Bayesian variable selection with a pleiotropic loss function in Mendelian randomization

Apostolos Gkatzionis[1*], Stephen Burgess[1,2], David V Conti[3], Paul J Newcombe[1],

1 MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom.
2 Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom.
3 Department of Preventive Medicine, Keck School of Medicine, University of Southern California, USA.

*apostolos.gkatzionis@mrc-bsu.cam.ac.uk

Abstract

Mendelian randomization is the use of genetic variants as instruments to assess the existence of a causal relationship between a risk factor and an outcome. A Mendelian randomization analysis requires a set of genetic variants that are strongly associated with the risk factor and only associated with the outcome through their effect on the risk factor. We describe a novel variable selection algorithm for Mendelian randomization that can identify sets of genetic variants which are suitable in both these respects. Our algorithm is applicable in the context of two-sample summary-data Mendelian randomization and employs a recently proposed theoretical extension of the traditional Bayesian statistics framework, including a loss function to penalize genetic variants that exhibit pleiotropic effects. One of our algorithm's main advantages is more robust inference through the use of model averaging, as we illustrate by running it on a wide range of simulation scenarios and comparing it against established pleiotropy-robust Mendelian randomization methods. In a real data application, we study the effect of systolic and diastolic blood pressure on the risk of suffering from coronary heart disease. Although this application has been studied in the past, we add to the literature by using for the first time a recent large-scale GWAS on blood pressure, allowing us to select 395 genetic variants for systolic and 391 variants for diastolic blood pressure. Both traits are shown to have significant risk-increasing effects on coronary heart disease risk.

Introduction

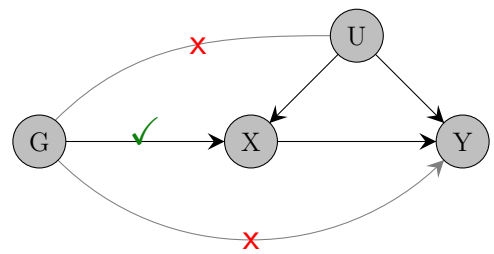
Mendelian randomisation provides a framework for probing questions of causality from observational data using genetic variants. It applies the theory of instrumental variable analysis from the causal inference literature, using genetic variants associated with the risk factor as instruments. Mendelian randomization relies on the idea that, since genetic variants are randomly inherited and fixed at conception, they should be uncorrelated with potential confounders of the relationship between the risk factor and outcome and are therefore suitable to use as instruments. This approach has received much attention since the seminal paper of Davey Smith and Ebrahim [1], and has led to a number of influential results over the last decade addressing a variety of aetiological questions [2]. For example, in coronary heart disease, Mendelian randomization has been used to strengthen the evidence for a causal role of lipoprotein(a) [3], but to weaken the case for C-reactive protein [4].

Formally, Mendelian randomization relies on three basic assumptions:

1. The genetic variants are independent of any confounders of the risk factor-outcome association.
2. The genetic variants are strongly associated with the risk factor of interest.
3. The genetic variants only influence the outcome via their association with the risk factor and not through alternative causal mechanisms.

The three assumptions are illustrated in Fig 1. Under these assumptions, valid causal inferences can be made as to whether the risk factor affects the outcome.

Fig 1. A causal diagram representation of the three assumptions of Mendelian randomization. Here, X represents the risk factor, Y the outcome, G the genetic instrument and U denotes confounders of the $X - Y$ relationship.



In practice, assumption (1) is usually justified on the basis of Mendelian inheritance: an individual's genotype is randomly assigned at conception and not influenced by external confounding factors. To ensure the validity of assumption (2) Mendelian randomization analyses typically rely on the results of large consortia meta-GWAS, which use sample sizes of tens or hundreds of thousands of individuals to identify genetic variants robustly associated with a trait. In particular, many recent Mendelian randomization studies have utilized a two-sample approach in which genetic associations with the target risk factor and with the outcome are assessed in separate datasets. However, these GWAS results are rarely available as individual-level data. Usually, only a set of summary statistics, such as univariate variant-trait associations and corresponding standard errors, are reported. As a result, a large number of recent Mendelian randomization investigations rely on summarized data.

Assumption (3), often called the exclusion restriction or no-pleiotropy assumption, has received much attention in the recent literature. In practical applications, we typically do not know which genetic variants exhibit pleiotropic effects and there is need for methods to perform Mendelian randomization in the presence of some potentially pleiotropic variants. Traditional approaches include MR-Egger regression [5] and median estimation [6], while several algorithms for pleiotropy-robust Mendelian randomization have been developed recently [7–14]. A recent review and comparison of the various methods can be found in [15].

In this paper, we add to the relevant literature by proposing a new method for variable selection and causal effect estimation in Mendelian randomization. Our method is derived as an extension of the JAM algorithm (Joint Analysis of Marginal Summary Statistics, [16]). JAM was originally proposed as an algorithm for fine-mapping genetic regions. Similar to other recently proposed fine-mapping algorithms [17, 18], JAM is designed to work with summary GWAS data. The algorithm performs variable selection to identify genetic variants robustly associated with the trait. Genetic correlations are

taken into account by estimating them from a reference dataset such as 1000 Genomes or the UK Biobank. Variable selection is performed according to a Bayesian stochastic search algorithm, which can explore the complete space of causal configurations. Consequently, JAM is able to explore complex models with large numbers of variants, as recently demonstrated while fine-mapping dense genotype data for prostate cancer risk [19].

We develop a novel model averaging variable selection algorithm for Mendelian randomization, which we call JAM-MR (JAM for Mendelian randomization). To do so, we modify JAM's variable selection to downweight genetic variants which exhibit pleiotropic effects. Using a recently proposed theoretical extension of the core principles of Bayesian inference citeBissiri2016, we augment JAM's likelihood with a loss function that penalizes models containing variants with heterogeneous univariate causal effect estimates. Our algorithm performs variable selection and returns variant-specific posterior inclusion probabilities, which can be interpreted as probabilities of each variant being a valid instrument, and posterior model probabilities, which can subsequently be used to estimate the causal effect of interest by averaging across model-specific estimates. Uncertainty in which variants should be excluded on the basis of pleiotropy is reflected by averaging estimates over competing selections of instruments; model averaged causal inference is an attractive feature and one of the key contributions of our method.

Further advantages of our algorithm compared to established approaches for pleiotropy-robust Mendelian randomization include its dependence on a tuning parameter that represents the strength of pleiotropy penalization; varying the value of this parameter can be a useful tool for sensitivity analysis. JAM-MR also offers a natural framework for incorporating genetic correlations, when conducting Mendelian randomization with several genetic variants coming from the same gene region. The use of the Bayesian paradigm allows us to incorporate prior information on the suitability of genetic variants as instruments into the analysis. It also allows us to model the uncertainty in genetic associations with the risk factor, which is often ignored by other approaches and can cause underestimation of standard errors. Finally, the use of the Bayesian loss function framework [20] means that our algorithm does not have to make any modelling assumptions for the risk factor-outcome relationship.

The performance of our algorithm is illustrated in a range of simulated datasets, as well as in a real data application where we investigate the causal effect of systolic and diastolic blood pressure on the risk of coronary heart disease. Although this application has been studied in the past, for the first time we instrument blood pressure using a recently published large scale meta-GWAS, which combined results across more than one million individuals, and therefore base our Mendelian randomization analysis on larger sample sizes and more genetic variants compared to previous studies in the literature. Our results strengthen the claim for a risk-increasing causal relationship between blood pressure traits and coronary heart disease risk.

Results

Simulation studies

Simulation setting

The JAM-MR algorithm's reliance on model averaging means that it is likely to yield robust causal inferences in a wide range of pleiotropic scenarios. To asess that, we conducted a large number of simulations comparing our algorithm to established approaches for Mendelian randomization. The Mendelian randomization model that we

used for the simulations was

$$
\begin{aligned}
U &= \sum_{j=1}^{P} \alpha_j G_j + \epsilon_U && (1)\\
X &= \sum_{j=1}^{P} \beta_{Xj} G_j + \alpha_X U + \epsilon_X && (2)\\
Y &= \theta X + \sum_{j=1}^{P} \delta_j G_j + \alpha_Y U + \epsilon_Y && (3)\\
\epsilon_U, \epsilon_X, \epsilon_Y &\sim N(0,1) \quad \text{independently of each other}
\end{aligned}
$$

where $G_1, \ldots, G_P$ are the genetic variants, X is the risk factor, Y the outcome and U denotes confounders of the risk factor-outcome association. The parameter θ denotes the causal effect to be estimated, while $\alpha_j, \beta_j, \delta_j$ are the effects of genetic variant G_j on U, X, Y respectively. For genetic variants that are valid instruments, $\delta_j = \alpha_j = 0$.

Our simulations were conducted using the statistical software R. In each simulation, we generated $P = 50$ independent genetic variants of which $P_1 = 35$ were valid and $P - P_1 = 15$ were pleiotropic. We considered either a null ($\theta = 0$) or a positive ($\theta = 0.5$) causal effect. Effect allele frequencies for the genetic variants were drawn uniformly in $(0.1, 0.9)$. Genetic effects β_{Xj} on the risk factor were simulated uniformly in $(0.1, 0.2)$; this loosely corresponds to genome-wide significant variants with p-values ranging from 10^{-40} to 10^{-7}. The proportion of variation in the risk factor explained by the genetic variants was fixed at 10%.

We considered four different simulation scenarios, corresponding to different patterns of pleiotropic behaviour. The first scenario was a "balanced pleiotropy" setting, in which the direct effects δ_j for the 15 pleiotropic variants were drawn uniformly at random in $(-0.2, 0.2)$ and the variant-confounder effects α_j were equal to zero. In the second scenario ("directional pleiotropy"), we simulated the direct effects δ_j in $(0, 0.2)$ instead. In the third scenario we fixed all pleiotropic effects δ_j at 0.1; this represents a situation where multiple pleiotropic variants act on the same causal pathway and have direct effects of similar magnitude. Finally, in the fourth scenario we generated direct effects uniformly in $(0, 0.2)$ and also allowed pleiotropic variants to influence the confounder, by simulating G-U effects α_j uniformly in $(-0.1, 0.1)$. This scenario represents a violation of the InSIDE assumption, which states that instrument strength is independent of the genetic variants' pleiotropic effects on the outcome. The four scenarios are summarized in Table 1. For each of the four scenarios, we replicated the simulation 1000 times.

Table 1. Different simulation scenarios.

Scenario	Type of Pleiotropy	Direct Effects	Confounder Effects
1	Balanced	$\delta_j \sim U(-0.2, 0.2)$	$\alpha_j = 0$
2	Directional	$\delta_j \sim U(0, 0.2)$	$\alpha_j = 0$
3	Directional	$\delta_j = 0.1$	$\alpha_j = 0$
4	InSIDE Violation	$\delta_j \sim U(0, 0.2)$	$\alpha_j \sim U(-0.1, 0.1)$

Since Mendelian randomization typically relies on summary data, we adapted our simulations accordingly. We used Eq (1), (2), (3) to generate individual-level data and then computed univariate estimates of association of each genetic variant with the risk factor and outcome, along with corresponding standard errors. These univariate estimates constitute a typical GWAS output. We based our implementations of JAM-MR and other Mendelian randomization algorithms on these estimates.

In each replication, we simulated two datasets: one from which to obtain genetic associations with the risk factor and one from which to obtain associations with the

outcome. This corresponds to a typical two-sample Mendelian randomization study, in which the exposure and outcome GWAS are conducted in different samples. We used sample sizes of $N_1 = N_2 = 50000$ for each dataset.

Competing methods

To assess the performance of the JAM-MR algorithm, we compared it against the following methods for Mendelian randomization.

- Standard inverse variance weighted (IVW) estimation.
- MR-Egger regression [5].
- Median estimation, weighted or unweighted [6].
- Mode-based estimation [7].
- A Lasso-type estimator [8, 9].
- MR-Presso [11].
- MR-Raps [12].

For further details on these approaches, the reader is referred to the relevant citations.

To illustrate the use of the JAM-MR algorithm as a tool for sensitivity analysis, we considered the following values for the tuning parameter: $w = 0$ (default JAM with no pleiotropy penalization), $w = 0.1N_1$, $w = 0.2N_1$, $w = 0.5N_1$, $w = N_1$, $w = 2N_1$ and $w = 5N_1$. In each instance, the algorithm was run for 1 million iterations. To implement the competing Mendelian randomization algorithms, we used available R packages - details are provided in the Methods section of the paper. For comparison purposes, we also computed "oracle" IVW causal effect estimates and standard errors using only the valid instruments.

Simulation results

The results of this simulation experiment are reported in Tables 2, 3, 4 and 5 for each of the four simulation scenarios. We report average causal effect estimates, their standard deviation across 1000 replications, estimated standard errors, average mean squared errors and Type I error rates at a 95% significance level. Mean squared errors and Type I error rates were computed using the estimated standard error values. In simulations with $\theta = 0$, Type I error rates were computed as the empirical power to reject the null hypothesis of no association between the risk factor and the outcome at a 95% significance level, while for $\theta = 0.5$ we interpret "Type I error" as falsely rejecting the true hypothesis $H_0 : \theta = 0.5$.

In the first simulation, all the Mendelian randomization methods provided nearly unbiased estimates of the causal effect of interest. This was the case even for the standard inverse-variance weighted estimator and the JAM implementation with $w = 0$, which are the only two methods which do not perform pleiotropy adjustments. In this "balanced pleiotropy" scenario, the pleiotropic effects cancel out and it is relatively easy to obtain an accurate estimate of the causal effect. The Lasso and MR-Raps methods resulted in the smallest estimated standard errors and mean squared errors, followed by JAM-MR, MR-Presso and the median. However, we note the substantial differences between estimated standard errors and the standard deviation of causal effect estimates. These differences can be observed for several already existing methods, as well as JAM-MR, and translate into confidence intervals with incorrect coverage and Type I error rates above nominal levels. The MR-Egger and median methods were

Table 2. Scenario 1: Balanced pleiotropy. Average causal effect estimates, standard deviation of estimates across replications, estimated standard errors, mean squared errors and empirical Type I error rates for a variety of MR methods.

Method	$\theta = 0$					$\theta = 0.5$				
	Mean	StDev	StdError	MSE	Type I	Mean	StDev	StdError	MSE	Type I
IVW	0.001	0.061	0.060	0.0074	0.063	0.495	0.064	0.061	0.0078	0.065
MR-Egger	-0.013	0.301	0.285	0.1738	0.068	0.397	0.302	0.291	0.1882	0.067
Median (Simple)	0.000	0.023	0.023	0.0011	0.048	0.500	0.034	0.032	0.0022	0.055
Median (Weighted)	0.000	0.024	0.023	0.0011	0.058	0.493	0.033	0.032	0.0022	0.066
Mode (Simple)	0.000	0.028	0.036	0.0021	0.012	0.495	0.041	0.053	0.0046	0.009
Mode (Weighted)	0.000	0.026	0.032	0.0017	0.011	0.480	0.038	0.047	0.0041	0.031
Lasso	0.000	0.020	0.015	0.0006	0.134	0.496	0.027	0.020	0.0012	0.163
MR-Presso	0.000	0.020	0.017	0.0007	0.095	0.496	0.029	0.025	0.0015	0.101
MR-Raps (Simple)	-0.005	0.253	0.014	0.0643	0.205	0.504	0.027	0.019	0.0011	0.165
MR-Raps (Overdispersed)	0.001	0.024	0.018	0.0010	0.132	0.503	0.037	0.031	0.0026	0.106
JAM-MR ($w = 0$)	0.001	0.061	0.060	0.0074	0.063	0.495	0.064	0.061	0.0078	0.065
JAM-MR ($w = 0.1N_1$)	0.001	0.048	0.034	0.0035	0.137	0.496	0.052	0.037	0.0041	0.147
JAM-MR ($w = 0.2N_1$)	0.000	0.034	0.024	0.0017	0.133	0.496	0.038	0.027	0.0022	0.150
JAM-MR ($w = 0.5N_1$)	-0.001	0.022	0.018	0.0008	0.126	0.493	0.031	0.022	0.0015	0.172
JAM-MR ($w = N_1$)	-0.001	0.021	0.017	0.0007	0.122	0.489	0.036	0.023	0.0019	0.220
JAM-MR ($w = 2N_1$)	0.000	0.024	0.017	0.0009	0.142	0.489	0.046	0.025	0.0029	0.269
JAM-MR ($w = 5N_1$)	0.001	0.034	0.020	0.0016	0.220	0.491	0.055	0.033	0.0042	0.263
Oracle	0.000	0.016	0.017	0.0005	0.043	0.495	0.023	0.023	0.0011	0.057

Table 3. Scenario 2: Directional pleiotropy. Average causal effect estimates, standard deviation of estimates across replications, estimated standard errors, mean squared errors and empirical Type I error rates for a variety of MR methods.

Method	$\theta = 0$					$\theta = 0.5$				
	Mean	StDev	StdError	MSE	Type I	Mean	StDev	StdError	MSE	Type I
IVW	0.191	0.036	0.053	0.0407	1.000	0.688	0.040	0.055	0.0402	0.993
MR-Egger	0.004	0.266	0.255	0.1370	0.059	0.401	0.271	0.261	0.1528	0.083
Median (Simple)	0.049	0.021	0.024	0.0034	0.525	0.570	0.032	0.035	0.0072	0.510
Median (Weighted)	0.045	0.022	0.024	0.0031	0.463	0.555	0.032	0.034	0.0052	0.347
Mode (Simple)	0.004	0.025	0.032	0.0017	0.006	0.506	0.038	0.049	0.0040	0.010
Mode (Weighted)	0.005	0.024	0.029	0.0015	0.015	0.490	0.036	0.043	0.0033	0.020
Lasso	0.022	0.019	0.015	0.0011	0.309	0.533	0.029	0.021	0.0023	0.386
MR-Presso	0.058	0.025	0.021	0.0045	0.770	0.562	0.032	0.029	0.0057	0.554
MR-Raps (Simple)	0.236	0.417	0.016	0.2298	0.439	0.530	0.028	0.019	0.0020	0.401
MR-Raps (Overdispersed)	0.056	0.066	0.025	0.0085	0.488	0.600	0.062	0.039	0.0156	0.731
JAM-MR ($w = 0$)	0.191	0.036	0.053	0.0407	1.000	0.688	0.040	0.055	0.0401	0.993
JAM-MR ($w = 0.1N_1$)	0.095	0.041	0.035	0.0121	0.813	0.592	0.046	0.037	0.0121	0.699
JAM-MR ($w = 0.2N_1$)	0.047	0.031	0.024	0.0038	0.464	0.543	0.039	0.028	0.0042	0.368
JAM-MR ($w = 0.5N_1$)	0.018	0.022	0.018	0.0012	0.218	0.512	0.031	0.022	0.0016	0.181
JAM-MR ($w = N_1$)	0.009	0.021	0.016	0.0008	0.165	0.501	0.035	0.023	0.0017	0.189
JAM-MR ($w = 2N_1$)	0.005	0.024	0.017	0.0009	0.160	0.498	0.047	0.026	0.0028	0.257
JAM-MR ($w = 5N_1$)	0.006	0.034	0.020	0.0016	0.233	0.505	0.057	0.033	0.0044	0.258
Oracle	0.000	0.016	0.017	0.0005	0.050	0.496	0.023	0.023	0.0011	0.051

well-calibrated, although in the case of MR-Egger this comes at the cost of much larger standard errors.

Regarding the values of w, our algorithm generally performed best for $w = 0.5N_1$, $w = N_1$ and $w = 2N_1$. When w was set equal to $5N_1$, JAM-MR started penalizing valid genetic variants and converging to fairly small models, exhibiting a small increase in standard errors as a result. The heuristic criterion of selecting the run with the smallest variance estimate suggests that on average the best w values are $w = N_1$ for $\theta = 0$ and $w = 0.5N_1$ for $\theta = 0.5$ (although there were variations in the best w value in individual

Table 4. Scenario 3: Directional pleiotropy with the same direct effect for all pleiotropic variants. Average causal effect estimates, standard deviation of estimates across replications, estimated standard errors, mean squared errors and empirical Type I error rates for a variety of MR methods.

Method	$\theta = 0$					$\theta = 0.5$				
	Mean	StDev	StdError	MSE	Type I	Mean	StDev	StdError	MSE	Type I
IVW	0.192	0.020	0.045	0.0392	1.000	0.686	0.027	0.048	0.0375	1.000
MR-Egger	-0.010	0.223	0.214	0.0963	0.073	0.380	0.225	0.224	0.1155	0.095
Median (Simple)	0.057	0.022	0.026	0.0044	0.602	0.581	0.035	0.039	0.0093	0.554
Median (Weighted)	0.051	0.024	0.025	0.0038	0.500	0.563	0.037	0.037	0.0066	0.381
Mode (Simple)	0.001	0.023	0.028	0.0013	0.018	0.496	0.034	0.047	0.0034	0.010
Mode (Weighted)	0.000	0.021	0.025	0.0011	0.020	0.481	0.031	0.040	0.0030	0.043
Lasso	0.017	0.020	0.016	0.0009	0.235	0.536	0.035	0.022	0.0030	0.404
MR-Presso	0.043	0.027	0.021	0.0030	0.501	0.594	0.041	0.036	0.0117	0.745
MR-Raps (Simple)	0.303	0.323	0.016	0.1958	0.553	0.523	0.035	0.019	0.0021	0.359
MR-Raps (Overdispersed)	0.170	0.018	0.048	0.0315	1.000	0.672	0.026	0.051	0.0328	0.998
JAM-MR ($w = 0$)	0.192	0.020	0.045	0.0392	1.000	0.686	0.027	0.048	0.0375	1.000
JAM-MR ($w = 0.1N_1$)	0.149	0.036	0.040	0.0252	0.979	0.638	0.041	0.042	0.0224	0.939
JAM-MR ($w = 0.2N_1$)	0.085	0.051	0.032	0.0110	0.674	0.578	0.054	0.035	0.0102	0.579
JAM-MR ($w = 0.5N_1$)	0.003	0.020	0.017	0.0007	0.085	0.499	0.036	0.023	0.0019	0.186
JAM-MR ($w = N_1$)	0.001	0.019	0.017	0.0006	0.089	0.487	0.035	0.024	0.0020	0.209
JAM-MR ($w = 2N_1$)	0.000	0.023	0.018	0.0008	0.129	0.487	0.046	0.027	0.0030	0.255
JAM-MR ($w = 5N_1$)	0.002	0.034	0.021	0.0016	0.215	0.497	0.055	0.034	0.0043	0.200
Oracle	0.001	0.016	0.017	0.0005	0.043	0.494	0.024	0.023	0.0012	0.074

Table 5. Scenario 4: InSIDE violation. Average causal effect estimates, standard deviation of estimates across replications, estimated standard errors, mean squared errors and empirical Type I error rates for a variety of MR methods.

Method	$\theta = 0$					$\theta = 0.5$				
	Mean	StDev	StdError	MSE	Type I	Mean	StDev	StdError	MSE	Type I
IVW	0.182	0.038	0.052	0.0372	0.994	0.661	0.039	0.058	0.0307	0.913
MR-Egger	0.014	0.253	0.180	0.0974	0.169	0.217	0.267	0.191	0.1885	0.369
Median (Simple)	0.049	0.021	0.024	0.0035	0.534	0.563	0.029	0.035	0.0061	0.440
Median (Weighted)	0.052	0.026	0.024	0.0039	0.578	0.547	0.034	0.033	0.0045	0.288
Mode (Simple)	0.009	0.091	0.045	0.0241	0.013	0.505	0.083	0.070	0.0518	0.008
Mode (Weighted)	0.005	0.090	0.043	0.0234	0.018	0.492	0.084	0.066	0.0509	0.026
Lasso	0.024	0.020	0.015	0.0012	0.377	0.526	0.029	0.020	0.0019	0.325
MR-Presso	0.058	0.027	0.021	0.0045	0.771	0.551	0.032	0.027	0.0044	0.465
MR-Raps (Simple)	0.358	0.745	0.022	0.6842	0.447	0.665	1.186	0.033	1.4927	0.370
MR-Raps (Overdispersed)	0.056	0.064	0.024	0.0081	0.488	0.572	0.056	0.034	0.0098	0.545
JAM-MR ($w = 0$)	0.179	0.038	0.051	0.0363	0.995	0.657	0.039	0.055	0.0293	0.921
JAM-MR ($w = 0.1N_1$)	0.117	0.046	0.036	0.0172	0.906	0.584	0.043	0.035	0.0101	0.677
JAM-MR ($w = 0.2N_1$)	0.087	0.044	0.030	0.0104	0.794	0.560	0.043	0.030	0.0064	0.506
JAM-MR ($w = 0.5N_1$)	0.040	0.033	0.021	0.0031	0.468	0.526	0.039	0.024	0.0028	0.313
JAM-MR ($w = N_1$)	0.022	0.026	0.017	0.0015	0.306	0.507	0.038	0.022	0.0020	0.247
JAM-MR ($w = 2N_1$)	0.012	0.026	0.017	0.0011	0.230	0.501	0.046	0.024	0.0027	0.288
JAM-MR ($w = 5N_1$)	0.009	0.034	0.020	0.0017	0.234	0.507	0.058	0.031	0.0044	0.296
Oracle	0.000	0.016	0.017	0.0005	0.037	0.495	0.022	0.024	0.0011	0.044

replications).

We also note that the performance of the various Mendelian randomization algorithms relative to each other was similar for $\theta = 0$ and for $\theta = 0.5$. A notable exception is the simple MR-Raps method, which was somewhat unstable for $\theta = 0$ but performed quite well for $\theta = 0.5$.

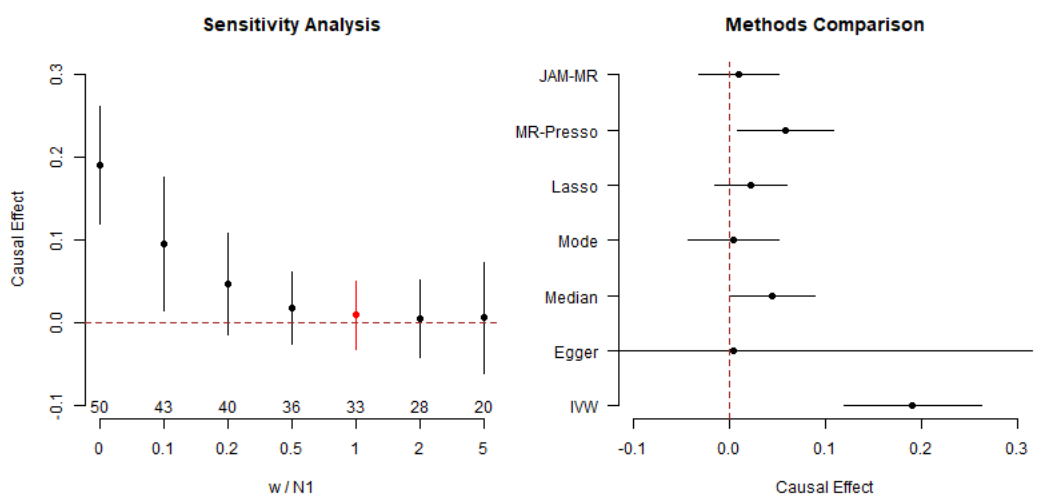
In the directional pleiotropy scenario 2 (Table 3), we observed deviations in the performance of the various methods. Bias in causal effect estimates was observed for the

majority of methods, with mode-based estimation and JAM-MR being the most accurate algorithms. Lasso, JAM-MR and MR-Presso exhibited the smallest standard errors, with simple MR-Raps having small average estimated standard errors, but large discrepancies in causal effect estimates between iterations for $\theta = 0$. Once again, deviations were observed between standard error estimates and the variance of causal effect estimates across simulations. JAM-MR exhibited the smallest mean squared errors. A similar pattern of results was observed in scenarios 3 and 4 (Tables 4 and 5 respectively).

In Fig 2 we visualize the causal effect estimates obtained from the various methods for the directional pleiotropy simulation with $\theta = 0$. We have plotted separately the estimates obtained from JAM-MR implementations using different values for the tuning parameter, to illustrate the use of the algorithm as a sensitivity analysis tool. Estimates for small values of w were much larger than those for moderate and large values, indicating that pleiotropic variants are present in the dataset and cause upwards bias in the causal effect estimate. Without accounting for pleiotropy, we would have to reject the null causal hypothesis at a 95% level. As we increase w towards the minimum-standard-error value (plotted in red), causal effect estimates become unbiased. However, for unnecessarily large values of w the estimates become more variable because the pleiotropic loss function removes some of the valid SNPs and causal effect estimation is based on fewer variants, increasing standard errors and reducing power to detect a causal association.

We have also included in the plot (above the x-axis) the average number of genetic variants assigned an inclusion probability higher than 0.5 for each JAM-MR implementation. Note that the true number of valid instruments in this simulation was 35, which is approximately the number of variants selected by the algorithm for w values with small standard errors.

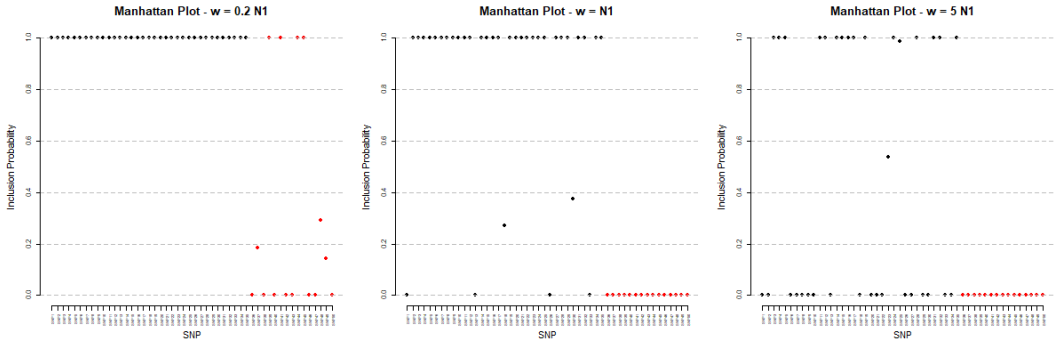
Fig 2. Simulation scenario 2: directional pleiotropy ($\theta = 0$). Causal effect estimates and 95% confidence intervals for JAM-MR implementations with a range of w values (left) and other Mendelian randomization algorithms (right).



The JAM-MR algorithm returns posterior inclusion probabilities for each genetic variant, which can be plotted in a Manhattan plot. In Fig 3 we have plotted the inclusion probabilities for a single implementation of simulation scenario 2 and three JAM-MR runs, with $w = 0.2N_1$, $w = N_1$ and $w = 5N_1$. We have coloured red the

pleiotropic genetic variants. The plot illustrates that our algorithm makes accurate selection of the valid instruments, when properly tuned (center). When a small value of w is used (left), the algorithm may retain some pleiotropic SNPs in the analysis. On the other hand, when w is large (right), the algorithm will correctly downweight pleiotropic variants but may also downweight some of the valid SNPs.

Fig 3. Manhattan plots illustrating the posterior inclusion probabilities assigned to each SNP by the JAM-MR algorithm with $w = 0..2N_1$ (left), $w = N_1$ (center) and $w = 5N_1$ (right), for a single implementation of simulation scenario 2.



Overall, the JAM-MR algorithm was one of the best-performing methods in our simulations. The algorithm proved to be robust to a variety of pleiotropic patterns and yielded accurate causal effect estimates. Together with mode-based estimation it exhibited the smallest bias in our simulations, and it was also one of the algorithms with smallest mean squared error. The main concern regarding the algorithm's performance is that it did not yield nominal Type I error rates. In the Methods section, we propose a simple adjustment for JAM-MR's standard errors to mitigate this issue.

In the appendix of the paper, we have included two additional sets of simulations. In the first set, we assessed the performance of the various Mendelian randomization methods using a larger sample size ($N_1 = N_2 = 200000$) and $P = 100$ genetic variants. In the second set of simulations, we tested how each method is affected by varying the proportion of invalid instruments. We considered 10%, 20% and 40% of the genetic variants being invalid. The results of these simulations exhibited many similarities to those already discussed. JAM-MR could compete with other established Mendelian randomization algorithms and when properly tuned, gave accurate causal effect estimates and low mean squared errors. Among competing approaches, the mode-based method was the most accurate in estimating causal effects, especially when a large proportion of instruments were invalid. Lasso, MR-Presso and MR-Raps attained the smallest MSE, but this sometimes came at the cost of inflated type I error rates.

In further simulations, not reported here, we have considered scenarios with varying SNP-risk factor associations, different numbers of genetic variants, diffferent magnitudes of pleiotropic effects and differnt instrument strength. In all these scenarios, our algorithm provided robust causal effect estimates and was among the best-performing Mendelian randomization methods.

Application: effect of blood pressure on CHD risk

Blood pressure traits and associated genetic variants

We now illustrate the use of the JAM-MR algorithm in a real-data application. We conduct a Mendelian randomization analysis to assess the effect of blood pressure on

the risk of coronary heart disease (CHD). This application has been studied in the past [5, 14, 21] and it is generally accepted that high blood pressure has an increasing effect on the risk of suffering from coronary heart disease, despite the fact that a previous Mendelian randomization analysis [5] did not identify a causal relationship.

Our analysis is novel not in the question it aims to answer but in the data sources it uses. We utilized a recently published meta-GWAS study [22] which identified hundreds of genetic variants associated with blood pressure. The authors meta-analyzed data from the International Consortium for Blood Pressure (ICBP), the UK BioBank, the US Million Veterans Project and the Estonian Genome Center Biobank (EGCUT). In total, a sample of approximately 1 million individuals of European descent was analyzed. The study confirmed previously reported findings about 258 known and 92 reported but not validated genetic variants associated with blood pressure. It also identified a total of 535 novel associations.

We used two blood pressure traits for our analysis, namely systolic and diastolic blood pressure. For each trait, we used all 258 genetic variants with an already established relation to blood pressure, as well as any of the reported-but-not-validated and novel variants reported to be associated with that trait. Among the novel findings, we excluded from consideration variants which were associated with blood pressure in the "discovery" GWAS but not in the "replication" GWAS in [22]. This resulted in a total of $P_1 = 395$ genetic variants for systolic and $P_2 = 391$ genetic variants for diastolic blood pressure; our analysis was therefore based on larger numbers of genetic variants than previous Mendelian randomization investigations.

Genetic associations with systolic and diastolic blood pressure were obtained from the Supplementary Tables of [22]. We used estimates based on the ICBP dataset of $N_1 = 299024$ individuals, as this was the only dataset for which genetic associations were reported for all variants. Since the genetic associations with blood pressure were replicated in independent datasets in [22], winner's curse bias is unlikely to have had a serious effect in our analysis. For the selected variants, we obtained genetic associations with coronary heart disease risk from the CARDIoGRAMplusC4D Consortium [23], based on a sample of $N_2 = 184305$ individuals. The variants were mostly independent, as a result of LD pruning in [22].

Mendelian randomization analysis

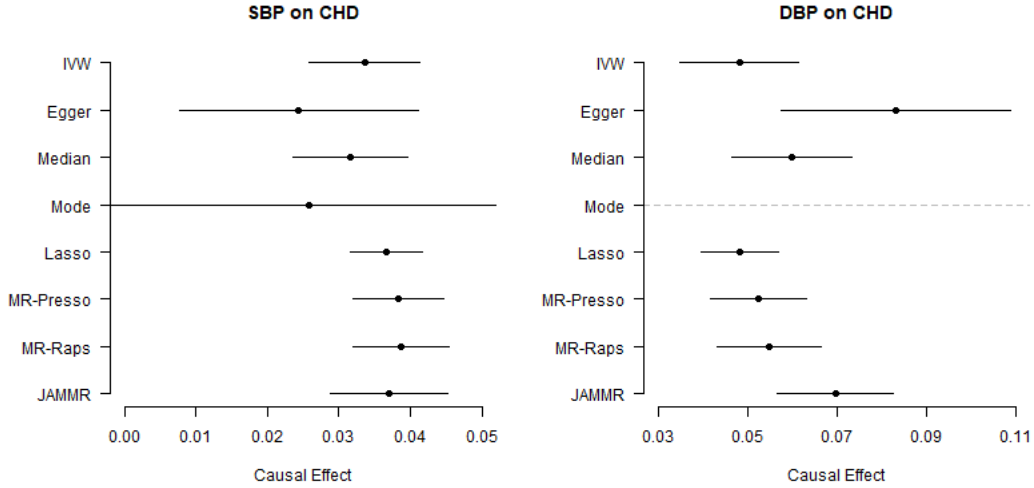
We implemented JAM-MR and the competing Mendelian randomization methods in this dataset to identify the causal effect of systolic and diastolic blood pressure on CHD risk. The results are listed in Table 6, where we report the estimated causal effect (log-odds ratio of increase in CHD risk per mmHg increase in blood pressure measurement) and its 95% confidence interval for each of the two traits and each Mendelian randomization method. Confidence intervals for the JAM-MR algorithm have been computed using the 1.3178 adjustment proposed in the Methods section. A graphical illustration of the results is provided in Fig 4-5. Tables of summary statistics and JAM-MR inclusion probabilities for each SNP can be found in supplementary material, and corresponding Manhattan plots are included in the Appendix.

The reported effects from the various methods confirm that increased blood pressure, both systolic and diastolic, has a deleterious effect on coronary heart disease risk. For systolic blood pressure, the causal effect reported by the various methods was in the region of $0.024 - 0.040$, corresponding to an odds ratio of $e^{0.024} = 1.024$ to $e^{0.04} = 1.041$. All methods were able to reject the null causal hypothesis at a 95% significance threshold, except the mode-based estimator whose standard error was unusually large. JAM-MR estimates were very close to those reported by other methods, and were quite consistent for different values of w, both in terms of causal effect estimates and in terms of which genetic variants received the largest inclusion probabilities. The minimum

Table 6. Log-odds ratios of increase in CHD risk per 1mmHg increase in the corresponding blood pressure measurement. Causal effect estimates and 95 % confidence intervals for a variety of MR methods.

Method	Systolic Blood Pressure			Diastolic Blood Pressure		
	Estimate	95% C.I.		Estimate	95% C.I.	
IVW	0.034	0.026	0.041	0.048	0.035	0.061
MR-Egger	0.024	0.008	0.041	0.083	0.058	0.109
Median (simple)	0.038	0.030	0.046	0.053	0.038	0.068
Median (weighted)	0.032	0.024	0.040	0.060	0.047	0.073
Mode (simple)	0.026	-0.090	0.142	-0.063	-21.735	21.610
Mode (weighted)	0.026	-0.090	0.141	-0.063	-21.698	21.573
Lasso	0.037	0.032	0.042	0.048	0.040	0.057
MR-Presso	0.038	0.032	0.045	0.052	0.042	0.063
MR-Raps (simple)	0.040	0.035	0.044	0.055	0.047	0.063
MR-Raps (overdispersed)	0.039	0.032	0.045	0.055	0.043	0.066
JAM-MR ($w = 0$)	0.031	0.018	0.044	0.055	0.037	0.073
JAM-MR ($w = 0.1N_1$)	0.033	0.022	0.043	0.058	0.043	0.073
JAM-MR ($w = 0.2N_1$)	0.031	0.021	0.041	0.060	0.045	0.074
JAM-MR ($w = 0.5N_1$)	0.031	0.022	0.041	0.064	0.047	0.080
JAM-MR ($w = N_1$)	0.031	0.022	0.041	0.070	0.057	0.083
JAM-MR ($w = 2N_1$)	0.037	0.029	0.045	0.067	0.052	0.081
JAM-MR ($w = 5N_1$)	0.022	0.007	0.038	0.075	0.057	0.094

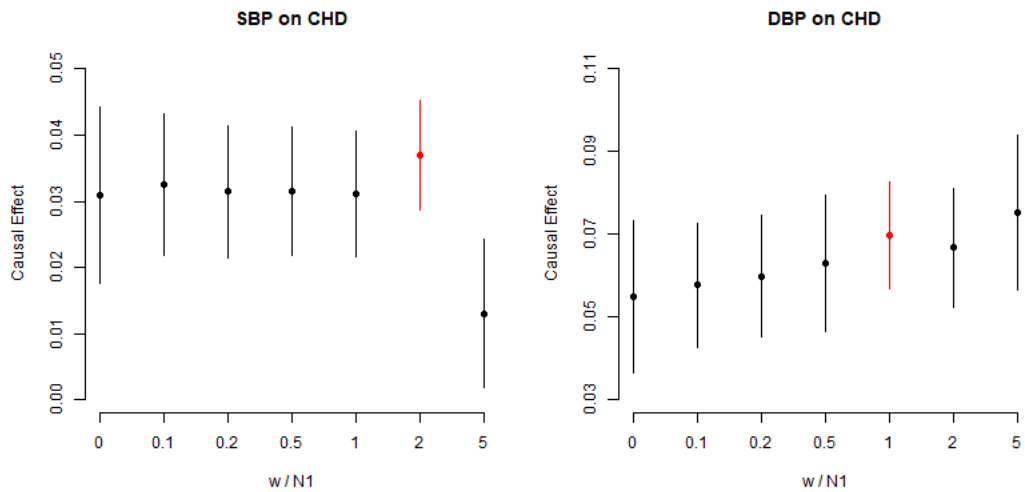
Fig 4. Log-odds ratios of increase in CHD risk per 1mmHg increase in systolic and diastolic blood pressure. Point estimates and 95% confidence intervals for various Mendelian randomization methods. The mode-based method for DBP is not plotted; it resulted in an estimate of -0.063 with 95% confidence interval $(-21.698, 21.573)$.



causal standard error criterion suggested using the implementation with $w = 2N_1$; the corresponding JAM-MR log-odds ratio estimate was 0.037 (95% confidence interval: $(0.029, 0.045)$) .

For diastolic blood pressure, most of the established Mendelian randomization methods reported causal effect estimates between 0.048 and 0.060, corresponding to odds ratios between 1.049 and 1.062. The JAM-MR estimates were slightly larger. For example, setting $w = N_1$ (the value with the smallest standard error) yields a log-odds

Fig 5. Log-odds ratios of increase in CHD risk per 1mmHg increase in systolic and diastolic blood pressure. Point estimates and 95% confidence intervals for JAM-MR implementations with various values of the tuning parameter.



ratio estimate of 0.070 (95% confidence interval $(0.057, 0.083)$) and an effect estimate of $e^{0.07} = 1.073$. Once again, all methods were able to reject the null causal hypothesis at a 95% level, except the mode-based estimator which did not give reasonable results.

Our analysis strongly suggests that both systolic and diastolic blood pressure have a causal role in increasing coronary heart disease risk.

Interpretation of results

It is worth noting that the MR-Egger method rejects the null hypothesis of no causal association and indicates the existence of a causal effect on CHD risk for both blood pressure traits. MR-Egger was the method used to generate the null findings in [5]. This inconsistency can be explained by the fact that the analysis in [5] was based only on a small set of 29 genetic variants and its power to detect a causal association was rather low. Our analysis is based on larger sample sizes and a much larger number of genetic variants associated with blood pressure traits, and these novel variants indicate that there is indeed a causal link between blood pressure and coronary heart disease.

The mode-based estimation method performed poorly, yielding much larger estimated standard errors than other methods. To implement mode-based estimation, we used the default function in the R package "MendelianRandomization". The poor performance was due to the presence of a few genetic variants with extreme outlying effects in our dataset, especially for diastolic blood pressure. These outlying effects were in turn the result of weak instrument bias.

In this example, JAM-MR's variable selection downweighted and removed a significant proportion of the available genetic instruments. For example, the JAM-MR implementation with $w = 2N_1$ for systolic blood pressure assigned posterior inclusion probability greater than 0.5 to 115 of the 395 genetic variants. Of the 280 variants that were assigned low probabilities, 125 had univariate causal effect estimates below the overall reported estimate of 0.037 and 155 had univariate estimates above that value. A similar performance was observed for diastolic blood pressure: for $w = N_1$, 127 of the 391 variants received inclusion probabilities higher than 0.5. Of the remaining 264

variants, 151 had lower and 113 had higher univariate causal effect estimates than the reported value of 0.070. Not all the downweighted variants were penalized because of pleiotropy; for many of them, the association with blood pressure was not deemed strong enough to justify their inclusion into the high-probability JAM-MR models. Nevertheless, the fact that several genetic variants with both small and large univariate causal effect estimates were assigned small inclusion probabilities suggests that this application resembles a "balanced pleiotropy" setting similar to that of simulation scenario 1. In line with that observation, the inverse-variance weighted method provided accurate causal effect estimates for both blood pressure traits but resulted in wider confidence intervals compared to some of the other Mendelian randomization methods due to the presence of pleiotropic variants.

In the simulations of Table 6, we assumed that the genetic variants are independent. While this is mostly true, there were still a few genetic variants in our dataset coming from the same DNA region. Since JAM can incorporate genetic correlations, we repeated the analysis using data from the UK Biobank as a reference dataset from which to estimate genetic correlations. The results of this JAM-MR implementation were very similar to those reported in Table 6 and are not reported here. Similar results were also obtained when we restricted our analysis only to the 258 genetic variants that had been known to be associated with blood pressure prior to the GWAS in [22]; the main difference was that effect estimates for diastolic blood pressure were slightly larger for many of the previously proposed Mendelian randomization methods and closer to the results reported by JAM-MR in Table 6. This could be an indication that some of the novel variants for diastolic blood pressure discovered in [22] exhibit pleiotropic effects on CHD in the risk-decreasing direction, and this may have caused a slight decrease in causal effect estimates from other methods. The sensitivity plot of Fig 5 illustrates the same pattern: for small values of the tuning parameter, the JAM-MR causal effect estimates for diastolic blood pressure are subject to slight downwards bias. The best-w causal effect estimate was less affected because the algorithm's variable selection in this example was rather harsh, downweighting a large proportion of the genetic variants included in the analysis; this included many of the "new" genetic variants.

Discussion

In this paper, we have developed a new algorithm for causal effect estimation in Mendelian randomization when some of the candidate instruments are pleiotropic. Our algorithm uses Bayesian variable selection to identify sets of genetic variants with homogeneous causal effect estimates, and model averaging among these sets to estimate the overall causal effect of exposure on outcome, while accounting for uncertainty in the selection of instruments to use. A wide range of simulation studies demonstrate how using model averaging to account for uncertainty in pleiotropic selections leads to more robsut inference than other current methods. The real data application is the first time a very recently published large scale meta-GWAS has been used to instrument blood pressure for CHD.

Compared to other approaches for Mendelian randomization with pleiotropic variants, the JAM-MR algorithm has a number of attractive features. The use of model averaging provides robust causal effect estimation, allows many variants to have small contributions to the overall causal effect estimate and offers uncertainty quantification for genetic associations with the risk factor. Unlike the competing approaches discussed in this paper, our algorithm uses the Bayesian framework which allows for incorporating prior information on the biological function of SNPs. JAM-MR's stochastic search procedure is quite flexible and can efficiently explore the large parameter space of causal configurations; as a result, the algorithm can be used with large numbers of genetic

variants.

Our algorithm also provides a natural framework for incorporating genetic correlations into a Mendelian randomization analysis and selecting the most relevant variants from a densely genotyped region. Common approaches for Mendelian randomization typically assume that genetic variants are independent. In related work, we are investigating the advantages of utilizing JAM's variable selection compared to pruning and other approaches for Mendelian randomization with correlated instruments, in order to incorporate multiple correlated effects from regions which harbour complex genetic signals for the trait of interest.

Limitations of our approach include the algorithm's issues with calibration of standard errors. The correction proposed in this paper behaves well in practical applications but does not come with any theoretical guarantees of good performance and further research is needed to fully address these issues. Alternatives could include directly fitting a truncated normal random-effects model to univariate causal effect estimates, or using bootstrap techniques. It would also be useful to devise an efficient automatic procedure for specifying the tuning parameter w. Running JAM-MR with several w values can help visualize how pleiotropic variants affect the causal effect estimate (as illustrated in Fig 2), but comes with an increased computational cost and it would be desirable to obtain an accurate causal effect estimate based on a single implementation of the algorithm. Finally, another extension would be to construct a fully Bayesian version of the JAM-MR algorithm, by fitting a model for SNP-outcome associations similar to (2)-(3) that contains the causal effect as a parameter.

In conclusion, JAM-MR performs pleiotropy-robust causal effect estimation for Mendelian randomization. Our algorithm has a number of desirable features, most notably model-averaged inference. It exhibits good performance in simulations and has been used to implement a Mendelian randomization analysis of the effect of blood pressure on CHD risk, using a recent large-scale blood pressure meta-GWAS. We therefore hope that JAM-MR will become a valuable addition to the Mendelian randomization literature.

The JAM-MR algorithm has been implemented in R as part of the GitHub package `R2BGLiMS`, available at https://github.com/pjnewcombe/R2BGLiMS.

Methods

The JAM algorithm

Introduction

Here, we provide a brief outline of the JAM algorithm. The reader is referred to [16] for a more detailed description.

The JAM algorithm is primarily a tool for fine-mapping densely genotyped DNA regions. Let X denote a trait of interest and let $G_1, \ldots, G_P$ be a set of (possibly correlated) genetic variants to be tested for association with X. The main purpose of JAM is to identify a subset of genetic variants that are robustly and independently associated with the trait.

For individual i, let $g_i = (g_{i1}, \ldots, g_{iP})$ and x_i be the allele counts and trait measurements respectively, and denote $G = (g_{ij})$ the genetic matrix. We assume that trait values and allele counts per variant have been centered. Typically, the "individual-level data" x_i, g_i are not available in practice. Instead, we only have access to a set of univariate association estimates $\hat{\beta}_{Xj}$ between each variant and the trait, as well as the corresponding standard errors $\hat{s}_{Xj}$.

The JAM model

JAM uses linear regression to model the trait: if all P genetic variants were assumed to be associated with X, the JAM algorithm would model the trait as

$$x_i|\beta, \sigma_X^2, G = \sum_{j=1}^{P} g_{ij}\beta_j + \epsilon_{Xi} \quad , \quad \epsilon_{Xi} \sim N(0, \sigma_X^2) \quad , \quad i = 1, \ldots, N \quad . \tag{4}$$

In practice JAM implements variable selection, reflecting that, in many applications, only a subset of the variants should be used to model the trait. Let $\gamma \in \{0,1\}^P$ index the selected subset, so that if $\gamma_j = 1$ variant G_j is included into the JAM model and if $\gamma_j = 0$ it is not. Using β_γ, G_γ to denote the subsets of β, G only for the variants G_j for which $\gamma_j = 1$, (4) becomes

$$x_i|\beta_\gamma, \sigma_X^2, \gamma, G_\gamma = \sum_{j:\gamma_j=1} g_{ij}\beta_j + \epsilon_{Xi} \quad , \quad \epsilon_{Xi} \sim N(0, \sigma_X^2) \quad . \tag{5}$$

Eq (5) can be used to build a likelihood for the individual-level data, $p(x|\beta_\gamma, \sigma_X^2, \gamma, G) = \prod_i p(x_i|\beta_\gamma, \sigma_X^2, \gamma, g_i)$. One can also obtain the marginal model likelihood, $p(x|\gamma, G) = \int p(x|\beta_\gamma, \sigma_X^2, \gamma, G) d\beta_\gamma d\sigma_X^2$. JAM works by constructing summary-data approximations to these two likelihoods (see the following subsections).

Prior specification

The likelihood $p(x|\beta_\gamma, \sigma_X^2, \gamma, G)$ is complemented with a set of priors in order to perform Bayesian inference. For the genetic associations β_γ, JAM uses a conjugate g-prior,

$$\beta_\gamma|\sigma_X^2, \gamma \quad \sim \quad N\left(0, \sigma_X^2 \tau (G_\gamma^T G_\gamma)^{-1}\right) \quad , \tag{6}$$

where τ is a constant. By default, the algorithm sets $\tau = \max(P^2, N)$, as has been previously recommended by various authors [24, 25]. The residual variance σ_X^2 is assigned its own conjugate prior, which is an Inverse-Gamma density,

$$\sigma_X^2 \quad \sim \quad IG(a_X, b_X) \quad , \tag{7}$$

for fixed a_X, b_X. Finally, JAM uses a Beta-Binomial prior on the space of all possible models,

$$p(\gamma) = \frac{B(a_\omega + P_\gamma, b_\omega + P - P_\gamma)}{B(a_\omega, b_\omega)} \quad , \tag{8}$$

where $B(a, b)$ denotes the Beta function and P_γ denotes the size of model γ. The relative sizes of the hyperparameters a_ω, b_ω reflect the proportion of genetic variants expected a priori to be associated with the trait; by default, $a_\omega = 1$, $b_\omega = P$, which correspond to an expectation that a single variant will be associated with the trait.

Posterior inference for the regression parameters

A Bayesian posterior distribution over models, genetic associations and residual variance can be obtained according to the standard principles of Bayesian inference:

$$p(\beta_\gamma, \sigma_X^2, \gamma | x, G) \propto p(x|\beta_\gamma, \sigma_X^2, \gamma, G) p(\beta_\gamma|\sigma_X^2, \gamma) p(\sigma_X^2|\gamma) p(\gamma) \quad . \tag{9}$$

Conditional on a particular model, i.e. combination of causal variants, posterior inference on the regression parameters β_γ, σ_X^2 can be conducted using known results for Bayesian linear regression with conjugate priors, which yield

$$\sigma_X^2 | x, \gamma \quad \sim \quad IG\left(a_X + \frac{N}{2} \ , \ b_X + \frac{s^2}{2} + \frac{\hat{\beta}_\gamma^T G_\gamma^T G_\gamma \hat{\beta}_\gamma}{2(\tau+1)}\right) \tag{10}$$

$$\beta_\gamma | x, \sigma_X^2, \gamma \quad \sim \quad N\left(\frac{\tau}{1+\tau}\hat{\beta}_\gamma \ , \ \frac{\tau}{\tau+1}\sigma_X^2 (G_\gamma^T G_\gamma)^{-1}\right) \tag{11}$$

where $\hat{\beta}_\gamma = (G_\gamma^T G_\gamma)^{-1} G_\gamma^T x$ and $s^2 = (x - G_\gamma \hat{\beta}_\gamma)^T (x - G_\gamma \hat{\beta}_\gamma)$.

Posterior model selection

An advantage of the linear regression setting is that it allows for fast and efficient variable selection. The regression coefficients β_γ and the residual variance σ_X^2 can be integrated out from the JAM likelihood $p(x|\beta_\gamma, \sigma_X^2, \gamma)$ to obtain the marginal model likelihood

$$p(x|\gamma) \propto (\tau+1)^{-\frac{P_\gamma}{2}} (2b_\sigma + S(\gamma))^{-a_\sigma - \frac{N}{2}} \quad , \tag{12}$$

where

$$S(\gamma) = x^T x - \frac{\tau}{\tau+1} x^T G_\gamma (G_\gamma^T G_\gamma)^{-1} G_\gamma^T x \quad . \tag{13}$$

This leads to the marginal model posterior, $p(\gamma|x) \propto p(x|\gamma)p(\gamma)$. The normalizing constant of that density can be difficult to evaluate, but this can be avoided by using reversible-jump MCMC [26]. JAM implements a standard reversible-jump algorithm with addition, deletion and swapping of genetic variants as possible moves. The stochastic search algorithm allows exploration of an unrestricted model space, without the need to set limits on the maximum number of causal variants. Consequently, JAM is able to efficiently explore complex causal configurations among large numbers of genetic variants. Posterior model probabilities can be estimated by the proportion of iterations JAM spends in each model.

JAM with summarized data

One of JAM's main advantages is that it does not require access to individual-level data. Variable selection is implemented according to Eq (12), (13), which depend on the observed data x_i, g_i only through the quantities $x^T x$, $G^T x$ and $G^T G$. These quantities are sufficient summary statistics for linear regression. In the setting of genome-wide

association studies, [16] describes a way of approximating $z = G^T x$ from the univariate variant-trait association estimates $\hat{\beta}_{Xj}$, $j = 1, \ldots, P$, as well as effect allele frequencies. In addition, note that $x^T x \approx (N_1 - 1)\hat{\sigma}_X^2$ since we have assumed that trait values have been centered before implementing JAM. Here, $\hat{\sigma}_X^2$ is an estimator of the trait variance $\mathrm{Var}(X)$ measured in the GWAS for the trait X (this is typically reported by genetic association studies). [16] uses a slightly different approximation to the JAM likelihood which avoids the need to estimate the trait variance, but providing the trait variance to the algorithm improves the accuracy of approximation, particularly when many genetic variants have significant effects on the trait. Finally, the matrix $G^T G$ models ($N - 1$ times) the genetic correlations between variants $G_1, \ldots, G_P$ and can be approximated if a reference dataset, such as the 1000 Genomes dataset or the UK Biobank, is available. This yields the following summary-data approximation for $S(\gamma)$:

$$S(\gamma) \approx (N-1)\hat{\sigma}_X^2 - \frac{\tau}{\tau+1}\hat{z}^T (G_{\gamma,\mathrm{ref}}^T G_{\gamma,\mathrm{ref}})^{-1} \hat{z} \quad . \tag{14}$$

The model-specific marginal posteriors (10) and (11) can be approximated by summary GWAS data in a similar way.

An extension of JAM for Mendelian randomization

Scope

We now describe our extension to the standard JAM algorithm that facilitates pleiotropy adjustment and accurate causal effect estimation in Mendelian randomization studies, which we call JAM-MR (JAM for Mendelian Randomization).

We start by pointing out a difference in the scope of our new algorithm compared to the original JAM algorithm. Traditionally, JAM is used to analyze correlated variants from one (fine-mapping) or multiple genetic regions. On the other hand, Mendelian randomization studies typically use variants from across the whole genome, often pruned for independence. Consequently, many practical implementations of JAM-MR will rely on independent, genome-wide significant variants. Mendelian randomization analyses using correlated variants are less common in practice, and assessing pleiotropy in these studies can be difficult because many genetic variants may share a common pleiotropic effect through correlation. Although JAM-MR is applicable in such studies too, we mainly consider the "traditional" Mendelian randomization framework in this paper.

For the JAM-MR algorithm, we work in the context of two-sample summary-data Mendelian randomization [27]. The two-sample summary-statistics framework is a common approach for Mendelian randomization, since it allows researchers to leverage the power and large sample sizes of large consortia GWAS studies that are already available for many important traits.

JAM takes as inputs the univariate genetic variant-risk factor associations $\hat{\beta}_X$, the trait variance $\hat{\sigma}_X^2$ and a reference matrix from which to compute genetic correlations. The two-sample summary-data Mendelian randomization framework further assumes the availability of univariate variant-outcome association estimates $\hat{\beta}_{Yj}$ and corresponding standard errors $\hat{s}_{Yj}$, obtained from a separate genetic association study. The coefficients $\hat{\beta}_{Xj}$ and $\hat{\beta}_{Yj}$ can be used to obtain variant-specific estimates of the causal effect θ according to a ratio formula [28],

$$\hat{\theta}_j = \frac{\hat{\beta}_{Yj}}{\hat{\beta}_{Xj}} \quad , \quad \mathrm{s.e.}\left(\hat{\theta}_j\right) = \frac{\hat{s}_{Yj}}{\hat{\beta}_{Xj}} \quad . \tag{15}$$

Sometimes a second-order formula is used to estimate the variant-specific standard

errors,

$$\text{s.e.}\left(\hat{\theta}_j\right) = \sqrt{\frac{\hat{s}_{Yj}^2}{\hat{\beta}_{Xj}^2} + \frac{\hat{\beta}_{Yj}^2 \hat{s}_{Xj}^2}{\hat{\beta}_{Xj}^4}} \quad . \tag{16}$$

Bayesian inference with loss functions

The JAM-MR algorithm implements a recently proposed framework for performing Bayesian inference using loss functions [20]. This new theoretical framework constitutes a generalization of the core Bayesian paradigm. For a dataset $\mathcal{D}$, a parameter vector θ and a prior distribution $\pi(\theta)$, the standard Bayesian updating scheme, $p(\theta|\mathcal{D}) \propto \pi(\theta)p(\mathcal{D}|\theta)$, is replaced with a loss-function update of the form

$$p_\ell(\theta|\mathcal{D}) \propto \pi(\theta) \exp^{-w\ell(\mathcal{D},\theta)} \quad , \tag{17}$$

where $\ell(\mathcal{D}|\theta)$ represents a loss function. If $w = 1$ and $\ell(\mathcal{D}, \theta) = -\log p(\mathcal{D}|\theta)$ is the negative log-likelihood, we obtain traditional Bayesian inference.

The motivation behind this approach is that the loss function can be used to tailor Bayesian inference towards specific objectives. For example, if the objective of interest is classification, the misclassification error can be used as a loss function [29]. In addition, using a loss function instead of a likelihood avoids the need for the Bayesian statistician to specify a full data-generating model; this is especially useful when the object of interest is a low-dimensional parameter of a complex, high-dimensional model. A rigorous decision-theoretic framework for the new updating scheme is provided in [20].

JAM with a pleiotropic loss function

In order to construct the JAM-MR algorithm, we use a slight modification of the Bayesian loss function framework. Specifically, we use both a likelihood and a loss function to construct the "loss-posterior" $p_\ell(\theta|\mathcal{D})$:

$$p_\ell(\theta|\mathcal{D}) \propto \pi(\theta)p(\mathcal{D}|\theta) \exp\left\{-w\ell(\mathcal{D},\theta)\right\} \quad . \tag{18}$$

This is equivalent to using $\ell(\mathcal{D}, \theta) - \frac{1}{w} \log p(\mathcal{D}|\theta)$ as a loss function in (17). The use of both a likelihood and a loss function in JAM-MR is justified because the objective of the model selection procedure is two-fold: we use the likelihood to select genetic variants strongly associated with the risk factor and the loss function to penalize variants which exhibit pleiotropic effects on the outcome. The parameter w can be interpreted as a tuning parameter that balances the impact of the pleiotropic loss on model selection relative to that of the JAM likelihood and the prior. Note that the loss function framework allows us to avoid making specific modelling assumptions for the SNP-outcome association, since that association is only modelled through the loss function.

In the context of JAM-MR, the data to be used are the univariate summary statistics along with the reference dataset from which to compute genetic correlations, so $\mathcal{D} = \{\hat{\beta}_X, \hat{\beta}_Y, \hat{s}_Y, G_{\text{ref}}\}$. For the purpose of model selection, and since β_γ and σ_X^2 can be integrated out, the parameter vector is simply the model indicator $\theta = \gamma$ and the likelihood in (18) is the marginal model likelihood (12).

We now discuss how to specify the loss function $\ell(\mathcal{D}, \gamma)$. It is common in the Mendelian randomization literature to use some measure of heterogeneity between univariate causal effect estimates as a proxy for pleiotropic behaviour [7, 10]. The intuition is that genetic variants which are valid instruments yield the same univariate causal effect estimates, up to some random variation. On the other hand, estimates based on pleiotropic variants can exhibit systematic differences, especially if the variants

operate on different causal pathways towards the outcome, because the estimated causal effects depend on the strength and direction of the pleiotropic G-Y association. This suggests that our loss function should upweight models with homogeneous univariate causal effect estimates, as such models are likely to contain valid instruments, and downweight models with heterogeneous estimates, as at least some of the genetic variants contained in them are likely to be pleiotropic.

Consequently, in order to penalize pleiotropic models, we use loss functions that measure heterogeneity of univariate estimates. A simple option is the variance of the univariate causal effect estimates,

$$\ell_1(\hat{\theta}, \gamma) = \mathrm{Var}\left(\hat{\theta}_j \ : \ \gamma_j = 1\right) = \frac{1}{P_\gamma - 1} \sum_{j:\gamma_j=1} \left(\hat{\theta}_j - \hat{\theta}_\gamma\right)^2 \quad , \tag{19}$$

where $\hat{\theta}_\gamma = \frac{1}{P_\gamma} \sum_{j:\gamma_j=1} \hat{\theta}_j$ is the mean of the univariate causal effect estimates in model γ. The loss function downweights (in terms of posterior probability) models with heterogeneous causal effect estimates, since the variance of such estimates is larger and the term $\exp\{-w\ell(\mathcal{D}, \theta)\}$ in (18) is smaller. Note that the loss function (19) is identically zero for models containing only one genetic variant. Such models carry no evidence as to whether the variant included is valid or pleiotropic. Therefore, we ignore these models and restrict JAM-MR to only consider models with at least two genetic variants.

An alternative loss function can be obtained by weighting the individual causal effect estimates in (19) by the inverse of their squared standard errors,

$$\ell_2(\hat{\theta}, \gamma) = \sum_{j:\gamma_j=1} \left(\hat{\theta}_j - \hat{\theta}_{\gamma,IVW}\right)^2 \mathrm{s.e.}\left(\hat{\theta}_j\right)^{-2} \quad .$$

where $\hat{\theta}_{\gamma,IVW}$ is the inverse-variance weighted causal effect estimate based on variants in model γ,

$$\hat{\theta}_{\gamma,IVW} = \frac{\sum_{j\,:\,\gamma_j=1} \hat{\theta}_j \mathrm{s.e.}\left(\hat{\theta}_j\right)^{-2}}{\sum_{j\,:\,\gamma_j=1} \mathrm{s.e.}\left(\hat{\theta}_j\right)^{-2}} \quad . \tag{20}$$

A similar function was used in [10] to obtain a pleiotropy-robust exhaustive-search model averaging procedure for Mendelian randomization. In practice, we have found that the two loss functions often yield similar results. In the simulations and the real-data application, we have used the variance loss (19).

In conclusion, the Bayesian loss-posterior for JAM-MR's variable selection is

$$\begin{aligned} p_\ell(\gamma|\hat{\beta}_X, \hat{\beta}_Y, G_{\mathrm{ref}}) \quad \propto \quad & \frac{B(a_\omega + P_\gamma, b_\omega + P - P_\gamma)}{B(a_\omega, b_\omega)} \ (\tau+1)^{-\frac{P_\gamma}{2}} (2b_\sigma + S(\gamma))^{-a_\sigma - \frac{N_1}{2}} \\ & \times \ \exp\left\{-w \frac{1}{P_\gamma - 1} \sum_{j:\gamma_j=1} \left(\hat{\theta}_j - \hat{\theta}_\gamma\right)^2\right\} \quad , \end{aligned} \tag{21}$$

where $S(\gamma)$ can be computed from (14). Note that N_1 represents the sample size from which the G-X associations are obtained (the "first sample" in two-sample Mendelian randomization).

The algorithm's stochastic search procedure scales well to large datasets containing hundreds of genetic variants, is quite flexible and can explore large parts of the model space. JAM-MR's probabilistic output (posterior probabilities for each model) also offers the potential to use Bayesian model averaging for causal effect estimation.

Since we are using heterogeneity as a proxy for pleiotropic behaviour, our algorithm implicitly makes a plurality assumption, similar to that made in [7, 10]. The algorithm will assign a high posterior probability to a large set of pleiotropic variants with consistent univariate causal effect estimates, especially if the size of that set is larger than the number of valid SNPs. As a result, model-averaged causal effect estimates will be biased. Our algorithm allows such cases to be detected by inspecting the list of posterior model probabilities: if two very different models are both assigned high posterior probabilities, there is evidence for the presence of a set of genetic variants with similar pleiotropic effects. This is typically the result of many pleiotropic variants acting on the same causal pathway to the outcome, and the biological interpretation of such sets can be an interesting question in applications. We note however that JAM-MR often requires a very large number of iterations in order to identify different sets of variants with homogeneous effects.

Causal effect estimation using JAM-MR

The JAM-MR posterior probabilities can subsequently be used to obtain an overall estimate of the causal effect of interest, according to a model averaging procedure. The use of Bayesian model averaging enables quantification of uncertainty related to the choice of instruments and allows many genetic variants to have small contributions to the overall causal effect estimate.

For each model γ visited by the algorithm, a model-specific inverse variance weighted estimate $\hat{\theta}_{\gamma,IVW}$ can be computed according to (20). The variance of this estimator is

$$\hat{s}^2_{\gamma,IVW} = \frac{1}{\sum_{j:\gamma_j=1} \text{s.e.}\left(\hat{\theta}_j\right)^{-2}} \quad . \tag{22}$$

This implicitly corresponds to a fixed-effects model for the variance. In practice, a multiplicative random-effects model is often preferred, where $\hat{\theta}_j \sim N(\theta, \phi^2 \hat{s}^2_j)$ instead of $\hat{\theta}_j \sim N(\theta, \hat{s}^2_j)$ and the IVW variance (22) is multiplied by ϕ^2. The overdispersion parameter ϕ can be estimated using weighted linear regression [30].

The model-specific IVW estimates and their standard errors can be combined into a single estimator,

$$\hat{\theta}_{\text{JAM-MR}} = \sum_{\gamma} p(\gamma|x)\hat{\theta}_{\gamma,IVW} \quad , \tag{23}$$

with variance

$$\hat{s}^2_{\text{JAM-MR}} = \sum_{\gamma} p(\gamma|x)\hat{s}^2_{\gamma,IVW} + \sum_{\gamma} p(\gamma|x)\hat{\theta}_{\gamma,IVW}\left(\hat{\theta}_{\gamma,IVW} - \hat{\theta}_{\text{JAM-MR}}\right) \quad , \tag{24}$$

where the summation is over all models γ assigned positive posterior probability by the variable selection procedure. Eq (24) can be derived as an approximation to the posterior variance $\text{Var}(\theta|\mathcal{D})$, by expressing it in terms of the model-specific posterior moments $\mathbb{E}(\theta|\mathcal{D},\gamma)$, $\text{Var}(\theta|\mathcal{D},\gamma)$ and approximating these by $\hat{\theta}_{\gamma,IVW}$, $\hat{s}^2_{\gamma,IVW}$.

This algorithm is not fully Bayesian. Although it utilizes JAM's Bayesian variable selection to obtain posterior model probabilities and facilitate model averaging, the process of causal effect estimation for each model is conducted using classical inverse-variance-weighted formulas. A fully Bayesian algorithm for Mendelian randomization would require further modelling assumptions for the causal effect parameter and return a posterior sample for that parameter. This is an interesting potential extension of JAM-MR, but is beyond the scope of this paper.

Tuning the algorithm

The tuning parameter w plays a crucial role in JAM-MR's variable selection. Tuning w is subject to a bias-variance tradeoff. For relatively small values, the pleiotropic loss function has limited effect on the variable selection procedure and JAM-MR tends to favour larger models. These models may still include some pleiotropic variants, and the resulting causal effect estimates may exhibit bias. On the other hand, with a large value of w the algorithm favours models that contain no pleiotropic variants but may also ignore some of the valid instruments. In this case JAM-MR yields unbiased causal effect estimates, but these estimates may have large standard errors.

Setting $w = 0$ is equivalent to the standard JAM algorithm with no pleiotropy adjustment; the algorithm is similar to a simple IVW estimator except it downweights genetic variants that are weakly associated with the risk factor. The other extreme case is to assign a very large value to w. In this case the algorithm becomes very selective and converges to models containing only two genetic variants.

Heuristic approaches for tuning w [20, 31] rely on the idea of "balancing" the magnitude of the loss function relative to the likelihood. For the JAM-MR algorithm, we have empirically observed in simulations that the optimal value for the tuning parameter w depends primarily on the sample size N_1, and secondarily on the number P of genetic variants in the analysis and the proportion of genetic variation in the risk factor. Note that these quantities appear in the JAM likelihood (12) - (14) (the proportion of genetic variation is a function of the risk factor variance, the genetic effects on the risk factor and effect allele frequencies).

The algorithm's dependence on the tuning parameter w can be potentially useful. Running JAM-MR with multiple values of the tuning parameter can help detect and visualize the effect of pleiotropy on the causal effect estimate, as illustrated in Fig 2. Comparisons with the results obtained from other Mendelian randomization approaches can also be made. We recommend running our algorithm with several w values similar to a grid search. Values used in the grid search should be expressed as multiples of the sample size N_1 of the GWAS from which the G-X associations are obtained. This is because the JAM likelihood scales with N_1 and the tuning process aims to balance the effect of the loss function against the likelihood.

A simple heuristic to determine the best value of the grid search and obtain a single causal effect estimate is to choose the implementation with the smallest causal standard error (24). Both the inclusion of genetic variants with pleiotropic effects and the removal of valid genetic variants is likely to increase the standard error, so the minimum standard error implementation is likely to include the valid instruments.

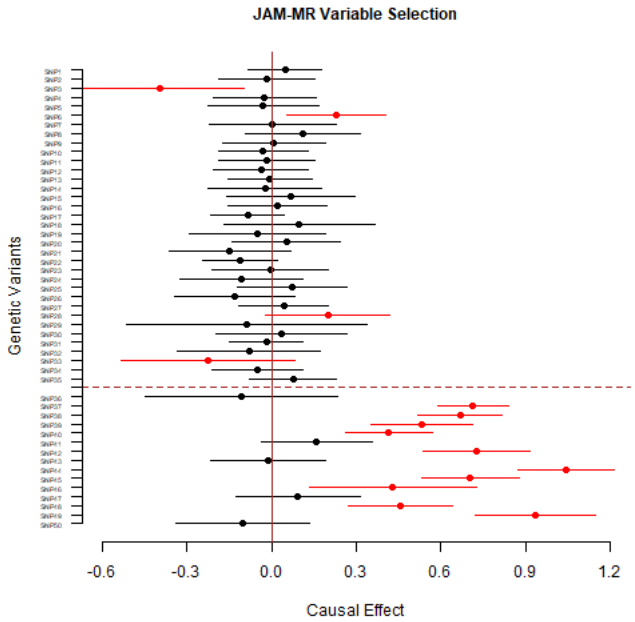
Accuracy of JAM-MR standard error estimates

As reflected in our simulations, the estimated JAM-MR standard errors (24) may yield confidence intervals with Type I error rates above nominal levels. The reason why this happens is illustrated in the funnel plot of Fig 6. The plot shows causal effect estimates and standard errors for 50 genetic variants in one replication of simulation scenario 2. Variants 1-35 were generated as valid instruments and variants 36-50 were generated to be pleiotropic. We have also coloured in red the variants that were assigned by JAM-MR a posterior inclusion probability lower than 0.5.

JAM-MR uses a random effects framework, $\hat{\theta}_j \sim N(\theta, \phi^2 \hat{s}_j^2)$, to derive the causal standard error for each model visited. The parameter ϕ models the spread of univariate causal effect estimates around the true causal effect θ, ignoring the finite-sample variations which are modelled by the $\hat{s}_j$.

Since JAM-MR uses heterogeneity of univariate causal effect estimates as a proxy for pleiotropic behaviour, it is likely to downweight genetic variants which are valid

Fig 6. JAM-MR variable selection in practice. A single implementation of a directional pleiotropy simulation (scenario 2) with $\theta = 0$. Causal effect estimates and 95% confidence intervals for each variant are plotted. Variants above the dotted line were simulated as valid and variants below the dotted line were simulated as pleiotropic. Variants assigned a posterior inclusion probability lower than 50% are coloured red.



instruments, but whose causal effect estimates happen to be far from θ due to random variation. This can be confirmed in Fig 6, where valid genetic variants with outlying causal effect estimates are assigned low posterior probabilities. As a result, the dispersion parameter ϕ may be underestimated. This problem is more likely to occur when JAM-MR identifies a large number of variants as pleiotropic. In fact, the opposite scenario is also possible. If a large number of mildly pleiotropic variants are present, JAM-MR may be unable to remove all of them from consideration and include some of them in its top models. In this case, overestimation of the parameter ϕ and the variance can occur.

A simple and somewhat naive, yet often useful, way of correcting the problem is to rescale the variance by a constant in order to improve coverage of confidence intervals. We have found empirically that rescaling by

$$\left(1 - 2\frac{z_{0.975}}{0.95\sqrt{2\pi}}e^{-\frac{1}{2}z_{0.975}^2}\right)^{-1} \approx 1.3178$$

can improve the coverage of confidence intervals; this value corresponds to estimating the variance based on a multiplicative random-effects model with truncated normal standard errors, with truncation at the 2.5-th and 97.5-th quantiles. This adjustment does not work well for all values of the tuning parameter w (this is the reason why we did not use it in the results of Tables 2-5) but is likely to work well for the smallest standard error values. In addition, when the variants deemed pleiotropic by JAM-MR are only a small proportion of the overall number of genetic instruments, there is often no need for adjustment.

The improvement in coverage of JAM-MR confidence intervals for the simulations

presented earlier is illustrated in Table 7.

Table 7. Standard deviation of causal effect estimates across replications, estimated and adjusted standard errors for the best (i.e. minimum standard error) JAM-MR implementations in Tables 2-5, along with corresponding Type I error rates.

Sim Scenario	Best Fit	StDev	StdError	Adj SE	Type I	Adj Type I
Scen 1 - $\theta = 0$	$w = N_1$	0.0207	0.0165	0.0217	0.122	0.038
Scen 1 - $\theta = 0.5$	$w = 0.5N_1$	0.0311	0.0223	0.0294	0.172	0.073
Scen 2 - $\theta = 0$	$w = N_1$	0.0213	0.0164	0.0216	0.165	0.068
Scen 2 - $\theta = 0.5$	$w = 0.5N_1$	0.0311	0.0225	0.0296	0.181	0.079
Scen 3 - $\theta = 0$	$w = N_1$	0.0190	0.0169	0.0222	0.089	0.020
Scen 3 - $\theta = 0.5$	$w = 0.5N_1$	0.0362	0.0234	0.0309	0.186	0.078
Scen 4 - $\theta = 0$	$w = 2N_1$	0.0260	0.0169	0.0223	0.230	0.119
Scen 4 - $\theta = 0.5$	$w = N_1$	0.0379	0.0222	0.0293	0.247	0.132

Pleiotropy-robust Mendelian randomization

Existing methods

Mendelian randomization in the presence of invalid instruments is a very active area of research. A wide range of statistical techniques have been used to either identify the pleiotropic variants and remove them from the analysis, or robustify the process of causal effect estimation. One of the most widely used approaches is MR-Egger regression [5], which yields consistent causal effect estimates under the assumption that instrument strength is independent of the genetic variants' direct effects on the outcome (InSIDE). Another common approach is to estimate the causal effect of interest by the (weighted or unweighted) median of univariate estimates for all available genetic variants. This median estimator is robust to outlying pleiotropic effects and is asymptotically unbiased if more than 50% of the available genetic variants are valid instruments.

More recent approaches have utilized a wide range of statistical techniques to address the problem of invalid instruments. Mode-based estimation [27] fits a kernel density to the univariate causal effect estimates and uses the mode of that density as the overall estimate; this approach weakens the majority assumption of the median and yields accurate causal effect estimates if only a plurality of genetic instruments are valid. Other authors have relied on variable selection to identify the pleiotropic variants: [8, 9] use Lasso regularization while [11] performs outlier detection and deletion. Finally, [12] relies on profile likelihood and robust regression techniques for causal effect estimation.

The list of methods we have considered in our simulation study is not exhaustive. For example, we have not implemented the heterogeneity penalization approach presented in [10]. Similar to JAM-MR, this approach relies on model averaging, but instead of our algorithm's stochastic search it uses exhaustive search over possible models, and is therefore only applicable for small numbers of genetic variants. We have also not considered the mixture modelling method of [13] and the modified Q-statistics approach of [14].

Several algorithms for Mendelian randomization in the presence of pleiotropic variants were compared in a recent review [15].

It is also worth mentioning that Reversible-Jump MCMC has been used for instrumental variables analysis in [32]. An important difference with our work is that the algorithms developed in [32] rely on the availability of individual-level data, while in this paper we only require access to summarized data for the instrument-risk factor and instrument-outcome associations. Some modelling assumptions are also different: [32] do not use the Bayesian loss function framework as JAM-MR does.

Implementation of Mendelian randomization methods in the simulation study

For the implementation of the various Mendelian randomization methods in our simulation study, we used available R packages. In particular, we used the packages "MendelianRandomization" (for the IVW, MR-Egger, median and mode-based methods), "MRPRESSO" (for MR-Presso) and "mr.raps" (for MR-Raps). For the Lasso method, we used the R code provided in the Appendix of [9].

For the median and mode-based methods we implemented both an unweighted and a weighted version. For the lasso method we used the "heterogeneity" approach described in [9] to specify a value for the tuning parameter. MR-Raps was implemented using the Tukey loss function and either a simple or an overdispersed model. For the other methods, we used the default settings in the corresponding R packages. For the "oracle" method, we estimated the causal effect of interest by IVW, using only the genetic variants that were valid instruments.

For the implementation of JAM-MR we did not simulate a reference dataset since genetic variants were assumed to be independent. Instead, the genetic correlation matrix $G^T G$ was approximated by setting the off-diagonal elements equal to zero and computing the diagonal elements as $(G^T G)_{jj} = 2N_1 f_j(1 - f_j)$ where f_j is the effect allele frequency.

The computational cost of implementing JAM-MR was slightly higher than that for most other methods. Nevertheless, the cost is not prohibitive in an absolute scale: running the algorithm for 10^6 iterations with 50 genetic variants and a single value of w only required a few seconds. When using multiple w values, the algorithm has to be run separately for each value; in the modern era, many researchers have access to high performance computing, in which case the algorithm can be run in parallel for several w values in the same amount of time as would be required for a single w. The cost increases when modelling correlated data, due to the need to perform matrix operations with the genetic correlation matrix.

Supporting information

S1 Appendix. Additional simulations. Contains the setting and results of additional simulations conducted, varying the sample size and the proportion of pleiotropic SNPs.

S1 Table. Blood Pressure SNPs Contains summary statistics and JAM-MR selection probabilities for the genetic variants associated with systolic and diastolic blood pressure in the real-data application.

Acknowledgements

This work was supported by the UK Medical Research Council (Core Medical Research Council Biostatistics Unit Funding Code: MC UU 00002/7). Apostolos Gkatzionis and Paul Newcombe were supported by a Medical Research Council Methodology Research Panel grant (Grant Number RG88311). Stephen Burgess was supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 204623/Z/16/Z).

References

1. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology. 2003;32(1):1 – 22. doi:10.1093/ije/dyg070.

2. Boef AGC, Dekkers OM, le Cessie S. Mendelian randomization studies: a review of the approaches used and the quality of reporting. International Journal of Epidemiology. 2015;44(2):496–511. doi:10.1093/ije/dyv071.

3. Kamstrup PR, Tybjærg-Hansen A, Steffensen R, Nordestgaard BG. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. JAMA. 2009;301(22):2331–2339. doi:10.1001/jama.2009.801.

4. CCGC (CRP CHD Genetics Collaboration). Association between C reactive protein and coronary heart disease: Mendelian randomisation analysis based on individual participant data. British Medical Journal. 2011;342. doi:10.1136/bmj.d548.

5. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International Journal of Epidemiology. 2015;44(2):512–525. doi:10.1093/ije/dyv080.

6. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genetic Epidemiology. 2016;40(4):301–314. doi:10.1002/gepi.21965.

7. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomisation via the zero modal pleiotropy assumption. International Journal of Epidemiology. 2017;46(6):1985–1998. doi:10.1093/ije/dyx102.

8. Kang H, Zhang A, Cai T, Small D. Instrumental variables estimation with some invalid instruments, and its application to Mendelian randomisation. Journal of the American Statistical Association. 2016;111(513):132–144. doi:10.1080/01621459.2014.994705.

9. Burgess S, Bowden J, Dudbridge F, Thompson SG. Assessing the effectiveness of robust instrumental variable methods using multiple candidate instruments with application to Mendelian randomization; 2018.

10. Burgess S, Zuber V, Gkatzionis A, Foley CN. Modal-based estimation via heterogeneity-penalized weighting: model averaging for consistent and efficient estimation in Mendelian randomization when a plurality of candidate instruments are valid. International Journal of Epidemiology. 2018;47(4):1242–1254. doi:10.1093/ije/dyy080.

11. Verbanck M, Chen CY, Neele B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nature Genetics. 2018;50(3):693–698.

12. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score; 2018.

13. Qi G, Chatterjee N. Mendelian Randomization Analysis Using Mixture Models (MRMix) for Genetic Effect-Size-Distribution Leads to Robust Estimation of Causal Effects; 2018.

14. Bowden J, Del Greco M F, Minelli C, Zhao Q, Lawlor DA, Sheehan NA, et al. Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. International Journal of Epidemiology. 2018;doi:10.1093/ije/dyy258.

15. Slob EAW, Burgess S. A Comparison Of Robust Mendelian Randomization Methods Using Summary Data. bioRxiv. 2019;doi:10.1101/577940.

16. Newcombe PJ, Conti DV, Richardson S. JAM: A Scalable Bayesian Framework for Joint Analysis of Marginal SNP Effects. Genetic Epidemiology. 2016;40(3):188 – 201. doi:10.1002/gepi.21953.

17. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics. 2016;32(10):1493–1501. doi:10.1093/bioinformatics/btw018.

18. Chen W, Larrabee BR, Ovsyannikova IG, Kennedy RB, Haralambieva IH, Poland GA, et al. Fine Mapping Causal Variants with an Approximate Bayesian Method Using Marginal Test Statistics. Genetics. 2015;200(3):719–736. doi:10.1534/genetics.115.176107.

19. Dadaev T, Saunders EJ, Newcombe PJ, Anokian E, Leongamornlert DA, Brook MN, et al. Fine-mapping of prostate cancer susceptibility loci in a large meta-analysis identifies candidate causal variants. Nature Communications. 2018;9(1):2256. doi:10.1038/s41467-018-04109-8.

20. Bissiri PG, Holmes CC, Walker SG. A General Framework for Updating Belief Distributions. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2016;78(5):1103–1130. doi:10.1111/rssb.12158.

21. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nature Genetics. 2017;49(3):403–415. doi:10.1038/ng.3768.

22. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nature Genetics. 2018;50:1412–1425. doi:10.1038/s41588-018-0205-x.

23. CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nature Genetics. 2015;47(10):1121–1130.

24. George EI, McCulloch RE. Approaches for Bayesian variable selection. Statistica Sinica. 1997;7(2):339–373.

25. Fernandez C, Ley E, Steel MF. Benchmark priors for Bayesian model averaging. Journal of Econometrics. 2001;100(2):381–427.

26. Green PJ. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. Biometrika. 1995;82(4):711–732. doi:10.1093/biomet/82.4.711.

27. Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. International Journal of Epidemiology. 2016;45(6):1717–1726. doi:10.1093/ije/dyx028.

28. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Statistical Methods in Medical Research. 2017;26(5):2333–2355. doi:10.1177/0962280215597579.

29. Jiang W, Tanner MA. Gibbs posterior for variable selection in high-dimensional classification and data mining. The Annals of Statistics. 2008;36(5):2207–2231.

30. Burgess S, Bowden J. Integrating summarized data from multiple genetic variants in Mendelian randomization: bias and coverage properties of inverse-variance weighted methods; 2016.

31. Holmes CC, Walker SG. Assigning a value to a power likelihood in a general Bayesian model. Biometrika. 2017;104(2):497–503. doi:10.1093/biomet/asx010.

32. Koop G, Leon-Gonzalez R, Strachan R. Bayesian model averaging in the instrumental variable regression model. Journal of Econometrics. 2012;171(2):237 – 250. doi:http://dx.doi.org/10.1016/j.jeconom.2012.06.005.

Appendix - additional simulations

Further to the simulations presented in the paper, we conducted two additional sets of simulations to investigate the performance of JAM-MR with varying sample sizes and different numbers of pleiotropic instruments. In the first set, we implemented simulations similar to those in Tables 2-5 with a larger sample size of $N_1 = N_2 = 200000$. We generated $P = 100$ genetic variants, of which $P_1 = 30$ were pleiotropic. Genetic effects were simulated in the interval $(0.025, 0.05)$, which loosely corresponds to genome-wide significant variants with p-values between 10^{-100} and 10^{-8}. Pleiotropic effects were drawn uniformly at random from the interval $(-0.05, 0.05)$ in the first scenario, $(0, 0.05)$ in the second and fourth scenario, and were fixed at 0.03 in the third scenario. Genetic effects on the confounder for pleiotropic variants in scenario 4 were drawn uniformly from the interval $(-0.05, 0.05)$. Otherwise, the simulations were the same as those in the main body of the paper.

The simulations were repeated 1000 times. As in Tables 2-5, we report the mean causal effect estimate, standard error, theoretical standard error (computed as the standard deviation of causal effect estimates among the 1000 replications), Mean Squared Error and Type I error rate (for $\theta = 0.5$, we define this to be the empirical probability of rejecting $H_0 : \theta = 0.5$) at a 95% significance level. The results are presented in Tables 8-11.

In the second set of simulations, we considered variations of scenarios 1-4 with different numbers of pleiotropic variants. We considered simulations with 10%, 20% and 40% of the genetic variants being pleiotropic (this corresponds to 5, 10 and 20 variants respectively). Three simulation experiments were performed for each of the four simulation scenarios: balanced pleiotropy, directional pleiotropy, directional pleiotropy with common direct effects and directional pleiotropy with InSIDE violation. This resulted in 12 sets of simulations, and 1000 replications were performed for each set. Except for the number of pleiotropic variants, the simulations were identical to those presented in the paper. Results are reported in Tables 12-15.

For both sets of simulations, the results were similar to those in the main part of the paper. JAM-MR was generally able to accurately estimate the causal effect of interest and, when properly tuned, was among the methods with the smallest bias across the simulations. This was the case even for a fairly large (40%) proportion of invalid instruments. The "smallest causal standard error" criterion performed well, typically selecting a JAM-MR implementation with small bias and mean squared error.

Among the established Mendelian randomization methods, mode-based estimation emerged as the main competitor to JAM-MR in terms of accuracy of causal effect estimates. This was especially the case in simulations with many pleiotropic variants, where several other methods exhibited severe biases. This can be explained by the fact that the mode-based method makes rather weak assumptions on the proportion of pleiotropic variants (it only requires a plurality of instruments to be valid, instead of the more common majority assumption) and hence it is more likely to work well when many instruments are invalid. We note, however, that the method's process of computing standard errors was rather unstable in simulation scenario 4 with $N_1 = N_2 = 200000$.

In terms of magnitude of Mean Squared Errors, the Lasso method competed with JAM-MR and was among the best methods, closely followed by MR-Presso and (in some simulations) MR-Raps. However, we note that these methods come with their own issues related to coverage of confidence intervals and Type I error rate inflation. On the other hand, the MR-Egger method had well-calibrated Type I error rates, although this came at the cost of excessively wide confidence intervals. The median method also had well-calibrated standard errors, especially in simple simulations (balanced pleiotropy and/or 10% invalid instruments), but Type I error rates were inflated due to bias in more challenging simulations.

We conclude by pointing out that the various Mendelian randomization methods make different assumptions and work well in different simulation scenarios. In practice, it is useful as a sensitivity analysis tool to implement multiple methods and compare their results.

Table 8. Simulation A1: Balanced pleiotropy. $P = 100$ genetic variants, $N_1 = N_2 = 200000$.

Method	$\theta = 0$					$\theta = 0.5$				
	Mean	StDev	StdError	MSE	Type I	Mean	StDev	StdError	MSE	Type I
IVW	-0.001	0.043	0.042	0.0036	0.068	0.496	0.045	0.043	0.0039	0.060
MR-Egger	-0.010	0.210	0.210	0.0889	0.040	0.455	0.216	0.214	0.0952	0.052
Median (Simple)	0.000	0.012	0.012	0.0003	0.052	0.500	0.017	0.017	0.0006	0.044
Median (Weighted)	0.000	0.012	0.012	0.0003	0.060	0.495	0.017	0.016	0.0006	0.056
Mode (Simple)	0.001	0.015	0.020	0.0006	0.008	0.498	0.022	0.030	0.0014	0.012
Mode (Weighted)	0.000	0.014	0.018	0.0005	0.008	0.490	0.020	0.027	0.0012	0.012
Lasso	0.000	0.009	0.008	0.0001	0.096	0.499	0.014	0.010	0.0003	0.128
MR-Presso	0.000	0.010	0.008	0.0002	0.068	0.498	0.014	0.012	0.0004	0.100
MR-Raps (Simple)	-0.020	0.517	0.008	0.2671	0.364	0.501	0.013	0.010	0.0003	0.148
MR-Raps (Overdispersed)	0.001	0.009	0.006	0.0001	0.128	0.501	0.017	0.011	0.0004	0.140
JAM-MR ($w = 0$)	-0.001	0.043	0.042	0.0036	0.068	0.496	0.045	0.043	0.0039	0.060
JAM-MR ($w = 0.1N_1$)	0.001	0.035	0.024	0.0018	0.180	0.499	0.037	0.025	0.0020	0.176
JAM-MR ($w = 0.2N_1$)	0.000	0.021	0.016	0.0007	0.140	0.499	0.023	0.017	0.0008	0.172
JAM-MR ($w = 0.5N_1$)	0.001	0.012	0.010	0.0002	0.072	0.497	0.016	0.012	0.0004	0.116
JAM-MR ($w = N_1$)	0.000	0.010	0.008	0.0002	0.108	0.496	0.015	0.011	0.0004	0.140
JAM-MR ($w = 2N_1$)	0.001	0.010	0.008	0.0002	0.104	0.495	0.016	0.011	0.0004	0.179
JAM-MR ($w = 5N_1$)	0.001	0.012	0.009	0.0002	0.165	0.494	0.024	0.013	0.0008	0.313
Oracle	0.001	0.008	0.008	0.0001	0.052	0.498	0.012	0.012	0.0003	0.060

Table 9. Simulation A2: Directional pleiotropy. $P = 100$ genetic variants, $N_1 = N_2 = 200000$.

Method	$\theta = 0$					$\theta = 0.5$				
	Mean	StDev	StdError	MSE	Type I	Mean	StDev	StdError	MSE	Type I
IVW	0.192	0.025	0.037	0.0387	1.000	0.689	0.023	0.038	0.0376	1.000
MR-Egger	0.001	0.194	0.184	0.0718	0.084	0.418	0.194	0.188	0.0802	0.084
Median (Simple)	0.037	0.012	0.012	0.0017	0.852	0.550	0.016	0.018	0.0031	0.812
Median (Weighted)	0.033	0.012	0.012	0.0014	0.792	0.540	0.016	0.017	0.0022	0.636
Mode (Simple)	0.004	0.014	0.018	0.0005	0.012	0.502	0.021	0.027	0.0012	0.012
Mode (Weighted)	0.004	0.014	0.016	0.0005	0.016	0.495	0.019	0.024	0.0010	0.008
Lasso	0.015	0.010	0.008	0.0004	0.484	0.522	0.015	0.011	0.0008	0.496
MR-Presso	0.065	0.019	0.013	0.0048	1.000	0.553	0.017	0.016	0.0033	0.940
MR-Raps (Simple)	0.535	0.466	0.010	0.5032	0.700	0.514	0.013	0.009	0.0005	0.364
MR-Raps (Overdispersed)	0.009	0.017	0.007	0.0004	0.312	0.539	0.050	0.015	0.0043	0.532
JAM-MR ($w = 0$)	0.192	0.025	0.037	0.0387	1.000	0.689	0.023	0.038	0.0376	1.000
JAM-MR ($w = 0.1N_1$)	0.101	0.028	0.024	0.0115	1.000	0.594	0.029	0.025	0.0103	0.988
JAM-MR ($w = 0.2N_1$)	0.050	0.020	0.016	0.0031	0.908	0.545	0.022	0.017	0.0028	0.692
JAM-MR ($w = 0.5N_1$)	0.019	0.012	0.010	0.0006	0.472	0.515	0.016	0.012	0.0006	0.268
JAM-MR ($w = N_1$)	0.010	0.011	0.008	0.0003	0.252	0.505	0.014	0.011	0.0004	0.136
JAM-MR ($w = 2N_1$)	0.006	0.010	0.008	0.0002	0.160	0.500	0.017	0.011	0.0004	0.198
JAM-MR ($w = 5N_1$)	0.004	0.014	0.009	0.0003	0.209	0.500	0.025	0.013	0.0008	0.310
Oracle	0.001	0.009	0.008	0.0001	0.068	0.498	0.011	0.012	0.0003	0.032

Table 10. Simulation A3: Directional pleiotropy with common direct effects for pleiotropic variants. $P = 100$ genetic variants, $N_1 = N_2 = 200000$.

Method	$\theta = 0$					$\theta = 0.5$				
	Mean	StDev	StdError	MSE	Type I	Mean	StDev	StdError	MSE	Type I
IVW	0.230	0.013	0.037	0.0545	1.000	0.728	0.016	0.038	0.0538	1.000
MR-Egger	-0.019	0.183	0.184	0.0678	0.044	0.454	0.180	0.189	0.0703	0.048
Median (Simple)	0.041	0.011	0.013	0.0020	0.932	0.556	0.016	0.020	0.0038	0.888
Median (Weighted)	0.036	0.012	0.013	0.0016	0.820	0.547	0.017	0.019	0.0028	0.724
Mode (Simple)	-0.001	0.013	0.015	0.0004	0.012	0.497	0.019	0.022	0.0008	0.016
Mode (Weighted)	-0.001	0.012	0.014	0.0003	0.004	0.491	0.017	0.020	0.0008	0.040
Lasso	0.004	0.010	0.008	0.0002	0.128	0.506	0.014	0.011	0.0004	0.148
MR-Presso	0.047	0.011	0.010	0.0024	1.000	0.538	0.019	0.015	0.0020	0.684
MR-Raps (Simple)	0.771	0.101	0.011	0.6052	0.992	0.499	0.012	0.009	0.0002	0.152
MR-Raps (Overdispersed)	0.096	0.038	0.024	0.0112	0.960	0.698	0.030	0.040	0.0417	0.996
JAM-MR ($w = 0$)	0.230	0.013	0.037	0.0545	1.000	0.728	0.016	0.038	0.0538	1.000
JAM-MR ($w = 0.1N_1$)	0.154	0.035	0.032	0.0259	1.000	0.653	0.035	0.032	0.0258	1.000
JAM-MR ($w = 0.2N_1$)	0.033	0.033	0.016	0.0025	0.418	0.535	0.036	0.019	0.0029	0.425
JAM-MR ($w = 0.5N_1$)	0.000	0.008	0.008	0.0001	0.044	0.494	0.013	0.011	0.0003	0.107
JAM-MR ($w = N_1$)	0.000	0.008	0.008	0.0001	0.044	0.493	0.014	0.011	0.0004	0.144
JAM-MR ($w = 2N_1$)	0.000	0.009	0.008	0.0002	0.088	0.493	0.016	0.012	0.0004	0.162
JAM-MR ($w = 5N_1$)	0.000	0.013	0.009	0.0002	0.129	0.495	0.025	0.014	0.0008	0.286
Oracle	0.000	0.008	0.008	0.0001	0.032	0.496	0.012	0.012	0.0003	0.076

Table 11. Simulation A4: InSIDE violation. $P = 100$ genetic variants, $N_1 = N_2 = 200000$.

Method	$\theta = 0$					$\theta = 0.5$				
	Mean	StDev	StdError	MSE	Type I	Mean	StDev	StdError	MSE	Type I
IVW	0.161	0.028	0.035	0.0281	1.000	0.592	0.032	0.043	0.0114	0.588
MR-Egger	0.245	0.134	0.089	0.0863	0.688	0.522	0.185	0.111	0.0473	0.236
Median (Simple)	0.027	0.012	0.012	0.0010	0.572	0.526	0.017	0.017	0.0012	0.340
Median (Weighted)	0.051	0.018	0.013	0.0031	0.964	0.525	0.022	0.017	0.0014	0.404
Mode (Simple)	0.023	0.168	1.180	106.6855	0.000	0.521	0.243	1.322	38.0741	0.004
Mode (Weighted)	0.013	0.143	1.180	106.6483	0.004	0.511	0.246	1.321	38.0731	0.008
Lasso	0.017	0.012	0.008	0.0005	0.512	0.511	0.016	0.010	0.0005	0.308
MR-Presso	0.053	0.021	0.011	0.0034	0.980	0.525	0.019	0.013	0.0011	0.452
MR-Raps (Simple)	1.881	9.251	0.594	118.0576	0.776	2.911	49.030	11.464	18561.3280	0.700
MR-Raps (Overdispersed)	0.009	0.019	0.006	0.0005	0.308	0.508	0.020	0.010	0.0006	0.316
JAM-MR ($w = 0$)	0.161	0.028	0.032	0.0278	1.000	0.592	0.032	0.037	0.0108	0.754
JAM-MR ($w = 0.1N_1$)	0.133	0.029	0.024	0.0192	1.000	0.558	0.028	0.020	0.0045	0.736
JAM-MR ($w = 0.2N_1$)	0.119	0.030	0.022	0.0156	1.000	0.547	0.027	0.019	0.0033	0.660
JAM-MR ($w = 0.5N_1$)	0.092	0.030	0.019	0.0097	0.992	0.532	0.026	0.016	0.0020	0.496
JAM-MR ($w = N_1$)	0.060	0.028	0.015	0.0046	0.923	0.519	0.024	0.014	0.0012	0.377
JAM-MR ($w = 2N_1$)	0.030	0.019	0.010	0.0014	0.687	0.506	0.026	0.012	0.0008	0.371
JAM-MR ($w = 5N_1$)	0.015	0.017	0.009	0.0006	0.459	0.497	0.030	0.012	0.0010	0.431
Oracle	0.000	0.008	0.008	0.0001	0.044	0.498	0.013	0.012	0.0003	0.080

Table 12. Simulation A5: Balanced pleiotropy. Various numbers of pleiotropic instruments.

Method	10% invalid				20% invalid				40% invalid			
	Mean	SE	MSE	Type I	Mean	SE	MSE	Type I	Mean	SE	MSE	Type I
IVW	-0.002	0.036	0.0027	0.059	0.001	0.049	0.0051	0.060	0.000	0.068	0.0098	0.067
MR-Egger	0.002	0.171	0.0636	0.067	0.006	0.236	0.1215	0.066	-0.019	0.326	0.2266	0.062
Median (Simple)	0.000	0.021	0.0008	0.043	0.001	0.022	0.0009	0.044	0.001	0.025	0.0014	0.081
Median (Weighted)	0.000	0.020	0.0008	0.044	0.001	0.021	0.0009	0.037	0.000	0.024	0.0013	0.072
Mode (Simple)	0.000	0.040	0.0025	0.006	0.002	0.038	0.0022	0.009	0.001	0.034	0.0019	0.010
Mode (Weighted)	0.000	0.036	0.0021	0.008	0.001	0.034	0.0019	0.009	0.000	0.030	0.0016	0.012
Lasso	0.000	0.014	0.0005	0.112	0.001	0.015	0.0005	0.096	0.001	0.016	0.0008	0.156
MR-Presso	0.000	0.014	0.0004	0.085	0.000	0.015	0.0005	0.075	0.001	0.019	0.0010	0.129
MR-Raps (Simple)	-0.002	0.013	0.0021	0.094	-0.010	0.014	0.0376	0.138	0.019	0.014	0.0967	0.289
MR-Raps (Overdispersed)	0.000	0.014	0.0004	0.080	0.001	0.014	0.0005	0.106	0.002	0.041	0.0048	0.095
JAM-MR ($w = 0$)	-0.002	0.036	0.0027	0.059	0.001	0.049	0.0051	0.060	0.000	0.068	0.0098	0.067
JAM-MR ($w = 0.1N_1$)	0.000	0.022	0.0012	0.081	0.001	0.028	0.0022	0.124	-0.001	0.040	0.0052	0.178
JAM-MR ($w = 0.2N_1$)	0.000	0.017	0.0007	0.080	0.002	0.020	0.0011	0.103	0.000	0.027	0.0021	0.137
JAM-MR ($w = 0.5N_1$)	0.000	0.015	0.0005	0.082	0.001	0.016	0.0006	0.086	0.001	0.019	0.0010	0.137
JAM-MR ($w = N_1$)	0.000	0.015	0.0005	0.092	0.001	0.015	0.0006	0.099	0.001	0.018	0.0009	0.131
JAM-MR ($w = 2N_1$)	0.000	0.015	0.0006	0.114	0.001	0.016	0.0007	0.107	0.001	0.019	0.0011	0.173
JAM-MR ($w = 5N_1$)	0.000	0.017	0.0011	0.213	0.001	0.019	0.0012	0.199	0.000	0.023	0.0019	0.227
Oracle	0.000	0.015	0.0004	0.058	0.001	0.016	0.0005	0.037	0.001	0.018	0.0006	0.042

Table 13. Simulation A6: Directional pleiotropy. Various numbers of pleiotropic instruments.

Method	10% invalid				20% invalid				40% invalid			
	Mean	SE	MSE	Type I	Mean	SE	MSE	Type I	Mean	SE	MSE	Type I
IVW	0.065	0.035	0.0061	0.416	0.127	0.046	0.0193	0.954	0.256	0.059	0.0707	1.000
MR-Egger	0.007	0.168	0.0594	0.058	-0.017	0.220	0.1008	0.056	-0.008	0.279	0.1627	0.071
Median (Simple)	0.013	0.021	0.0009	0.064	0.028	0.022	0.0017	0.214	0.081	0.028	0.0080	0.864
Median (Weighted)	0.012	0.020	0.0009	0.059	0.025	0.022	0.0015	0.195	0.073	0.027	0.0069	0.784
Mode (Simple)	0.002	0.040	0.0024	0.002	0.002	0.036	0.0020	0.006	0.008	0.030	0.0015	0.018
Mode (Weighted)	0.001	0.036	0.0020	0.006	0.002	0.032	0.0018	0.006	0.008	0.027	0.0013	0.029
Lasso	0.009	0.014	0.0005	0.133	0.012	0.015	0.0007	0.177	0.045	0.017	0.0029	0.703
MR-Presso	0.007	0.014	0.0005	0.091	0.023	0.017	0.0012	0.294	0.116	0.030	0.0157	0.987
MR-Raps (Simple)	0.011	0.014	0.0098	0.096	0.131	0.015	0.1420	0.250	0.371	0.017	0.3419	0.674
MR-Raps (Overdispersed)	0.004	0.014	0.0004	0.081	0.008	0.014	0.0006	0.145	0.191	0.053	0.0442	0.963
JAM-MR ($w = 0$)	0.065	0.035	0.0061	0.415	0.127	0.046	0.0193	0.954	0.256	0.059	0.0707	1.000
JAM-MR ($w = 0.1N_1$)	0.028	0.022	0.0018	0.208	0.058	0.028	0.0052	0.542	0.139	0.040	0.0236	0.935
JAM-MR ($w = 0.2N_1$)	0.015	0.017	0.0009	0.133	0.029	0.021	0.0019	0.275	0.067	0.027	0.0069	0.650
JAM-MR ($w = 0.5N_1$)	0.006	0.015	0.0005	0.082	0.011	0.016	0.0007	0.136	0.026	0.019	0.0018	0.310
JAM-MR ($w = N_1$)	0.004	0.015	0.0005	0.102	0.006	0.015	0.0006	0.110	0.013	0.018	0.0011	0.206
JAM-MR ($w = 2N_1$)	0.002	0.015	0.0006	0.110	0.003	0.016	0.0007	0.119	0.007	0.019	0.0012	0.186
JAM-MR ($w = 5N_1$)	0.002	0.017	0.0010	0.197	0.002	0.019	0.0013	0.219	0.008	0.023	0.0021	0.253
Oracle	0.001	0.015	0.0004	0.038	0.000	0.016	0.0005	0.047	0.000	0.018	0.0007	0.046

Table 14. Simulation A7: Directional pleiotropy with common direct effects for all pleiotropic variants. Various numbers of pleiotropic instruments.

Method	10% invalid				20% invalid				40% invalid			
	Mean	SE	MSE	Type I	Mean	SE	MSE	Type I	Mean	SE	MSE	Type I
IVW	0.064	0.031	0.0053	0.559	0.127	0.040	0.0181	0.997	0.256	0.048	0.0682	1.000
MR-Egger	-0.001	0.148	0.0436	0.052	-0.005	0.190	0.0739	0.055	-0.002	0.229	0.1057	0.061
Median (Simple)	0.014	0.021	0.0009	0.066	0.031	0.023	0.0018	0.231	0.098	0.034	0.0116	0.903
Median (Weighted)	0.012	0.021	0.0009	0.060	0.027	0.022	0.0016	0.197	0.091	0.032	0.0109	0.837
Mode (Simple)	0.000	0.038	0.0023	0.005	0.000	0.033	0.0017	0.006	0.001	0.026	0.0011	0.017
Mode (Weighted)	0.001	0.034	0.0019	0.008	-0.001	0.029	0.0014	0.006	0.001	0.023	0.0010	0.023
Lasso	0.006	0.014	0.0005	0.095	0.007	0.015	0.0005	0.092	0.054	0.020	0.0042	0.682
MR-Presso	0.002	0.014	0.0004	0.060	0.011	0.015	0.0006	0.131	0.129	0.037	0.0201	0.942
MR-Raps (Simple)	0.000	0.014	0.0004	0.055	0.030	0.014	0.0237	0.136	0.611	0.018	0.3811	0.993
MR-Raps (Overdispersed)	0.000	0.014	0.0004	0.050	0.074	0.034	0.0080	0.668	0.246	0.053	0.0639	1.000
JAM-MR ($w = 0$)	0.064	0.031	0.0053	0.560	0.127	0.040	0.0181	0.997	0.256	0.048	0.0682	1.000
JAM-MR ($w = 0.1N_1$)	0.042	0.025	0.0030	0.355	0.089	0.034	0.0101	0.817	0.219	0.046	0.0512	0.998
JAM-MR ($w = 0.2N_1$)	0.024	0.020	0.0015	0.177	0.049	0.026	0.0044	0.436	0.138	0.039	0.0256	0.815
JAM-MR ($w = 0.5N_1$)	0.002	0.015	0.0005	0.052	0.003	0.016	0.0006	0.068	0.003	0.019	0.0010	0.082
JAM-MR ($w = N_1$)	0.000	0.015	0.0005	0.056	0.000	0.016	0.0005	0.071	0.000	0.019	0.0008	0.093
JAM-MR ($w = 2N_1$)	0.000	0.015	0.0005	0.079	0.000	0.016	0.0007	0.108	0.001	0.020	0.0011	0.143
JAM-MR ($w = 5N_1$)	0.001	0.017	0.0010	0.185	0.000	0.019	0.0012	0.201	0.002	0.024	0.0019	0.194
Oracle	0.000	0.015	0.0004	0.029	0.000	0.016	0.0005	0.039	0.000	0.018	0.0007	0.042

Table 15. Simulation A8: InSIDE Violation. Various numbers of pleiotropic instruments.

Method	10% invalid				20% invalid				40% invalid			
	Mean	SE	MSE	Type I	Mean	SE	MSE	Type I	Mean	SE	MSE	Type I
IVW	0.064	0.034	0.0059	0.420	0.124	0.045	0.0185	0.893	0.242	0.058	0.0635	1.000
MR-Egger	0.005	0.143	0.0835	0.271	0.018	0.170	0.0982	0.217	0.014	0.185	0.0921	0.132
Median (Simple)	0.013	0.021	0.0010	0.069	0.028	0.022	0.0016	0.205	0.081	0.028	0.0080	0.854
Median (Weighted)	0.014	0.020	0.0010	0.082	0.029	0.022	0.0018	0.242	0.085	0.027	0.0095	0.843
Mode (Simple)	0.000	0.046	0.0121	0.006	0.001	0.046	0.0236	0.002	0.013	0.045	0.0290	0.031
Mode (Weighted)	0.001	0.042	0.0117	0.006	0.001	0.043	0.0234	0.010	0.016	0.043	0.0284	0.053
Lasso	0.009	0.014	0.0006	0.148	0.012	0.015	0.0006	0.168	0.046	0.017	0.0030	0.721
MR-Presso	0.008	0.014	0.0005	0.112	0.024	0.017	0.0012	0.291	0.111	0.028	0.0147	0.978
MR-Raps (Simple)	0.038	0.014	0.0672	0.123	0.229	0.022	0.6480	0.266	0.441	0.021	0.6798	0.668
MR-Raps (Overdispersed)	0.004	0.014	0.0005	0.092	0.009	0.014	0.0006	0.146	0.179	0.050	0.0398	0.955
JAM-MR ($w = 0$)	0.063	0.033	0.0058	0.436	0.122	0.044	0.0181	0.905	0.239	0.056	0.0621	1.000
JAM-MR ($w = 0.1N_1$)	0.039	0.023	0.0029	0.344	0.078	0.031	0.0085	0.709	0.161	0.041	0.0300	0.979
JAM-MR ($w = 0.2N_1$)	0.028	0.020	0.0020	0.273	0.056	0.026	0.0051	0.550	0.119	0.034	0.0181	0.897
JAM-MR ($w = 0.5N_1$)	0.014	0.016	0.0009	0.168	0.026	0.019	0.0017	0.298	0.056	0.024	0.0053	0.587
JAM-MR ($w = N_1$)	0.008	0.015	0.0007	0.138	0.013	0.016	0.0009	0.190	0.028	0.019	0.0020	0.377
JAM-MR ($w = 2N_1$)	0.004	0.015	0.0006	0.146	0.007	0.016	0.0008	0.163	0.017	0.018	0.0016	0.287
JAM-MR ($w = 5N_1$)	0.002	0.017	0.0011	0.216	0.003	0.018	0.0013	0.230	0.014	0.022	0.0025	0.308
Oracle	0.001	0.015	0.0004	0.046	0.000	0.016	0.0005	0.046	0.001	0.018	0.0007	0.055

Fig 7. Manhattan plot of posterior inclusion probabilities for SNPs associated with systolic blood pressure, based on the best JAM-MR run ($w = 2N_1$).

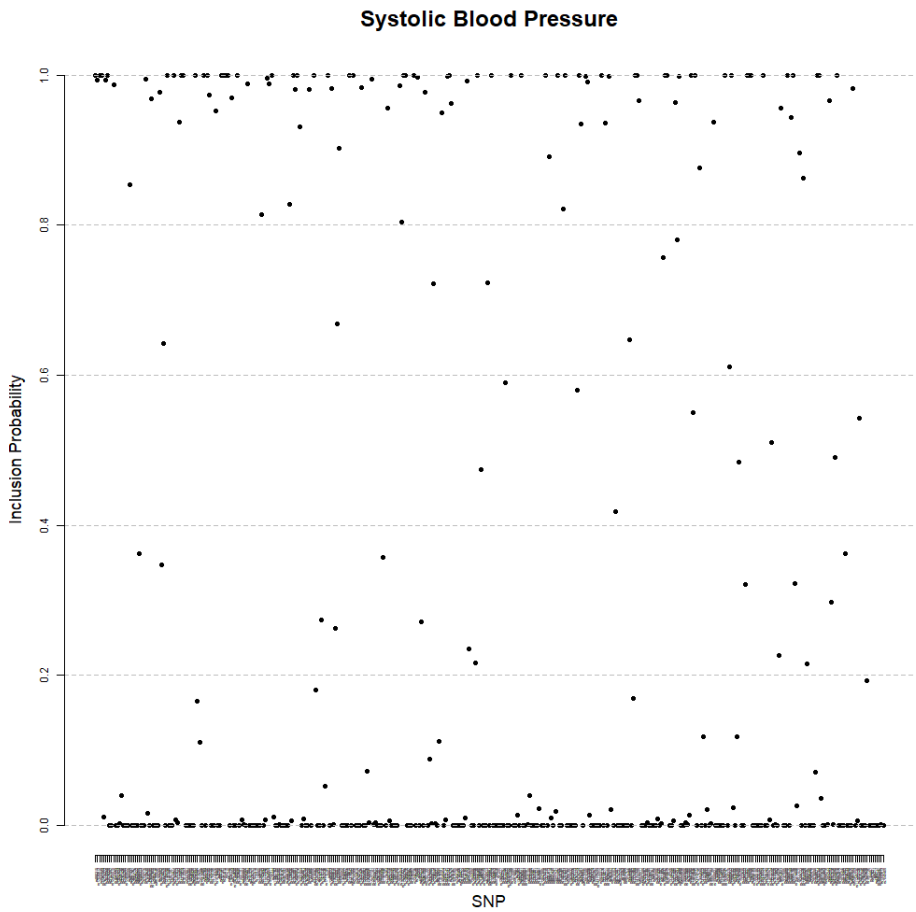
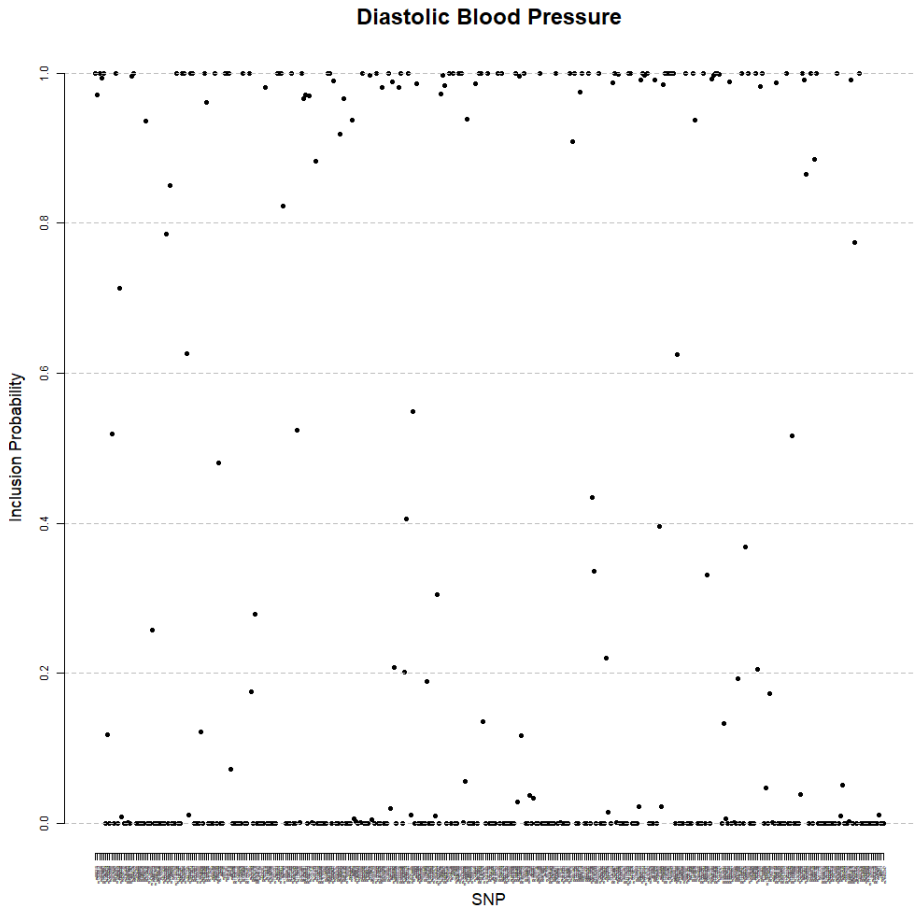



Fig 8. Manhattan plot of posterior inclusion probabilities for SNPs associated with diastolic blood pressure, based on the best JAM-MR run ($w = N_1$).