

Inference of causal and pleiotropic effects with unlinked multi-SNP genetic instruments and multiple exposures

Athina Spiliopoulou ^{1 2}, Andrii Iakovlev ², Buddhiprabha Erabadda ^{1 2}, Joseph Mellor ^{1 2}, Paul M McKeigue ^{1 2}

1 Usher Institute, College of Medicine and Veterinary Medicine, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, Scotland.

2 Institute of Genetics and Cancer, College of Medicine and Veterinary Medicine, University of Edinburgh, Western General Hospital Campus, Crewe Road, Edinburgh EH4 2XUC, Scotland.

Abstract

Current methods for Mendelian randomization (MR) analyses are restricted to using a single genetic variant (typically a single nucleotide polymorphism; SNP) to construct each unlinked instrument. This reduces the robustness of findings as results can change substantially for different choice of which SNP is chosen in each genomic region, and does not exploit all the information available in the data. We describe methods to overcome this limitation by using all SNPs associated with each exposure to construct unlinked scalar instruments from summary statistics on SNP-exposure associations. Furthermore we describe methods for joint modelling of effects of multi-SNP instruments on multiple exposures to distinguish causal from direct (pleiotropic) effects on the outcome by allowing for sharing of pleiotropy between overlapping instruments. We demonstrate the approach with simulated data, showing that joint analysis of multiple exposures achieves a modest gain in statistical power, compared with modelling exposures one at a time. We demonstrate an application to real data to identify proteins with causal effects on rheumatoid arthritis in UK Biobank participants, using instruments constructed from summary statistics for SNP-protein associations.

Introduction

Two-sample Mendelian randomization (2SMR) has been widely used to infer causal effects of exposure on outcome.^{1,2} The most widely-used methods for 2SMR use unlinked genetic instruments, and model the relationship of instrument-outcome coefficients to instrument-exposure coefficients. These coefficients and their standard errors can be extracted from summary statistics of genome-wide association studies (GWAS) of the exposure and of the outcome in non-overlapping samples. For each unlinked instrument, the instrument-exposure and instrument-outcome coefficients are effectively modeled as an independent observation for this relationship. If there were no direct effects of the instruments on the outcome and no sampling errors in estimating the coefficients, a scatter plot of instrument-outcome coefficients against instrument-exposure coefficients would show points lying on a straight line passing through the origin, with slope equal to the causal effect parameter. In the real world, where the points on the scatter plot are dispersed by the direct effects of instruments on outcome and by sampling errors in coefficient estimates, inference about the causal effect parameter requires marginalizing over these unobserved random effects.

A limitation of existing methods for Mendelian randomization is that the requirement for unlinked genetic instruments necessitates that each instrument comprises only one variant from each genomic region in which genetic associations with the exposure have been detected. This does not use all the available information about the effects of variants in the region on the exposure, and the results may depend upon which variants are selected.

Another limitation of the most widely-used methods for Mendelian randomization is that exposures are modelled one at a time. In principle it should be possible to learn more about the direct effects of the instruments on the outcome by modelling multiple exposures jointly, as instruments with pleiotropic effects on multiple exposures are also likely to have direct effects. Statistical methods for modelling the effects of a set of instruments on multiple exposures (“multivariable MR”) in the presence of pleiotropic effects have been described; most of these methods are generalizations of single-variable methods for 2SMR that construct “estimators” based on coefficient ratios.^{3–5} As shown in the accompanying paper, defects in these methods may not be obvious because they are not based on the data likelihood. A specific problem with using coefficient ratios as estimators is that when the denominator (the SNP-exposure coefficient) is small the sampling distribution of the ratio is non-Gaussian, giving rise to “weak instrument bias”. It is possible to exclude weak instruments when only one exposure is being considered, but not with multivariable MR as instruments that have strong effects on one exposure may have weak effects on other exposures. More recently likelihood-based methods for multivariable MR have been described.^{6,7} As these likelihood-based methods do not rely on the sampling properties of ratios, weak instrument bias does not arise. However these methods assume that the same SNP is selected at each locus, so that the direct effects on outcome can be modelled as the same across all exposures.

This paper describes methods that overcome these limitations. The paper is organized as follows. First we describe a method for constructing scalar instruments from multiple SNPs; this allows us to select the best SNP or set of SNPs as instruments for each exposure. Second, we describe a method for modelling multiple exposures in which scalar instruments for different exposures are grouped into clumps: pleiotropic loci are represented by clumps containing instruments for multiple exposures. We examine the Type 1 and Type 2 error rates of this multivariable method using simulated data, and demonstrate its application to a real dataset on putative core genes for rheumatoid arthritis, where we use *trans* protein quantitative trait loci (*trans*-pQTLs) to construct instrumental variables for the protein levels of 10 previously detected core genes⁸. The broader implications for the design and analysis of 2SMR studies are discussed in an accompanying paper.

Methods

Constructing scalar instruments from multiple SNPs

Standard methods for 2SMR require as input data an estimate of two scalar quantities for each unlinked instrument: the instrument-exposure coefficient and the instrument-outcome coefficient. We describe below how to combine multiple SNPs from a given region into a single instrument and calculate a scalar coefficient for the effect of this new instrument on the exposure given univariate summary statistics for the association of each SNP with the exposure.

From summary statistics we obtain a table of exposure-associated SNPs with regression coefficients and standard errors, filtered by p -value. After ordering by genomic position, we can group the SNPs into regions such that no gap between adjacent SNPs in the same region exceeds L Mb, and each gap greater than L Mb defines a new regions. We use the term “locus” for a region that contains SNPs associated with the exposure at a given p -value threshold. We use p -value $< 10^{-5}$ unless explicitly stated otherwise. For each locus, we calculate multivariable coefficient estimates $\hat{\alpha}_m$ by multiplying the vector of univariate coefficient estimates $\hat{\alpha}_u$ by the inverse of the correlation matrix Σ_g between the SNP genotypes.

$$\hat{\alpha}_m = \Sigma_g^{-1} \hat{\alpha}_u$$

The correlation matrix Σ_g is obtained from a reference panel such as 1000 Genomes. A shrinkage penalty (equivalent to ridge regression) can be imposed by adding a penalty factor to the diagonal elements of Σ_g . Where the correlation matrix is singular or ill-conditioned, a pseudo-inverse solution is used to calculate the multivariable coefficients. For an individual with genotypes \mathbf{g} at the exposure-associated SNPs in a locus, a locus-specific score S predicting the exposure is calculated as $\mathbf{g}^\top \cdot \hat{\alpha}_m$.

Because the score S is calculated from the genotypes and the genotype-exposure coefficients, we cannot use it as a genotypic instrument for the exposure (the instrument-exposure coefficient is incorporated in the score). We can however construct a genotypic instrument by factoring the dot product $\mathbf{g}^\top \cdot \hat{\alpha}_m$ as the product of two scalars (using the definition of the dot product of two Euclidean vectors): the magnitude of the multivariable coefficient vector $\|\hat{\alpha}_m\|$ and a pseudo-genotype $\|\mathbf{g}\|\rho_{\mathbf{g}, \hat{\alpha}_m}$, where $\rho_{\mathbf{g}, \hat{\alpha}_m}$ is the correlation between \mathbf{g} and $\hat{\alpha}_m$, geometrically equivalent to the cosine of the angle between these vectors. The pseudo-genotype $\|\mathbf{g}\|\rho_{\mathbf{g}, \hat{\alpha}_m}$ can be thought of as the projection of the genotype vector \mathbf{g} on to the exposure effects vector $\hat{\alpha}_m$, while $\|\hat{\alpha}_m\|$ is the magnitude of the effect on the exposure. We can then substitute the pseudo-genotype $\|\mathbf{g}\|\rho_{\mathbf{g}, \hat{\alpha}_m} = S/\|\hat{\alpha}_m\|$ for the scalar instrument Z and $\|\hat{\alpha}_m\|$ for the scalar coefficient estimate $\hat{\alpha}$ in the statistical model. The derived coefficient for the estimated effect of the instrument Z on the exposure X is always positive; flipping the coding of the alleles would flip the sign of $\rho_{\mathbf{g}, \hat{\alpha}_m}$. We can calculate the standard error of $\hat{\alpha}$ as $1/\sqrt{\mathcal{I}_\alpha}$ where \mathcal{I}_α is the Fisher information about the slope of the linear regression of X on Z , given by

$$\mathcal{I}_\alpha = N_\alpha \frac{\text{Var}(Z)}{\sigma_X^2 - \hat{\alpha}^2 \text{Var}(Z)}$$

where N_α is the sample size of the study from which the estimated univariate coefficients $\hat{\alpha}_u$ were obtained, $\text{Var}(Z)$ is the variance of the scalar instrument Z estimated in a reference panel, and σ_X^2 is the variance of the exposure X . If an estimate of σ_X^2 is not given with the summary statistics for SNP-exposure associations, it can be calculated from the allele frequency p_j of the j th SNP, the standard error s_{g_j} of the j th univariate coefficient estimate $\hat{\alpha}_{u_j}$, and the sample size N_α as

$$\sigma_X^2 = 2p_j(1-p_j) \left(N_\alpha s_{g_j}^2 + \hat{\alpha}_{u_j}^2 \right)$$

Estimation of instrument-outcome coefficients from summary level data

Where an individual-level dataset for genotype-outcome associations is available, the instrument-outcome coefficients for scalar instruments constructed from multiple SNPs can be estimated by fitting regression models for the effect of each instrument on the outcome.

Where only summary-level data are available for the genotype-outcome associations, it is possible to obtain an estimate $\hat{\gamma}$ of the coefficient for the effect of the scalar instrument Z on the outcome and its standard error, using individual-level genotype data from a reference panel of individual with similar ancestry to that in the genotype-outcome dataset, such as 1000 Genomes. Although we have not used this method, we give a derivation briefly for completeness. First, the vector $\hat{\gamma}_m$ of estimated multivariable coefficients is calculated by premultiplying the estimates $\hat{\gamma}_u$ of the univariate coefficients by Σ_G^{-1} . The coefficient $\hat{\gamma}$ for the scalar instrument Z is then estimated by minimizing the sum of the squared differences between the predictor calculated from Z and the linear predictor calculated from the vector of genotypes G , where the sum is taken over the genotypes of all n individuals in the reference panel.

$$\sum_{i=1}^n (\gamma Z_i - G_i \cdot \hat{\gamma}_m)^2$$

Equating to zero the derivative of this expression with respect to γ , substituting $\|G_i\|\rho_{G_i, \hat{\alpha}_m}$ for Z_i and factoring the dot product $G_i \cdot \hat{\gamma}_m$ as $\|G_i\|\rho_{G_i, \hat{\gamma}_m}\|\hat{\gamma}_m\|$ we obtain

$$\hat{\gamma} = \|\hat{\gamma}_m\| \frac{\sum_{i=1}^n \rho_{G_i, \hat{\gamma}_m} \|G_i\|^2 \rho_{G_i, \hat{\alpha}_m}}{\sum_{i=1}^n \|G_i\|^2 \rho_{G_i, \hat{\alpha}_m}^2}$$

The ratio in this expression can be recognized as a weighted average of the ratio $\rho_{G_i, \hat{\gamma}_m}/\rho_{G_i, \hat{\alpha}_m}$, with weights Z_i^2 . The standard error of $\hat{\gamma}$ is $1/\sqrt{I_\gamma}$ where I_γ is the Fisher information on the slope of the regression of Y on Z , given (for a logistic regression model) by

$$I_\gamma = N_\gamma \text{Var}(Z) p(1-p)$$

where N_γ and p are respectively the total sample size and the proportion of cases in the dataset from which the coefficient estimates $\hat{\gamma}_u$ were obtained.

The regularized horseshoe prior on direct effects

For a Mendelian randomization study of a single exposure with J unlinked genetic instruments, we specify a model with three parameters:

- α vector of coefficients of effects of the instruments on exposure X ,
- β vector of coefficients of direct (pleiotropic) effects of the instruments on outcome Y ,
- θ causal effect of X on Y .

The crude effect of the j th instrument on the outcome is the sum of the direct effect and the causal effect:

$$\gamma_j = \beta_j + \theta \alpha_j$$

As the instruments are unlinked, we can model the coefficient estimates $\hat{\gamma}_j, \hat{\alpha}_j$ as independent Gaussian variables conditional on the true values:

$$\hat{\alpha}_j \sim \mathcal{N}(\alpha_j, s_{\alpha(j)}^2)$$

$$\hat{\gamma}_j \sim \mathcal{N}(\gamma_j, s_{\gamma(j)}^2)$$

With priors on θ and β , we can specify a full Bayesian probability model, and sample from the joint posterior distribution over the direct effects β_j , the instrument-exposure effects α_j , and the causal effect θ , given the maximum likelihood estimates $\hat{\alpha}_j, \hat{\gamma}_j$ and their standard errors $s_{\gamma_j}, s_{\alpha_j}$.

As the form of the distribution of pleiotropic effects over loci is unknown, any realistic statistical model has to specify a prior on these effects that encompasses a broad family of symmetric distributions ranging from a spike-and-slab mixture to a Gaussian. A convenient model for the direct effects is the horseshoe prior, which is equivalent to a mixture of spike and slab distributions.⁹ This eliminates the computational inconvenience of having to average over all possible partitions of the variables into two disjoint sets (spike-and-slab) as in the contamination mixture model¹⁰ or over different settings of the number of nonzero effect parameters.¹¹ The horseshoe prior specifies a global shrinkage parameter τ and J local shrinkage parameters λ_j . The regularized horseshoe prior modifies this by specifying a global regularization parameter η that controls the size of the largest nonzero effects.

The prior for the direct effects β_1, \dots, β_J is

$$\beta_j \sim \mathcal{N}\left(0, \tau^2 \tilde{\lambda}_j^2\right), \quad \tilde{\lambda}_j^2 = \frac{\eta^2 \lambda_j^2}{\eta^2 + \tau^2 \lambda_j^2}$$

Half-Cauchy priors are specified on the local scale parameters λ_j and the global scale parameter τ :

$$\lambda_j \sim C^+(0, 1), \quad \tau \sim C^+(0, s_{\text{global}})$$

The heavy tail of the half-Cauchy priors on λ_j allows some of the regression coefficients to escape the shrinkage imposed by small values of the global parameter τ . These nonzero coefficients are the slab component of the spike-and-slab distribution. The value specified for the scale of the global shrinkage parameter τ encodes an expectation of the fraction of coefficients that are nonzero. A weakly informative prior is specified for the regularization parameter η , which encodes the scale of the slab component.

$$\eta^2 \sim \text{Inverse-Gamma}\left(\frac{\nu}{2}, \frac{\nu s_{\text{slab}}^2}{2}\right)$$

This translates to a Student t prior with ν degrees of freedom and scale s_{slab} on the largest direct effects.

Encoding a prior expectation of the effective number of nonzero effects

The shrinkage coefficient κ_j for the j th regression coefficient β_j , with prior $\sim \mathcal{N}(0, \tau^2 \lambda_j^2)$ and Gaussian likelihood with Fisher information \mathcal{I}_j about β_j can be defined as the fraction by which the information in the prior shrinks the posterior mean of β_j from the maximum likelihood estimate:

$$\kappa_j = 1 - \frac{\tau^2 \lambda_j^2}{\tau^2 \lambda_j^2 + \mathcal{I}_j} = \frac{1}{1 + \tau^2 \lambda_j^2 / \mathcal{I}_j}$$

The shrinkage coefficients κ_j can take values from 0 (no shrinkage) to 1 (complete shrinkage). The prior on each shrinkage coefficient has a horseshoe shape. Calculation of the shrinkage coefficients and the effective number of nonzero parameters does not affect the results of statistical modelling, but is helpful for interpretation of what is being learned from the data.

The effective fraction f of nonzero coefficients is

$$f = \frac{1}{J} \sum_{j=1}^J (1 - \kappa_j)$$

The prior expectation of f is

$$\mathbb{E}\langle f | \tau \rangle = 1 - \frac{1}{J} \sum_{j=1}^J \frac{1}{1 + \tau^2 \mathcal{I}_j}$$

The information \mathcal{I}_j about the coefficient β_j depends on the sample sizes in the studies from which the summary statistics $\hat{\alpha}_j$ and $\hat{\gamma}_j$ were obtained. At $\theta = 0$, $\beta_j = \gamma_j$ and \mathcal{I}_j is asymptotically equivalent to the inverse variance of the maximum likelihood estimate of γ_j .

With more information, a smaller value of τ is required to encode the same prior expectation of f . Piironen and Vehtari recommend setting the scale of the prior on τ to be consistent with this prior expectation.¹² For these analyses we have set the scale of the Cauchy prior on τ so that the prior median of τ is the value at which the prior expectation of f given $\mathcal{I}_1, \dots, \mathcal{I}_J$ is 0.4, and we have set the scale of the Inverse-Gamma prior on η^2 as 0.05, based on prior knowledge that effects of any single genomic region on a complex disease are usually of modest size. Varying these settings had only slight effects on the results.

Modelling multiple exposures with a grouped regularized horseshoe prior

When the procedure for constructing multi-SNP scalar instruments is applied to multiple exposures, a pleiotropic locus will be represented by several instruments constructed from SNPs with overlapping or nearly overlapping positions on the genome. We group these instruments together and refer to this grouping as a “clump”. Within a clump, pleiotropic SNPs will contribute to the scalar instruments for multiple exposures, and these instruments will be correlated. In this situation we observe pleiotropy directly. By jointly modelling multiple exposures that share such pleiotropic loci, we can incorporate information about direct effects that is not available from modelling exposures one at a time.

To model K exposures, we group the instruments into J clumps of overlapping or nearly-overlapping instruments. We then format the instrument-exposure coefficient estimates $\hat{\alpha}_{j,k}$ and instrument-outcome coefficient estimates $\hat{\gamma}_{j,k}$ and the corresponding standard errors $s_{\alpha_{j,k}}, s_{\gamma_{j,k}}$ as arrays with J rows and K columns. The instrument-exposure effects $\alpha_{j,k}$ and the direct effects $\beta_{j,k}$ are specified as $J \times K$ arrays, and the causal effects of each exposure θ_k are specified as a vector of length K .

The $\alpha_{j,k}$ parameters are sampled as independent Gaussian variables, while the $\beta_{j,k}$ parameters are sampled from a regularized horseshoe prior as follows. The local scale parameters λ_j are shared by instruments for all exposures in each clump; this is a regularized version of the grouped horseshoe.¹³ Where a clump contains a locus that has direct effects on the outcome and pleiotropic effects on multiple exposures, these direct effects will tend to escape shrinkage for all exposures, captured in the model by learning a large λ_j parameter shared by all instruments in the j -th clump.

Fig 1 shows the grouped regularized horseshoe prior model as a directed acyclic graph in plate notation.

Computational methods

In principle, the grouped regularized horseshoe prior can be implemented in any probabilistic programming language. We chose to use `NumPyro`, which supports masking of uninformative cells in arrays. This allows us to encode the instrument-exposure and instrument-outcome coefficients as arrays of size $J \times K$, where J is the number of clumps and K is the number of

exposures. Because most clumps contain instruments for only a few exposures, these arrays are sparse. In practice, using masking slows down the sampler considerably. We can obtain the same result without masking, by specifying the estimated instrument-exposure coefficients for the cells containing missing values to be close to zero with very large standard errors: this ensures that these cells do not contribute appreciably to the likelihood.

Posterior samples are generated by the NUTS (No U-Turn Sampling) algorithm, a self-tuning Hamiltonian Monte Carlo sampler. To regularize posterior geometry, the half-Cauchy distributions for the local scale parameters λ_j and the global shrinkage parameter τ are parameterized as mixture distributions (half-Gaussian with inverse gamma distribution of scale parameter) as shown in Fig 1. Even with this parameterization, a computational problem known as funnelling arises: the Hamiltonian Monte Carlo sampler cannot explore regions of the posterior where the curvature varies markedly. The diagnostics provided with the NUTS algorithm show that divergent transitions occur where the regularization parameter η takes large values and the global shrinkage parameter τ takes very small values. This problem is fixed by specifying $\nu = 4$ in the Inverse-Gamma prior on η^2 , encoding a prior belief that the direct effects of the instruments are unlikely to take very large values.

To obtain the marginal likelihood as a function of the causal effect parameter, we fitted a kernel density to the posterior samples of θ , weighting each observation by the inverse of the prior. A quadratic function was fitted by least-squares to the logarithm of this likelihood function. The maximum likelihood estimate of θ and its standard error are obtained from this quadratic function.

Example: putative core genes for rheumatoid arthritis

As an example, we compare the ungrouped (single exposure) and grouped (multiple exposures) models using a set of putative core genes for rheumatoid arthritis as exposures. The study of rheumatoid arthritis is described in detail elsewhere.⁸ In brief, 10 putative core genes for rheumatoid arthritis were identified by testing for association of genetically predicted protein levels with rheumatoid arthritis case/control status in the UK Biobank. *Trans-* effects of SNPs on protein levels (*trans*-pQTLs) were extracted from the UK Biobank proteomics study, which evaluated SNP-associations with circulating plasma protein levels measured on the Olink Explore panel in 45,037 UK Biobank participants of European ancestry¹⁴. Predicted protein levels were computed from aggregated *trans*- effects and tested for association with the disease in 4,702 cases and 402,926 non-cases of primary rheumatoid arthritis who were not included in the proteomics study.

For the analysis described here, for each exposure (protein encoded by a core gene) and for each *trans*-pQTL contributing to the aggregated *trans*-effects score of predicted protein level, a scalar instrument was constructed using all exposure-associated SNPs in the locus.

Cis-pQTLs were excluded. The *trans*-pQTLs from the 10 exposures were grouped into 152 clumps. The instrument-outcome and instrument-exposure coefficients were encoded as arrays with 152 rows and 10 columns. In each of these arrays, 199 of the 1520 cells contained nonmissing values, corresponding to a clump containing an instrument for the exposure. Coefficients for instrument-outcome associations were computed using a logistic regression model, with sex and the first five genetic principal components as covariates, fitted on the 4,702 cases and 402,926 non-cases of primary rheumatoid arthritis in the UK Biobank as before. Each instrument was standardized to unit variance; thus the instrument-exposure coefficients correspond to the predicted effect on the exposure of an increase of one standard deviation in the scalar instrument.

Results

Simulation comparing multi-exposure and single-exposure models

We simulated 20 datasets each with 100 instruments, 12 exposures each perturbed by 20 instruments randomly chosen from the 100 instruments, and 4 of the exposures having a causal effect on the outcome. The absolute size of the causal effect was set as 0.25. A random 40 of the 100 instruments were simulated to have direct effects on the outcome, drawn from a Student t distribution with scale 0.05 and 4 degrees of freedom. To simulate a realistic level of pleiotropy, 8 of the instruments with direct effects were simulated to perturb 10 of the 12 exposures.

For exposures simulated as having zero causal effect, the variance of the test statistic was 0.8 and the Type 1 error rate at a threshold of $p < 0.001$ was 0. For exposures simulated as having a causal effect of ± 0.25 , the power to detect a causal effect at a threshold of $p < 0.001$ was 0.76.

For comparison, with the ungrouped analysis of the same simulated datasets, the variance of the test statistic was 1.11 and the Type 1 error rate at a threshold of $p < 0.001$ was 0.01. For exposures simulated as having a causal effect of ± 0.25 , the power to detect a causal effect at a threshold of $p < 0.001$ was 0.70.

Application to real data

Table 1 shows the maximum likelihood estimates of the causal effect parameters and their standard errors. The results of the grouped analysis, with local shrinkage parameters shared across instruments for different exposures that are located in the same clump, are very similar to the results of the ungrouped analysis in which the local shrinkage parameters are specified for the direct effect of each instrument on each outcome.

Fig 2 plots the effect of each instrument on outcome (rheumatoid arthritis) against the effect on the exposure (circulating protein levels), for the proteins encoded by the 10 genes previously identified as putative core genes for rheumatoid arthritis. The shrinkage of direct effects by the regularized horseshoe prior is quantified by the shrinkage coefficient κ , which can be defined as the relative weight of the prior mean of zero to the maximum likelihood estimate of the direct effect. A shrinkage coefficient of 1 indicates that the effect size has been shrunk to near zero, and a shrinkage coefficient of zero implies no shrinkage by the prior. Clumps of instruments with a shrinkage coefficient < 0.95 , implying that they have escaped shrinkage to zero, are labelled on the plot. These include the regions containing *PTPN22* and *ICAM1*.

Discussion

The methods described in this paper allow unlinked scalar genetic instruments to be constructed from all SNPs associated with each exposure without having to restrict to one SNP from each genomic region. We also extend existing likelihood-based methods for 2SMR to allow joint modelling of multiple exposures with multi-SNP instruments, where the instruments for each exposure are unlinked but instruments for different exposures are grouped into clumps so that inference of direct (pleiotropic) effects of an instrument on the outcome “borrows strength” from instruments for other exposures in the same clump.

We demonstrate with simulated data based on plausible assumptions about the frequency and size of pleiotropic effects that while the Type 1 error rate is controlled with or without modelling the instruments as grouped, the grouped analysis has slightly higher statistical power. The advantages of using the grouped model may be greater in situations where the effects of the pleiotropic instruments are large compared with the causal effects of the exposures under study. We demonstrate with an application to real data on rheumatoid

arthritis that the method identifies loci such as the *PTPN22* gene region, that are known to have pleiotropic effects on immune function,¹⁵ as having large direct effects on the outcome, compared with other loci that do not have pleiotropic effects. Other evidence for the causal role of these exposures (proteins) in rheumatoid arthritis has been discussed elsewhere⁸

A key difference between the multi-exposure model described here and the MVMR-Horse method described recently⁷ is that our method does not require the scalar instruments to be the same for each exposure: the local shrinkage factors on the direct effects on the outcome are shared by all instruments in each clump, but the effects themselves are allowed to vary in magnitude and direction. Other differences are that our method uses a regularized horseshoe prior to encode prior knowledge that the direct effects are usually of modest size, and uses a gradient-based sampling algorithm that updates all variables simultaneously, rather than a Gibbs sampler that updates variables one at a time. Gradient-based sampling algorithms can scale to large datasets. Broader implications for the design and analysis of Mendelian randomization studies are discussed in the accompanying paper.

Declarations

Data and code availability

A Jupyter notebook containing Python code using the NumPyro library to fit the statistical model to fit the MR-Hevo model, together with the summary-level data on rheumatoid arthritis used in this paper, is provided as a supplementary file.

Acknowledgements

No specific funding was received for this work. This research has been conducted using the UK Biobank Resource under application number 23652. The development of the GENOSCORES platform was supported by a Springboard Award (SBF006/1109) to AS from the Academy of Medical Sciences, supported in turn by the Wellcome Trust, the UK Government Department of Business, Energy and Industrial Strategy, the British Heart Foundation, and Diabetes UK. AI was supported by the Medical Research Council Cross-Disciplinary Fellowship (XDF) Programme (MC_FE_00035).

Declaration of interests

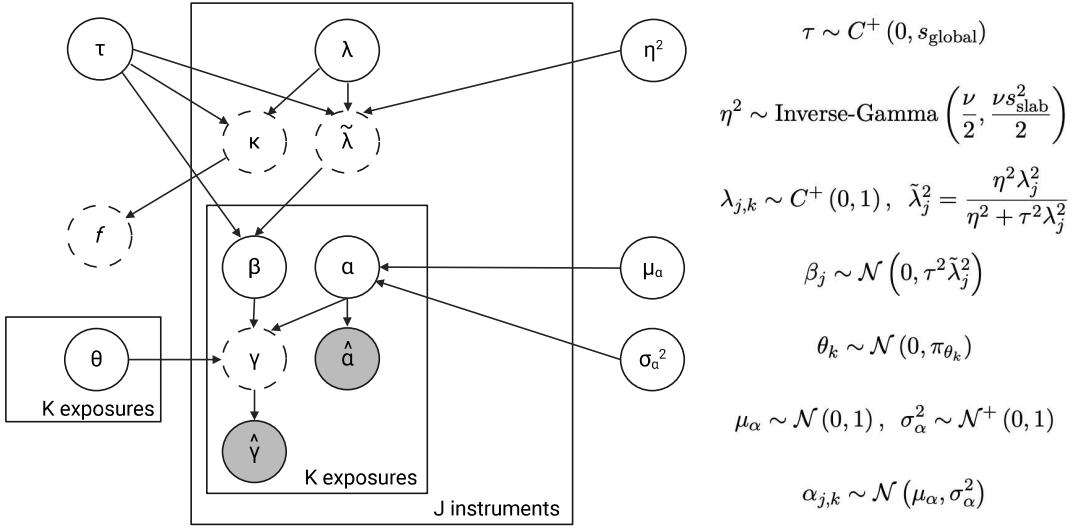
The authors declare no competing interests.

References

1. Smith, G.D., and Ebrahim, S. (2024). Mendelian randomisation at 20 years: How can it avoid hubris, while achieving more? *Lancet Diabetes Endocrinol* *12*, 14–17.
[https://doi.org/10.1016/S2213-8587\(23\)00348-0](https://doi.org/10.1016/S2213-8587(23)00348-0).
2. Stender, S., Gellert-Kristensen, H., and Smith, G.D. (2024). Reclaiming mendelian randomization from the deluge of papers and misleading findings. *Lipids in Health and Disease* *23*, 286. <https://doi.org/10.1186/s12944-024-02284-w>.
3. Sanderson, E., Davey Smith, G., Windmeijer, F., and Bowden, J. (2019). An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int J Epidemiol* *48*, 713–727.
<https://doi.org/10.1093/ije/dyy262>.

4. Grant, A.J., and Burgess, S. (2021). Pleiotropy robust methods for multivariable Mendelian randomization. *Stat Med* *40*, 5813–5830. <https://doi.org/10.1002/sim.9156>.
5. Lorincz-Comi, N., Yang, Y., Li, G., and Zhu, X. (2024). MRBEE: A bias-corrected multivariable Mendelian randomization method. *HGG Adv* *5*, 100290. <https://doi.org/10.1016/j.xhgg.2024.100290>.
6. Lin, Z., Xue, H., and Pan, W. (2023). Robust multivariable Mendelian randomization based on constrained maximum likelihood. *Am J Hum Genet* *110*, 592–605. <https://doi.org/10.1016/j.ajhg.2023.02.014>.
7. Grant, A.J., and Burgess, S. (2024). A Bayesian approach to Mendelian randomization using summary statistics in the univariable and multivariable settings with correlated pleiotropy. *Am J Hum Genet* *111*, 165–180. <https://doi.org/10.1016/j.ajhg.2023.12.002>.
8. Spiliopoulou, A., Iakovlev, A., Plant, D., Sutcliffe, M., Sharma, S., Cubuk, C., Lewis, M., Pitzalis, C., Barton, A., and McKeigue, P.M. (2025). Genome-Wide Aggregated Trans Effects Analysis Identifies Genes Encoding Immune Checkpoints as Core Genes for Rheumatoid Arthritis. *Arthritis Rheumatol*. <https://doi.org/10.1002/art.43125>.
9. Carvalho, C.M., Polson, N.G., and Scott, J.G. (2009). Handling Sparsity via the Horseshoe. In Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (PMLR), pp. 73–80.
10. Burgess, S., Foley, C.N., Allara, E., Staley, J.R., and Howson, J.M.M. (2020). A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nat Commun* *11*, 376. <https://doi.org/10.1038/s41467-019-14156-4>.
11. Xue, H., Shen, X., and Pan, W. (2021). Constrained maximum likelihood-based Mendelian randomization robust to both correlated and uncorrelated pleiotropic effects. *Am J Hum Genet* *108*, 1251–1269. <https://doi.org/10.1016/j.ajhg.2021.05.014>.
12. Piironen, J., and Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics* *11*, 5018–5051.
13. Xu, Z., Schmidt, D.F., Makalic, E., Qian, G., and Hopper, J.L. (2016). Bayesian Grouped Horseshoe Regression with Application to Additive Models. In AI 2016: Advances in Artificial Intelligence, B. H. Kang and Q. Bai, eds. (Springer International Publishing), pp. 229–240. https://doi.org/10.1007/978-3-319-50127-7_19.
14. Sun, B.B., Chiou, J., Traylor, M., Benner, C., Hsu, Y.-H., Richardson, T.G., Surendran, P., Mahajan, A., Robins, C., Vasquez-Grinnell, S.G., et al. (2023). Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* *622*, 329–338. <https://doi.org/10.1038/s41586-023-06592-6>.
15. Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J.G., Zgaga, L., Manolio, T., Rudan, I., McKeigue, P., Wilson, J.F., and Campbell, H. (2011). Abundant pleiotropy in human complex diseases and traits. *Am. J. Hum. Genet.* *89*, 607–618. <https://doi.org/10.1016/j.ajhg.2011.10.004>.

Figures



Data: $\hat{\alpha}_j \sim \mathcal{N}(\alpha_j, s_{\alpha_j}^2)$, $\hat{\gamma}_j \sim \mathcal{N}(\gamma_j, s_{\gamma_j}^2)$, with $\gamma_j = \beta_j + \theta\alpha_j$.

Hyperparameters: $\nu = 4$, $s_{\text{slab}} = 0.05$, $\pi_{\theta_k} = 1 \forall k \in K$, s_{global} : set to encode prior expectation of effective fraction of nonzero direct effects $f = 0.4$.

Fig 1. MR-Hevo model as a directed acyclic graph in plate notation. Stochastic nodes are shown as circles with continuous borders, deterministic nodes as circles with dashed borders. Observed nodes are shaded. The half-Cauchy priors on the global shrinkage parameter τ and the local shrinkage parameters λ_j are encoded by multiplying auxiliary variables sampled from a half-normal distribution, $\mathcal{N}^+(0, 1)$, and another auxiliary variable sampled from an inverse-gamma distribution, Inverse-Gamma(0.5, 0.5). The global shrinkage parameter τ is scaled by s_{global} . We use auxiliary variables to encode the regularization parameter η and the direct effects $\beta_{j,k}$, sampling from an unscaled distribution followed by multiplication with the corresponding scale value

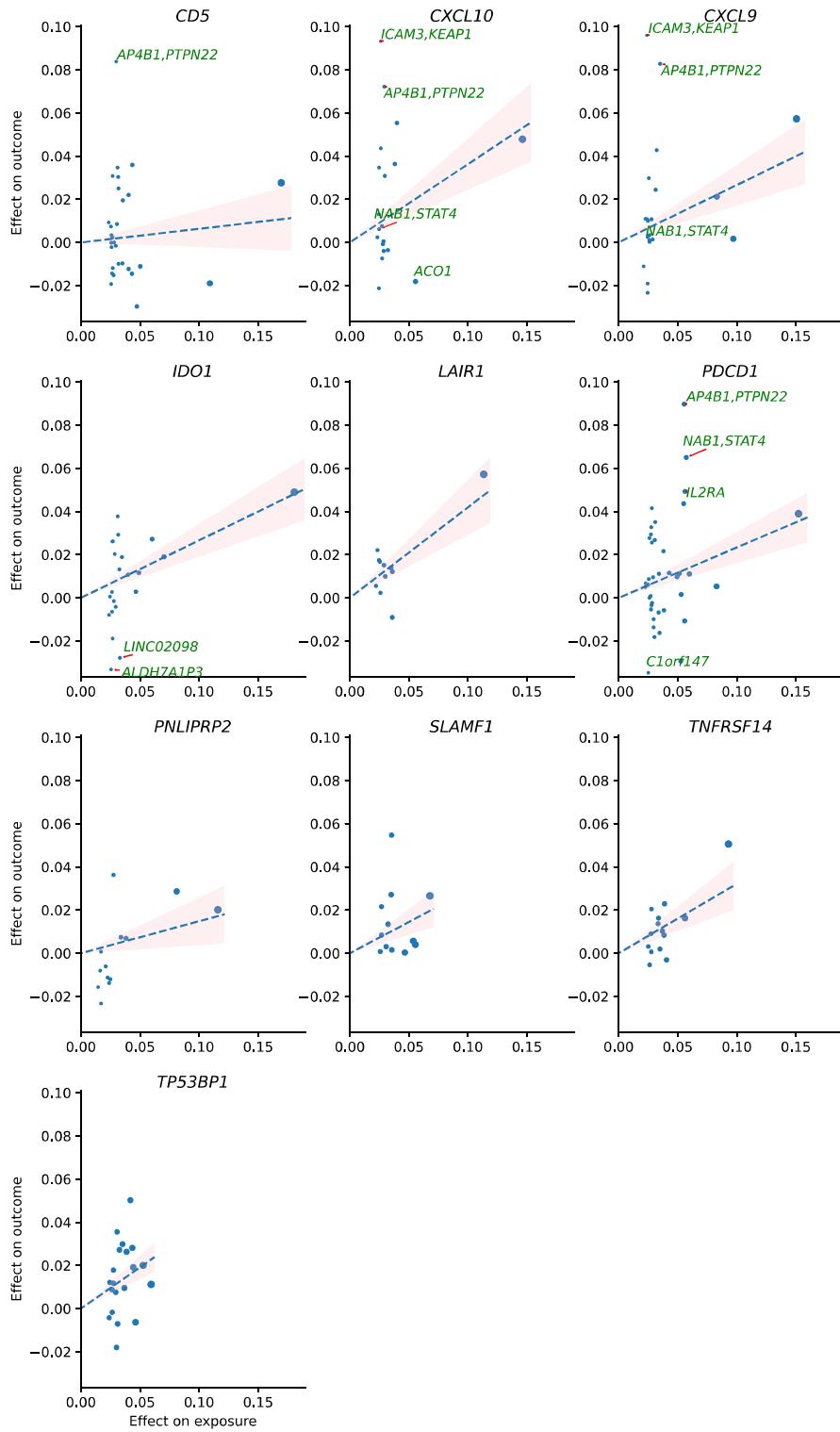


Fig 2. Plot of coefficients of regression of rheumatoid arthritis on each instrument against coefficients of regression of protein levels on each instrument. Size of each data point is inversely proportional to the standard error of the ratio estimate. The maximum likelihood estimate is shown as the slope of a line passing through the origin. *Cis*-pQTLs are excluded from these estimates. Instruments with a shrinkage coefficient < 0.95 are labelled on the plot by nearby genes.

Tables

Table 1. Mendelian randomization analysis of putative core genes for rheumatoid arthritis: comparison of models based on ungrouped and grouped instruments

Exposure (gene)	Number of instruments	Ungrouped		Grouped	
		Maximum likelihood estimate	p-value	Maximum likelihood estimate	p-value
<i>CD5</i>	29	0.065	0.4	0.064	0.5
<i>CXCL10</i>	19	0.355	0.002	0.363	0.002
<i>CXCL9</i>	20	0.292	0.002	0.266	0.005
<i>IDO1</i>	22	0.274	2×10^{-4}	0.267	5×10^{-4}
<i>LAIR1</i>	11	0.437	4×10^{-4}	0.419	0.001
<i>PDCD1</i>	39	0.240	7×10^{-4}	0.234	0.001
<i>PNLIPRP2</i>	13	0.147	0.2	0.148	0.2
<i>SLAMF1</i>	12	0.270	0.03	0.290	0.01
<i>TNFRSF14</i>	14	0.340	0.003	0.322	0.005
<i>TP53BP1</i>	20	0.382	8×10^{-4}	0.382	4×10^{-4}