobserved confounders among d phenotypes. GWAS summary data for these d phenotypes are available. Generally, for continuous phenotypes, beta coefficients and their standard errors can be obtained from linear regression; for binary phenotypes, log(OR) coefficients and their standard errors can be obtained from logistic regression. Firstly, MRSL initializes the target causal graph with an empty graph and then obtains a marginal causal graph \mathcal{G}_M , using bi-directional MR in pairs of variables. The marginal causal graph \mathcal{G}_M includes all the edges in the true causal graph \mathcal{G} , but may add extra edges and spurious colliders. At the second step, we find the topological sorting of the nodes in the marginal causal graph using Depth First Search (DFS) algorithm [33-34]. The order of the topological sorting for the true causal graph \mathcal{G} and the marginal causal graph \mathcal{G}_M are the same. Based on this order, MVMR is performed to remove extra edges in \mathcal{G}_M . For each edge $X_p \rightarrow X_q$ in the marginal causal graph \mathcal{G}_M , we search a sufficient separating set $SSS_{X_p \rightarrow X_q} = \{ADJ_{X_p \rightarrow X_q}, U_{pq}\}$ and adjust for genetic associations with $ADJ_{X_p \rightarrow X_q}$ using MVMR to detect whether an edge is extra. The marginal causal graph is updated after each edge's test. $ADJ_{X_p \rightarrow X_q}$ can be searched in three ways: (1) all variables on the open paths from X_p to X_q ; (2) minimal sufficient adjustment set [1-2] and all the mediators from X_p to X_q ; (3) $V \setminus \{X_p, X_q \text{ and } S^d\}$, where S^d refers to the colliders where X_p and X_q have direct edges on them. For each edge $X_p \rightarrow X_q$ pointing to them in the marginal causal graph \mathcal{G}_M , if there is a sufficient separating set $SSS_{X_p \rightarrow X_q}$ such that $X_p \perp X_q | SSS_{X_p \rightarrow X_q}$ using MVMR, we delete the extra edge $X_p \rightarrow X_q$ in the true causal graph \mathcal{G} . In addition, adjusting for the nodes in $SSS_{X_p \rightarrow X_q}$ cannot unlock any blocked

pathways in the true causal graph \mathcal{G} . Thus the third step removes the extra edges in the graph \mathcal{G}_M . For the fourth step, we add an iteration step to perform MVMMR in step 2 and step 3 again, using the graph obtained by step 3 as the initialization, until this graph converges. Detailed illustration of MRSL is shown in the Methods section.

We provide a motivating example to illustrate the workflow of MRSL (Figure 1 A-F). The true causal diagram is Figure 1 (A). The input are the GWAS summary datasets of five phenotypes. Firstly, bidirectional MR in pairs of five variables are performed to obtain a marginal causal graph (Figure 1 (B)). We find that its topological sorting is $\{X_1, X_3, X_5, X_2, X_4\}$. Then we perform MVMMR varying across each edge in Figure 1 (B) to detect whether the edge is extra. In this stage, we adjust for the genetic associations with phenotypes in $ADJ_{X_p \rightarrow X_q}$ for each MVMMR. We firstly focus on the edge $X_1 \rightarrow X_3$. The other three nodes are not included in $ADJ_{X_1 \rightarrow X_3}$ thus this edge retains. Then we are interested in the edge $X_1 \rightarrow X_5$, and $ADJ_{X_1 \rightarrow X_5}^{(1)} = ADJ_{X_1 \rightarrow X_5}^{(2)} = \{X_3\}$ and $ADJ_{X_1 \rightarrow X_5}^{(3)} = \{X_2, X_3, X_4\}$. MVMMR is performed by adjusting for the genetic associations with phenotypes in $ADJ_{X_p \rightarrow X_q}$ and the result reveals a null direct causal relationship between X_1 and X_5 . Thus the edge $X_1 \rightarrow X_5$ is removed. The rest edges are tested by the same ways (Figure 1 C-J). After all the edges in Figure 1 (B) are tested once, we obtained Figure 1 (J). An iteration for step 2 and step 3 are performed using this graph and stop when the causal graph converges. Finally MRSL output the target causal diagram.

Simulations: Firstly we conducted a simulation study to evaluate the performance of MVMMR to estimate the direct causal effect of an exposure (X) on an outcome (Y) when adjusting for a collider (S), a mediator (M) or a measured confounder (C), respectively (Figure 2 A-C). We considered seven kinds of available SNPs as IVs: (1) G_1 : SNPs only associated with X; (2) G_2 : SNPs associated with X and adjusting variable; (3) G_3 : SNPs only associated with the adjusting variable; (4) $G_1 + G_3$; (5) $G_1 + G_2$; (6) $G_2 + G_3$; (7) $G_1 + G_2 + G_3$. Details of data generation are shown in Methods section and Supplementary Notes.

[Please insert Figure 2 here]

[Please insert Figure 3 here]

Figure 3 shows the results of MVMMR when there are 100 IVs. When adjusting for the collider (S), the causal estimation is biased whatever IVs are used, and this bias becomes larger as the increasing of other edges' effects. When adjusting for the mediator (M), causal

estimation is unbiased when the IVs are G_2 only, G_1+G_3 or $G_1+G_2+G_3$. The type I error rates of causal estimations are stable around 0.05 under these three kinds of IV selection. The causal estimations under G_1+G_3 or $G_1+G_2+G_3$ have higher power than that under G_2 only when the causal effect is 0.1. G_3 are terrible IVs with large biased estimations whatever variables adjusting for. When adjusting for the measured confounder (C), G_1 only, G_2 only, G_1+G_2 and $G_1+G_2+G_3$ are good choices for IVs in MVMR with unbiased estimations. When G_2 only and G_1+G_2 are IVs, the type I error rates of causal estimations are stable around 0.05, whereas when G_1 only and $G_1+G_2+G_3$ are IVs, the type I error rates are a little bit inflated. The power of causal estimation is high whatever kinds of IVs are selected except G_2 only. The simulation results of 6, 20 and 60 IVs are shown in the Supplementary Fig.1-4. In practice, practitioners always don't know the roles of the adjusting variables. Consider the above three graphs together, $G_1+G_2+G_3$ is the best choice of IVs when performing MVMR.

For the simulation study 2, we conducted a simulation study for continuous and binary phenotypes to learn the structures of random graphs, respectively. We generated the random graphs with 5, 10 and 15 nodes, respectively. Considering the different complexity of network, we set the probability of each edge to be present in a graph as 0.2, 0.5 and 0.8. The effect of each edge follows uniform distribution with four settings: $U(0,0.25)$, $U(0.25,0.5)$, $U(0.5,0.75)$ and $U(0.75,1)$. We varied across the number of SNPs g_1 and g_2 as 5, 10, 20, 30, 40 and 50, respectively. We compared MRSL with eight published methods: BIMMER [28], cGAUGE based on IVW, MR Egger and MR PRESSO [25], HC algorithm incorporating genetic anchors [23] (based on genetic risk score or the most significant SNP) and MRPC algorithm [24] (based on genetic risk score or the most significant SNP). Details of data generation are shown in Methods section.

[Please insert Figure 4 here]

[Please insert Figure 5 here]

[Please insert Table 1 here]

Simulation results of 10 continuous nodes are shown in Figure 4-5 and Table 1. Figure 4 demonstrates the F1 score with different edges' effects and complexity of network. Figure 5 shows the mean of precision and recall when there are 20 IVs. Results of precision and recall when there are 5, 10, 30, 40, 50 IVs are shown in Supplementary Fig. 5-8. When the network

is simple (prob=0.2), F1 score of MRSL is the highest and the performance of three adjustment categories are similar. As the network become more complex, the F1 score of MRSL when adjusting for all nodes on the open paths and minimum separated set is decreasing, whereas MRSL when adjusting for $V \setminus \{X_p, X_q, S^d \text{ and } U\}$ still has the highest F1 score. The recall of the former is smaller than the latter as the edges' effects and the complexity of graph rising. When the edges' effects are small, F1 score of MRSL rises as the number of IVs increasing. When the edges' effects are large, F1 score of MRSL decreases as the number of IVs increasing due to the reduction of precision. This may because in simulation study 1, as the number of IVs increasing, the type I error rate of $G_1 + G_2 + G_3$ is even more inflated and lead to the increase of false negative rate. Besides, the power of causal estimation using MVMR is decreasing as other edges' effects increasing. In addition, the number of adjusting variables is increasing as the network become more complex, then the accuracy of causal estimation using MVMR is reducing. Table 1 shows the computing time of MRSL and other eight methods when there are 5, 20 and 50 IVs. MRSL has the fastest computing time among these methods. Computing time of all the methods with 10, 30 and 40 IVs are listed in Supplementary Table 1. The results of MRSL with 10 binary nodes are similar as that with continuous nodes (Supplementary Fig. 9-13 and Supplementary Table 2). As the number of nodes increasing in the network, the F1 score of MRSL is reducing especially when the network is complex. Results of 5 and 15 nodes are shown in Supplementary Fig. 14-33 and Supplementary Table 3-6.

[Please insert Figure 6 here]

[Please insert Figure 7 here]

For the simulation study 3, in order to evaluate the performance of MRSL in practical application, we choose three fixed networks (Figure 6), which are representative examples in practice, including MAGIC-NIAB [35], ASIA (lung cancer network) [36] and Healthcare Cost [2], with continuous, binary and mixed nodes, respectively. Details of these three networks are illustrated in the Method section. Figure 7 shows the F1 scores of MRSL and eight methods when learning three networks. MRSL has the best performance among all methods. The performances of ASIA (binary) and MAGIC-NIAB (continuous) are similar as that in simulation study 2. For the mixed variables network Healthcare Cost, F1 score is lower than that of ASIA and MAGIC-NIAB. And when the edges' effects are larger, MRSL when adjusting for adjusting for all nodes on the open paths and minimum separated set is a little bit

larger F1 score than MRSL when adjusting for $V \setminus \{X_p, X_q, S^d \text{ and } U\}$ because of higher precision. The precision and recall are shown in Supplementary Fig. 34-45.

Applied example: network of 44 diseases and 26 biomarkers. We applied MRSL to learn the network of 44 diseases with ICD 10 coding and 26 biomarkers using GWAS summary data in UK Biobank. The list of these 70 traits are shown in the Supplementary Table 7. Figure 8 A) shows the marginal causal graph by MRSL step 1, resulting in 70 nodes and 388 edges. Figure 8 B) shows the conditional causal graph obtained by MVMR adjusting for $V \setminus \{X_p, X_q, S^d \text{ and } U\}$, resulting in 69 nodes and 192 edges. This result was obtained by removing 196 direct edges induced by mediation pathways after bonferroni correction. Figure 8 C) shows the causal mediation pathways from biomarkers on each diseases. Vitamin D, Total protein, Urate and Urea are root causes for nearly all the mediation pathways of diseases [37-39]

[Please insert Figure 8 here]

Most of causal links are expected and have clear interpretation of biological pathways or have been confirmed by experiments. Such as B37 Candidiasis, Vitamin D [40], K40 Inguinal hernia [41] and G81 Hemiplegia [42] are direct risk factors; Phosphate [43] and Glycated haemoglobin [44] are direct protective factors. For F33 Recurrent depressive disorder, Testosterone have a positive effect on F33 [45]. Other biomarkers affect F33 through F43 Reaction to severe stress, and adjustment disorders. IGF-1 [46] directly influence the risk of D04 Carcinoma in situ of skin. Vitamin D directly affect G81 Hemiplegia with a protective effect [47]. Glucose [48] and Urate [49] are risk factors of K74 Fibrosis and cirrhosis of liver. Biomarkers have causal effects on K90 Intestinal malabsorption through R14 Flatulence [50] and related conditions and L43 Lichen planus [51].

Several novel causal links are founded and supported by clinical case report or observational studies. For example, for C16 Malignant neoplasm of stomach, Urate [52], T17 Foreign body in respiratory tract [53-54], F31 Bipolar affective disorder [55] and K41 Femoral hernia [56] have negative causal effects on the Malignant neoplasm of stomach. Urate [57] positively affects the risk of Carcinoma in situ of skin. IGF-1 [58-59] and K07 Dentofacial anomalies [including malocclusion] directly increase the risk of G81 Hemiplegia. For H60 Otitis externa, Glycated haemoglobin [60] is protective factor, J03 Acute tonsillitis is a risk factor of Otitis externa. For J03 Acute tonsillitis, IGF-1 [61], HDL cholesterol [62], total protein [63] and total bilirubin [64] positively affect J03 Acute tonsillitis. Urate [65] is a

risk factor for K12 Stomatitis and related lesions. Urate negatively affects the risk of M81 Osteoporosis without pathological fracture [66] whereas R25 Abnormal involuntary movements is a risk factor [67].

Discussion

In this work, we present a novel algorithm called MRSL based on UVMR and MVMR for structural learning. Our method is flexible as it requires only summarized genetic data. Besides, MRSL relaxes the causal sufficiency assumption and can be implemented with fast computing speed and outputs a conditional causal graph with directed causal effects. MRSL consists of four steps: step 1 outputs a marginal causal graph using bi-directional MR in pairs, step 2 find the topological sorting of marginal causal graph, step 3 outputs a conditional causal graph by MVMR and step 4 is an iteration process of step 2 and 3. The marginal causal graph reveals the total causal relationships of each pair of variables and the conditional causal graph identifies the mediation pathways then discloses the direct effects of each pair of variables. The application to 26 biomarkers and 44 ICD10-defined diseases cover lots of expected causal links which have biological interpretations and several new links supported by clinical case reports or previous observational literatures.

The core of MRSL is MR analysis, thus how well MRSL performs depends on the performance of MR decides. The first point we need to focus on is the selection of IVs. For the bi-directional MR, an alternative choice is the SNPs only associated with the exposure but not associated with other variables in the network. To some extent, this can block nearly all pleiotropic pathways. For MVMR, we firstly conducted a simulation study to choose the most valid IVs. Here we should consider the valid IVs which makes the MVMR perform best when adjusting for the collider, mediator and confounder simultaneously. From the results of simulation 1, $G_1 + G_2 + G_3$ is the best choice. This is supported by previous literature [68-69]. It is necessary to have as many IVs as possible, that is, including genetic variables that are associated with at least one exposure and removing the instruments that only strong with one exposure will lead to a loss of precision in the estimation or other potential bias. We only focus on the effect of particular exposure on the outcome using MVMR each time, thus we only force positive association with respect to the exposure we are interested in [70]. It has no influence on our results although this may change the sign of the association with respect to the adjusting variables. In this work, we use univariable and multivariable IVW as the main methods. MRSL can be extend to use other UVMR methods, such as pleiotropy-robust

methods (e.g. MR-Egger [71], the weighted median method [72], the mode-based estimate method [73], MR-RAPS [74] and contamination mixture method [75], etc), and MVMR methods (e.g. MVMR-Egger, MVMR-Robust, MVMR-Median, and MVMR-Lasso [70], etc) instead of IVW. However, combination of these methods in MRSL is a time consuming process and may lose precision due to the low accuracy of methods themselves.

In the second step of MRSL, we present three strategies for adjusting variables in MVMR with the complement of graph theory in causal inference. Because MR overcome the influence of unobserved confounding, we exclude U in the three sets of adjusting variables. Another point we pay attention to is that whether these three sets of adjusting variables are the same in the marginal causal graph and the true causal graph, or in other words, we have two questions: does adjusting these variables in the marginal causal graph unlock the blocked pathways in the true causal graph? or not completely block the mediation pathways in the true causal graph? For these two questions, we proposes Lemma 1-3 and Theorem 2. The first way is adjusting for all nodes on the open paths in the marginal causal graph. This enables all the open paths between two variables can be blocked, whether mediation pathways or confounding pathways. This adjusting set doesn't include the spurious colliders in the marginal causal graph. For the second way, minimal separating set may include spurious colliders at the cost of include other confounders or mediators to ensure the separation of two variables. This not only block the pathways in the true causal graph, but also block all the pathways include spurious pathways in the marginal causal graph. The third adjusting set is the most conservative set, that is, adjusting for all the variables excluding colliders, which are particular colliders that must have direct edges on the two variables we are interested in. In all, the second step of MRSL is a process to remove extra edges in marginal causal graph, and obtain a conditional causal graph.

Combination of graph theory and MVMR is a unique property of our algorithm, and we utilize this novel property into causal discovery to improve the precision and recall. Our method can be easily implemented only using GWAS summary data, which is public available for the most phenotypes as the emergence of a large number GWAS studies with huge sample size. Published MR-based algorithm such as cGAUGE, requires the individual-level data and are thus not as easily available as the GWAS summary statistics, and is very time consuming [25]; BIMMER are implemented based on the complex inverse sparse regression and obtained an approximately estimation of DCE matrix, this require time roughly $\mathcal{O}(\kappa d^4)$ for d phenotypes [28]. In the result of simulation study 2-3, we found that the computing time of

MRSL is only around 1/100 of BIMMER, and 1/1000 of cGAUGE, respectively. MRSL has two-fold higher F1 score than other eight methods when the network is simple. Also, MRSL outputs the unbiased direct effect of each pair of variables. Moreover, MRSL can be applied into the structure with feedback loops between any two variables, because our main MR method IVW can powerfully deal with the case of bi-directional causal relationship between two variables [28,76]. Similar to MR analysis, GWAS summary data of d phenotypes should from the homogenous population. We also need to pay more attention to other issues, such as measurement error, selection bias and missing data, etc, in the future.

In conclusion, we proposed a novel algorithm, utilizing the combination of graph theory and MR into causal discovery to learn the conditional causal graph. We look forward to offer constructive suggestions for disease diagnostic and apply our method beyond the scope considered here.

Methods

MRSL. We consider an algorithm MRSL for structural learning based on summarized genetic data. An illustration diagram of MRSL is displayed in Figure 1. Assume a DAG $\mathcal{G}=\langle V, E \rangle$ with unobserved confounders, where V is a set of nodes and E is a set of pairs of nodes. Assume we are interested in d phenotypes $\{X_1, X_2, \dots, X_d\}$. U represents unobserved confounders among d phenotypes. $E_{\mathcal{G}}$ and $E_{\mathcal{G}_M}$ denote all the pairs of nodes for directed edges in the true causal graph \mathcal{G} and the marginal causal graph \mathcal{G}_M , respectively. $S_{\mathcal{G}}$ and $S_{\mathcal{G}_M}$ denote all the colliders in the true causal graph \mathcal{G} and the marginal causal graph \mathcal{G}_M , respectively. For convenience, the unobserved confounders among phenotypes in Figure 1 are omitted. GWAS summary data for these d phenotypes are available. Generally, for a continuous phenotype, beta coefficient and its standard error can be obtained from linear regression; for a binary phenotype, log(OR) coefficient and its standard error can be obtained from logistic regression. MRSL initializes with an empty graph. The first step of MRSL is to obtain a marginal causal graph, denoted by \mathcal{G}_M , using bi-directional MR in pairs. Here, we choose the univariable inverse-variance weighted (IVW) method as the main method because of its high accuracy.

Assumption 1. For an exposure and an outcome, the independent IVs in UVMR must satisfy (**Relevance**) IVs are strongly associated with the exposure; (**Exchangeability**) IVs are independent with unobserved confounders between the exposure and the outcome; (**Exclusion restriction**) IVs affect the outcome only through the exposure.

For a pair of phenotypes X_p and X_q , we firstly focus on the causal effect of X_p on

X_q . We select J_p IVs for X_p to perform weighted regression:

$$\hat{\beta}_{X_{qj}} = b_{p \rightarrow q} \hat{\beta}_{X_{pj}} + \varepsilon_{X_{qj}}, \varepsilon_{X_{qj}} \sim N(0, se(\hat{\beta}_{X_{qj}})^2), j=1, \dots, J_p, \quad (1)$$

where $\hat{\beta}_{X_{pj}}$ and $\hat{\beta}_{X_{qj}}$ are genetic associations for X_p and X_q on j -th IVs in linear regression in GWAS, respectively. $se(\hat{\beta}_{X_{qj}})$ is the standard error of $\hat{\beta}_{X_{qj}}$. J_p IVs for X_p must satisfy the Assumption 1. For the reverse direction, we select J_q IVs for X_q to perform weighted regression:

$$\hat{\beta}_{X_{pj}} = b_{q \rightarrow p} \hat{\beta}_{X_{qj}} + \varepsilon_{X_{pj}}, \varepsilon_{X_{pj}} \sim N(0, se(\hat{\beta}_{X_{pj}})^2), j=1, \dots, J_q. \quad (2)$$

$b_{p \rightarrow q}$ or $b_{q \rightarrow p}$ represents the total effect of X_p on X_q or X_q on X_p , respectively. Similarly, J_q IVs for X_q must satisfy the Assumption 1. Wald tests for the total effect estimations can be used to test whether there are causal pathways from X_p to X_q or X_q to X_p , respectively.

Assumption 2 (Causal Markov condition). Each variable is independent of its non-descendants given its parents in graph \mathcal{G} .

Assumption 3 (Faithfulness assumption). All independencies embedded in the observed distribution \mathbb{P} are stable and are invariant to changes in parameterization. Thus, it implies (together with d-separation) that $(X \perp Y | Z)_{\mathbb{P}} \Leftrightarrow (X \perp Y | Z)_{\mathcal{G}}$.

Lemma 1. For the true causal graph \mathcal{G} and the marginal causal graph \mathcal{G}_M , $E_{\mathcal{G}} \subseteq E_{\mathcal{G}_M}$ and $S_{\mathcal{G}} \subseteq S_{\mathcal{G}_M}$.

For the second step, we find the topological sorting $T_{\mathcal{G}_M}$ in marginal causal graph \mathcal{G}_M using DFS [33-34]. Topological sorting for a DAG is a linear ordering of vertices such that for every directed edge $X_p \rightarrow X_q$, vertex X_p comes before X_q in the ordering. This ensures that parent nodes will be ordered before their child nodes, and honors the forward direction of edges in the ordering. The DFS algorithm loops through each node of the graph, in an arbitrary order. DFS terminates when it hits any node that has already been visited since the beginning of the topological sort or the node has no outgoing edges. Each node X gets prepended to the output list only after considering all other nodes which depend on X (all descendants of X in the graph).

Lemma 2 (Topological sorting invariance). The topological sorting of the true causal graph \mathcal{G} and the marginal causal graph \mathcal{G}_M are the same $T_{\mathcal{G}} = T_{\mathcal{G}_M}$.

In the third step, MVMR is performed to remove $E_{\mathcal{G}_M} \setminus E_{\mathcal{G}}$ in \mathcal{G}_M , and obtain the true causal graph. For each edge in \mathcal{G}_M (e.g. $X_p \rightarrow X_q$), we detect whether this edge exists after adjusting for the genetic associations with the phenotypes in a sufficient separating set using MVMR.

Assumption 4. For an exposure, a set of covariates and an outcome, the independent IVs in MVMR must satisfy (**Relevance**) IVs are strongly associated with the exposure and covariates; (**Exchangeability**) IVs are independent with any unobserved confounders among the exposure, adjusting variables and the outcome; (**Exclusion restriction**) IVs affect the outcome only through the exposure or covariates.

For an edge $X_p \rightarrow X_q$ in \mathcal{G}_M , $SSS_{X_p \rightarrow X_q} = \{ADJ_{X_p \rightarrow X_q}, U_{pq}\}$ denotes the sufficient separating set. Then we can perform multivariable IVW by following weighted regression of X_q on X_p adjustment for $ADJ_{X_p \rightarrow X_q}$:

$$\beta_{X_{qj}} = a_{p \rightarrow q} \beta_{X_{pj}} + a_{ADJ} \beta_{X_{ADJ_{X_p \rightarrow X_q} j}} + \varepsilon_{X_{qj}}, \varepsilon_{X_{qj}} \sim N(0, se(\beta_{X_{qj}})^2), \quad (3)$$

where $a_{p \rightarrow q}$ is the direct effect of X_p on X_q not through mediators and confounders in $ADJ_{X_p \rightarrow X_q}$. The Wald test for estimation of $a_{p \rightarrow q}$ can be used to test whether there is a direct edge from X_p to X_q . IVs for the regression (3) above must satisfy the Assumption 4. This kind of IV can be obtained by statistical filtering criteria (see application examples) or ImpIV filter and ExSep test [25].

For MVMR, the sufficient adjustment set $ADJ_{X_p \rightarrow X_q}$ can be made up of three ways: (1) all nodes on the open paths from X_p to X_q ; (2) minimal sufficient adjustment set for confounders and all the mediators from X_p to X_q ; (3) $V \setminus \{X_p, X_q\}$ and S^d . S^d refers to a set of colliders where the two interested nodes have direct edges on them, e.g. for two nodes X_p and X_q , the collider S_1 in $X_p \rightarrow S_1 \leftarrow X_q$ is included in S^d but the collider S_2 in $X_p \rightarrow S_2 \leftarrow C \rightarrow X_q$ is not included in S^d . S^d in the graph \mathcal{G}_M includes the colliders and the nodes not on the pathway from X_p to X_q , but not includes any mediators, confounders or the nodes which are both mediators and confounders on the pathways from X_p to X_q in the true causal graph \mathcal{G} . Note that one node can be a mediator and a collider simultaneously. For example, a graph \mathcal{D} with edges $X_1 \rightarrow X_3, X_3 \rightarrow X_2, X_4 \rightarrow X_2$ and $X_4 \rightarrow X_3$, X_3 is a mediator on the path $X_1 \rightarrow X_3 \rightarrow X_2$ and also a collider on the path $X_1 \rightarrow X_3 \leftarrow X_4$. This

kind of node is included in the sufficient separating set $ADJ_{X_p \rightarrow X_q}^*$ by including other nodes (e.g. X_4) to ensure X_p and X_q are sufficiently separated.

We use $ADJ_{X_p \rightarrow X_q}^{(i)}$ ($i=1,2,3$) to denote that adjusting variables $ADJ_{X_p \rightarrow X_q}$ obtained by the i -th way. The order of MVMR varying across each edge is the same as the topological sorting in step 2. The topological sorting avoids the case in Supplementary Fig. 46, in which we show the MVMR adjusting for the genetic associations with the phenotypes in $ADJ_{X_p \rightarrow X_q}^{(3)}$ with and without topological sorting. For the latter, when we perform MVMR to test whether the edge $X_4 \rightarrow X_1$ exists, the genetic association with X_3 is adjusted in MVMR. However, in A3, X_3 is a collider but is not included in S^d , and X_2 is included in S^d . Then we perform MVMR $\beta_{X_{1j}} \sim \alpha_{41} \beta_{X_{4j}} + \alpha_{31} \beta_{X_{3j}}$ and the estimation for direct effect of X_4 on X_1 is biased because X_3 is a collider. The bias formula of causal estimation when adjusting for a collider using MVMR are shown in Supplementary Notes. For the former, this kind of problem can be avoided and the process of removing edges is more accurate and faster after topological sorting.

Theorem 1: Under the Assumptions 2-4, for each edge $X_p \rightarrow X_q$ in the marginal causal graph \mathcal{G}_M , given a sufficient separating set $SSS_{X_p \rightarrow X_q} = \{ADJ_{X_p \rightarrow X_q}, U_{pq}\}$ such that $X_p \perp X_q | SSS_{X_p \rightarrow X_q}$, which can be tested by adjusting for genetic associations with $ADJ_{X_p \rightarrow X_q}$ using MVMR,

$$\beta_{X_{qj}} = a_{p \rightarrow q} \beta_{X_{pj}} + a_{ADJ} \beta_{X_{ADJ_{X_p \rightarrow X_q} j}} + \varepsilon_{X_{qj}}, \varepsilon_{X_{qj}} \sim N(0, se(\beta_{X_{qj}})^2),$$

where $a_{p \rightarrow q}$ determines the existence of edge from X_p to X_q in the true causal graph \mathcal{G} , then there is no direct edge from X_p to X_q in the true causal graph \mathcal{G} .

Based on Theorem 1, the edges $E_{\mathcal{G}_M} \setminus E_{\mathcal{G}}$ in the graph \mathcal{G}_M are removed. After the third step, we add an iteration step to perform step 2 and 3 again, using the graph obtained in the previous step 3, until the graph is convergence. The aim of this step is to eliminate the random error and statistical test error in the UVMR and MVMR and increase the precision of MRSL.

Simulation. We conduct three simulation studies to evaluate the performance of MRSL. Firstly, we conduct a simulation study to choose the optimal selection strategy of IVs for

MVMR. Secondly, we evaluate the performance of MRSL and other eight published methods for structure discovery, when learning the structure of random graph with continuous and binary nodes, respectively. Thirdly, we select three representative graphs from previous literatures and access the performance of MRSL and other eight published methods.

Simulation study 1 on IVs selection in MVMR. The basis of MRSL is MR, thus it is vital to select valid IVs. We firstly conduct a simulation study to evaluate the performance of MVMR when estimate the causal effect of an exposure (X) on an outcome (Y). We consider three roles of adjusting variables in MVMR: a collider (S), a mediator (M) or a measured confounder (C) in the causal pathway from X to Y (Figure 2 A-C). Based on the three figures, there are seven kinds of available SNPs as IVs: (1) G_1 : SNPs only associated with X; (2) G_2 : SNPs associated with X and adjusting variable; (3) G_3 : SNPs only associated with the adjusting variable; (4) $G_1 + G_3$; (5) $G_1 + G_2$; (6) $G_2 + G_3$; (7) $G_1 + G_2 + G_3$. When the adjusting variable is a confounder, G_3 is also associated with X, and it may be selected as instrumental variable because practitioners don't know the true roles of the adjusting variables. The process of data generation is shown in Supplementary Notes. We generate 10,000 individuals and 1,000 repeated datasets. To access the performance of MVMR, we plot a boxplot to evaluate the estimation of causal effect of X on Y, and calculate the type I error rate for null causal effect and statistical power to detect the non-zero causal effect. The nominal level is set to 0.05.

Simulation study 2 on MRSL with random graphs. To valid the utility of the MRSL method for learning structures, we conduct a simulation study for continuous and binary variables, respectively. Genetic IVs are generated from binomial distribution $B(2,0.3)$. Let Y denote the $N \times d$ matrix of d phenotypes and G denote a $N \times J$ matrix of J SNPs. For continuous variables, d phenotypes are generated from the following model

$$Y = P_Y \beta + G \alpha + U + \varepsilon ,$$

where P_Y represents the parent nodes of Y , β are the effects of P_Y on Y and generated from uniform distribution, α is a $d \times J$ matrix of effects of SNPs on phenotypes, U represents the confounding factors among d phenotypes and ε is the residual term following normal distribution $N(0,1)$. For binary variables, d nodes are generated from the following model

$$\log it[P(Y = 1)] = P_Y \beta + G \alpha + U .$$

We generate the random graphs with 5, 10 and 15 nodes, respectively. Considering the different complexity of network, we set the probability of each edge to be present in a graph as 0.2, 0.5 and 0.8. In practice, there may be effects of different magnitude between traits, thus we consider β follows uniform distribution with four parameter settings: U(0,0.25), U(0.25,0.5), U(0.5,0.75) and U(0.75,1) for continuous nodes, odd ratio (OR) U(1,1.5), U(1.5,2), U(2,2.5) and U(2.5,3) for binary nodes. The IVs are assumed uncorrelated, and subdivided into two categories: (1) g_1 SNPs that only predict one phenotype; (2) g_2 SNPs that predict all the phenotypes simultaneously. The variance of each phenotype explained by all the SNPs is around 10%. We vary across the number of SNPs g_1 and g_2 with 5, 10, 20, 30, 40 and 50, respectively.

We compared our method with eight published methods: BIMMER [28], cGAUGE based on IVW, MR Egger and MR PRESSO [25], HC algorithm incorporating genetic anchors [23] (based on genetic risk score or the most significant SNP) and MRPC algorithm [24] (based on genetic risk score or the most significant SNP). BIMMER can be implemented using GWAS summary data whereas other seven methods need individual genetic and phenotypic data. To access the performance of algorithm, we calculate the mean of F1 score, recall, precision and computing time across 100 data sets with 10,000 individuals for each method. Recall (i.e., power, or sensitivity) measures how many edges from the true causal graph a method can recover, whereas precision (i.e., 1-FDR) measures how many correct edges are recovered in the inferred graph, and F1 score is a combined index of recall and precision. Details of calculation formula are shown in Supplementary Notes.

Simulation study 3 on MRS� with fixed graphs. In order to evaluate the performance of MRS� in practical application, we choose three representative examples (Figure 3 A-C): A) MAGIC-NIAB, a network based on the Multiparent Advanced Generation Inter-Cross (MAGIC) winter wheat population produced by the UK National Institute of Agricultural Botany (NIAB), seven continuous traits were measured: yield (YLD), flowering time (FT), height (HT), yellow rust in the glasshouse (YR.GLASS) and in the field (YR.FIELD), Fusarium (FUS), and mildew (MIL). Such a scheme is designed to produce a mapping population from several generations of intercrossing among eight founders and has the potential to improve quantitative trait loci (QTL) mapping precision. B) ASIA, also called lung cancer network, consists of eight binary variables. Lauritzen and Spiegelhalter (1988) [36] motivate this example as follows: “Shortness-of-breath (dyspnoea) may be due to tuberculosis, lung cancer or bronchitis, or none of them, or more than one of them. A recent

visit to Asia increases the chances of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis. The results of a single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnea.” C) Healthcare Cost, is a recurring theme in most countries’ public discourse due to the combination of an ageing population and the availability of more advanced (read, expensive) treatments. We use the simple example in Marco et al. (2021) [77], which modeled an individual’s yearly medical expenditure by seven mixed variables: age, pre-existing conditions, outpatient expenditure, inpatient expenditure, any hospital stay, days of hospital stay and taxes.

The data generation process and parameters’ settings of these three networks are similar with that in simulation study 2. We also use F1 score, recall, precision and computing time to evaluate the performance of MRSL.

Applied example. We apply MRSL to learn the network of 26 biomarkers and 44 ICD10-defined diseases using GWAS summary data in UK Biobank. For quantitative phenotypes, we use the phenotypes that have been inverse rank normalized.

For MRSL, we first clumped the UK Biobank summary statistics to $p < 5 \times 10^{-8}$ for 26 biomarkers and 44 diseases, with $r^2 < 0.0001$ and distance 10000 kilobases using the European reference panel in mrbase (<https://www.mrbase.org/>). To avoid the selection bias, we choose the IVs in the male population and use the summarized statistics in female population. For step 1 in Figure 1, in order to obtain a marginal causal graph using bi-directional MR, for each MR analysis, we select the SNPs associated with the exposure but not associated with other phenotypes (except exposure and outcome) as IVs. For example, if we perform MVMR $X_3 \sim X_1 + X_4$, SNPs associated with X_1 but not associated with X_4 are selected as IVs. Next we perform MVMR using three adjustment strategies to obtain the true graph. We select SNPs associated with at least one phenotype of exposure and the adjusting variables as IVs. For example, if we perform MVMR $X_3 \sim X_1 + X_4$, SNPs associated with at least one of X_1 , X_3 and X_4 but not associated with X_2 are selected as IVs. For each MVMR, we also need to filter out the SNPs with linkage disequilibrium ($r^2 > 0.0001$).

Data and code availability

The GWAS summary data in UK Biobank are publicly available at <http://www.nealelab.is/uk-biobank>. All the analysis in our article were implemented by R software. MRSL can be implemented by <https://github.com/hhoulei/MRSL>. BIMMER were implemented using R packages *bimmer*. MRPC were implemented using R packages *MRPC*. HC algorithm were implemented using R packages *bnlearn*. cGAUGE were implemented using functions in <https://github.com/david-dd-amar/cGAUGE> and R packages *MendelianRandomization*, *MRPRESSO*. All the networks were plotted using R packages *igraph*.

Conflicts of Interest

None declared

Ethics approval and consent to participate

Ethical approval was not sought, because this study involved analysis of publicly available summary-level data from GWASs, and no individual-level data were used.

Source of Funding

FX was supported by the National Natural Science Foundation of China (Grant 82173625) and the Shandong Provincial Key Research and Development project (2018CXGC1210). HL was supported by the National Natural Science Foundation of China (Grant 82003557).

Authors' contributions

HL and FX conceived the study. LH contributed to theoretical derivation with assistance from HL and ZG. LH contributed to the data simulation. LH and CW contributed to the application. LH, ZG, XS and HL wrote the manuscript with input from all other authors. All authors reviewed and approved the final manuscript.

Acknowledgements

None.

Reference

- [1]. Pearl, J. *Causality: Models, Reasoning and Inference* (Cambridge University Press, 2009).
- [2]. Spirtes, P., Glymour, C. & Scheines, R. *Causation, Prediction, and Search* 2nd edn, Vol. 1 (The MIT Press, 2001).
- [3]. Isci, S., H. Dogan, C. Ozturk, and H. H. Otu. 2013. Bayesian network prior: Network analysis of biological data using external knowledge. *Bioinformatics* 30 (6):860–67. doi:10.1093/bioinformatics/btt660.
- [4]. Friedman, N., M. Linial, I. Nachman, and D. Pe'er. 2000. Using Bayesian networks to analyze expression data. *Journal of Computational Biology* 7 (3–4):601–20. doi:10.1089/106652700750050961.
- [5]. Zagorecki, A., P. Orzechowski, and K. Holownia. 2013. A system for automated general medical diagnosis using Bayesian networks. *MedInfo* 192:461–65.
- [6]. Suchánek, P., F. Marecki, and R. Bucki. 2014. Self-learning Bayesian networks in diagnosis. *Procedia Computer Science* 35:1426–35. doi:10.1016/j.procs.2014.08.200.
- [7]. Druzdzel, M. J. (2009). The role of assumptions in causal discovery.
- [8]. Spirtes P. Introduction to Causal Inference. *Journal of Machine Learning Research*. 2010; 11:1643–1662.
- [9]. Scutari M, Denis JB. *Bayesian Networks with Examples in R*. Texts in Statistical Science, Chapman & Hall/CRC (US); 2014.
- [10]. Colombo D, Maathuis MH (2014). "Order-Independent Constraint-Based Causal Structure Learning". *Journal of Machine Learning Research*, 15:3921–3962.
- [11]. Margaritis D (2003). *Learning Bayesian Network Model Structure from Data*. Ph.D. thesis, School of Computer Science, Carnegie-Mellon University, Pittsburgh, PA.
- [12]. Tsamardinos I, Aliferis CF, Statnikov A (2003). "Algorithms for Large Scale Markov Blanket Discovery". *Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference*, 376–381.
- [13]. Yaramakala S, Margaritis D (2005). "Speculative Markov Blanket Discovery for Optimal Feature Selection". *Proceedings of the Fifth IEEE International Conference on Data Mining*, 809–812.
- [14]. Tsamardinos, I., Brown, L. E., & Aliferis, C. F. (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Machine learning*, 65(1), 31-78.
- [15]. Prosser, P. (1993). Hybrid algorithms for the constraint satisfaction problem.

Computational intelligence, 9(3), 268-299.

[16]. Tsamardinos I, Brown LE, Aliferis CF (2006). "The Max-Min Hill-Climbing Bayesian Network Structure Learning Algorithm". *Machine Learning*, 65(1):31–78.

[17]. Friedman N, Nachman I, Pe'er D (1999). "Learning Bayesian Network Structure from Massive Datasets: the Sparse Candidate Algorithm." *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, 206–215.

[18]. Spirtes, P., Glymour, C., Scheines, R., Kauffman, S., Aimale, V., & Wimberly, F. (2000). Constructing Bayesian network models of gene expression networks from microarray data.

[19]. Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in genetics*, 10, 524.

[20]. Ogarrio, J. M., Spirtes, P., & Ramsey, J. (2016, August). A hybrid causal search algorithm for latent variable models. In *Conference on probabilistic graphical models* (pp. 368-379). PMLR.

[21]. Emdin, C. A., Khera, A. V., & Kathiresan, S. (2017). Mendelian randomization. *Jama*, 318(19), 1925-1926.

[22]. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiology*. 2003; 32:1–22. <https://doi.org/10.1093/ije/dyg070>

[23]. Howey, R., Shin, S. Y., Relton, C., Davey Smith, G., & 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 genetics*, 16(3), e1008198.

[24]. Badsha, M., & Fu, A. Q. (2019). Learning causal biological networks with the principle of Mendelian randomization. *Frontiers in genetics*, 10, 460.

[25]. Amar, D., Sinnott-Armstrong, N., Ashley, E. A., & Rivas, M. A. (2021). Graphical analysis for phenome-wide causal discovery in genotyped population-scale biobanks. *Nature communications*, 12(1), 1-11.

[26]. Lin, B. D., Alkema, A., Peters, T., Zinkstok, J., Libuda, L., Hebebrand, J., ... & Luykx, J. J. (2019). Assessing causal links between metabolic traits, inflammation and schizophrenia: a univariable and multivariable, bidirectional Mendelian-randomization study. *International journal of epidemiology*, 48(5), 1505-1514.

[27]. Rees, J. M., Wood, A. M., & Burgess, S. (2017). Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured

- pleiotropy. *Statistics in medicine*, 36(29), 4705-4718.
- [28]. Brown, B. C., & Knowles, D. A. (2020). Phenome-scale causal network discovery with bidirectional mediated Mendelian randomization. *bioRxiv*.
- [29]. Sanderson, E. (2021). Multivariable Mendelian randomization and mediation. *Cold Spring Harbor perspectives in medicine*, 11(2), a038984.
- [30]. Sanderson, E., Davey Smith, G., Windmeijer, F., & Bowden, J. (2019). An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *International journal of epidemiology*, 48(3), 713-727.
- [31]. Zhao, J. V., & Schooling, C. M. (2020). The role of testosterone in chronic kidney disease and kidney function in men and women: a bi-directional Mendelian randomization study in the UK Biobank. *BMC medicine*, 18(1), 1-10.
- [32]. Darrous, L., Mounier, N., & Kutalik, Z. (2021). Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics. *Nature communications*, 12(1), 1-15.
- [33]. Tarjan, Robert E. (1976), "Edge-disjoint spanning trees and depth-first search", *Acta Informatica*, 6 (2): 171–185, doi:10.1007/BF00268499, S2CID 12044793
- [34]. King, D. J., & Launchbury, J. (1995, January). Structuring depth-first search algorithms in Haskell. In *Proceedings of the 22nd ACM SIGPLAN-SIGACT symposium on Principles of programming languages* (pp. 344-354).
- [35]. M. Scutari, P. Howell, D. J. Balding and I. Mackay (2014). Multiple Quantitative Trait Analysis Using Bayesian Networks. *Genetics*, 198(1), 129-137.
- [36]. S. Lauritzen, D. Spiegelhalter. Local Computation with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 50(2):157-224, 1988.
- [37]. Meng, X., Li, X., Timofeeva, M. N., He, Y., Spiliopoulou, A., Wei, W. Q., Gifford, A., Wu, H., Varley, T., Joshi, P., Denny, J. C., Farrington, S. M., Zgaga, L., Dunlop, M. G., McKeigue, P., Campbell, H., & Theodoratou, E. (2019). Phenome-wide Mendelian-randomization study of genetically determined vitamin D on multiple health outcomes using the UK Biobank study. *International journal of epidemiology*, 48(5), 1425–1434. <https://doi.org/10.1093/ije/dyz182>
- [38]. Li, X., Meng, X., Spiliopoulou, A., Timofeeva, M., Wei, W. Q., Gifford, A., Shen, X., He, Y., Varley, T., McKeigue, P., Tzoulaki, I., Wright, A. F., Joshi, P., Denny, J. C., Campbell, H., & Theodoratou, E. (2018). MR-PheWAS: exploring the causal effect of SUA level on multiple

- disease outcomes by using genetic instruments in UK Biobank. *Annals of the rheumatic diseases*, 77(7), 1039–1047. <https://doi.org/10.1136/annrheumdis-2017-212534>
- [39]. Li, X., Meng, X., He, Y., Spiliopoulou, A., Timofeeva, M., Wei, W. Q., Gifford, A., Yang, T., Varley, T., Tzoulaki, I., Joshi, P., Denny, J. C., Mckeigue, P., Campbell, H., & Theodoratou, E. (2019). Genetically determined serum urate levels and cardiovascular and other diseases in UK Biobank cohort: A phenome-wide mendelian randomization study. *PLoS medicine*, 16(10), e1002937. <https://doi.org/10.1371/journal.pmed.1002937>
- [40]. Amegah, A. K., Baffour, F. K., Appiah, A., Adu-Frimpong, E., & Wagner, C. L. (2020). Sunlight exposure, consumption of vitamin D-rich foods and vulvovaginal candidiasis in an African population: a prevalence case-control study. *European journal of clinical nutrition*, 74(3), 518–526. <https://doi.org/10.1038/s41430-019-0517-7>
- [41]. Luhmann, A., & Moses, A. (2015). Successful conservative treatment of a candida albicans intraperitoneal mesh infection following laparoscopic ventral hernia repair. *Hernia : the journal of hernias and abdominal wall surgery*, 19(5), 845–847. <https://doi.org/10.1007/s10029-013-1183-7>
- [42]. Li, C. S., Huang, C. R., Lu, C. H., Lui, C. C., Chien, C. C., & Chang, W. N. (2004). Concomitant stroke and Candida parapsilosis native valve endocarditis: report of one case and literature review. *Acta neurologica Taiwanica*, 13(3), 131–135.
- [43]. Köhler, J. R., Acosta-Zaldívar, M., & Qi, W. (2020). Phosphate in Virulence of Candida albicans and Candida glabrata. *Journal of fungi (Basel, Switzerland)*, 6(2), 40. <https://doi.org/10.3390/jof6020040>
- [44]. Hill, L. V., Tan, M. H., Pereira, L. H., & Embil, J. A. (1989). Association of oral candidiasis with diabetic control. *Journal of clinical pathology*, 42(5), 502–505. <https://doi.org/10.1136/jcp.42.5.502>
- [45]. Walther, A., Breidenstein, J., & Miller, R. (2019). Association of Testosterone Treatment With Alleviation of Depressive Symptoms in Men: A Systematic Review and Meta-analysis. *JAMA psychiatry*, 76(1), 31–40. <https://doi.org/10.1001/jamapsychiatry.2018.2734>
- [46]. Wang, X., Wang, R., Jiang, K., Zhu, M., Guo, L., & Wu, C. (2022). PINCH-1 promotes IGF-1 receptor expression and skin cancer progression through inhibition of the GRB10-NEDD4 complex. *Theranostics*, 12(6), 2613–2630. <https://doi.org/10.7150/thno.70744>
- [47]. Sari, A., Durmus, B., Karaman, C. A., Ogut, E., & Aktas, I. (2018). A randomized, double-blind study to assess if vitamin D treatment affects the outcomes of rehabilitation and

balance in hemiplegic patients. *Journal of physical therapy science*, 30(6), 874–878. <https://doi.org/10.1589/jpts.30.874>

[48]. Abdelkader, R. Y., Abdelrazek, M. A., Attallah, A., Farid, K., & El-Far, M. (2021). High blood glucose levels are associated with fibrosis/cirrhosis progression in chronic hepatitis C. *Journal of immunoassay & immunochemistry*, 42(6), 559–570. <https://doi.org/10.1080/15321819.2021.1911813>

[49]. Sari, D., Soetoko, A. S., Soetoko, A. S., Romi, M. M., Tranggono, U., Setyaningsih, W., & Arfian, N. (2020). Uric acid induces liver fibrosis through activation of inflammatory mediators and proliferating hepatic stellate cell in mice. *The Medical journal of Malaysia*, 75(Suppl 1), 14–18.

[50]. "Malabsorption Syndrome". MedlinePlus. Retrieved 29 April 2018.

[51]. Lauritano, D., Boccalari, E., Di Stasio, D., Della Vella, F., Carinci, F., Lucchese, A., & Petruzzi, M. (2019). Prevalence of Oral Lesions and Correlation with Intestinal Symptoms of Inflammatory Bowel Disease: A Systematic Review. *Diagnostics (Basel, Switzerland)*, 9(3), 77. <https://doi.org/10.3390/diagnostics9030077>

[52]. Yang, S., He, X., Liu, Y., Ding, X., Jiang, H., Tan, Y., & Lu, H. (2019). Prognostic Significance of Serum Uric Acid and Gamma-Glutamyltransferase in Patients with Advanced Gastric Cancer. *Disease markers*, 2019, 1415421. <https://doi.org/10.1155/2019/1415421>

[53]. Garcia, I., Varon, J., & Surani, S. (2016). Airway Complications from an Esophageal Foreign Body. *Case reports in pulmonology*, 2016, 3403952. <https://doi.org/10.1155/2016/3403952>

[54]. Jeon, S. Y., Choe, Y. H., Song, E. K., Yim, C. Y., & Lee, N. R. (2021). Foreign body removal using flexible bronchoscopy in terminal cancer: A case report. *Medicine*, 100(43), e27620. <https://doi.org/10.1097/MD.00000000000027620>

[55]. Chen, M. H., Tsai, S. J., Su, T. P., Li, C. T., Lin, W. C., Cheng, C. M., Chen, T. J., & Bai, Y. M. (2022). Cancer risk in patients with bipolar disorder and unaffected siblings of such patients: A nationwide population-based study. *International journal of cancer*, 150(10), 1579–1586. <https://doi.org/10.1002/ijc.33914>

[56]. Qin, R., Zhang, Q., Weng, J., & Pu, Y. (2014). Incidental finding of a malignant tumour in an inguinal hernia sac. *Contemporary oncology (Poznan, Poland)*, 18(2), 130–133. <https://doi.org/10.5114/wo.2014.42728>

[57]. Yiu, A., Van Hemelrijck, M., Garmo, H., Holmberg, L., Malmström, H., Lambe, M., Hammar, N., Walldius, G., Jungner, I., & Wulaningsih, W. (2017). Circulating uric acid levels

- and subsequent development of cancer in 493,281 individuals: findings from the AMORIS Study. *Oncotarget*, 8(26), 42332–42342. <https://doi.org/10.18632/oncotarget.16198>
- [58]. Shaheen, H., Sobhy, S., El Mously, S., Niazi, M., & Gomaa, M. (2018). Insulin-Like Growth Factor-1 in Acute Ischemic Stroke. *The Egyptian journal of neurology, psychiatry and neurosurgery*, 54(1), 42. <https://doi.org/10.1186/s41983-018-0042-y>
- [59]. Silva-Couto, M., Prado-Medeiros, C. L., Oliveira, A. B., Alcântara, C. C., Guimarães, A. T., Salvini, T., Mattioli, R., & de Russo, T. L. (2014). Muscle atrophy, voluntary activation disturbances, and low serum concentrations of IGF-1 and IGFBP-3 are associated with weakness in people with chronic stroke. *Physical therapy*, 94(7), 957–967. <https://doi.org/10.2522/ptj.20130322>
- [60]. Chin, R., Roche, P., Sigston, E., & Valance, N. (2012). Malignant otitis externa: an Australian case series. *The surgeon : journal of the Royal Colleges of Surgeons of Edinburgh and Ireland*, 10(5), 273–277. <https://doi.org/10.1016/j.surge.2011.09.004>
- [61]. Farmarzi, M., Shishegar, M., Heydari, S. T., Haghghi, A., & Sharouny, H. (2016). Effects of Adenotonsillectomy on Serum Levels of IGF-1 and IGFBP-3 and Growth Indices in Children with Adenotonsillar Hypertrophy or Recurrent Tonsillitis. *Iranian journal of otorhinolaryngology*, 28(88), 329–335.
- [62]. Rader, D. J., & deGoma, E. M. (2012). Approach to the patient with extremely low HDL-cholesterol. *The Journal of clinical endocrinology and metabolism*, 97(10), 3399–3407. <https://doi.org/10.1210/jc.2012-2185>
- [63]. SZPUNAR, J., & RYBAKOWA, M. (1961). Electrophoretic serum studies. Studies of children with frequently recurring acute tonsillitis. *Archives of otolaryngology (Chicago, Ill. : 1960)*, 74, 267–271. <https://doi.org/10.1001/archotol.1961.00740030274007>
- [64]. Bechtel-Grosch, U., Beguelin, C., Berezowska, S., Dufour, J. F., Takala, J., & Schefold, J. C. (2016). Fulminant hepatic and multiple organ failure following acute viral tonsillitis: a case report. *Journal of medical case reports*, 10, 7. <https://doi.org/10.1186/s13256-015-0777-3>
- [65]. Bakhtiari, S., Toosi, P., Samadi, S., & Bakhshi, M. (2017). Assessment of Uric Acid Level in the Saliva of Patients with Oral Lichen Planus. *Medical principles and practice : international journal of the Kuwait University, Health Science Centre*, 26(1), 57–60. <https://doi.org/10.1159/000452133>
- [66]. Yan, D. D., Wang, J., Hou, X. H., Bao, Y. Q., Zhang, Z. L., Hu, C., & Jia, W. P. (2018). Association of serum uric acid levels with osteoporosis and bone turnover markers in a Chinese population. *Acta pharmacologica Sinica*, 39(4), 626–632.

<https://doi.org/10.1038/aps.2017.165>

[67]. Bhatnagar, N., Lingaiah, P., Lodhi, J. S., & Karkhur, Y. (2017). Pathological Fracture of Femoral Neck Leading to a Diagnosis of Wilson's Disease: A Case Report and Review of Literature. *Journal of bone metabolism*, 24(2), 135–139.

<https://doi.org/10.11005/jbm.2017.24.2.135>

[68]. Sanderson, E., Davey Smith, G., Windmeijer, F., & Bowden, J. (2019). An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *International journal of epidemiology*, 48(3), 713-727.

[69]. Zuber, V., Colijn, J. M., Klaver, C., & Burgess, S. (2020). Selecting likely causal risk factors from high-throughput experiments using multivariable Mendelian randomization. *Nature communications*, 11(1), 1-11.

[70]. Grant, A. J., & Burgess, S. (2021). Pleiotropy robust methods for multivariable Mendelian randomization. *Statistics in medicine*, 40(26), 5813-5830.

[71]. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015; 44: 512–525. <https://doi.org/10.1093/ije/dyv080> PMID: 26050253

[72]. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016; 40: 304– 314. <https://doi.org/10.1002/gepi.21965> PMID: 27061298

[73]. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. 2017; 46:1985–1998. <https://doi.org/10.1093/ije/dyx102> PMID: 29040600

[74]. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data mendelian randomization using robust adjusted profile score. *Ann Stat*. 2020; 48: 1742–1769.

[75]. Burgess, S., Foley, C. N., Allara, E., Staley, J. R., & Howson, J. M. (2020). A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nature communications*, 11(1), 1-11.

[76]. Sun, D., Zhou, T., Heianza, Y., Li, X., Fan, M., Fonseca, V. A., & Qi, L. (2019). Type 2 diabetes and hypertension: a study on bidirectional causality. *Circulation research*, 124(6), 930-937.

[77]. Scutari, M., & Denis, J. B. (2021). *Bayesian networks: with examples in R*. Chapman and Hall/CRC.

Figure legends

Figure 1. Workflow and the motivating example of MRS� algorithm. Confounders of d phenotypes are omitted. The input data is GWAS summary data for each phenotype. Initialization with an empty graph. For the Step 1, a marginal causal graph can be obtained using bi-directional MR in pairs. For the Step 2, the topological sorting of marginal causal graph should be found using Depth First Search (DFS). For the Step 3, MVMR is performed to remove extra edges in the marginal causal graph by adjusting for the genetic associations with phenotypes of three strategies of sufficient separating sets. The in the Step 4, iteration for step 2 and step 3 is performed until the graph converges. Finally MRS� outputs a conditional causal graph. (A-J) Motivating example with five nodes. (A) The true causal graph. (B) Marginal causal graph obtained by step 1. (C-J) Perform MVMR for each edge in graph (F) based on its topological sorting. Blue nodes denote the exposure and outcome we are interested in. MRS� outputs the graph (J). $SSS_{X_p \rightarrow X_q} = \{ADJ_{X_p \rightarrow X_q}, U_{pq}\}$ denotes the sufficient separating set from X_p to X_q . $ADJ_{X_p \rightarrow X_q}^{(1)}$ includes all nodes on the open paths from X_p to X_q ; (2) $ADJ_{X_p \rightarrow X_q}^{(2)}$ includes the elements in the minimal sufficient adjustment set and all the mediators from X_p to X_q ; (3) $ADJ_{X_p \rightarrow X_q}^{(3)} = V \setminus \{X_p, X_q \text{ and } S^d\}$. S^d refers to the colliders where X_p and X_q have direct edges on them. In the motivating example, we omit the unobserved confounders U and the instrumental variables for each phenotypes used in MR.

Figure 2. Diagrams for simulation study 1. X , exposure; Y , outcome; S , collider; M , mediator; C , measured confounder; U , unobserved confounder. β is the causal effect of X on Y . α_1 is the causal effect of X on S/M or C on X . α_2 is the causal effect of C/M on Y or Y on S . G_1 are SNPs only associated with X . G_2 are SNPs associated with X and adjusting variable. G_3 are SNPs only associated with the adjusting variable

Figure 3. Simulation results of MVMR in simulation study 1. (A-C) Causal effect estimation of X on Y when causal effect $\beta=0$; (D-F) Type I error rates of causal effect estimation of X on Y when causal effect $\beta=0$; (G-I) Causal effect estimation of X on Y when causal effect $\beta=0.1$; (J-L) Statistical power of causal effect estimation of X on Y when causal effect $\beta=0.1$. The x-axis represent the other edges' effect (α_1 and α_2 in Figure 2).

Figure 4. F1 score with 10 continuous nodes. The x axis represents the number of IVs.

Considering the different complexity of network, we set the probability of each edge to be present in a graph as 0.2, 0.5 and 0.8. The effects of any two traits β follows uniform distribution with four parameter settings: U(0,0.25), U(0.25,0.5), U(0.5,0.75) and U(0.75,1) for continuous nodes. MRSL_min_sep_set indicates the MRSL adjusting for minimal sufficient adjustment set and all the mediators; MRSL_open_path indicates the MRSL adjusting for all the nodes on the open paths; MRSL_remove_collider indicates the MRSL adjusting for $\forall \{ X_p, X_q \text{ and } S^d \}$.

Figure 5. Precision and recall with 10 continuous nodes when there is 20 IVs.

Figure 6. Diagrams of practical example in the simulation study 3. Green and yellow circles denote continuous and binary variables, respectively. MAGIC, Multiparent Advanced Generation Inter-Cross; NIAB, National Institute of Agricultural Botany; YLD, yield; FT, flowering time; HT, height; YR.GLASS, yellow rust in the glasshouse; YR.FIELD, yellow rust in the field; FUS, Fusarium; MIL, mildew.

Figure 7. F1 score of MRSL when learning the structure of MAGIC-NIAB, ASIA and Healthcare Cost.

Figure 8. Structures of 26 biomarkers and 44 diseases in applied example 2. A) the marginal causal graph obtained by MRSL-step1; B) the conditional causal graph obtained by MRSL-step2; C) mediation pathways of several diseases. The color of an edge indicates whether the effect is positive or negative. Yellow edges represent the positive effects and green edges represent the negative effects. Blue nodes represent the ICD10-defiend diseases. Pink nodes represent the biomarkers.

Table 1. Computing time with network of 10 continuous nodes in simulation study 2 (seconds).

Edge effect	g	prob	MRS� ¹	MRS� ²	MRS� ³	MRPC ¹	MRPC ²	HC ¹	HC ²	BIMMER	cGAUGE ¹	cGAUGE ²	cGAUGE ³
0-0.25	5	0.2	0.44	0.63	0.01	10.18	9.98	1.38	1.09	9.63	89.86	93.17	153.23
		0.5	1.16	1.55	0.01	12.42	11.49	2.24	1.62	10.06	104.54	106.87	167.74
		0.8	2.86	2.49	0.01	11.00	11.73	2.70	2.27	9.62	107.74	105.53	162.45
	20	0.2	0.25	0.33	0.01	318.35	6.99	2.00	0.86	3.47	187.00	180.95	374.32
		0.5	0.79	0.89	0.02	524.53	5.94	2.47	1.08	4.07	188.95	185.82	375.04
		0.8	1.97	1.95	0.03	225.19	5.92	2.44	1.54	5.02	189.15	186.96	377.61
	50	0.2	0.43	0.54	0.08	454.50	13.33	2.69	1.45	6.33	769.21	755.44	1177.71
		0.5	1.62	1.62	0.15	620.90	11.57	3.31	1.86	7.64	761.88	716.36	1161.77
		0.8	3.58	3.66	0.24	556.61	11.41	3.33	2.51	11.42	740.36	716.61	1130.72
0.25-0.5	5	0.2	1.17	1.74	0.01	7.78	11.09	2.06	1.33	12.31	112.22	108.91	141.70
		0.5	6.29	4.16	0.02	5.34	10.03	2.65	2.99	8.63	115.11	112.67	139.95
		0.8	8.35	6.30	0.03	3.90	8.87	2.25	3.50	6.98	122.92	111.16	122.10
	20	0.2	1.10	1.39	0.03	110.26	9.65	2.77	1.21	8.42	338.80	332.62	610.84
		0.5	6.15	5.02	0.07	33.04	8.77	2.73	3.06	11.34	362.36	353.72	571.79
		0.8	8.54	10.07	0.08	5.60	7.53	2.00	3.34	7.43	344.11	339.55	523.73
	50	0.2	1.31	1.47	0.14	56.75	8.11	1.90	1.17	8.64	743.79	757.10	960.77
		0.5	7.11	7.50	0.28	25.80	8.48	2.41	2.76	14.58	791.12	733.85	942.95
		0.8	15.77	18.18	0.11	4.00	7.47	1.77	3.15	7.90	773.35	755.31	895.70

0.5-0.75	5	0.2	1.28	1.81	0.01	4.77	10.17	2.06	1.53	10.40	110.09	105.68	132.13
		0.5	6.99	4.42	0.02	2.86	5.44	2.43	3.53	6.77	116.25	108.11	115.02
		0.8	9.75	8.56	0.03	1.91	3.25	1.85	3.15	5.42	114.58	108.85	117.62
	20	0.2	1.52	1.83	0.04	24.12	7.52	2.41	1.50	9.56	344.63	323.82	551.96
		0.5	8.27	8.21	0.06	4.20	4.94	2.04	3.21	7.29	351.35	350.21	442.85
		0.8	12.58	17.88	0.03	1.52	2.70	1.51	2.73	5.55	354.66	347.40	377.69
	50	0.2	1.20	1.23	0.08	15.27	3.37	1.05	0.78	5.15	417.83	416.32	517.33
		0.5	8.42	12.04	0.10	2.00	3.54	1.53	2.63	5.72	589.59	651.33	864.82
		0.8	15.49	23.46	0.03	1.43	3.10	1.41	2.77	5.86	779.40	762.43	789.32
0.75-1	5	0.2	1.38	1.68	0.01	2.50	5.55	1.86	1.55	8.89	105.81	99.97	113.52
		0.5	8.34	6.12	0.02	1.71	3.30	2.30	3.33	6.15	119.01	112.45	113.43
		0.8	10.01	10.06	0.02	1.22	1.30	1.41	2.20	4.24	101.53	92.05	103.46
	20	0.2	1.86	2.07	0.04	11.28	4.41	2.05	1.49	8.57	334.46	335.41	429.89
		0.5	11.23	13.83	0.05	1.61	2.68	1.87	2.98	6.15	359.54	347.45	374.20
		0.8	14.08	20.04	0.02	1.24	1.38	1.26	2.25	4.97	345.71	338.63	342.01
	50	0.2	3.44	3.46	0.16	40.28	7.01	1.82	1.64	9.37	758.43	758.60	928.54
		0.5	15.67	19.07	0.09	1.38	3.44	1.74	3.26	6.63	780.50	776.38	821.59
		0.8	13.79	20.74	0.02	0.99	1.55	1.13	2.19	4.56	629.16	589.81	590.85

MRS^{L1}, MRS^L when adjusting for all nodes on the open paths; MRS^{L2}, MRS^L when adjusting for minimal sufficient adjustment set and all the mediators; MRS^{L3}, MRS^L when adjusting for $D \setminus \{X_p, X_q \text{ and } S^d\}$. MRPC¹, MRPC algorithm based on the most significant SNP; MRPC²,

MRPC algorithm based on genetic risk score. HC¹, HC algorithm incorporating genetic anchors based on the most significant SNP; HC², HC algorithm incorporating genetic anchors based on genetic risk score. cGAUGE¹, cGAUGE based on IVW; cGAUGE², cGAUGE based on MR Egger; cGAUGE³, cGAUGE based on MR PRESSO.

Workflow of MRSL

Input GWAS summary data for d phenotypes.

Initialization with an empty graph.

Step 1

Obtain a marginal causal diagram using bi-directional MR in pairs.

Bi-directional MR

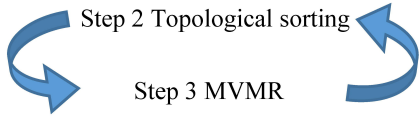
$$\begin{aligned}
 & U \\
 & \swarrow \quad \searrow \\
 & IV_p \rightarrow X_p \rightleftarrows X_q \leftarrow IV_q \\
 & X_p \rightleftarrows X_q: \hat{\beta}_{X_{qj}} \sim b_{pq} \hat{\beta}_{X_{pj}} \\
 & X_q \rightleftarrows X_p: \hat{\beta}_{X_{pj}} \sim b_{qp} \hat{\beta}_{X_{qj}}
 \end{aligned}$$

Step 2

Find **topological sorting** of the nodes in marginal causal graph using Depth First Search (DFS).

Step 4 Iteration

Initialization: (marginal causal diagram obtained in step 1)



Stop when the diagram converges.

Step 3 For each edge in marginal causal diagram, perform MVMR by adjusting genetic associations with $ADJ_{X_p \rightarrow X_q}$, which can be made up in three ways:

- (1) All nodes in the **open paths** from X_p to X_q ;
- (2) **Minimal sufficient adjustment set and all the mediators** from X_p to X_q ;
- (3) $\forall \{X_p, X_q \text{ and } S^d\}$.

The order of MVMR for each edge refers to the topological sorting in step 2.

$$\begin{aligned}
 & U \\
 & \swarrow \quad \searrow \\
 & IV_1 \rightarrow X_p \rightleftarrows X_q \\
 & IV_2 \rightarrow ADJ_{X_p \rightarrow X_q} \\
 & IV_3 \rightarrow ADJ_{X_p \rightarrow X_q} \\
 & \hat{\beta}_{X_{qj}} \sim \alpha_{pq} \hat{\beta}_{X_{pj}} + \alpha_{ADJ} \hat{\beta}_{ADJ_{X_p \rightarrow X_q}}
 \end{aligned}$$

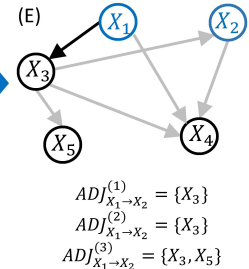
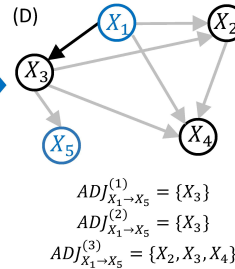
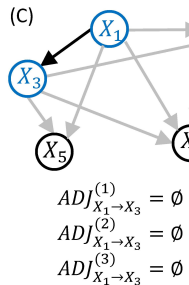
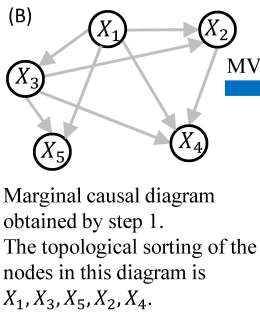
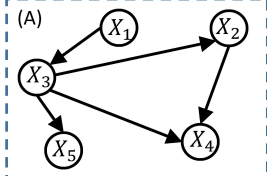
Multivariable MR (MVMR)

Output

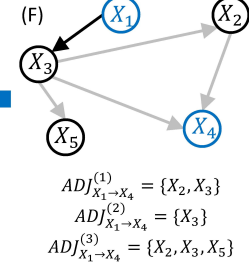
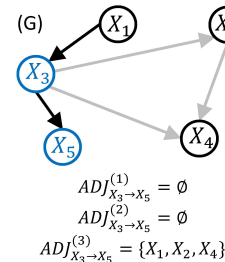
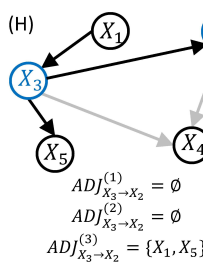
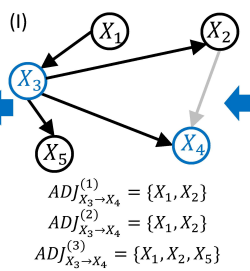
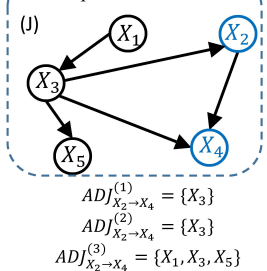
Conditional causal diagram of d phenotypes.

Motivating example

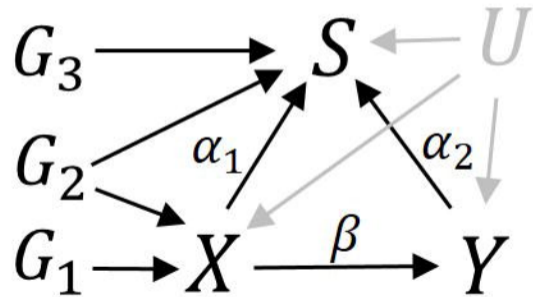
True diagram



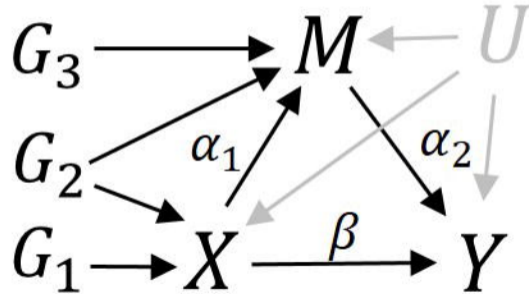
Output after iteration



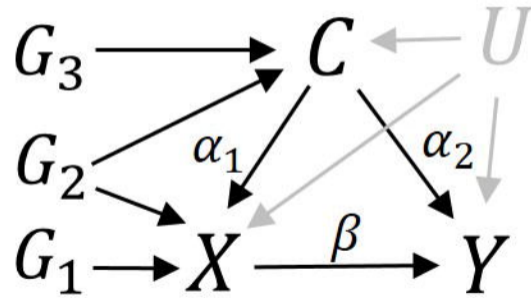
A)



B)

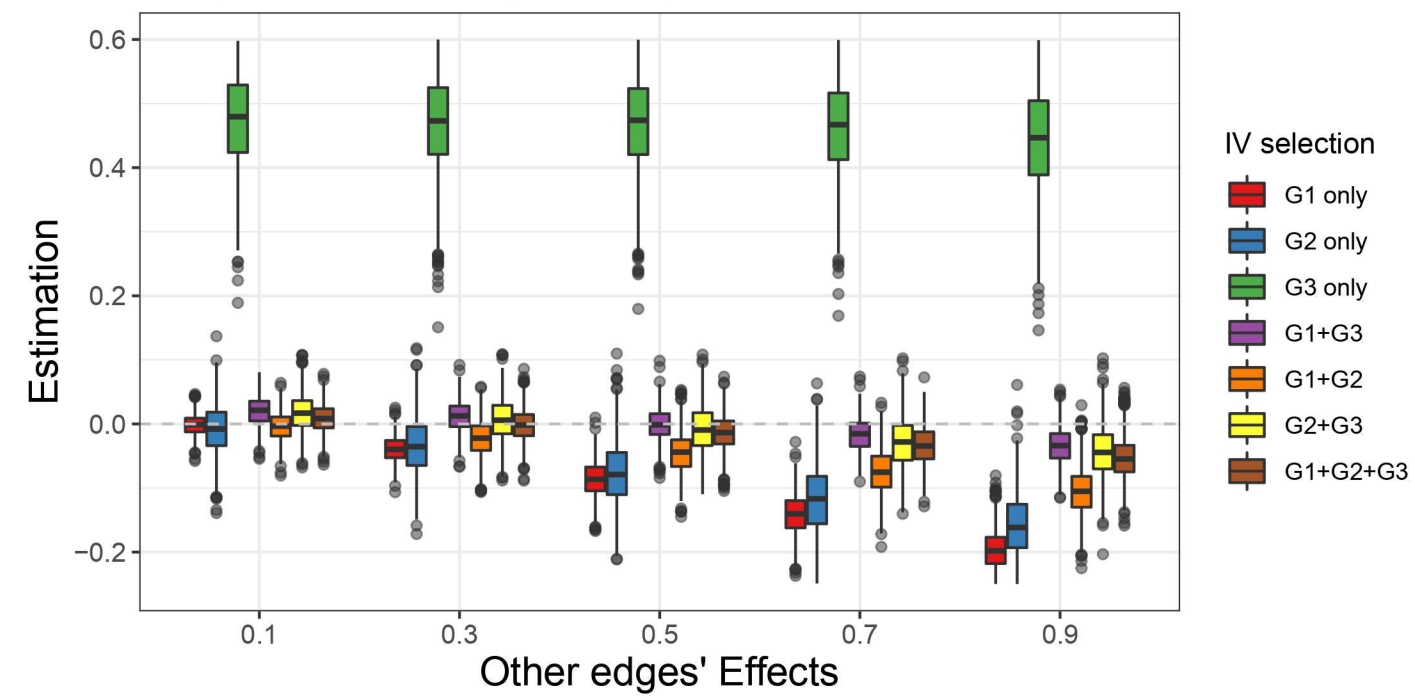


C)



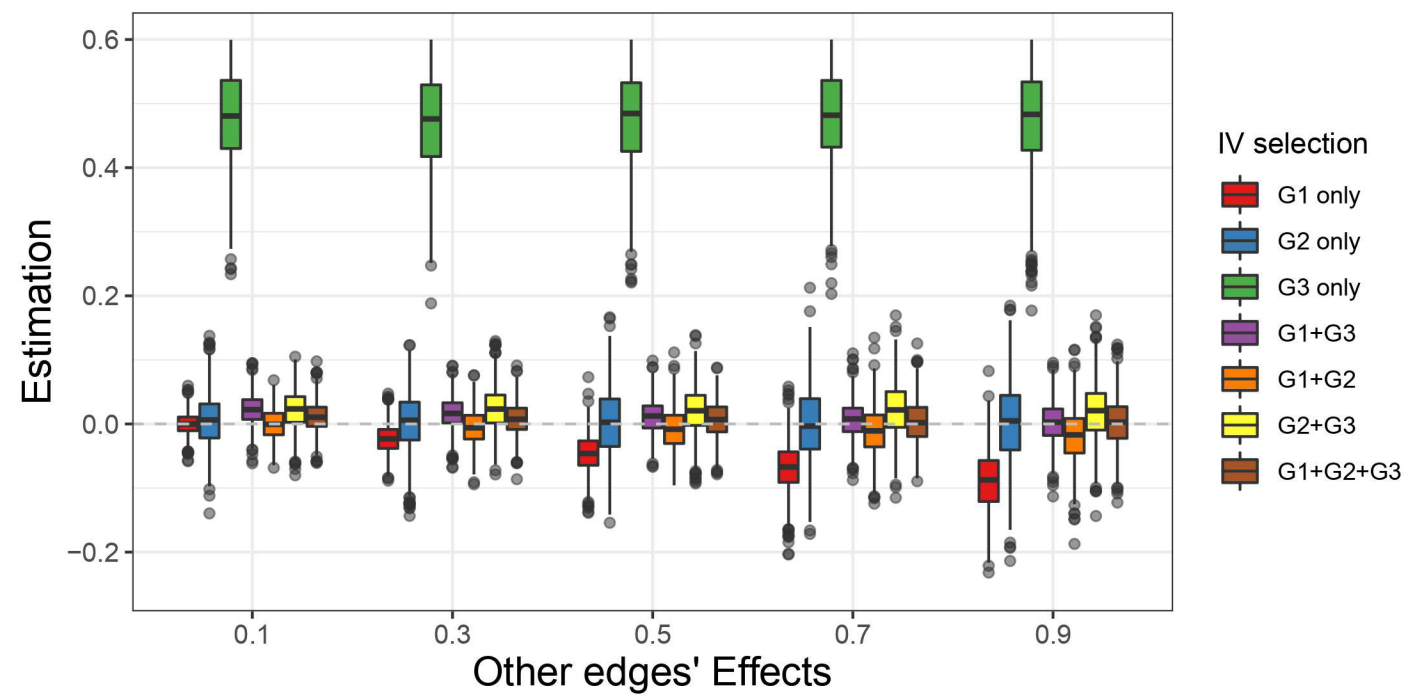
A)

Graph A collider (beta=0)



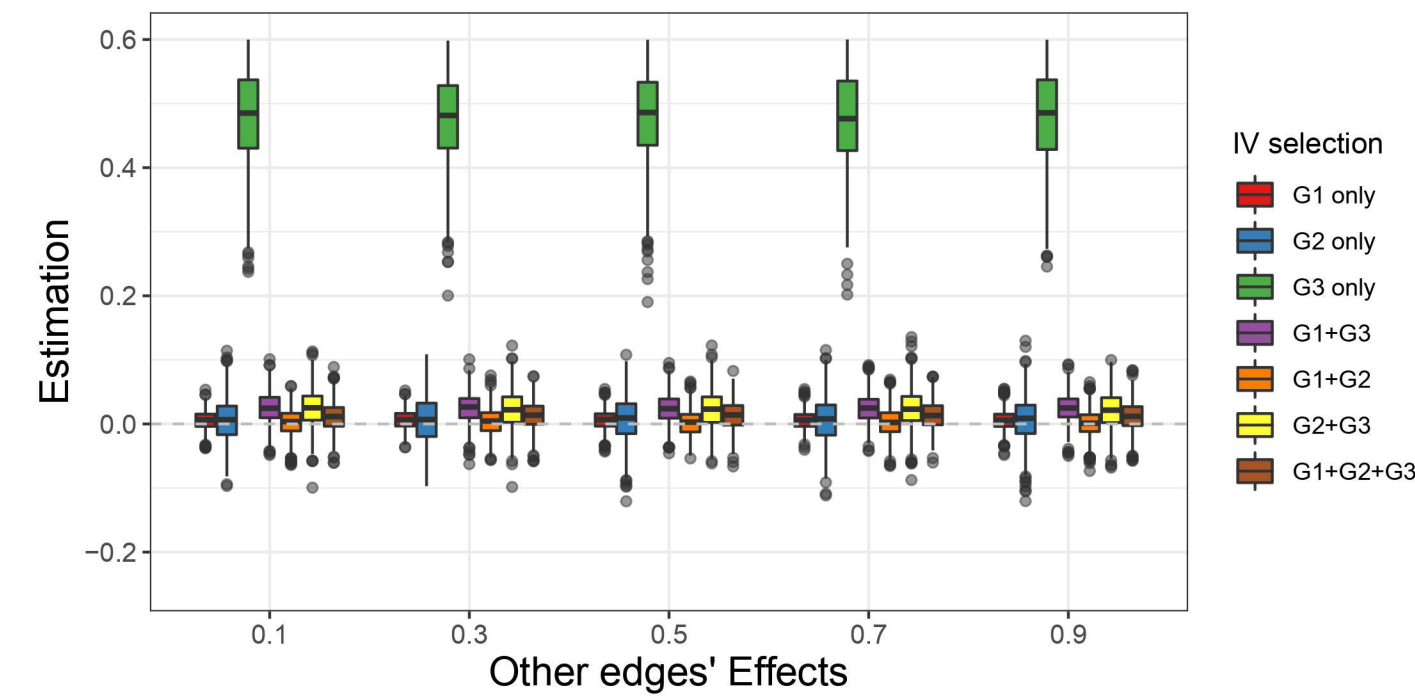
B)

Graph B mediator (beta=0)



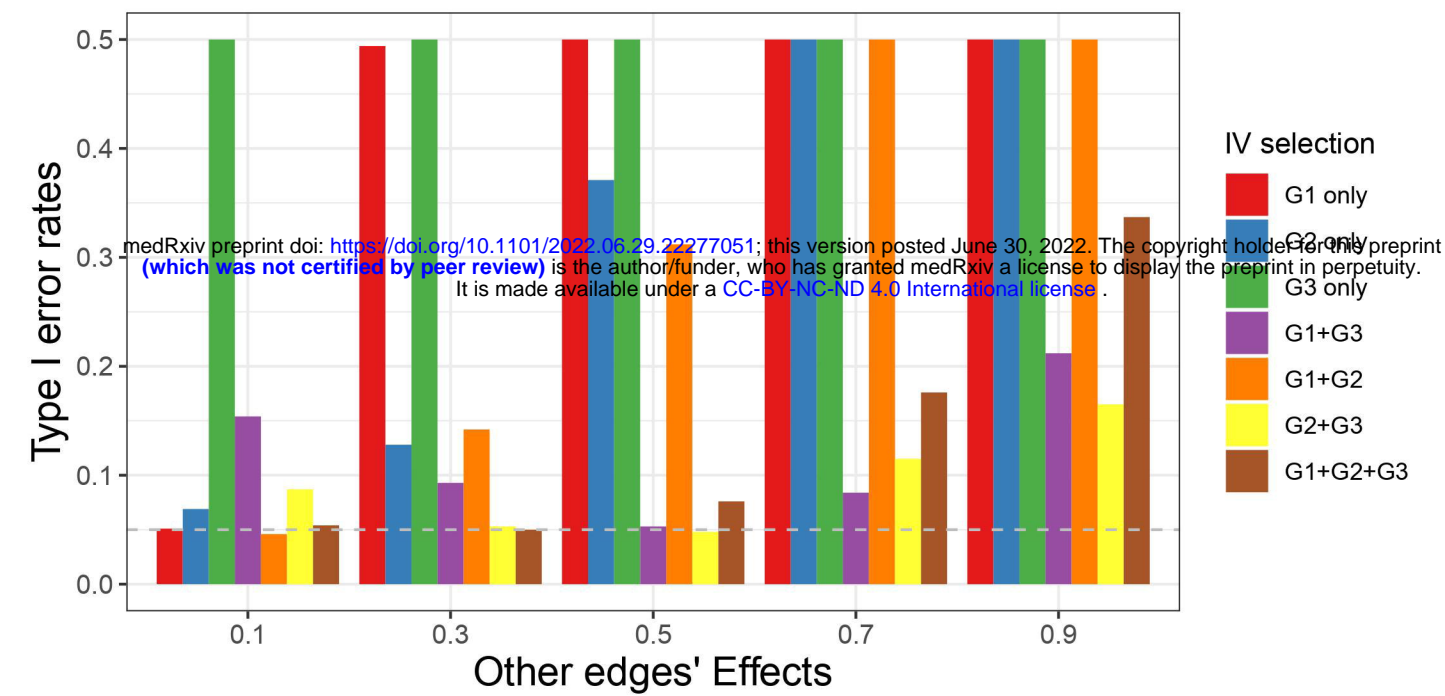
C)

Graph C confounder (beta=0)



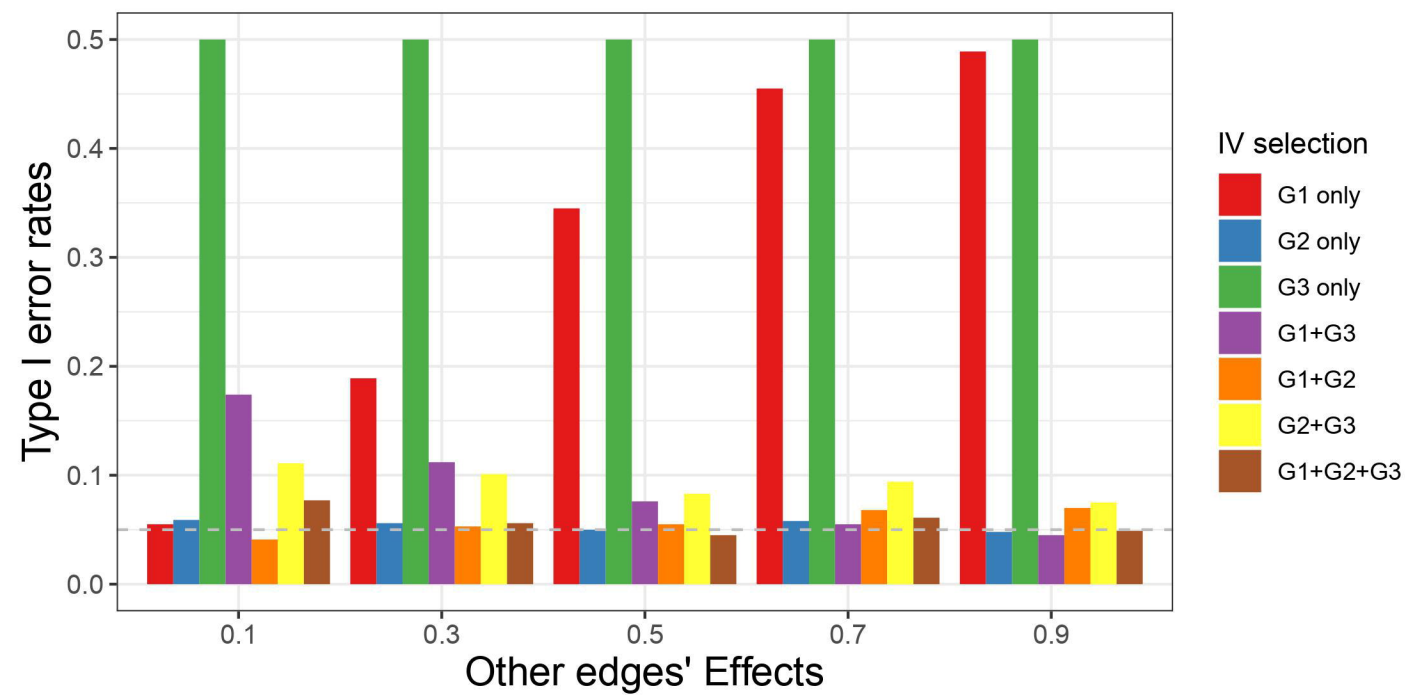
D)

Graph A collider (beta=0)



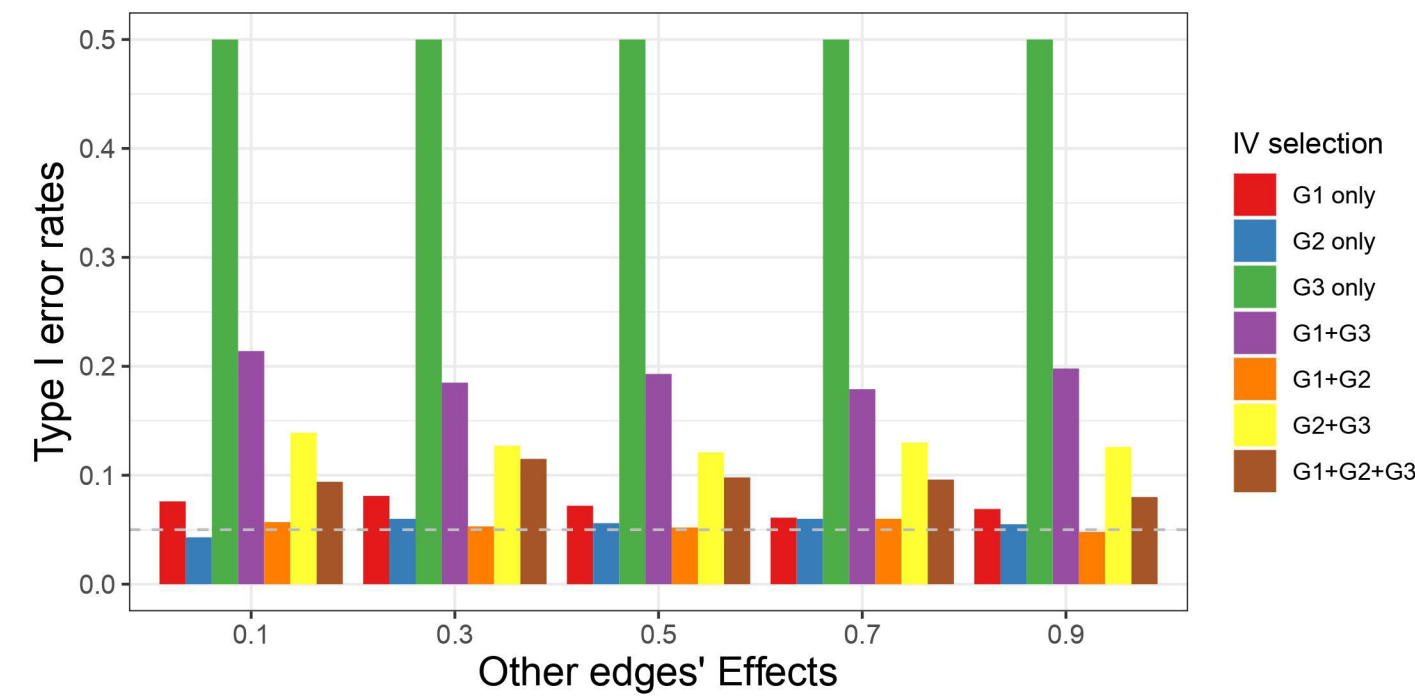
E)

Graph B mediator (beta=0)



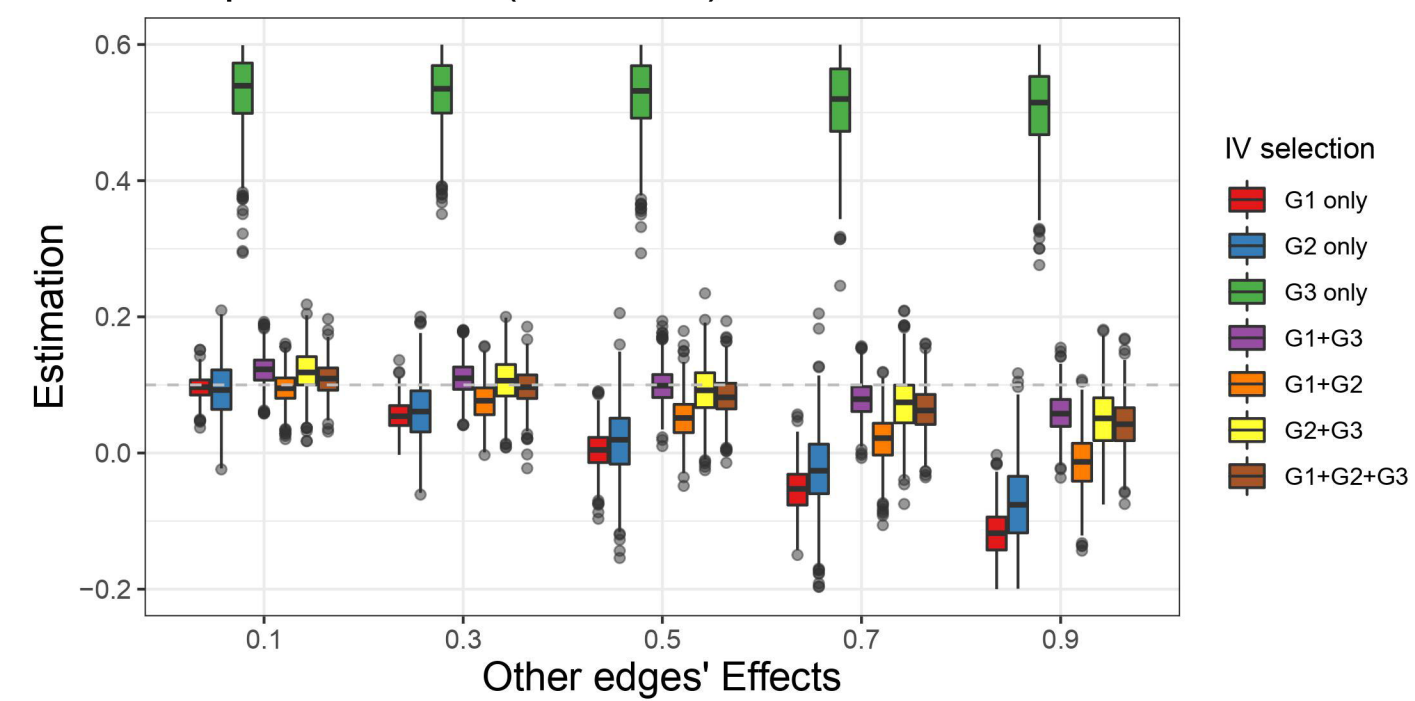
F)

Graph C confounder (beta=0)



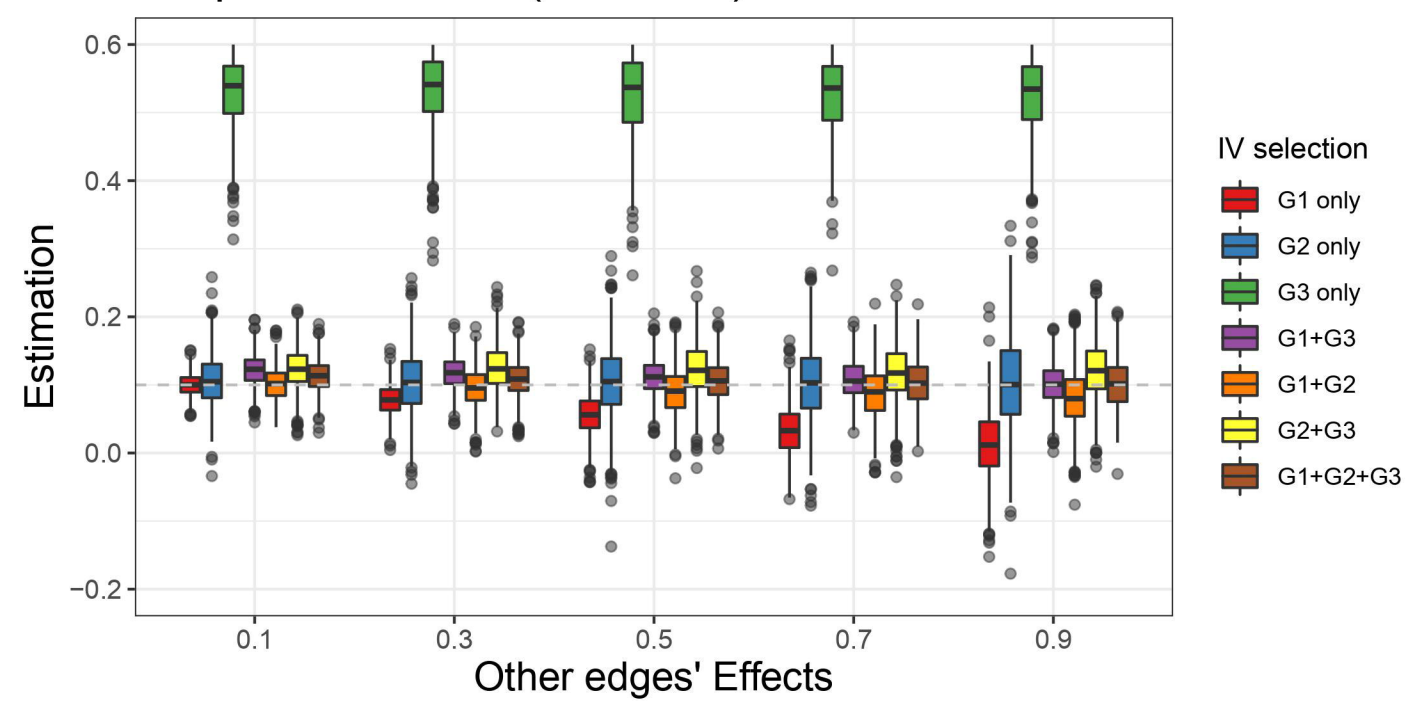
G)

Graph A collider (beta=0.1)



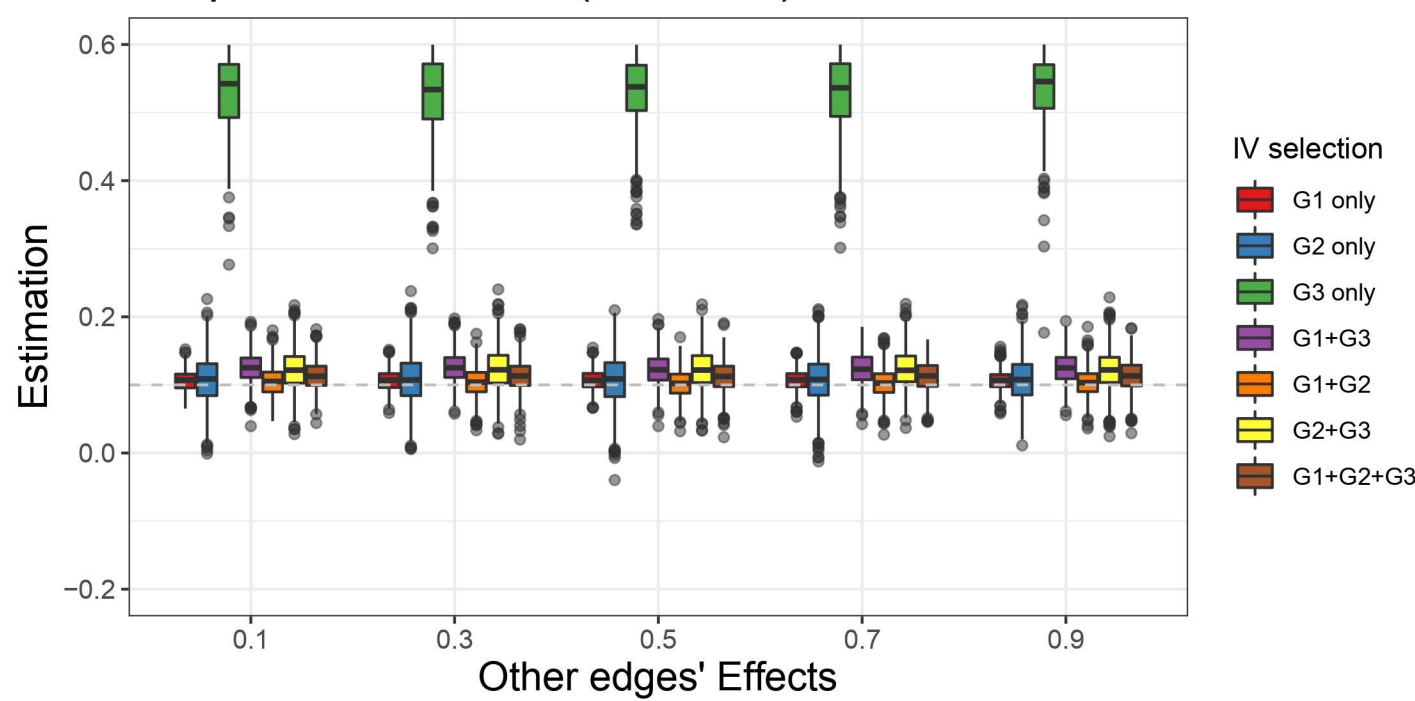
H)

Graph B mediator (beta=0.1)



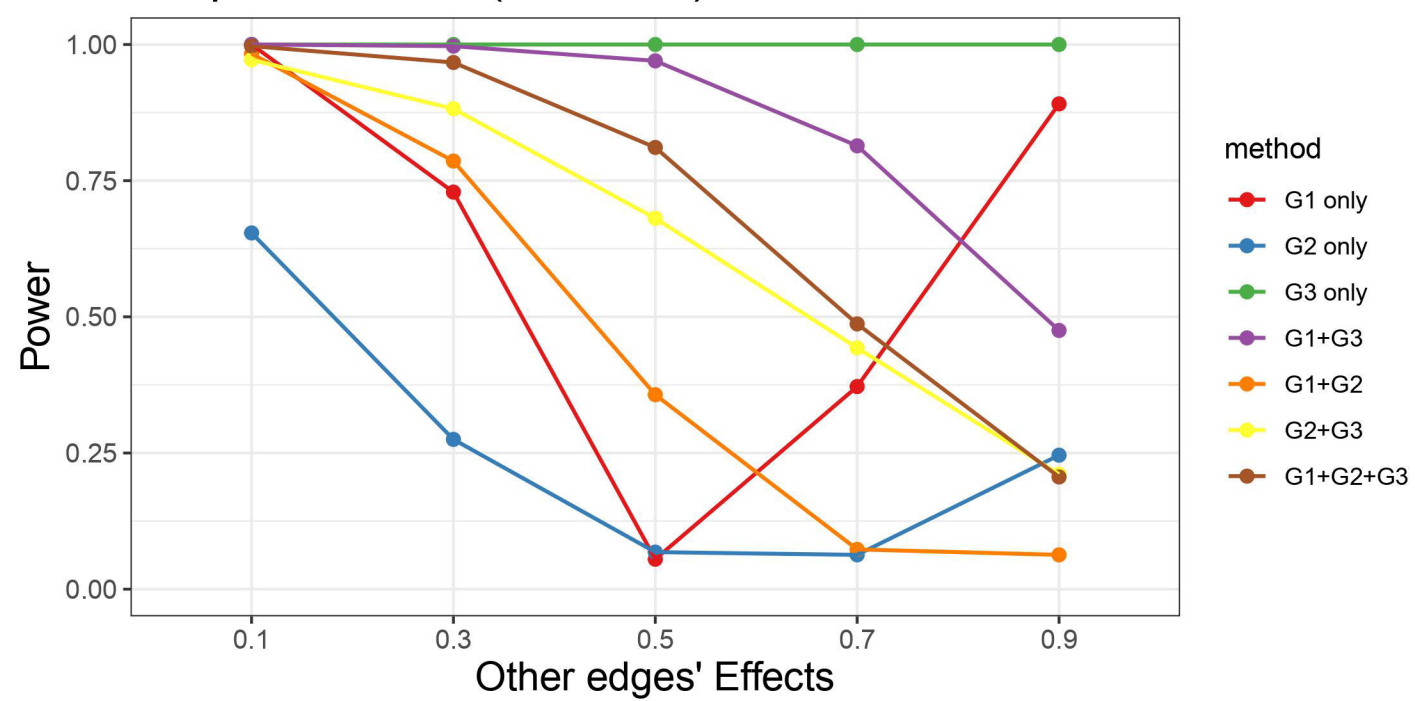
I)

Graph C confounder (beta=0.1)



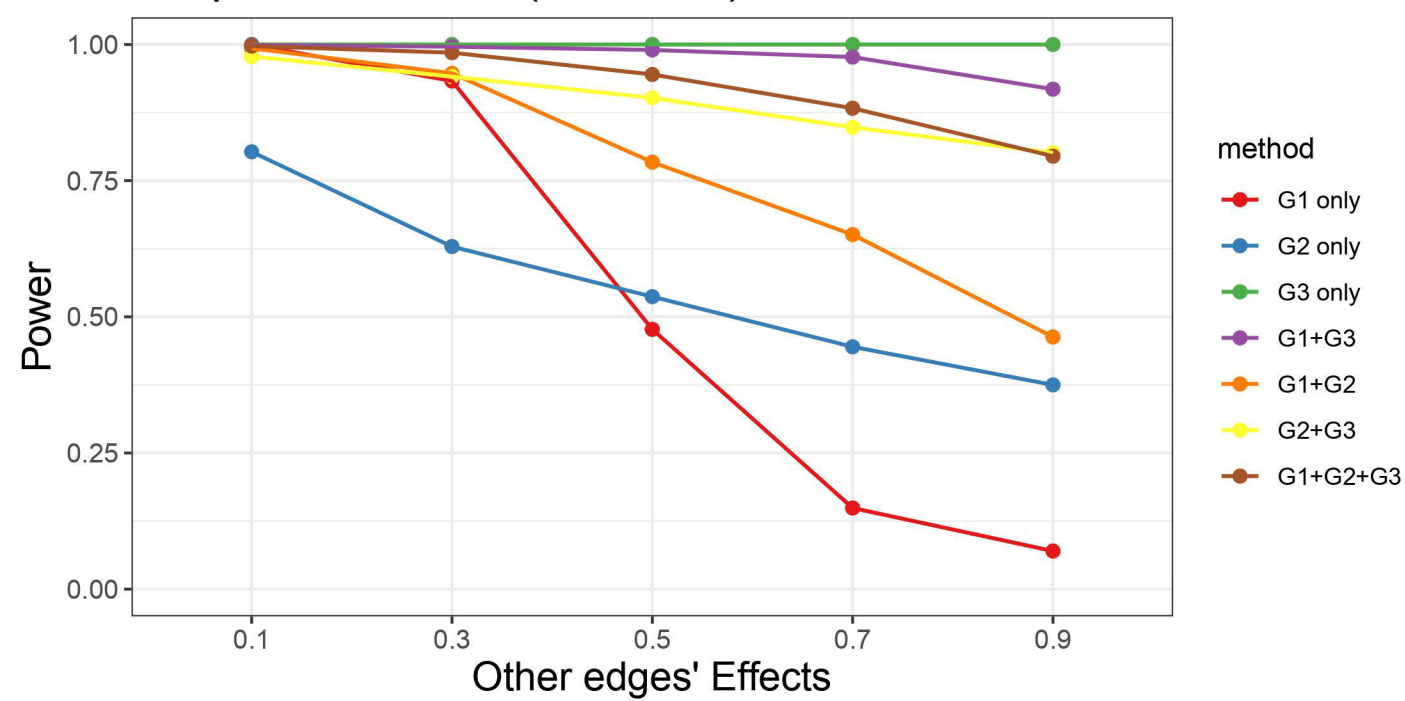
J)

Graph A collider (beta=0.1)



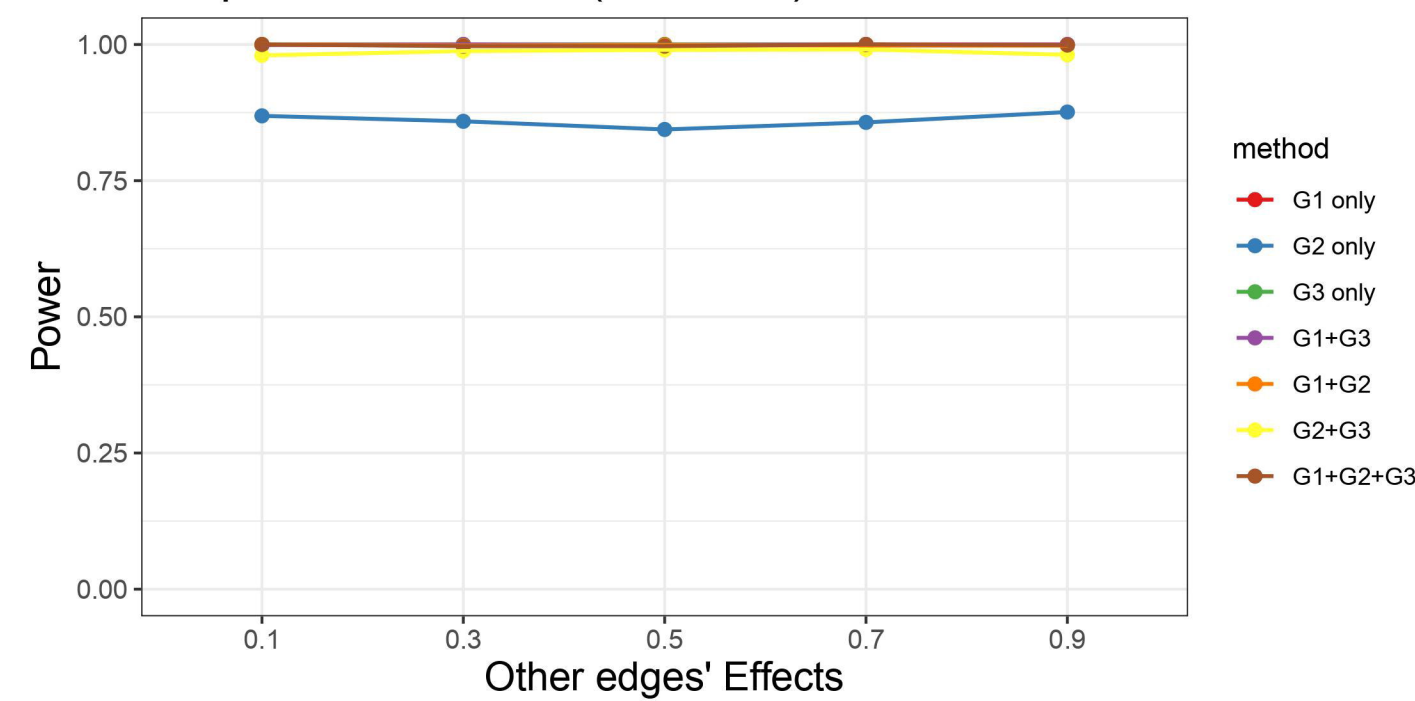
K)

Graph B mediator (beta=0.1)

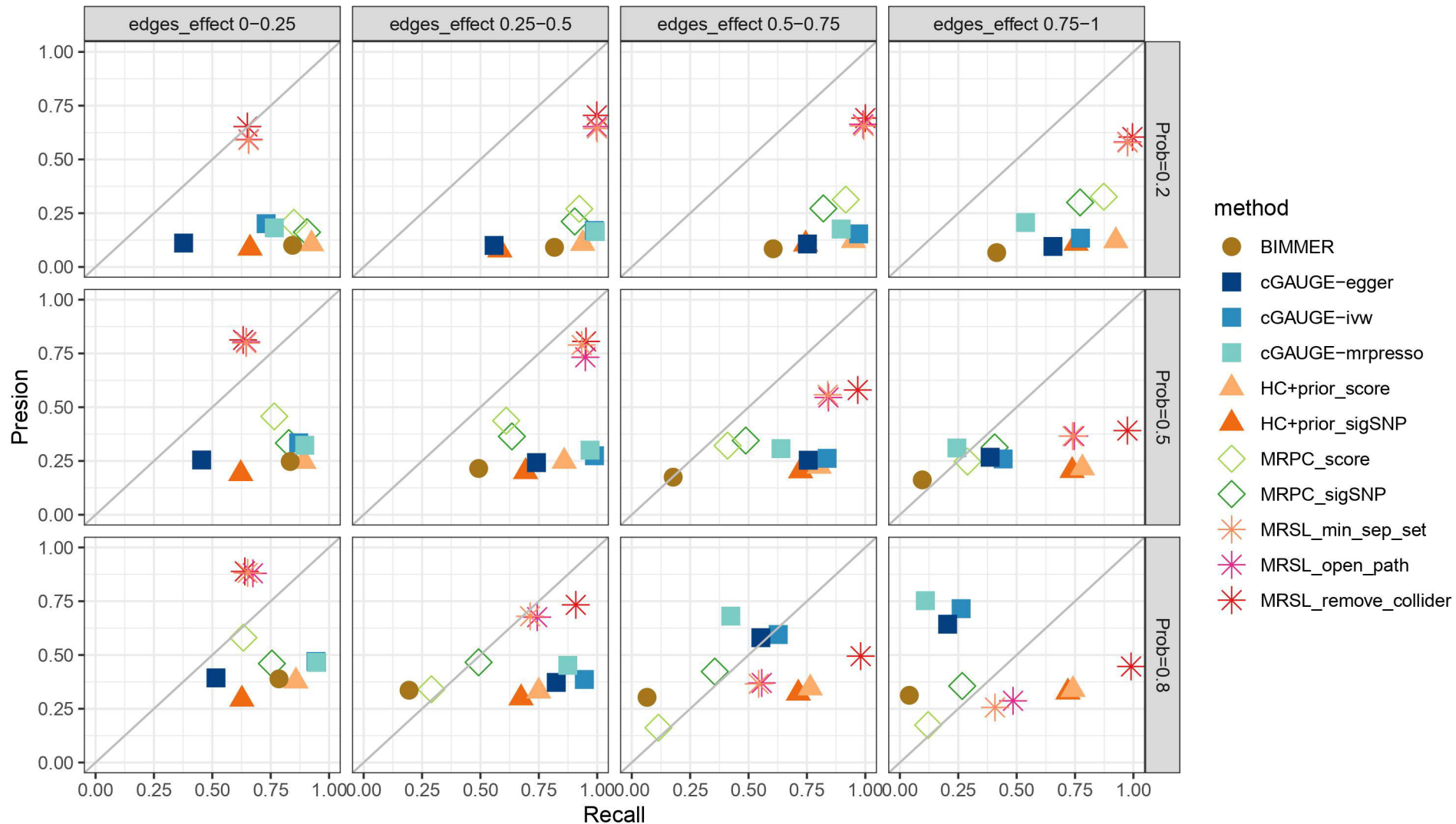


L)

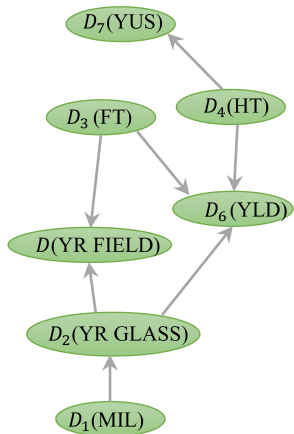
Graph C confounder (beta=0.1)



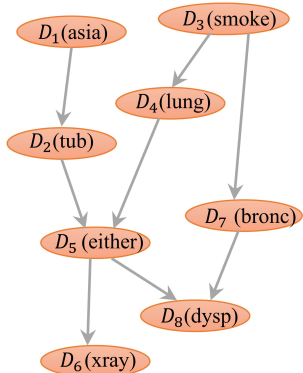
Precision-Recall with 10 continuous nodes (20 IVs)



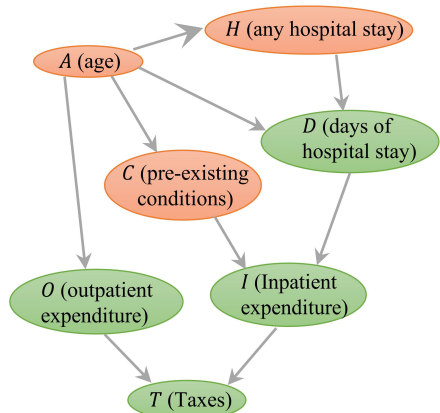
A) MAGIC-NIAB (continuous)



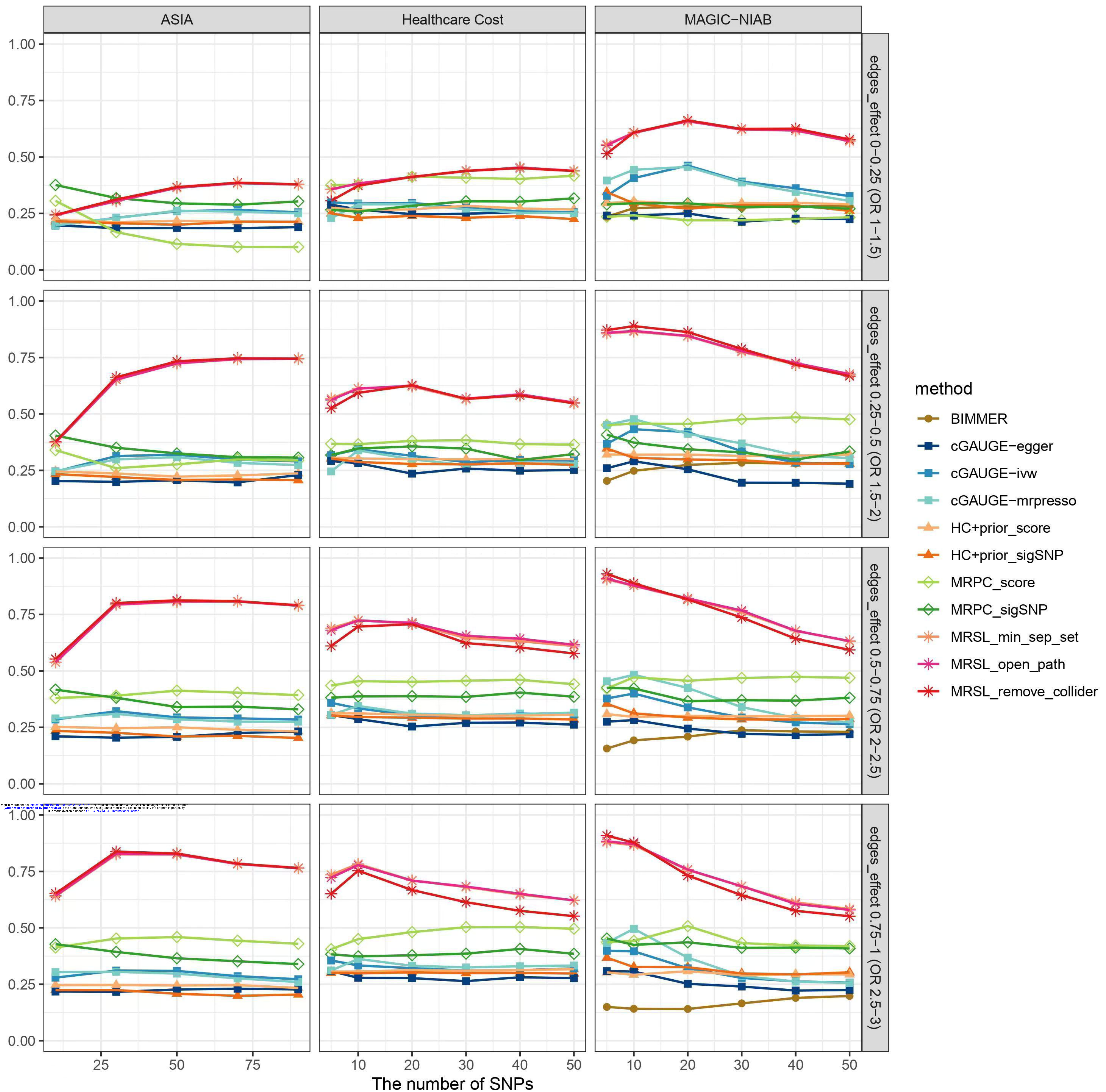
B) ASIA (binary)



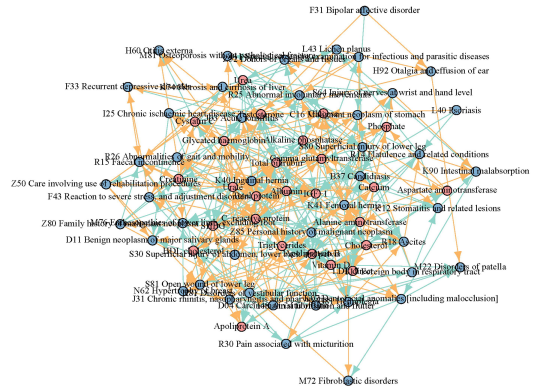
C) HEALTHCARE(mixed)



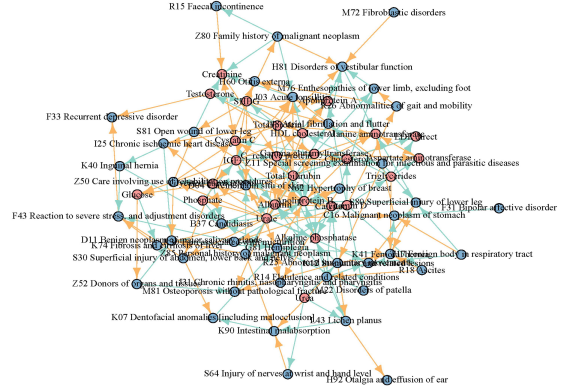
F1 score with fixed graph



A) Marginal causal graph

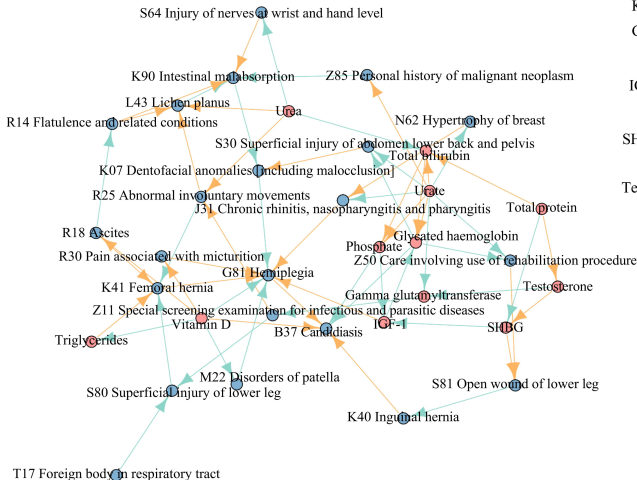


B) Conditional causal graph

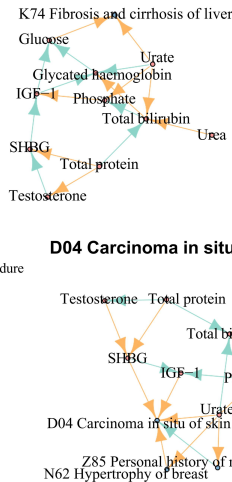


C) Mediation pathways for several diseases

B37 Candidiasis



K74 Fibrosis and cirrhosis of liver



F33 and F44 Mental disorders

