Received: 17 August 2022 | Revised: 14 December 2022 | Accepted: 1 February 2023

DOI: 10.1002/gepi.22520

RESEARCH ARTICLE

Genetic Epidemiology

OFFICIAL JOURNAL INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY www.geneticepi.org

WILEY

MR-BOIL: Causal inference in one-sample Mendelian randomization for binary outcome with integrated likelihood method

Dapeng Shi[1] | **Yuquan Wang**[2] | **Ziyong Zhang**[3] | **Yunlong Cao**[2] | **Yue-Qing Hu**[1,2]

[1]Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, China

[2]State Key Laboratory of Genetic Engineering, Human Phenome Institute, Institute of Biostatistics, School of Life Sciences, Fudan University, Shanghai, China

[3]Department of Statistics and Data Science, School of Management, Fudan University, Shanghai, China

Correspondence
Yue-Qing Hu, State Key Laboratory of Genetic Engineering, Human Phenome Institute, Institute of Biostatistics, School of Life Sciences, Fudan University, Shanghai, China.
Email: yuehu@fudan.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 11971117, 11571082

Abstract

Mendelian randomization is a statistical method for inferring the causal relationship between exposures and outcomes using an economics-derived instrumental variable approach. The research results are relatively complete when both exposures and outcomes are continuous variables. However, due to the noncollapsing nature of the logistic model, the existing methods inherited from the linear model for exploring binary outcome cannot take the effect of confounding factors into account, which leads to biased estimate of the causal effect. In this article, we propose an integrated likelihood method MR-BOIL to investigate causal relationships for binary outcomes by treating confounders as latent variables in one-sample Mendelian randomization. Under the assumption of a joint normal distribution of the confounders, we use expectation maximization algorithm to estimate the causal effect. Extensive simulations demonstrate that the estimator of MR-BOIL is asymptotically unbiased and that our method improves statistical power without inflating type I error rate. We then apply this method to analyze the data from Atherosclerosis Risk in Communications Study. The results show that MR-BOIL can better identify plausible causal relationships with high reliability, compared with the unreliable results of existing methods. MR-BOIL is implemented in R and the corresponding R code is provided for free download.

KEYWORDS

binary outcome, causal relationship, expectation maximization algorithm, instrumental variable, logistic model, Mendelian randomization, noncollapsing

1 | INTRODUCTION

One common goal of epidemiology is to investigate how unit changes in exposures affect the ones in outcome variables. When actual data is acquired from observational studies rather than randomized experiments, confounding issues frequently hinder us from finding a causal explanation between exposures and outcome variables. The instrumental variables (IVs) approach originated in economics, helps resolve this difficulty. In general, an instrumental variable Z is only related to exposure X, is independent of any other unobserved confounding factor U, and only impacts the outcome variable Y through the exposure X (Anderson & Rubin, 1949; Angrist et al., 1996; Sargan, 1958). As a result, we can summarize these core assumptions for the IV Z as follows:

1. Relevance: Z is associated with the exposure X.
2. Effective Random Assignment: Z is independent of any unobserved confounder U.

3. Exclusion Restriction: Z is independent of outcome Y conditional on X and confounder U.

Variables that satisfy assumptions (2) and (3) are also known as valid instrument variables in the literature (Holland, 1988; Wooldridge, 2010), while their counterparts are considered invalid instrument variables (Kang et al., 2016; Kolesár et al., 2015; Small, 2007), which represent gene pleiotropy in biological terms (Solovieff et al., 2013; Verbanck et al., 2018; Zhao et al., 2020). As for the validity test of IVs (satisfying the above assumptions), there have been many established tools in econometrics (Hahn & Hausman, 2002; Kitagawa, 2015; Stock et al., 2002).

An important application of the IV approach in epidemiology is Mendelian randomization (MR) (Davey Smith & Ebrahim, 2003). MR has recently been validated as a powerful technique in which genetic variants, that is, single-nucleotide polymorphisms (SNPs), are used as IVs to infer the causal relationship between a modifiable exposure and the interested disease outcome (Burgess et al., 2012; Burgess & Thompson, 2015; Hernán & Robins, 2006; Lawlor et al., 2008). We assume that the genetic variants selected always satisfy the assumptions (1)–(3), that is, they are valid IVs, though this approach can be extended to the invalid IVs case.

The MR analysis of continuous outcome variables is relatively complete due to the linear structure of the linear model (Burgess & Thompson, 2012; Kang et al., 2016). However, when the outcome variable is binary, the probability of the event occurring and the exposure are usually nonlinearly related. Fortunately, the log-linear model (Dai & Zhang, 2015), which is no more sophisticated than the linear model, can be used if the outcome is rare. But the situation gets more complicated when the logistic model is utilized for general binary outcome variables (Allman et al., 2021; Burgess et al., 2017; Burgess & Thompson, 2012, 2013; Cai et al., 2011; Palmer et al., 2008).

Logistic regression of the outcome variable on the exposure in the form of a measurement error model (Burgess & Thompson, 2012) is a good place to start. For simplicity, we pick out only one genetic variant and disregard the impact of confounding factors on outcome,

$$\begin{aligned} X_i &\sim \mathcal{N}\left(\eta_i, \sigma_x^2\right), \\ Y_i &\sim \text{Bernoulli}(\pi_i), \\ \eta_i &= \alpha_0 + \alpha_1 Z_i, \\ \text{logit}(\pi_i) &= \beta_0 + \beta_1 \eta_i, \end{aligned}$$

where $i = 1, ..., n$, is the index of the subject. A standard maximum likelihood approach can be adopted to estimate the parameter β_1 (Burgess & Thompson, 2012). However, the second stage model (the regression model of Y_i on X_i) ignores the fact that confounders might influence the outcome variable Y_i in tandem with the exposure X_i. When confounders are taken into account, the existing methods tend to fail due to the so-called noncollapsibility of the logistic model (Burgess & CRP CHD Genetics Collaboration, 2013; Greenland et al., 1999; Pang et al., 2016). Furthermore, in the existing summary statistics data studies, no one appears to be concerned with the accurate estimation problem for the causal effect β_1 when the confounders are involved in the second stage model (Xu et al., 2021; Xue et al., 2021; Zhao et al., 2020). The model they used is almost entirely the logistic model after approximation. This is the primary motivation for this paper.

It is worth mentioning that our work is only concerned with the statistical inference of the causal effect β_1, ignoring whether β_1 has a real causal meaning similar to that in the linear model (Angrist & Imbens, 1994; Clarke & Windmeijer, 2012; Harbord et al., 2013). Indeed, a considerable body of literature has focused on this topic and proposed some more general semi-parametric models (Bowden & Vansteelandt, 2011; Burgess et al., 2014; Vansteelandt et al., 2011) to explain it.

To address the limitations of the existing approaches, in this one-sample Mendelian randomization study for binary outcome we propose an integrated likelihood method (MR-BOIL) to identify the causal relationship between the exposure and outcome. In particular, we investigate the joint effects of confounding factors on the two-stage model, assuming that they follow a bivariate normal distribution. Next, based on this model, we develop the data's joint integrated likelihood function and construct a new form by probability density transformation, which motivates us to use the expectation-maximization (EM) algorithm to seek the optimal value. In addition, the likelihood ratio test (LRT) is used to test the null hypothesis. The assessment of estimation performance, type I error rate and power for various methods is conducted by extensive simulations. Finally, we apply MR-BOIL to the data of Atherosclerosis Risk in Communications Study (ARIC) and validate the significant causal effect of systolic blood pressure (SBP) on hypertension. As a supplement, we also assess the null causal effect estimation of diastolic blood pressure (DBP) on the negative control outcome (Hu et al., 2021; Sanderson et al., 2020) hypertension for different methods. We conclude that even analyzing the simplest causal relationships, our approach outperforms traditional plain methods in real-world data analysis. MR-BOIL is implemented in R and the corresponding R code is provided for free download from https://github.com/YQHuFD/MR-BOIL.git.

2 | METHODS

2.1 | The logistic structural model

Consider the potential causal effect of a continuous exposure X on a binary disease outcome Y in an MR study. For n individuals, let $\mathbf{Z} = (Z_1, ..., Z_n)^T$ denote the measurement of the genetic variant (i.e., SNP) selected as the instrumental variable for inferring the causality, and $\mathbf{X} = (X_1, ..., X_n)^T$, $\mathbf{Y} = (Y_1, ..., Y_n)^T$ are measurements of the exposure and outcome, respectively. Denote the collection of confounding factors U, V as the latent variables that hide the causal effect of X on Y and Z on X, respectively. Then, under the independent and identically distributed sampling scheme, subjects are assumed to be generated from the following structural model (Harbord et al., 2013),

$$\begin{aligned} X_i &= \gamma Z_i + V_i, \\ Y_i &\sim \text{Bernoulli}(\pi_i), \\ \text{logit}(\pi_i) &= \beta_0 + \beta_1 X_i + U_i, \\ \begin{pmatrix} U_i \\ V_i \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right), \end{aligned} \tag{1}$$

implicitly assuming that the exposure X_i and genetic variant Z_i in the first stage model have been centralized.

2.2 | Existing methods used in MR for binary outcome

The following are the most often used approaches for estimating causal effect β_1 in a logistic model for the binary outcome in one-sample MR: Wald-type method (Burgess & Thompson, 2012; Wald, 1940), two-stage method (Cai et al., 2011) and adjusted-IV method (Palmer et al., 2008; Terza et al., 2008). We will briefly describe the basic principles of these methods and the concern associated with their application. Note in advance that Wald-type method and two-stage method are consistent when only one genetic variant is incorporated.

2.2.1 | Wald-type method

The Wald or ratio of coefficients method is one of the oldest and simplest IV methods (Burgess & Thompson, 2012; Wald, 1940). In the above model setting, we plug expression of X into Y to get

$$\text{logit}\mathbb{E}[Y_i|Z_i, U_i, V_i] = \beta_0 + \beta_1(\gamma Z_i + V_i) + U_i,$$

which motivates us to estimate β_1 through

$$\hat{\beta}_1 = \frac{\hat{\Gamma}}{\hat{\gamma}},$$

where $\hat{\gamma}$ is the regression coefficient of X_i on Z_i, and $\hat{\Gamma}$ is the regression coefficient of Y_i on Z_i by ignoring the term $\beta_1 V_i + U_i$.

2.2.2 | Two-stage method

Another popular method to estimate β_1 is two-stage approach (Cai et al., 2011). As the name implies, this method divides the process of estimation into two stages. In the first stage, through linear regression, we can estimate

$$\widehat{X}_i = \hat{\gamma} Z_i,$$

and in the second stage, $\hat{\beta}_1$ can be estimated by regressing Y_i on $\widehat{X}_i$ by plugging $\widehat{X}_i$ into X_i and ignoring term U_i, that is,

$$\text{logit}\mathbb{E}[Y_i|\widehat{X}_i] = \beta_0 + \beta_1 \widehat{X}_i.$$

2.2.3 | Adjusted-IV method

The above two strategies both introduce bias by omitting the terms U_i and V_i. The adjusted-IV method (also known as a control function approach or two-stage residual inclusion) is widely used to adjust the bias (Palmer et al., 2008; Terza et al., 2008). In short, this method estimates the residuals from the first stage and treats U_i and V_i as the same thing, that is, plugging residual U_i estimates from the first stage

$$\widehat{U}_i = X_i - \widehat{X}_i$$

into the second stage

$$\text{logit}\mathbb{E}[Y_i|\widehat{X}_i, \widehat{U}_i] = \beta_0 + \beta_1 \widehat{X}_i + \beta_U \widehat{U}_i.$$

Then, to obtain a debiased $\hat{\beta}_1$ estimator, we regress Y_i on $\widehat{X}_i$ and $\widehat{U}_i$. However, the method is only a rough estimate of V_i in the second-stage model, and it fails when the difference between U_i and V_i is large, as shown in Section 3.

2.3 | Integrated likelihood method for binary outcome in MR

Given the drawbacks of the above mentioned approaches, in this Mendelian randomization study for binary outcome we propose an integrated likelihood method (MR-BOIL) to identify the causal relationship between the exposure and outcome. Denote $\boldsymbol{\Theta} = (\beta_0, \beta_1, \gamma, \sigma_1, \sigma_2, \rho)$ as the whole parameters. Using some algebra (see Appendix A.1), we can reform the conditional integrated log-likelihood based on the samples $\{(x_i, y_i, z_i), 1 \le i \le n\}$ as

$$\begin{aligned}\ell(\boldsymbol{\Theta}) &= \sum_{i=1}^{n} \log \int_R \tilde{p}(y_i|x_i, z_i, u; \boldsymbol{\Theta}) \cdot \tilde{p}(x_i|z_i, \boldsymbol{\Theta}) \cdot \tilde{p}(u) du \\ &= \sum_{i=1}^{n} \log \int_R \tilde{p}(x_i, y_i|z_i, u; \boldsymbol{\Theta}) \cdot \tilde{p}(u) du,\end{aligned} \tag{2}$$

where

$$\begin{aligned}K_i &= \frac{\sigma_1}{\sigma_2}\rho \cdot (x_i - \gamma z_i), \tilde{p}(y_i|x_i, z_i, u; \boldsymbol{\Theta}) \\ &= \frac{\exp(y_i(\beta_0 + x_i\beta_1 + K_i + \sqrt{1-\rho^2}\sigma_1 u))}{1 + \exp(\beta_0 + x_i\beta_1 + K_i + \sqrt{1-\rho^2}\sigma_1 u)}, \\ \tilde{p}(u) &= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{u^2}{2}\right\}, \tilde{p}(x_i|z_i; \boldsymbol{\Theta}) \\ &= \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\frac{(x_i - \gamma z_i)^2}{2\sigma_2^2}\right\},\end{aligned}$$

and

$$\tilde{p}(x_i, y_i|z_i, u; \boldsymbol{\Theta}) = \tilde{p}(y_i|x_i, z_i, u; \boldsymbol{\Theta}) \cdot \tilde{p}(x_i|z_i, \boldsymbol{\Theta}).$$

Then our MR-BOIL estimator can be defined as the respective component of the maximizer of the log-likelihood function (2).

2.3.1 | Algorithm

Note that it is difficult to get the optimal value by calculating the integrated likelihood directly. Hence, the form of Equation (2) suggests us adopt the EM algorithm (Dempster et al., 1977). Denote $\mathbf{x} = (x_1, ..., x_n)^T$, $\mathbf{y} = (y_1, ..., y_n)^T$, $\mathbf{z} = (z_1, ..., z_n)^T$. Next, we show the algorithm flow to seek the optimal value (detailed derivation can be found in Appendix A.2).

For an initial parameter value $\boldsymbol{\Theta}^0$, and iteration $j = 1, ..., K$,

- E-Step: for fixed $\boldsymbol{\Theta}^j$, calculate the posterior expectation

$$\begin{aligned}&\mathbb{E}_{\mathbf{U}|\mathbf{y},\mathbf{x}\sim\tilde{p}(u|\mathbf{y},\mathbf{x},\mathbf{z};\boldsymbol{\Theta}^j)}[\log\tilde{p}(\mathbf{y}, \mathbf{x}, u|\mathbf{z}; \boldsymbol{\Theta})] \\ &= \sum_{i=1}^{n}\int_R \tilde{p}(u|y_i, x_i, z_i; \boldsymbol{\Theta}^j) \\ &\quad \cdot \log(\tilde{p}(y_i, x_i|z_i, u; \boldsymbol{\Theta}) \cdot \tilde{p}(u))du.\end{aligned}$$

- M-Step: maximize the posterior expectation:

$$\boldsymbol{\Theta}^{j+1} = \arg\max_{\boldsymbol{\Theta}} \sum_{i=1}^{n} \frac{\int_R \tilde{p}(y_i|x_i, z_i, u; \boldsymbol{\Theta}^j) \cdot \tilde{p}(u) \cdot \log(\tilde{p}(y_i|x_i, z_i, u; \boldsymbol{\Theta}) \cdot \tilde{p}(x_i|z_i; \boldsymbol{\Theta}))du}{\int_R \tilde{p}(y_i|x_i, z_i, u; \boldsymbol{\Theta}^j) \cdot \tilde{p}(u)du},$$

  where the integral here can be approximated using numerical methods or Monte-Carlo methods.
- Stop Criterion: the algorithm stops when $||\boldsymbol{\Theta}^j - \boldsymbol{\Theta}^{j+1}||_2 < \varepsilon$, where ε is usually taken as 10^{-4} or 10^{-5}.

2.4 | LRT

Following the estimation of causal effects, hypothesis testing is required. The variance is difficult to calculate due to the complexity of the integrated likelihood. As a result, the LRT statistic is chosen to test our interested statistical hypothesis: the association between the exposure and the outcome, that is,

$$H_0 : \beta_1 = 0, H_1 : \beta_1 \neq 0.$$

The LRT statistic is given as follows:

$$\Lambda = 2\left(\ell(\widehat{\boldsymbol{\Theta}}^{MLE}) - \ell\left(\widehat{\boldsymbol{\Theta}}_0^{MLE}\right)\right), \tag{3}$$

where $\widehat{\boldsymbol{\Theta}}^{MLE}$ and $\widehat{\boldsymbol{\Theta}}_0^{MLE}$ correspond to the maximizers obtained from the integrated likelihood function optimized in the global parameter space and in the null hypothesis parameter space, respectively. Moreover, type I error rate and power can be obtained through the asymptotical $\chi^2(1)$ property of the LRT statistic (Van der Vaart, 2000) under the null hypothesis.

3 | SIMULATION

3.1 | Estimation

To compare the performance of our MR-BOIL approach with other methods, we set up a series of scenarios with

one genetic variant. In addition to the two methods mentioned in Section 2 (two-stage is consistent with the Wald-type method in this case), we add the direct method, in which Y_i is regressed directly on X_i without taking instrument variables into account.

We simulate a cohort of 1000 subjects (Y_i, X_i, Z_i) with independent distributed modes from model (1), that is,

$$\begin{aligned} X_i &= \gamma Z_i + V_i, \\ Y_i &\sim \text{Bernoulli}(\pi_i), \\ \text{logit}(\pi_i) &= \beta_0 + \beta_1 X_i + U_i, \end{aligned}$$

where U_i and V_i follow a joint normal distribution with means 0 and 0, variances σ_1^2 and σ_2^2, respectively, and the correlation coefficient between U_i, V_i is ρ. Besides, Z_i stands for the carefully selected SNP coded as 0, 1, 2. We assume that each Z_i is in Hardy-Weinberg equilibrium with a minor allele frequency of 30%. V_i and U_i represent the confounding factors affecting X_i and Y_i, respectively, and we assume that their effects are homogeneous for all subjects with parameter ρ to measure their correlation. Moreover, the outcome for every subject is drawn from the Bernoulli distribution with probability $\pi_i = \mathbb{E}[Y_i|X_i, U_i] = \text{expit}(\beta_0 + \beta_1 X_i + U_i)$, where expit is the inverse of the logit function.

Next we first consider the influence of confounding factors on the estimation of causal effect β_1 by changing the values of σ_1 and σ_2. Fixing the nuisance parameters at $\beta_0 = 2$, $\gamma = 1$, $\rho = 0.8$, we assume the true causal effect $\beta_1 = 1$ and -1, respectively. In particular, consider three confounding scenarios increasingly affecting the exposure X_i by adjusting σ_2 to 1, 2, 3, and in each scenario, σ_1 is set separately to 0.5, 1, 2, 3 representing the increasing effect of the confounding factor U_i on Y_i.

In addition, we also consider the impact of the correlation coefficient ρ between V_i and U_i, and the sample size n on the estimation of causal effect β_1. Here, β_0 and γ are set as before. In the former case we set $\sigma_1 = 5$ and $\sigma_2 = 5$. Then ρ is set to -0.8, -0.4, 0, 0.4, 0.8 representing different degrees of correlation between confounding factors U_i and V_i. In the latter case, we set $\sigma_1 = 1.5$, $\sigma_2 = 1$, and $\rho = 0.8$, while letting sample size n change to 1000, 2000, 4000. The true value for causal effect in both cases is $\beta_1 = 1$. For each set of parameter values, 1000 replications are performed.

We mainly use mean, bias, and root mean square error (RMSE) of 1000 estimates of β_1 to evaluate the performance of different methods. As shown in Figures 1 and 2 and Tables 1 and 2, when the impact of the confounders V_i, U_i on Y_i, X_i, respectively, is minor, that is, σ_1, σ_2 are small, the adjusted-IV and MR-BOIL methods perform almost the same on average, and both are relatively robust compared with the directed and two-stage methods. When σ_1 and σ_2 are relatively large, the other three approaches show bias, whereas our MR-BOIL method remains asymptotically unbiased, although the variance is large to some extent. Therefore, as shown in Figure 3 and Table 3, we increase the sample size, which reduces the variance of our method and enhances its robustness. Figure 4 and Table 4 further show that when changing ρ from negative to positive, that is, when the influence of confounders on exposure X_i and outcome Y_i changes from negatively to positively correlated, our MR-BOIL method always remains asymptotically unbiased compared with other three methods.

3.2 | LRT

We then evaluate the type I error rates of different methods for testing the null hypothesis $H_0 : \beta_1 = 0$. Here, we will not compare the direct method, which, as shown in the supplementary, has a completely out-of-control type I error rate. Imitating the setup scenario of the previous parameter estimation, we fix the nuisance parameters at $\beta_0 = 2$, $\gamma = 1$, $\rho = 0.8$ and still adjust σ_1 and σ_2 to investigate the impact of confounding factors on the different methods. We set σ_2 to 1, 2, 3, 4, 5, and let σ_1 vary among 0.5, 1, 2, 3, 4, 5. As demonstrated in Figure 5 and Table 5, the type I error rates of the two-stage method and our MR-BOIL method are consistently controlled at the nominal level of 0.05 (ones with nominal 0.01 are provided in Appendix A.3) as σ_1 and σ_2 rise, that is, the influence of confounding factors increases. However, the type I error rate of the adjusted-IV method maintains an increasing trend. Similarly, the impact of ρ on the type I error rates with fixed parameters $\beta_0 = 2$, $\gamma = 1$, $\sigma_1 = 5$, $\sigma_2 = 5$ is also investigated. Figure 6 and Table 6 show that when the absolute value of $|\rho|$ is big, the adjusted-IV method fails, whereas the two-stage method and our MR-BOIL approach succeed, in controlling the type I error rate.

3.3 | Power

The formula in Appendix A.4 (similar formula appears in Zhao et al., 2020) can explain why the type I error rate of the two-stage method is controlled. The parameter β_1^{TSLS} (abbreviation for the parameter in two-stage estimation) can be approximated as a contraction of the true parameter β_1, and the contraction coefficient is only related to β_1, ρ, σ_1, and σ_2. Thus, if the null hypothesis holds, the true

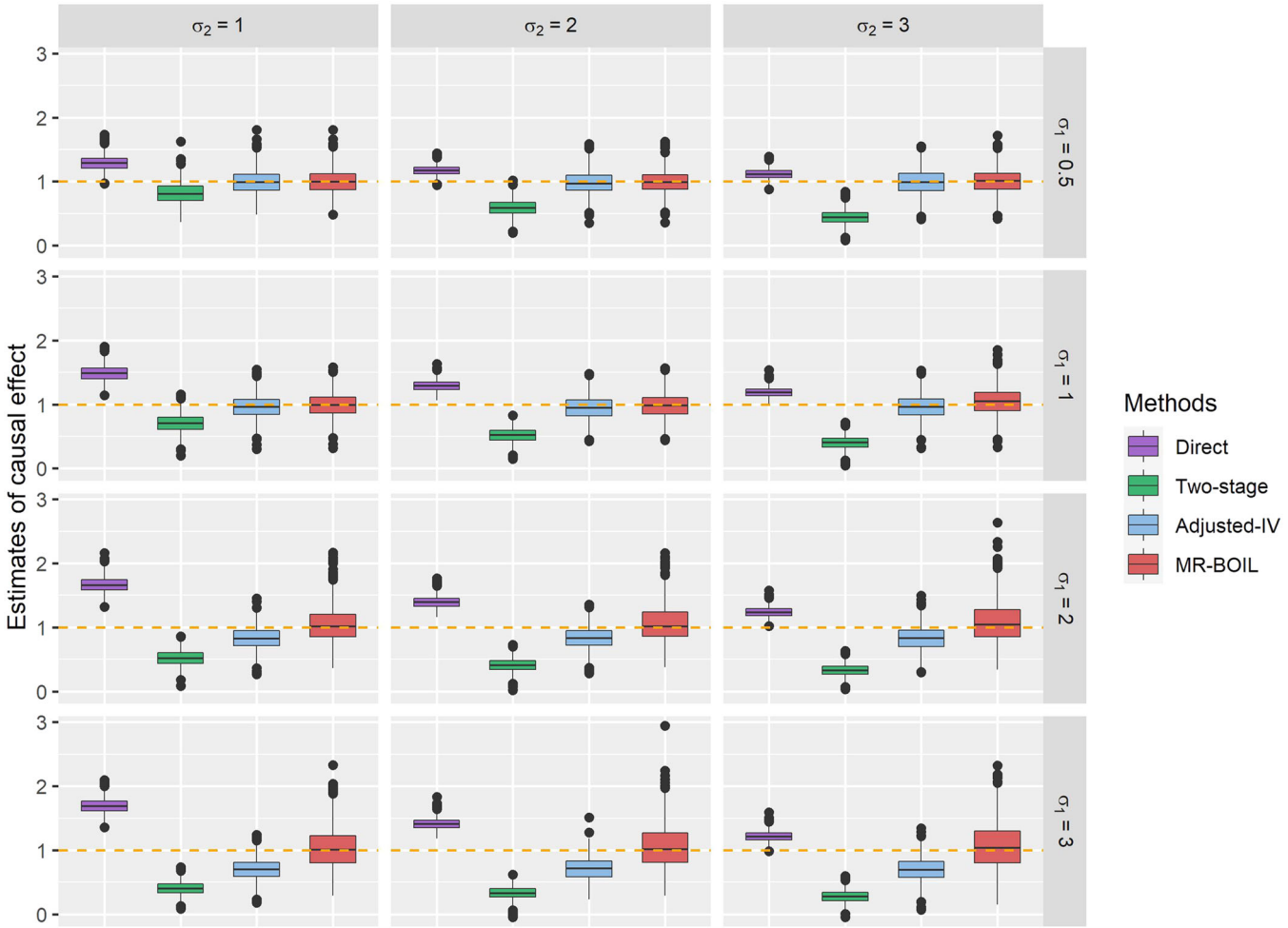

FIGURE 1 Boxplot of point estimation for the causal effect β_1 with direct method, two-stage method, adjusted-IV method and MR-BOIL method. The true value of causal effect is $\beta_1 = 1$. Sample size is $n = 1000$, and other nuisances are $\beta_0 = 2$, $\gamma = 1$, $\rho = 0.8$. Twelve confounding scenarios are set by adjusting the effect of confounder on the exposure with σ_2 to 1, 2, 3, and for each σ_2, σ_1 is set separately to 0.5, 1, 2, 3 representing the increasing effect of confounder on the outcome. For each set of parameter values, 1000 replications are performed. The orange dashed line indicates the true value of β_1.

value of β_1^{TSLS} equals 0, providing control of the type I error rate. This motivates us to set up scenarios in which we may compare the power of the two-stage method to that of the MR-BOIL method. According to the formula in Appendix A.4, when σ_1, σ_2, ρ and β_1 are large, β_1^{TSLS} will be a seriously biased estimate with a larger deviation from the true value β_1. Therefore we still fix $\beta_0 = 2$, $\gamma = 1$ as before, and define the alternative hypothesis as $H_1 : \beta_1 = 1$. Then we choose $\sigma_2 = 4, 5$, and set σ_1 to 1, 2, 3, 4, 5, respectively. Figure 7 and Table 7 indicate that, as expected, the power of the two-stage method is always lower than that of MR-BOIL method, and the difference is larger with growing σ_1. Furthermore, our method always maintains a relatively high power (>0.7) regardless of the influence of confounding factors.

3.4 | Sensitivity analysis

To investigate the robustness of MR-BOIL method compared with other methods, two types of violation of the assumptions are considered: (1) MR with pleiotropy, (2) MR with confounders following a multivariate t distribution.

3.4.1 | MR with pleiotropy

The data generation mechanism here is similar to that in the previous subsection, except for the added pleiotropic effect of the IV in the second stage of Equation (1), that is, violation of exclusion restriction assumption,

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 X_i + \alpha Z_i + U_i.$$

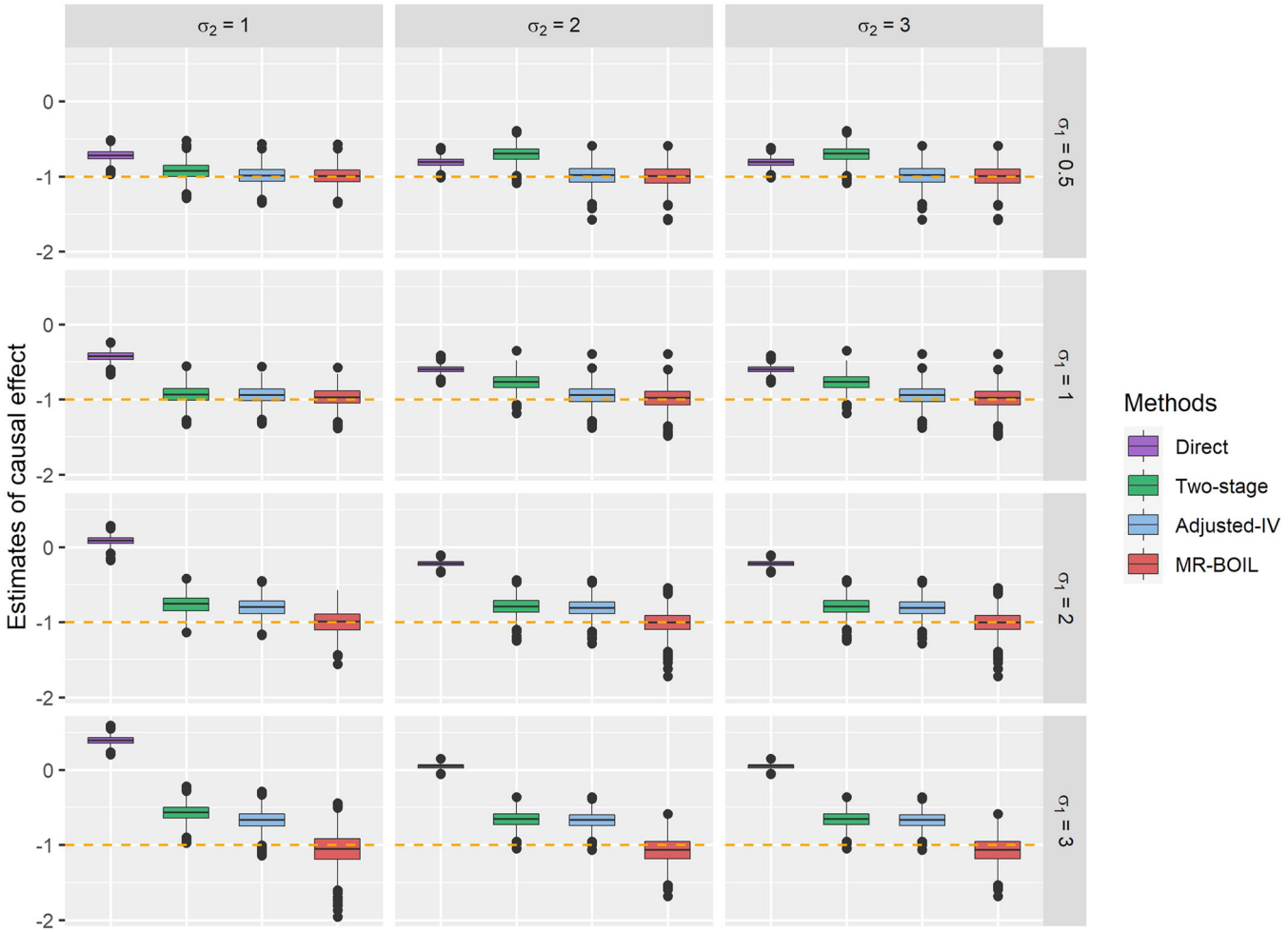

FIGURE 2 Boxplot of point estimation for the causal effect β_1 with direct method, two-stage method, adjusted-IV method and MR-BOIL method. The true value of causal effect is $\beta_1 = -1$. Sample size is $n = 1000$, and other nuisances are $\beta_0 = 2$, $\gamma = 1$, $\rho = 0.8$. Twelve confounding scenarios are set by adjusting the effect of confounder on the exposure with σ_2 to 1, 2, 3, and for each σ_2, σ_1 is set separately to 0.5, 1, 2, 3 representing the increasing effect of confounder on the outcome. For each set of parameter values, 1000 replications are performed. The orange dashed line indicates the true value of β_1.

Then we consider the influence of pleiotropic effect on the estimation of causal effect β_1 by changing the values of α. With the nuisance parameters set to $\beta_0 = 2, \gamma = 1, \sigma_1 = 2, \sigma_2 = 1, \rho = 0.8$, the true causal effect is $\beta_1 = 1$. In particular, consider three pleiotropic scenarios that affect the outcome by adjusting α to $-0.1, 0.01, 0.1$. As shown in Figure 8 and Table 8, all methods show bias, while our MR-BOIL method has smaller bias and performs relatively more robust.

3.4.2 | MR with confounders following a multivariate t distribution

We also consider the impact on estimation of causal effect β_1 under mis-specified distribution of confounders in model (1) by assuming that confounding factors U_i and V_i follow a multivariate t distribution with mean $(0, 0)^T$, scale matrix V and k degrees of freedom (detailed description shown in Appendix A.5). The true causal effect is $\beta_1 = 1$, with the nuisance parameters set to $\beta_0 = 2, \gamma = 1, \sigma_2 = 1, \rho = 0.8$. In particular, consider four confounding scenarios that increasingly affect the outcome by adjusting σ_1 to $1, 2, 3, 4$. As shown in Figure 9 and Table 9, with σ_1 changing, the difference between the t-distribution and the normal distribution becomes larger, while MR-BOIL method has smaller bias compared with other methods and is therefore more robust.

4 | REAL DATA ANALYSIS

In this section, we use real data to demonstrate the validity of our method. When both exposure and outcome are continuous variables, it is convenient to

TABLE 1 Performance of point estimation for the causal effect β_1 with direct method, two-stage method, adjusted-IV method and MR-BOIL method.

Setting		Direct			Two-stage			Adjusted-IV			MR-BOIL		
σ_2	σ_1	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
1	0.5	1.287	0.287	0.311	0.817	−0.183	0.248	0.994	−0.006	0.189	1.004	0.004	0.188
	1	1.489	0.489	0.504	0.706	-0.294	0.327	0.960	−0.040	0.188	0.991	−0.009	0.192
	2	1.665	0.665	0.676	0.519	−0.481	0.497	0.828	−0.172	0.243	1.044	0.044	0.280
	3	1.693	0.693	0.703	0.400	−0.600	0.610	0.698	−0.302	0.345	1.037	0.037	0.321
2	0.5	1.170	0.170	0.189	0.590	−0.410	0.430	0.985	−0.015	0.185	0.999	−0.001	0.181
	1	1.292	0.292	0.306	0.516	−0.484	0.497	0.951	−0.049	0.187	0.985	−0.015	0.188
	2	1.395	0.395	0.406	0.408	−0.592	0.602	0.835	−0.165	0.240	1.061	0.061	0.293
	3	1.412	0.412	0.422	0.331	−0.669	0.676	0.711	−0.289	0.340	1.053	0.053	0.339
3	0.5	1.120	0.120	0.144	0.442	−0.558	0.569	0.988	−0.012	0.192	1.005	0.005	0.188
	1	1.192	0.192	0.208	0.402	−0.598	0.607	0.961	−0.039	0.196	1.045	0.045	0.223
	2	1.240	0.240	0.254	0.328	−0.672	0.679	0.831	−0.169	0.253	1.076	0.076	0.330
	3	1.218	0.218	0.233	0.277	−0.723	0.730	0.700	−0.300	0.351	1.064	0.064	0.367

Note: The true value of causal effect is $\beta_1 = 1$. Sample size is $n = 1000$, and other nuisances are $\beta_0 = 2$, $\gamma = 1$, $\rho = 0.8$. Twelve confounding scenarios are set by adjusting the effect of confounder on the exposure with σ_2 to 1, 2, 3, and for each σ_2, σ_1 is set separately to 0.5, 1, 2, 3 representing the increasing effect of the confounder on the outcome. For each set of parameter values, 1000 replications are performed.

TABLE 2 Performance of point estimation for the causal effect β_1 with direct method, two-stage method, adjusted-IV method and MR-BOIL method. The true value of causal effect is $\beta_1 = -1$.

Setting		Direct			Two-stage			Adjusted-IV			MR-BOIL		
σ_2	σ_1	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
1	0.5	−0.718	0.282	0.291	−0.929	0.071	0.136	−0.989	0.011	0.122	−0.995	0.005	0.122
	1	−0.427	0.573	0.577	−0.935	0.065	0.133	−0.943	0.057	0.131	−0.970	0.030	0.126
	2	0.086	1.086	1.088	−0.765	0.235	0.264	−0.801	0.199	0.234	−1.003	−0.003	0.159
	3	0.393	1.393	1.394	−0.573	0.427	0.442	−0.673	0.327	0.352	−1.063	−0.063	0.225
2	0.5	−0.812	0.188	0.196	−0.699	0.301	0.319	−0.987	0.013	0.136	−0.995	0.005	0.139
	1	−0.601	0.399	0.401	−0.772	0.228	0.253	−0.948	0.052	0.141	−0.984	0.016	0.143
	2	−0.219	0.781	0.781	−0.793	0.207	0.239	−0.810	0.190	0.226	−1.007	−0.007	0.155
	3	0.047	1.047	1.047	−0.662	0.338	0.355	−0.673	0.327	0.345	−1.076	−0.076	0.194
3	0.5	−0.866	0.134	0.146	−0.521	0.479	0.488	−0.992	0.008	0.154	−1.005	−0.005	0.163
	1	−0.706	0.294	0.298	−0.576	0.424	0.434	−0.957	0.043	0.146	−1.005	−0.005	0.166
	2	−0.393	0.607	0.608	−0.660	0.340	0.358	−0.816	0.184	0.229	−1.017	−0.017	0.190
	3	−0.156	0.844	0.844	−0.653	0.347	0.365	−0.676	0.324	0.345	−1.076	−0.076	0.204

Note: Sample size is $n = 1000$, and other nuisances are $\beta_0 = 2$, $\gamma = 1$, $\rho = 0.8$. Twelve confounding scenarios are set by adjusting the effect of confounder on the exposure with σ_2 to 1, 2, 3, and for each σ_2, σ_1 is set separately to 0.5, 1, 2, 3 representing the increasing effect of the confounder on the outcome. For each set of parameter values, 1000 replications are performed.

test the validity of any MR method by taking the outcome as the exposure, for example, both as BMI (Wang et al., 2022; Zhao et al., 2020) factor. Moreover, any valid method should be able to identify the causal effect of the exposure on itself as one. When the outcome is the binary variable, one analogous idea is to consider exposure X as SBP and outcome Y as hypertension indicator ($Y = 1$ if SBP $\geq$140 mmHg, $Y = 0$ otherwise). It

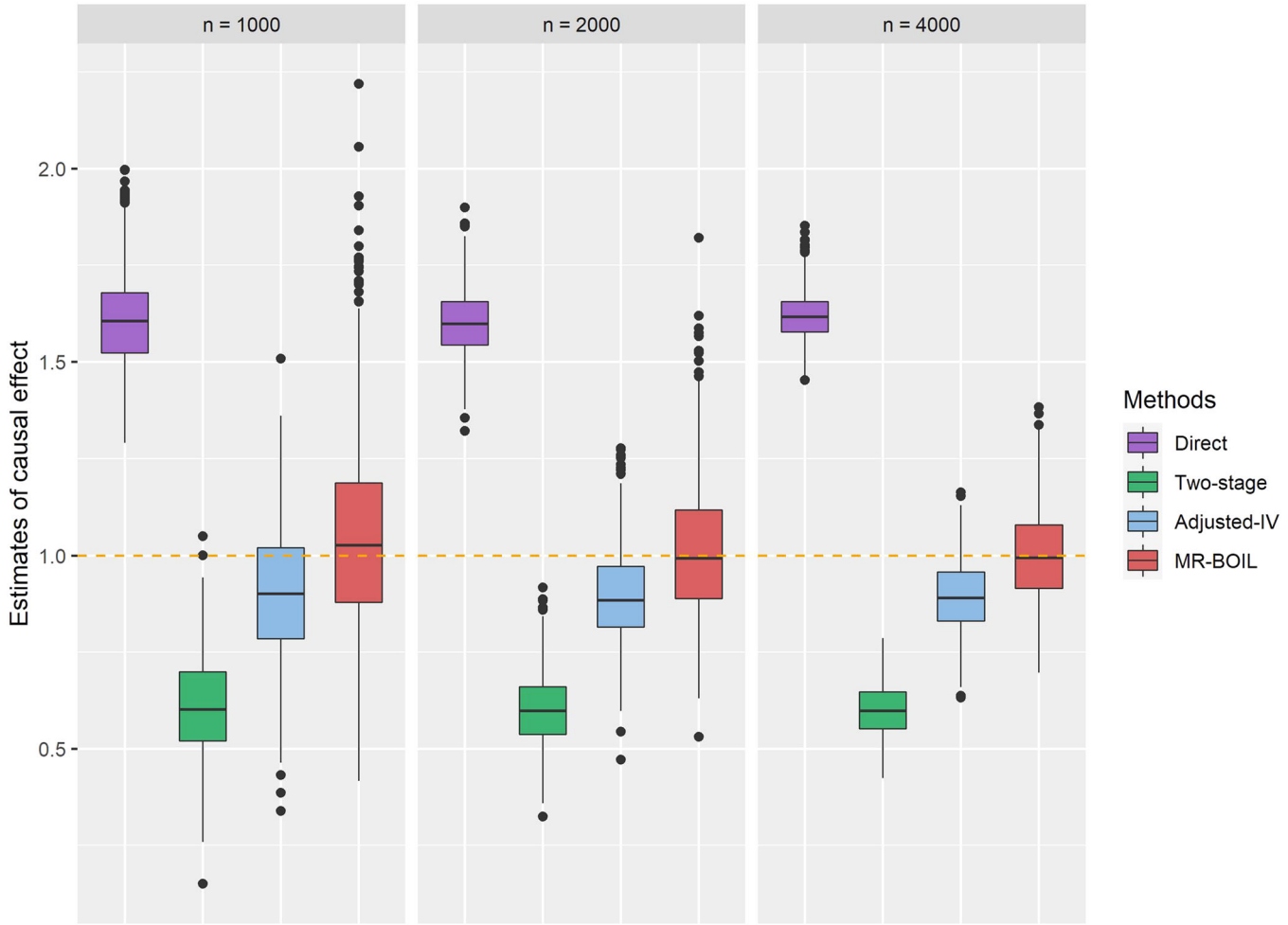

FIGURE 3 Boxplot of point estimation for the causal effect β_1 with direct method, two-stage method, adjusted-IV method and MR-BOIL method. The true value of causal effect is $\beta_1 = 1$. The nuisances are $\beta_0 = 2, \gamma = 1, \rho = 0.8, \sigma_1 = 1.5, \sigma_2 = 1$. Three scenarios are set by letting sample size n change to 1000, 2000, 4000. For each set of parameter values, 1000 replications are performed. The orange dashed line indicates the true value of β_1.

TABLE 3 Performance of point estimation for the causal effect β_1 with direct method, two-stage method, adjusted-IV method and MR-BOIL method.

Setting	Direct			Two-stage			Adjusted-IV			MR-BOIL		
n	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
1000	1.605	0.605	0.617	0.608	−0.392	0.413	0.901	−0.099	0.196	1.045	0.045	0.246
2000	1.602	0.602	0.608	0.599	−0.401	0.411	0.891	−0.109	0.163	1.012	0.012	0.170
4000	1.619	0.619	0.622	0.599	−0.401	0.406	0.894	−0.106	0.140	1.000	0.000	0.118

Note: The true value of causal effect is $\beta_1 = 1$. The nuisances are $\beta_0 = 2, \gamma = 1, \rho = 0.8, \sigma 1 = 1.5, \sigma 2 = 1$. Three scenarios are set by letting sample size n change to 1000, 2000, 4000. For each set of parameter values, 1000 replications are performed.

is known that there should be a significant causal relationship between them.

To evaluate the causal relationship between SBP and hypertension, we use data of 13,113 individuals from the ARIC study, which is one of the largest multi-ethnic sampling frame studies in the United States. ARIC data set contains 870,275 SNPs and over 450 phenotypes, with nearly 70% of people being European-American and the remaining being African American.

Considering SBP as an exposure, individuals and valid SNPs are selected based on the following strategy. Firstly, only individuals of white origin are included in

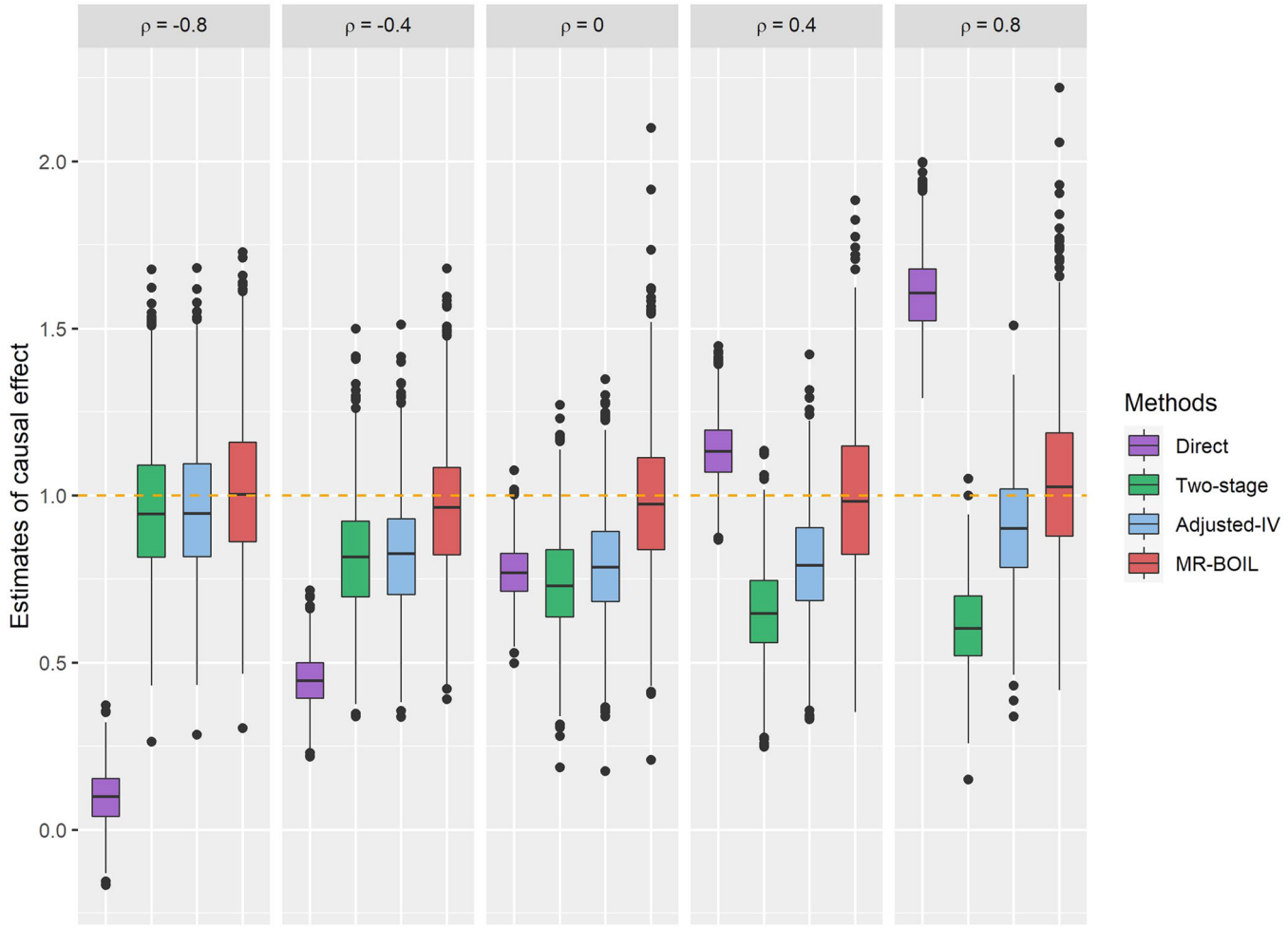

FIGURE 4 Boxplot of point estimation for the causal effect β_1 with direct method, two-stage method, adjusted-IV method and MR-BOIL method. The true value of causal effect is $\beta_1 = 1$. Sample size is $n = 1000$, and other nuisances are $\beta_0 = 2$, $\gamma = 1$, $\sigma_1 = 5$, $\sigma_2 = 5$. Five scenarios are set by taking ρ to −0.8, −0.4, 0, 0.4, 0.8, respectively to measure the degree of correlation between confounders. For each set of parameter values, 1000 replications are performed. The orange dashed line indicates the true value of β_1.

TABLE 4 Performance of point estimation for the causal effect β_1 with direct method, two-stage method, adjusted-IV method and MR-BOIL method.

Setting	Direct			Two-stage			Adjusted-IV			MR-BOIL		
ρ	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
−0.8	0.098	−0.902	0.906	0.955	−0.045	0.220	0.959	−0.041	0.220	1.016	0.016	0.229
−0.4	0.447	−0.553	0.559	0.815	−0.185	0.253	0.825	−0.175	0.247	0.962	−0.038	0.204
0	0.771	−0.229	0.245	0.735	−0.265	0.308	0.789	−0.211	0.268	0.982	−0.018	0.221
0.4	1.134	0.134	0.165	0.654	−0.346	0.374	0.795	−0.205	0.262	0.991	−0.009	0.229
0.8	1.605	0.605	0.617	0.608	−0.392	0.413	0.901	−0.099	0.196	1.045	0.045	0.246

Note: The true value of causal effect is $\beta_1 = 1$. Sample size is $n = 1000$, and other nuisances are $\beta_0 = 2$, $\gamma = 1$, $\sigma_1 = 5$, $\sigma_2 = 5$. Five scenarios are set by taking ρ to −0.8, −0.4, 0, 0.4, 0.8, respectively to measure the degree of correlation between confounders. For each set of parameter values, 1000 replications are performed.

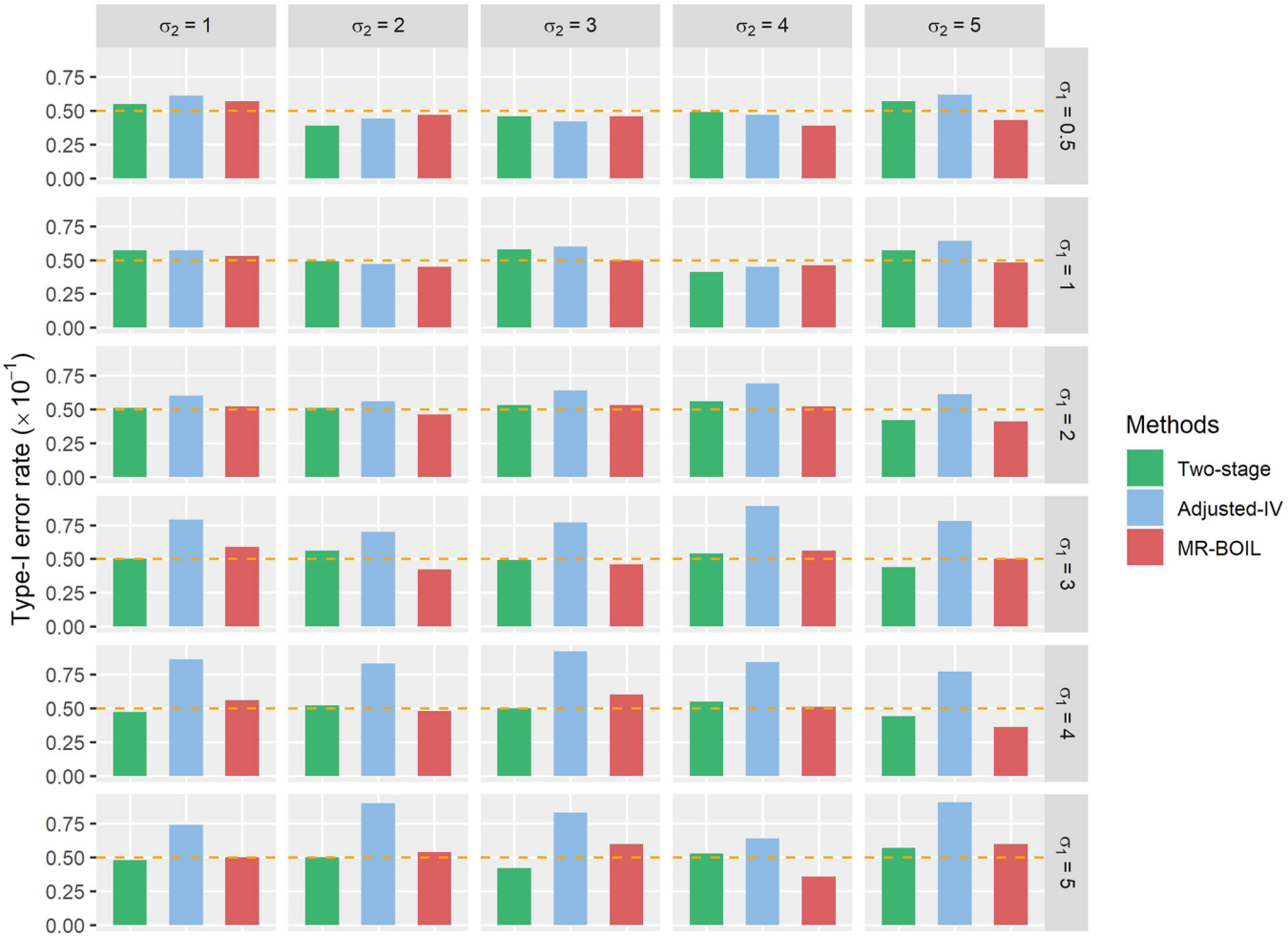

FIGURE 5 Bar plot of type I error rates of two-stage method, adjusted-IV method and MR-BOIL method under the null hypothesis $H_0 : \beta_1 = 0$. The sample size is $n = 1000$, and other nuisances are $\beta_0 = 2$, $\gamma = 1$, $\rho = 0.8$. Thirty confounding scenarios are set by adjusting the effect of the confounder on the exposure with σ_2 to 1, 2, 3, 4, 5, and for each σ_2, σ_1 is set separately to 0.5, 1, 2, 3, 4, 5 representing the increasing effect of the confounder on the outcome. For each set of parameter values, 1000 replications are performed. The orange dashed horizontal line indicates the nominal significance of 0.05.

the following analysis. We extract European ancestry individuals from ARIC genome data set and impute the data set on Michigan Imputation Center (Das et al., 2016) with EUR population from 1000 Genomes Phase 3 v5 (GRCh37/hg19) reference panel (Chang et al., 2015). Second, We conduct LD prune procedure (with r^2 below 0.001 and MAF above 0.01) along with Hardy–Weinberg equilibrium test (with p value above 1e−6) to select 10,421 independent common SNPs from the imputed genome data set in PLINK (version 2.0, The 1000 Genomes Project Consortium, 2015). Moreover, since the ARIC data set has insufficient number of significant SNPs associated with SBP when screening through the whole genome with simple linear regression model, we choose 16,098 candidate SNPs with reference to the SNPs whose p values less than 1e−3 in the UK Biobank GWAS round 2 results (Nealelab http://www.nealelab.is/uk-biobank). and then take the intersection of the above two sets. Finally, 8 SNPs and 8708 individuals are selected for analysis.

After preprocessing the raw data, we start the MR study. The eight chosen SNPs are used as IVs, and it is assumed that they all satisfy the assumption of valid IV. The corresponding SBP-hypertension causal effect value is then determined for each IV using the model and method in this paper. As the comparative methods, we choose the two-stage method and the adjusted-IV method (the direct method is excluded due to its inflated

TABLE 5 Performance of type I error rates of two-stage method, adjusted-IV method and MR-BOIL method under the null hypothesis $H_0 : \beta_1 = 0$.

Setting		Type I error rate		
σ_2	σ_1	Two-stage	Adjusted-IV	MR-BOIL
1	0.5	0.055	0.061	0.057
	1	0.057	0.057	0.053
	2	0.051	0.060	0.052
	3	0.050	0.079	0.059
	4	0.047	0.086	0.056
	5	0.048	0.074	0.050
2	0.5	0.039	0.044	0.047
	1	0.049	0.047	0.045
	2	0.051	0.056	0.046
	3	0.056	0.070	0.042
	4	0.052	0.083	0.048
	5	0.050	0.090	0.054
3	0.5	0.046	0.042	0.046
	1	0.058	0.060	0.050
	2	0.053	0.064	0.053
	3	0.049	0.077	0.046
	4	0.050	0.092	0.060
	5	0.042	0.083	0.060
4	0.5	0.049	0.047	0.039
	1	0.041	0.045	0.046
	2	0.056	0.069	0.052
	3	0.054	0.089	0.056
	4	0.055	0.084	0.051
	5	0.053	0.064	0.036
5	0.5	0.057	0.062	0.043
	1	0.057	0.064	0.048
	2	0.042	0.061	0.041
	3	0.044	0.078	0.050
	4	0.044	0.077	0.036
	5	0.057	0.091	0.060

Note: The sample size is $n = 1000$, and other nuisances are $\beta_0 = 2$, $\gamma = 1$, $\rho = 0.8$ Thirty confounding scenarios are set by adjusting the effect of confounder on the exposure with σ_2 to 1, 2, 3, 4, 5, and for each σ_2, $\sigma 1$ is set separately to 0.5, 1, 2, 3, 4, 5 representing the increasing effect of the confounder on the outcome. For each set of parameter values, 1000 replications are performed.

type I error rate). Together with the fact that the objective function of the M-step is a nonconvex function, the choice of the initial value might easily lead to a locally optimum solution, which is an unavoidable problem in the implementation of the EM algorithm. Therefore, we use a warm-up strategy to determine the appropriate global initial values. First, a random one-eighth subsample of the full sample is selected each time to run the EM algorithm, and then the solution after 100 iterations is used as our warm-up initial value alternative. Second, repeat the above procedure for 200 times, with the following beginning value for the optimization of the LRT statistic (3) in each warm-up process as $\beta_0 \sim Uniform(-5, 5)$, $\sigma_1 \sim Uniform(0.1, 5)$, $\rho \sim Uniform(-0.5, 0.5)$ and γ, σ_2 as the regression coefficients of SBP to SNPs. Furthermore, we select the set of initial values that maximize the likelihood of complete samples as our global initial values from the above 200 potential warm-up beginning values (some details of runtime and memory usage for both simulation study and real data analysis are provided in Appendix A.6). In addition, bootstrap method can be adopted to estimate the variance of MR-BOIL method (details see Appendix A.6). Table 10 shows the results of the casual relationship between SBP and hypertension for two-stage method, adjusted-IV method and MR-BOIL method. We can clearly see that except for the SNP rs7356321 (Two-stage: $\hat{\beta}_1 = 2.456$, p value $= 0.004$; Adjusted-IV: $\hat{\beta}_1 = 3.728$, p value; MR-BOIL: $\hat{\beta}_1 = 2.662$, p value $= 6.329 \times 10^{-8}$), neither two-stage method nor adjusted-IV method could estimate the causal effect of SBP on hypertension when any of the remaining seven SNPs is used as an IV. In contrast, our method consistently estimates that significant causal relationship for all eight SNPs. On the other hand, our method correctly estimates the causal relationship of positive sign between SBP and hypertension for each IV, while the other two methods do not guarantee this, for example, SNP rs11884411 (Two-stage: $\hat{\beta}_1 = -1.330$, p value $= 0.802$; adjusted-IV: $\hat{\beta}_1 = -4.900$, p value $= 0.436$; MR-BOIL: $\hat{\beta}_1 = 3.271$, p value $= 4.244 \times 10^{-3}$) and SNP rs7100920 (Two-stage: $\hat{\beta}_1 = -0.975$, p value $= 0.822$; adjusted-IV: $\hat{\beta}_1 = -0.362$, p value $= 0.944$; MR-BOIL: $\hat{\beta}_1 = 3.439$, p value $= 2.748 \times 10^{-2}$).

Another strategy to verify the soundness of our approach is to use the negative control outcome (Hu et al., 2021; Sanderson et al., 2020). The characters of the negative control outcome can be summarized as: (a) they

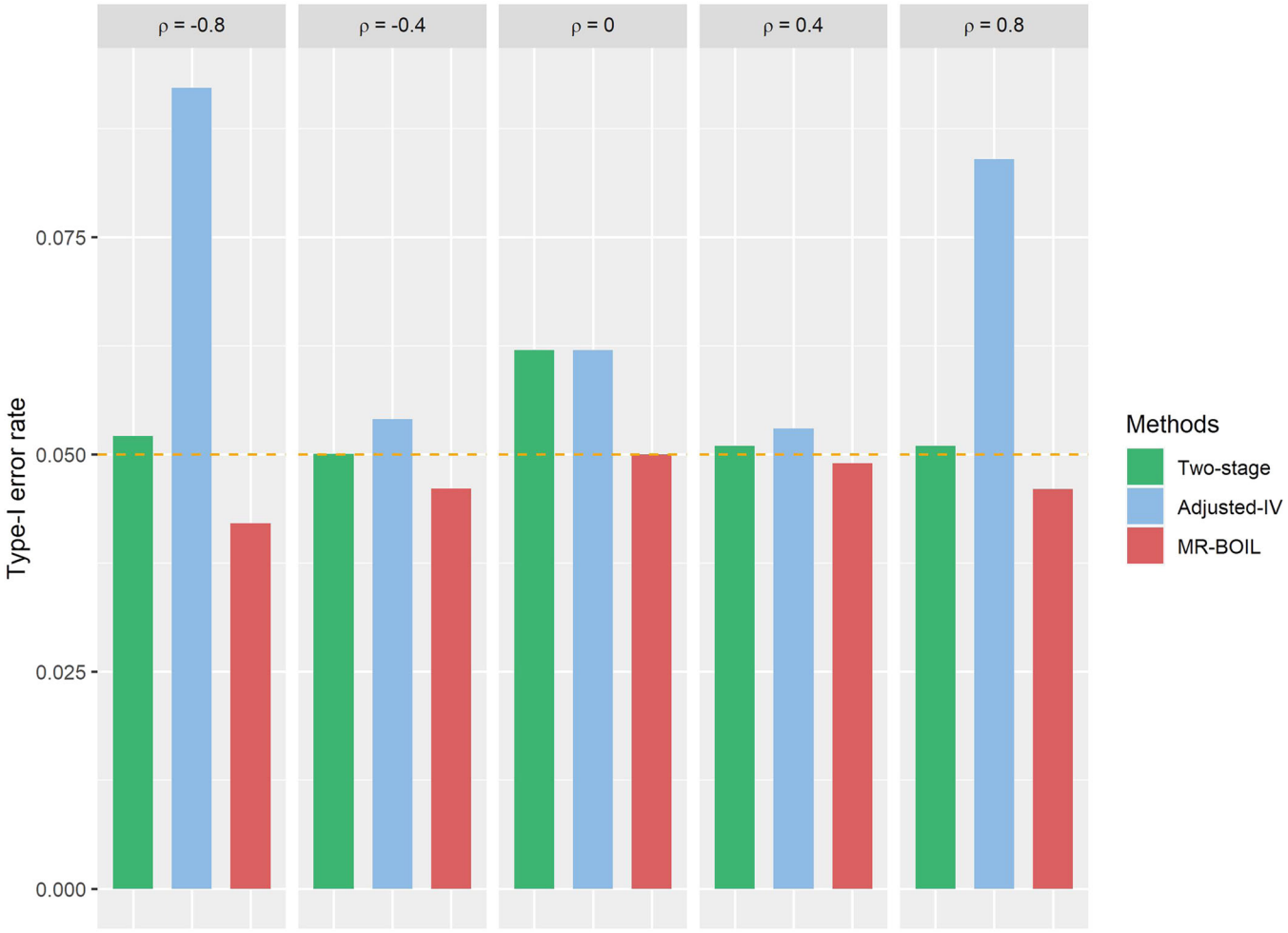

FIGURE 6 Bar plot of type I error rates of two-stage method, adjusted-IV method and MR-BOIL method under the null hypothesis $H_0 : \beta_1 = 0$. Sample size is $n = 1000$, and other nuisances are $\beta_0 = 2, \gamma = 1, \sigma_1 = 5, \sigma_2 = 5$. Five scenarios are set by taking ρ to $-0.8, -0.4, 0, 0.4, 0.8$, respectively to measure the degree of correlation between confounders. For each set of parameter values, 1000 replications are performed. The orange dashed horizontal line indicates the nominal significance of 0.05.

TABLE 6 Performance of type I error rates of two-stage method, adjusted-IV method and MR-BOIL method under the null hypothesis $H_0 : \beta_1 = 0$.

Setting	Type I error rate		
ρ	Two-stage	Adjusted-IV	MR-BOIL
−0.8	0.052	0.092	0.042
−0.4	0.050	0.054	0.046
0	0.062	0.062	0.050
0.4	0.051	0.053	0.049
0.8	0.051	0.084	0.046

Note: Sample size is $n = 1000$, and other nuisances are $\beta_0 = 2, \gamma = 1, \sigma_1 = 5, \sigma_2 = 5$. Five scenarios are set by taking ρ to $-0.8, -0.4, 0, 0.4, 0.8$, respectively to measure the degree of correlation between confounders. For each set of parameter values, 1000 replications are performed.

should not be causally affected by the chosen exposure; (b) exposure and outcome could be affected by some unmeasured confounders, for example, population stratification. Therefore, we choose DBP and hypertension as our exposure and outcome, respectively. The expected value of the causal effect between them is zero. Similar to the above data preprocess and MR-BOIL method implementation flow, twelve SNPs are selected to be IVs. Table 11 presents the estimates and p-values of the DBP-hypertension causal effects for two-stage method, adjusted-IV method and MR-BOIL method. It can be seen that for the SNP rs12534379, two-stage method and adjusted-IV method significantly estimate the causal effect (Two-stage: $\hat{\beta}_1 = -8.129$, p value $= 0.031$; adjusted-IV: $\hat{\beta}_1 = -9.572$, p value $= 0.020$), which contradicts the

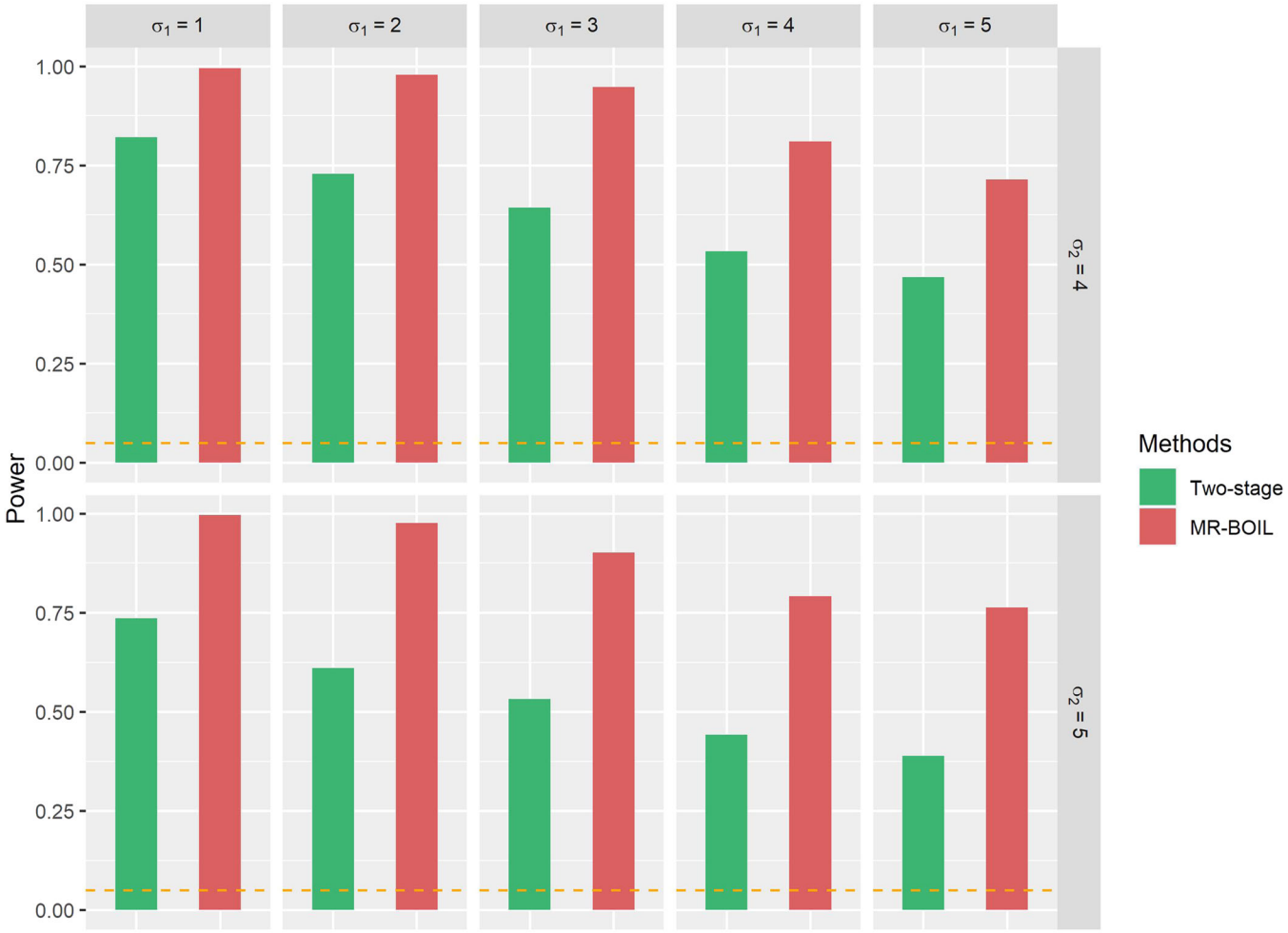

FIGURE 7 Bar plot of power of two-stage method and MR-BOIL method under the alternative hypothesis $H_1 : \beta_1 = 1$. The sample size is $n = 1000$, and other nuisances are $\beta_0 = 2$, $\gamma = 1$, $\rho = 0.8$. Ten scenarios are set by choosing $\sigma_2 = 4$, 5, and σ_1 to 1, 2, 3, 4, 5, respectively to represent the increasing effect of the confounder on the exposure and outcome. For each set of parameter values, 1000 replications are performed. The orange dashed horizontal line indicates the nominal significance of 0.05.

TABLE 7 Performance of power of two-stage method and MR-BOIL method under the alternative hypothesis $H_1 : \beta_1 = 1$.

Setting		Power	
σ_2	σ_1	Two-stage	MR-BOIL
4	1	0.821	0.995
	2	0.728	0.979
	3	0.643	0.947
	4	0.533	0.810
	5	0.468	0.714
5	1	0.736	0.997
	2	0.611	0.976
	3	0.532	0.902
	4	0.442	0.792
	5	0.389	0.764

Note: The sample size is $n = 1000$, and other nuisances are $\beta_0 = 2$, $\gamma = 1$, $\rho = 0.8$. Ten scenarios are set by choosing $\sigma_2 = 4$, 5, and σ_1 to 1, 2, 3, 4, 5, respectively to represent the increasing effect of the confounder on the exposure and outcome. For each set of parameter values, 1000 replications are performed.

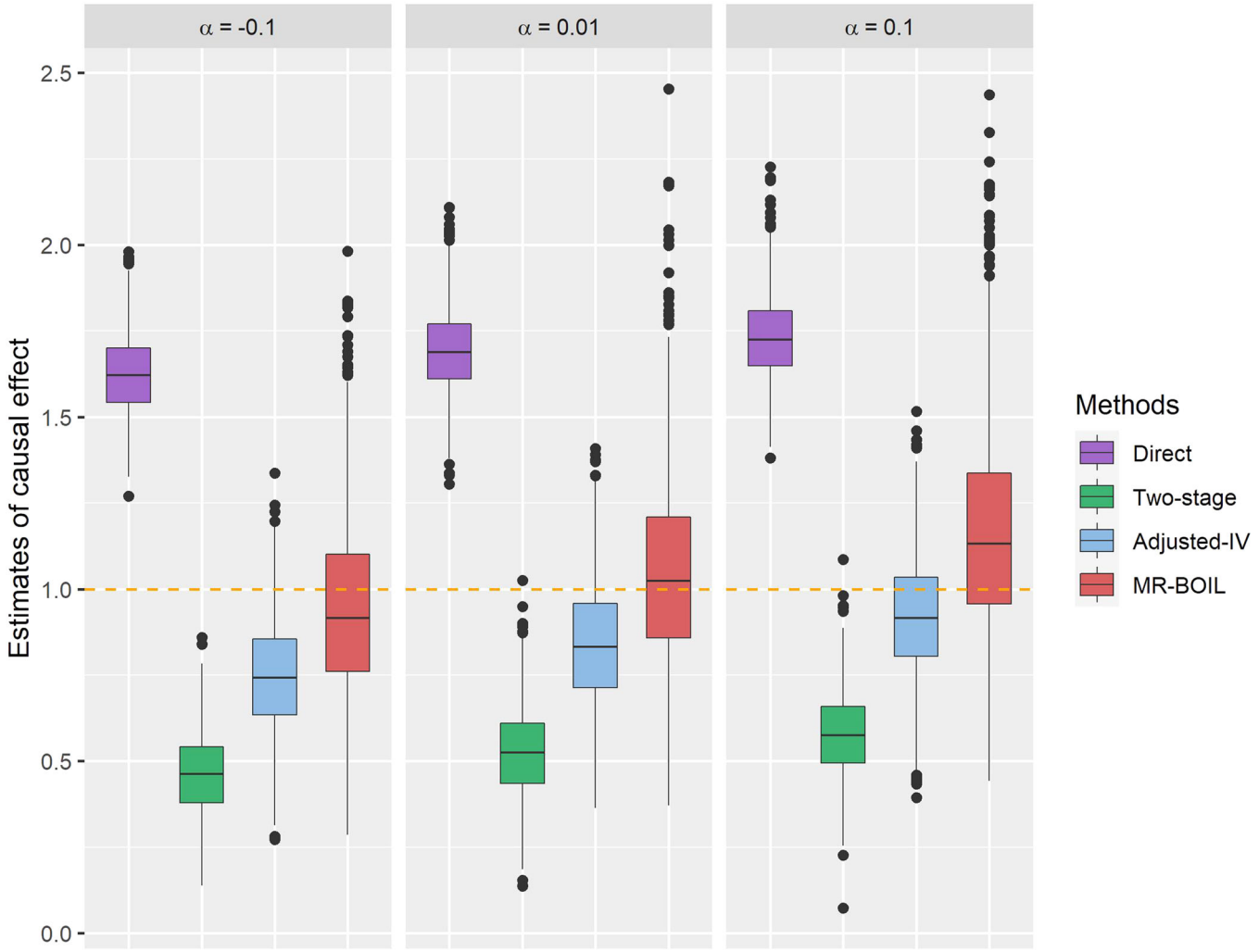

FIGURE 8 (MR with pleiotropy) Boxplot of point estimation for the causal effect β_1 with direct method, two-stage method, adjusted-IV method and MR-BOIL method. The true value of causal effect is $\beta_1 = 1$. Sample size is $n = 1000$, and other nuisances are $\beta_0 = 2, \gamma = 1, \sigma_1 = 2, \sigma_2 = 1, \rho = 0.8$. Three pleiotropic scenarios are set by taking α to −0.1, 0.01, 0.1, respectively, to measure the pleiotropic effect on outcome. For each set of parameter values, 1000 replications are performed. The orange dashed line indicates the true value of β_1.

TABLE 8 (MR with pleiotropy) Performance of point estimation for the causal effect β_1 with direct method, two-stage method, adjusted-IV method and MR-BOIL method.

Setting	Direct			Two-stage			Adjusted-IV			MR-BOIL		
α	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
−0.1	1.624	0.624	0.635	0.463	−0.537	0.550	0.745	−0.255	0.304	0.942	−0.058	0.270
0.01	1.693	0.693	0.704	0.522	−0.478	0.495	0.836	−0.164	0.246	1.050	0.050	0.297
0.1	1.728	0.728	0.739	0.577	−0.423	0.441	0.916	−0.084	0.196	1.168	0.168	0.351

Note: The true value of causal effect is $\beta 1 = 1$. Sample size is $n = 1000$, and other nuisances are $\beta_0 = 2, \gamma = 1, \sigma_1 = 2, \sigma_2 = 1, \rho = 0.8$. Three scenarios are set by taking α to −0.1, 0.01, 0.1, respectively, to measure the pleiotropic effect on outcome. For each set of parameter values, 1000 replications are performed.

property (a) of the negative control outcome DBP-hypertension. In contrast, our MR-BOIL method satisfies this property at this point (MR-BOIL: $\hat{\beta}_1 = -0.716$, p value = 0.336, nonsignificant). In addition, for the SNP rs1531475, two-stage method and adjusted-IV method estimate larger values for the causal effect compared with our MR-BOIL method (Two-stage: $\hat{\beta}_1 = 723.712$, p value = 0.293; adjusted-IV: $\hat{\beta}_1 = 885.760$, p value = 0.238;

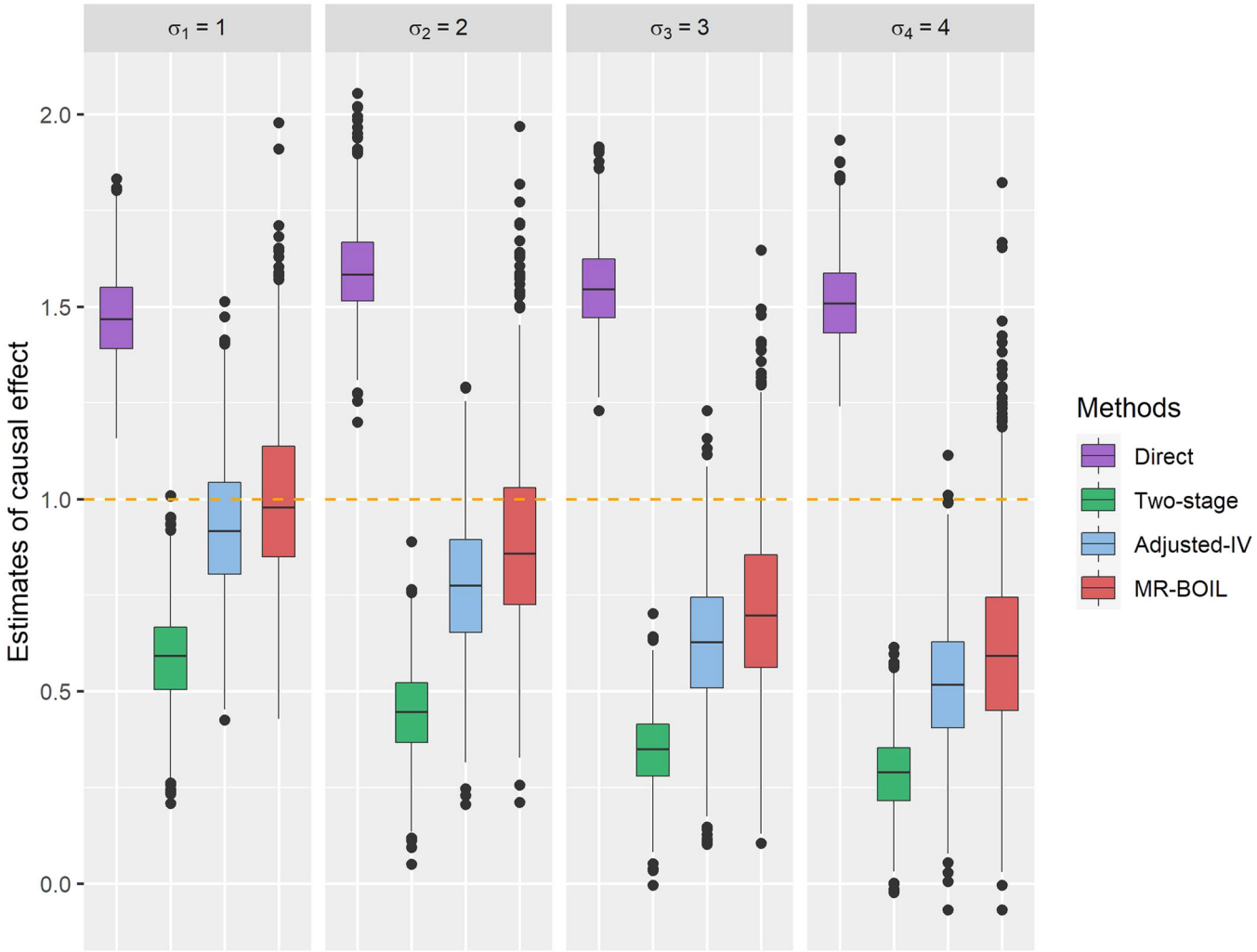

FIGURE 9 (MR with confounders following a multivariate*t* distribution) Boxplot of point estimation for the causal effect β_1 with direct method, two-stage method, adjusted-IV method and MR-BOIL method. The true value of causal effect is $\beta_1 = 1$. Sample size is $n = 1000$, and other nuisances are $\beta_0 = 2, \gamma = 1, \sigma_1 = 2, \sigma_2 = 1, \rho = 0.8$. Four confounding scenarios that increasingly affect the outcome by adjusting σ_1 to 1, 2, 3, 4. For each set of parameter values, 1000 replications are performed. The orange dashed line indicates the true value of β_1.

TABLE 9 (MR with confounders following a multivariate *t* distribution) Performance of point estimation for the causal effect β_1 with direct method, two-stage method, adjusted-IV method and MR-BOIL method.

Setting	Direct			Two-stage			Adjusted-IV			MR-BOIL		
σ_1	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
1	1.473	0.473	0.487	0.590	−0.410	0.429	0.926	−0.074	0.196	1.004	0.004	0.223
2	1.592	0.592	0.604	0.443	−0.557	0.570	0.773	−0.227	0.291	0.885	−0.115	0.275
3	1.548	0.548	0.559	0.349	−0.651	0.659	0.626	−0.374	0.413	0.716	−0.284	0.365
4	1.513	0.513	0.525	0.286	−0.714	0.721	0.520	−0.48	0.509	0.613	−0.387	0.457

Note: The true value of causal effect is $\beta_1 = 1$. Sample size is $n = 1000$, and other nuisances are $\beta_0 = 2, \gamma = 1, \sigma_2 = 1, \rho = 0.8$. Four senarios are set by taking σ_1 to 1, 2, 3, 4 representing the increasing effect of the confounder on the outcome. For each set of parameter values, 1000 replications are performed.

TABLE 10 The point estimates and p values of the SBP-hypertension causal effect for two-stage method, adjusted-IV method and MR-BOIL method.

	Two-stage		Adjusted-IV		MR-BOIL	
SNP Location	Estimate	p Value	Estimate	p Value	Estimate	p Value
rs11695862	−0.239	0.974	0.802	0.926	0.724	2.715E−6
rs11884411	−1.330	0.802	−4.900	0.436	3.271	4.244E−3
rs16854522	5.392	0.302	6.508	0.299	1.098	2.901E−6
rs7356321	2.456	0.004	3.278	0.001	2.662	6.329E−8
rs4140431	2.015	0.091	2.548	0.073	2.497	1.475E−6
rs10237149	3.486	0.345	4.859	0.268	1.102	1.016E−6
rs7100920	−0.975	0.822	−0.362	0.944	3.439	2.748E−2
rs1896341	6.874	0.625	4.413	0.792	1.824	3.774E−6

TABLE 11 The point estimates and p values of the DBP-hypertension causal effect for two-stage method, adjusted-IV method and MR-BOIL method.

	Two-stage		Adjusted-IV		MR-BOIL	
SNP Location	Estimate	p Value	Estimate	p Value	Estimate	p Value
rs7555399	0.226	0.931	0.793	0.780	2.506	0.873
rs6706087	4.375	0.487	5.764	0.397	0.200	0.483
rs7650936	0.079	0.941	−0.110	0.924	3.104	0.674
rs823729	1.657	0.165	1.945	0.138	2.491	0.155
rs1577636	13.369	0.253	16.907	0.184	2.504	0.239
rs507929	5.241	0.241	5.002	0.302	2.357	0.757
rs12534379	−8.129	0.031	−9.572	0.020	−0.716	0.336
rs10897694	2.348	0.768	2.380	0.785	−0.593	0.141
rs1531475	723.712	0.293	885.760	0.238	0.002	0.278
rs1444581	−4.165	0.607	−4.754	0.591	2.531	0.237
rs11075445	−9.813	0.522	−10.218	0.539	3.046	0.217
rs6108789	3.773	0.124	3.597	0.179	−0.597	0.353

MR-BOIL: $\hat{\beta}_1 = 0.002$, p value $= 0.278$), which can be considered as anomalous results. The remaining SNPs behave similarly for the estimation of the DBP-hypertension causal effect under the three methods.

In conclusion, as validation examples, it is found that our MR-BOIL approach can discover simple causality for the positive control outcome and identify the null causal effect for the negative control outcome, as opposed to the other two methods, which may even establish the exact opposite causality. Therefore, we need to be careful with the results found by the other two methods. In addition to the validation examples, some other exploratory example analysis are also conducted and the results are attached in Appendix A.6.

5 | DISCUSSION

This paper focuses on the situation when the outcome and exposure are binary and continuous variables in one-sample MR, respectively, to establish the so-called integrated likelihood method based on the expectation maximization algorithm for estimating the causal effect β_1. Specifically, we model the confounders affecting exposures and outcomes by a joint normal distribution. Then, based on the probability density transformation, we construct the integrated likelihood function, which motivates us to use the EM algorithm to find the optimal value.

Our main contributions are twofold. First, an asymptotically unbiased estimate of the causal effect is given when the confounders are normally distributed.

Second, the simulation reveals that when confounders have a large influence, our method not only controls the type I error rate better than the existing methods, but also has decent power.

Moreover, even when analyzing the simplest causal relationships in real data analysis, our approach outperforms traditional plain methods. In addition, when only one IV is included, many popular meta-analysis methods developed today to analyze summary data for the binary outcome variables in MR degenerate into two-stage method or adjusted-IV method (Burgess et al., 2013; Xu et al., 2021; Zhao et al., 2020). Therefore, when dealing with many IVs, these approaches can be regarded as generalizations of these two plain methods, and thus may inherit the drawback they show in this paper to some extent. This phenomenon deserves our attention.

There are flaws in our approach as well. First, we use only one IV at a time to infer the causality between the exposure and outcome. Second, we assume that all IVs are valid IVs, which has been widely extended to more complex situations in the presence of correlated and idiosyncratic pleiotropy (Xu et al., 2021; Xue et al., 2021; Zhao et al., 2020). Third, the proposed method MR-BOIL performs well under a strong parametric assumption of normal distribution for confounders. Although our method is found to be robust under model misspecification in some simulations, it is still a challenge to further generalize our method to a wider range of situations including multiple IVs case, invalid IVs case and summary statistics case. These will be a part of our future work.

In addition, there are still many optimization details to be explored in the implementation of the EM algorithm. For example, using the warm-up method to pick the initial value requires a sufficient number of replications if we want to get as close to the global optimum as possible. So it is still a challenge to effectively remove the uninformative replication. In the next place, in the simulation, we find that the variance of the estimate of the causal effects becomes larger when the influence of confounders is large. If we increase the sample size to reduce the variance, computational inefficiency occurs in both E-step and M-step. On the one hand, in the E-step, more repetitions are requisite to approximate the integration using Monte Carlo methods, which can be cumbersome and less accurate when the sample size is larger. On the other hand, the ordinary optimization algorithm in M-step is less efficient. Therefore, we will further combine the current popular stochastic gradient descent method (Bottou, 2012; Zinkevich et al., 2010), and the stochastic expectation maximization algorithm (Celeux et al., 1996; Nielsen, 2000) to improve the computational efficiency.

ACKNOWLEDGMENTS

We wish to thank ARIC for providing data. This work was supported in part by the National Natural Science Foundation of China (Nos. 11971117, 11571082).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data sets ARIC for this study can be applied from the dbGaP Study https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000090.v1.p1. GWAS catalogue is available at https://www.ebi.ac.uk/gwas/.

ORCID

Yue-Qing Hu http://orcid.org/0000-0002-5730-0317

REFERENCES

Allman, P. H., Aban, I., Long, D. M., Bridges Jr., S. L., Srinivasasainagendra, V., MacKenzie, T., Cutter, G., & Tiwari, H. K. (2021). A novel Mendelian randomization method with binary risk factor and outcome. *Genetic Epidemiology*, *45*(5), 549–560.

Anderson, T. W., & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, *20*(1), 46–63.

Angrist, J., & Imbens, G. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455.

Bottou, L. (2012). Stochastic gradient descent tricks. In: G. Montavon, G. B. Orr, & K. R. Müller (Eds.), *Neural networks: Tricks of the trade* (pp. 421–436). Springer.

Bowden, J., & Vansteelandt, S. (2011). Mendelian randomization analysis of case-control data using structural mean models. *Statistics in Medicine*, *30*(6), 678–694.

Burgess, S., Butterworth, A., Malarstig, A., & Thompson, S. G. (2012). Use of Mendelian randomisation to assess potential benefit of clinical intervention. *BMJ*, *345*, e7325.

Burgess, S., Butterworth, A., & Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology*, *37*(7), 658–665.

Burgess, S., CRP CHD Genetics Collaboration. (2013). Identifying the odds ratio estimated by a two-stage instrumental variable analysis with a logistic regression model. *Statistics in Medicine*, *32*(27), 4726–4747.

Burgess, S., Granell, R., Palmer, T. M., Sterne, J. A., & Didelez, V. (2014). Lack of identification in semiparametric instrumental variable models with binary outcomes. *American Journal of Epidemiology*, *180*(1), 111–119.

Burgess, S., Small, D. S., & Thompson, S. G. (2017). A review of instrumental variable estimators for Mendelian randomization. *Statistical Methods in Medical Research*, *26*(5), 2333–2355.

Burgess, S., & Thompson, S. G. (2012). Improving bias and coverage in instrumental variable analysis with weak instruments for continuous and binary outcomes. *Statistics in Medicine*, *31*(15), 1582–1600.

Burgess, S., & Thompson, S. G. (2013). Use of allele scores as instrumental variables for Mendelian randomization. *International Journal of Epidemiology*, *42*(4), 1134–1144.

Burgess, S., & Thompson, S. G. (2015). *Mendelian randomization: Methods for using genetic variants in causal estimation*. CRC Press.

Cai, B., Small, D. S., & Have, T. R. T. (2011). Two-stage instrumental variable methods for estimating the causal odds ratio: Analysis of bias. *Statistics in Medicine*, *30*(15), 1809–1824.

Celeux, G., Chauveau, D., & Diebolt, J. (1996). Stochastic versions of the em algorithm: An experimental study in the mixture case. *Journal of Statistical Computation and Simulation*, *55*(4), 287–314.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), s13742-015-0047-8.

Clarke, P. S., & Windmeijer, F. (2012). Instrumental variable estimators for binary outcomes. *Journal of the American Statistical Association*, *107*(500), 1638–1652.

Dai, J. Y., & Zhang, X. C. (2015). Mendelian randomization studies for a continuous exposure under case-control sampling. *American Journal of Epidemiology*, *181*(6), 440–449.

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P. R., Iacono, W. G., Swaroop, A., Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., ... Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nature genetics*, *48*(10), 1284–1287.

DaveySmith, G., & Ebrahim, S. (2003). mendelian randomization: Can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, *32*(1), 1–22.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22.

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, *1*(1), 54–75.

Greenland, S., Pearl, J., & Robins, J. M. (1999). Confounding and collapsibility in causal inference. *Statistical Science*, *14*(1), 29–46.

Hahn, J., & Hausman, J. (2002). A new specification test for the validity of instrumental variables. *Econometrica*, *70*(1), 163–189.

Harbord, R. M., Didelez, V., Palmer, T. M., Meng, S., Sterne, J. A., & Sheehan, N. A. (2013). Severity of bias of a simple estimator of the causal odds ratio in Mendelian randomization studies. *Statistics in Medicine*, *32*(7), 1246–1258.

Hernán, M. A., & Robins, J. M. (2006). Instruments for causal inference: An epidemiologist's dream? *Epidemiology*, *17*(4), 360–372.

Holland, P. W. (1988). Causal inference, path analysis and recursive structural equations models. *ETS Research Report Series*, *1988*(1), i–50.

Hu, X., Zhao, J., Lin, Z., Wang, Y., Peng, H., Zhao, H., Wan, X., & Yang, C. (2021). MR-APSS: A unified approach to Mendelian Randomization accounting for pleiotropy and sample structure using genome-wide summary statistics. *BioRxiv*.

Kang, H., Zhang, A., Cai, T. T., & Small, D. S. (2016). Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *Journal of the American Statistical Association*, *111*(513), 132–144.

Kitagawa, T. (2015). A test for instrument validity. *Econometrica*, *83*(5), 2043–2063.

Kolesár, M., Chetty, R., Friedman, J., Glaeser, E., & Imbens, G. W. (2015). Identification and inference with many invalid instruments. *Journal of Business & Economic Statistics*, *33*(4), 474–484.

Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N., & DaveySmith, G. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, *27*(8), 1133–1163.

Nielsen, S. F. (2000). The stochastic em algorithm: Estimation and asymptotic results. *Bernoulli*, *6*(3), 457–489.

Palmer, T. M., Thompson, J. R., Tobin, M. D., Sheehan, N. A., & Burton, P. R. (2008). Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. *International Journal of Epidemiology*, *37*(5), 1161–1168.

Pang, M., Kaufman, J. S., & Platt, R. W. (2016). Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Statistical Methods in Medical Research*, *25*(5), 1925–1937.

Sanderson, E., Richardson, T. G., Hemani, G., & Smith, G. D. (2020). The use of negative control outcomes in Mendelian Randomisation to detect potential population stratification or selection bias. *BioRxiv*.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society*, *26*(3), 393–415.

Small, D. S. (2007). Sensitivity analysis for instrumental variables regression with overidentifying restrictions. *Journal of the American Statistical Association*, *102*(479), 1049–1058.

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: Challenges and strategies. *Nature Reviews Genetics*, *14*(7), 483–495.

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, *20*(4), 518–529.

Terza, J. V., Basu, A., & Rathouz, P. J. (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*, *27*(3), 531–543.

The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68.

Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge university press.

Vansteelandt, S., Bowden, J., Babanezhad, M., & Goetghebeur, E. (2011). On instrumental variables estimation of causal odds ratios. *Statistical Science*, *26*(3), 403–422.

Verbanck, M., Chen, C. -Y., Neale, B., & Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nature Genetics*, *50*(5), 693–698.

Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics*, *11*(3), 284–300.

Wang, A., Liu, W., & Liu, Z. (2022). A two-sample robust Bayesian Mendelian randomization method accounting for linkage disequilibrium and idiosyncratic pleiotropy with applications to the COVID-19 outcomes. *Genetic Epidemiology*, *46*(3–4), 159–169.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Xu, S., Fung, W. K., & Liu, Z. (2021). MRCIP: A robust Mendelian randomization method accounting for correlated and idiosyncratic pleiotropy. *Briefings in Bioinformatics*, *22*(5), bbab019.

Xue, H., Shen, X., & Pan, W. (2021). Constrained maximum likelihood-based Mendelian randomization robust to both correlated and uncorrelated pleiotropic effects. *The American Journal of Human Genetics*, *108*(7), 1251–1269.

Zhao, Q., Wang, J., Hemani, G., Bowden, J., & Small, D. S. (2020). Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *The Annals of Statistics*, *48*(3), 1742–1769.

Zinkevich, M., Weimer, M., Li, L., & Smola, A. (2010). Parallelized stochastic gradient descent. *Advances in Neural Information Processing Systems*, *23*(1), 2595–2603.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Shi, D., Wang, Y., Zhang, Z., Cao, Y., & Hu, Y.-Q. (2023). MR-BOIL: Causal inference in one-sample Mendelian randomization for binary outcome with integrated likelihood method. *Genetic Epidemiology*, 47, 332–357. https://doi.org/10.1002/gepi.22520

APPENDIX A

A.1 | Derivation of likelihood function

Note that

$$\begin{pmatrix} U_i \\ X_i \end{pmatrix} | Z_i = z_i \sim N\left(\begin{pmatrix} 0 \\ z_i\gamma \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right),$$

and the conditional distriution of $U_i|X_i, Z_i$ is

$$U_i|X_i = x_i, Z_i = z_i \sim N\left(\frac{\sigma_1}{\sigma_2}\rho \cdot (x_i - \gamma z_i), \sigma_1^2(1-\rho^2) \right).$$

Denote

$$p(y_i|x_i, u, z_i; \boldsymbol{\Theta}) = \frac{\exp(y_i(\beta_0 + x_i\beta_1 + u))}{1 + \exp(\beta_0 + x_i\beta_1 + u)},$$

and

$$p(u|x_i, z_i; \boldsymbol{\Theta}) = \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left\{ -\frac{(u-K)^2}{2(1-\rho^2)\sigma_1^2} \right\},$$

and

$$p(x_i|z_i; \boldsymbol{\Theta}) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{ -\frac{(x_i - \gamma z_i)^2}{2\sigma_2^2} \right\}.$$

Then the integrated log-likelihood function can be formulated as

$$\begin{aligned}
\ell(\boldsymbol{\Theta}) &= \sum_{i=1}^{n} \log \int_R p(x_i, y_i|z_i, u; \boldsymbol{\Theta}) \cdot p(u|z_i; \boldsymbol{\Theta}) du \\
&= \sum_{i=1}^{n} \log \int_R p(y_i|x_i, u, z_i; \boldsymbol{\Theta}) \cdot p(x_i|u, z_i; \boldsymbol{\Theta}) \cdot p(u|z_i; \boldsymbol{\Theta}) du \\
&= \sum_{i=1}^{n} \log \int_R p(y_i|x_i, u, z_i; \boldsymbol{\Theta}) \cdot p(u|x_i, z_i; \boldsymbol{\Theta}) \cdot p(x_i|z_i; \boldsymbol{\Theta}) d \\
&\quad u \\
&= \sum_{i=1}^{n} \log \int_R \frac{\exp(y_i(\beta_0 + x_i\beta_1 + u))}{1 + \exp(\beta_0 + x_i\beta_1 + u)} \\
&\quad \cdot \exp\left\{ -\frac{(u-K)^2}{2(1-\rho^2)\sigma_1^2} \right\} \\
&\quad \cdot \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{(x_i - \gamma z_i)^2}{2\sigma_2^2} \right\} du \\
&= \sum_{i=1}^{n} \log \int_R \frac{\exp(y_i(\beta_0 + x_i\beta_1 + K_i + \sqrt{1-\rho^2}\sigma_1 u))}{1 + \exp(\beta_0 + x_i\beta_1 + K_i + \sqrt{1-\rho^2}\sigma_1 u)} \\
&\quad \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{u^2}{2} \right\} \\
&\quad \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{ -\frac{(x_i - \gamma z_i)^2}{2\sigma_2^2} \right\} du,
\end{aligned}$$

where $K_i = \frac{\sigma_1}{\sigma_2}\rho \cdot (x_i - \gamma z_i)$. So the transformated log-likelihood function is

$$\begin{aligned}
\ell(\boldsymbol{\Theta}) &= \sum_{i=1}^{n} \log \int_R p(y_i, x_i|z_i, u; \boldsymbol{\Theta}) \cdot p(u; \boldsymbol{\Theta}) du \\
&= \sum_{i=1}^{n} \log \int_R \widetilde{p}(y_i|x_i, z_i, u; \boldsymbol{\Theta}) \cdot \widetilde{p}(x_i|z_i, \boldsymbol{\Theta}) \cdot \widetilde{p}(u) du \\
&= \sum_{i=1}^{n} \log \int_R \widetilde{p}(x_i, y_i|z_i, u; \boldsymbol{\Theta}) \cdot \widetilde{p}(u) du,
\end{aligned}$$

where

$$K_i = \frac{\sigma_1}{\sigma_2}\rho \cdot (x_i - \gamma z_i),$$

$$\tilde{p}(y_i|x_i, z_i, u; \boldsymbol{\Theta}) = \frac{\exp(y_i(\beta_0 + x_i\beta_1 + K_i + \sqrt{1-\rho^2}\sigma_1 u))}{1 + \exp(\beta_0 + x_i\beta_1 + K_i + \sqrt{1-\rho^2}\sigma_1 u)},$$

$$\tilde{p}(x_i|z_i; \boldsymbol{\Theta}) = \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\frac{(x_i - \gamma z_i)^2}{2\sigma_2^2}\right\},$$

$$\tilde{p}(u) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{u^2}{2}\right\},$$

$$\tilde{p}(x_i, y_i|z_i, u; \boldsymbol{\Theta}) \triangleq \tilde{p}(y_i|x_i, z_i, u; \boldsymbol{\Theta}) \cdot \tilde{p}(x_i|z_i, \boldsymbol{\Theta}).$$

A.2 | Derivation of EM algorithm

For an initial parameter value $\boldsymbol{\Theta}^0$, and iteration $j = 1, ..., K$,

- E-Step: For fixed $\boldsymbol{\Theta}^j$, we compute the posterior distribution

$$\begin{aligned}\tilde{p}(u|\mathbf{y}, \mathbf{x}, \mathbf{z}; \boldsymbol{\Theta}^j) &= \frac{\tilde{p}(u, \mathbf{y}, \mathbf{x}|\mathbf{z}; \boldsymbol{\Theta}^j)}{\tilde{p}(\mathbf{y}, \mathbf{x}|\mathbf{z}; \boldsymbol{\Theta}^j)} \\ &= \frac{\tilde{p}(\mathbf{y}, \mathbf{x}|\mathbf{z}, u; \boldsymbol{\Theta}^j) \cdot \tilde{p}(u|\mathbf{z}; \boldsymbol{\Theta}^j)}{\tilde{p}(\mathbf{y}, \mathbf{x}|\mathbf{z}; \boldsymbol{\Theta}^j)} \\ &= \frac{\tilde{p}(\mathbf{y}, \mathbf{x}|\mathbf{z}, u; \boldsymbol{\Theta}^j) \cdot \tilde{p}(u)}{\tilde{p}(\mathbf{y}, \mathbf{x}|\mathbf{z}; \boldsymbol{\Theta}^j)} \\ &= \frac{\tilde{p}(\mathbf{y}, \mathbf{x}|\mathbf{z}, u; \boldsymbol{\Theta}^j) \cdot \tilde{p}(u)}{\int_R \tilde{p}(\mathbf{y}, \mathbf{x}|\mathbf{z}, u; \boldsymbol{\Theta}^j) \cdot \tilde{p}(u)du}\end{aligned}$$

and the posterior expectation

$$\begin{aligned}\mathbb{E}_{\mathbf{U}|\mathbf{y},\mathbf{x}\sim\tilde{p}(u|\mathbf{y},\mathbf{x},\mathbf{z};\boldsymbol{\Theta}^j)}[\log\tilde{p}(\mathbf{y}, \mathbf{x}, u|\mathbf{z}; \boldsymbol{\Theta})] = \sum_{i=1}^{n} &\int_R \tilde{p}(u|y_i, x_i, z_i; \boldsymbol{\Theta}^j) \\ &\cdot \log\tilde{p}(y_i, x_i, u|z_i; \boldsymbol{\Theta})du,\end{aligned}$$

where

$$\tilde{p}(u|y_i, x_i, z_i; \boldsymbol{\Theta}^j) = \frac{\tilde{p}(y_i, x_i|z_i, u; \boldsymbol{\Theta}^j) \cdot \tilde{p}(u)}{\tilde{p}(y_i, x_i|z_i; \boldsymbol{\Theta}^j)}.$$

- M-Step: Maximize the posterior expectation:

$$\begin{aligned}\boldsymbol{\Theta}^{j+1} &= \arg\max_{\boldsymbol{\Theta}} \mathbb{E}_{\mathbf{U}|\mathbf{y},\mathbf{x}\sim\tilde{p}(u|\mathbf{y},\mathbf{x},\mathbf{z};\boldsymbol{\Theta}^j)}[\log\tilde{p}(\mathbf{y}, \mathbf{x}, u|\mathbf{z}; \boldsymbol{\Theta})] \\ &= \arg\max_{\boldsymbol{\Theta}} \sum_{i=1}^{n} \int_R \tilde{p}(u|y_i, x_i, z_i; \boldsymbol{\Theta}^j) \\ &\quad \cdot \log\tilde{p}(y_i, x_i, u|z_i; \boldsymbol{\Theta})du \\ &= \arg\max_{\boldsymbol{\Theta}} \sum_{i=1}^{n} \int_R \frac{\tilde{p}(y_i, x_i|z_i, u; \boldsymbol{\Theta}^j) \cdot \tilde{p}(u)}{\tilde{p}(y_i, x_i|z_i; \boldsymbol{\Theta}^j)} \\ &\quad \cdot \log(\tilde{p}(y_i, x_i|z_i, u; \boldsymbol{\Theta}) \cdot \tilde{p}(u))du \\ &\qquad \int_R \tilde{p}(y_i, x_i|z_i, u; \boldsymbol{\Theta}^j) \cdot \tilde{p}(u) \cdot \log \\ &= \arg\max_{\boldsymbol{\Theta}} \sum_{i=1}^{n} \frac{(\tilde{p}(y_i, x_i|z_i, u; \boldsymbol{\Theta}) \cdot \tilde{p}(u))du}{\int_R \tilde{p}(y_i, x_i|z_i, u; \boldsymbol{\Theta}^j) \cdot \tilde{p}(u)du} \\ &\qquad \int_R \tilde{p}(y_i|x_i, z_i, u; \boldsymbol{\Theta}^j) \cdot \tilde{p}(u) \cdot \log \\ &= \arg\max_{\boldsymbol{\Theta}} \sum_{i=1}^{n} \frac{(\tilde{p}(y_i|x_i, z_i, u; \boldsymbol{\Theta}) \cdot \tilde{p}(x_i|z_i; \boldsymbol{\Theta}))du}{\int_R \tilde{p}(y_i|x_i, z_i, u; \boldsymbol{\Theta}^j) \cdot \tilde{p}(u)du}.\end{aligned}$$

Detailly, we can treat the numerator as an expectation

$$\begin{aligned}&\mathbb{E}_{\mathbf{U}\sim\varphi(u)}[\tilde{p}(y_i|x_i, z_i, \mathbf{U}; \boldsymbol{\Theta}^j) \\ &\quad \cdot \log(\tilde{p}(y_i|x_i, z_i, \mathbf{U}; \boldsymbol{\Theta}) \cdot \tilde{p}(x_i|z_i; \boldsymbol{\Theta}))] \\ &\quad \simeq \frac{1}{n_i}\sum_{k=1}^{n_i} \tilde{p}(y_i|x_i, z_i, \mathbf{u}_k; \boldsymbol{\Theta}^j) \cdot \log \\ &\qquad (\tilde{p}(y_i|x_i, z_i, \mathbf{u}_k; \boldsymbol{\Theta}) \cdot \tilde{p}(x_i|z_i; \boldsymbol{\Theta}))\end{aligned}$$

to avoid direct integration being infinite.

- Stop Criterion: We stop the algorithm when $||\boldsymbol{\Theta}^j - \boldsymbol{\Theta}^{j+1}||_2 < \varepsilon$, where ε is usually taken as 10^{-4} or 10^{-5}.

A.3 | Validity assessment with nominal significance level of 0.01

To assess the validity of the proposed method, nominal significance level 0.01 is also adopted. In Figures A1 and A2, we present Q-Q plot of distribution of p values versus uniform distribution $Uniform(0, 1)$ and bar plot of type I error rates for direct method, two-stage method, adjusted-IV method and MR-BOIL method under the null hypothesis $H_0 : \beta_1 = 0$. The sample size is $n = 1000$, and other nuisances are $\beta_0 = 2$, $\gamma = 1$, $\rho = 0.8$, $\sigma_1 = 1$, $\sigma_2 = 3$. For this scenario of parameter setting, 10, 000

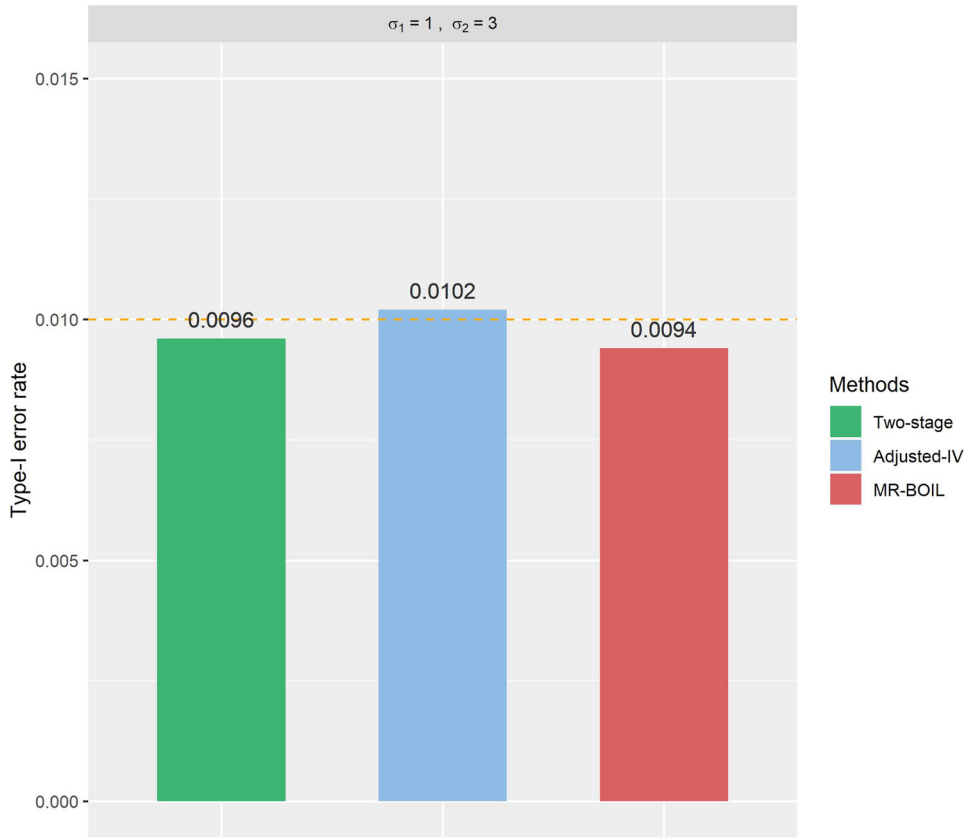

FIGURE A1 Bar plot of type I error rates of two-stage method, adjusted-IV method and MR-BOIL method under the null hypothesis $H_0 : \beta_1 = 0$. The sample size is $n = 1000$, and other nuisances are $\beta_0 = 2$, $\gamma = 1$, $\rho = 0.8$, $\sigma_1 = 1$, $\sigma_2 = 3$. For this scenario of parameter setting, 10, 000 replications are performed. The orange dashed horizontal line indicates the nominal significance of 0.01.

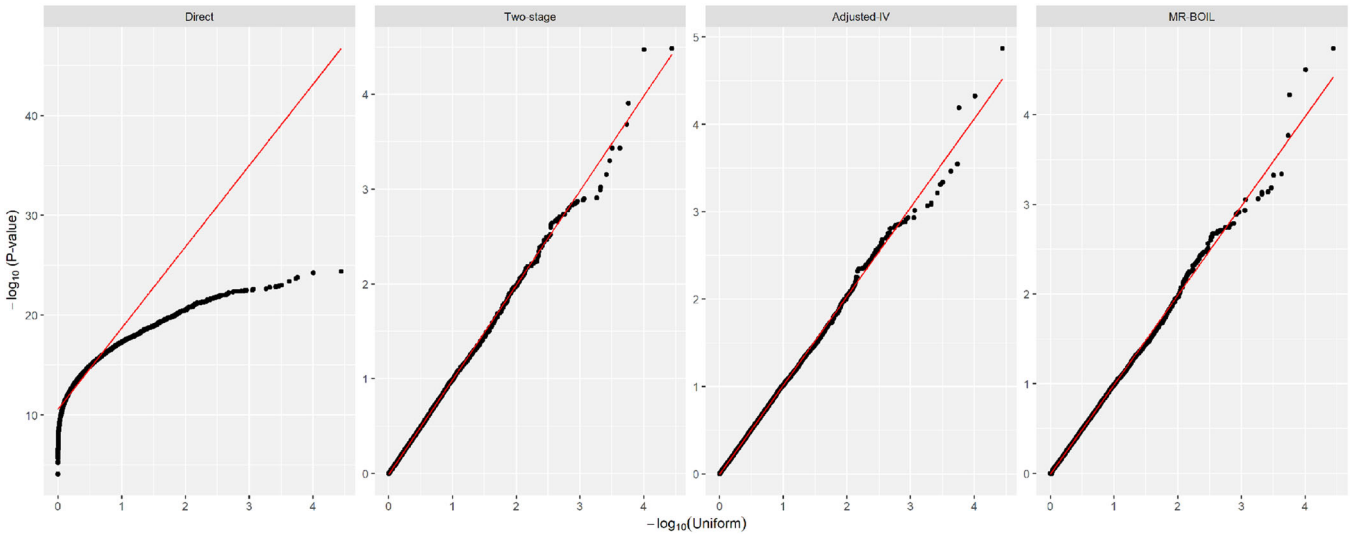

FIGURE A2 Q-Q plot of distribution of p-values versus uniform distribution *Uniform*(0, 1) for direct method, two-stage method, adjusted-IV method and MR-BOIL method under the null hypothesis $H_0 : \beta_1 = 0$. The sample size is $n = 1000$, and other nuisances are $\beta_0 = 2$, $\gamma = 1$, $\rho = 0.8$, $\sigma_1 = 1$, $\sigma_2 = 3$. For this scenario of parameter setting, 10,000 replications are performed. The red solid line is $y = x$.

replications are performed. It can be concluded that the type I error rates of two-stage method, adjusted IV and our MR-BOIL method are consistently controlled at the nominal level of 0.01 in this parameter setting.

A.4 | Approximated expression for two-stage method

In this section, we will derive the approximated expression of the two stage method. Under the following model,

$$\begin{aligned} X_i &= Z_i\gamma + V_i, \\ Y_i &\sim \text{Bernoulli}(H(X_i\beta_1 + U_i + \beta_0)), \\ \begin{pmatrix} U_i \\ V_i \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right), \end{aligned}$$

where $H(t) = 1/(1 + \exp(-t))$. Then

$$\begin{aligned} \mathbb{P}(Y_i = 1|Z_i = z_i) &= \mathbb{E}[Y_i|Z_i = z_i] \\ &= \int_{-\infty}^{\infty} H(z\beta_1\gamma + ce)\varphi(e)de, \end{aligned}$$

where $c^2 = \text{Var}(\beta_1 V_i + U_i + \beta_0) = \beta_1^2\sigma_2^2 + 2\rho\beta\sigma_1\sigma_2 + \sigma_1^2$. Using one approximating equation

$$H(t) \approx \Phi(t/1.7),$$

and another equality

$$\int_{-\infty}^{\infty} \Phi(a + bt)\phi(t)dt = \Phi\left(\frac{a}{\sqrt{1 + b^2}}\right),$$

we get

$$\begin{aligned} \mathbb{P}(Y_i = 1|Z_i = z_i) &= \int_{-\infty}^{\infty} H(z\beta_1\gamma + ce)\varphi(e)de \\ &\approx \int_{-\infty}^{\infty} \Phi\left(\frac{z\beta\gamma + ce}{1.7}\right)\varphi(e)de \\ &= \Phi\left(\frac{z\beta_1\gamma}{1.7\sqrt{1 + (c/1.7)^2}}\right) \\ &\approx H\left(\frac{z\beta_1\gamma}{\sqrt{1 + (c/1.7)^2}}\right). \end{aligned}$$

Hence

$$\Gamma \approx \frac{\beta_1\gamma}{\sqrt{1 + (c/1.7)^2}},$$

and

$$\beta_1^{TSLS} = \frac{\Gamma}{\gamma} \approx \frac{\beta_1}{\sqrt{1 + (c/1.7)^2}}.$$

A.5 | Sensitivity analysis

We simulate independently distributed 1000 subjects (Y_i, X_i, Z_i) from the following model,

$$\begin{aligned} X_i &= \gamma Z_i + V_i, \\ Y_i &\sim \text{Bernoulli}(\pi_i), \\ \text{logit}(\pi_i) &= \beta_0 + \beta_1 X_i + U_i, \end{aligned}$$

where confounding factors U_i and V_i follow a multivariate t distribution with mean $(0, 0)^T$, scale matrix V and k degrees of freedom, that is

$$f(u_i, v_i) = c\left(1 + \frac{1}{k}(u_i, v_i)V^{-1}(u_i, v_i)^T\right)^{-(k+2)/2},$$

where

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = 2V.$$

is the covariance matrix of $(U_i, V_i)^T$, and

$$c = \frac{1}{2\pi}|\det(V)|^{-1/2}, k = 4.$$

Besides, Z_i stands for the carefully selected SNP coded as 0, 1, 2. Moreover, the outcome for every subject is drawn from the Bernoulli distribution with probability $\pi_i = \mathbb{E}[Y_i|X_i, U_i] = \text{expit}(\beta_0 + \beta_1 X_i + U_i)$, where expit is the inverse of the logit function. Then we consider the influence of confounding factors on the estimation of causal effect β_1 by changing the values of σ_1. With the nuisance parameters set to $\beta_0 = 2, \gamma = 1, \sigma_2 = 1, \rho = 0.8$, the true causal effect is $\beta_1 = 1$. In particular, consider four confounding scenarios that increasingly affect the outcome by adjusting σ_1 to 1, 2, 3, 4.

A.6 | Real data analysis

Bootstrap for variance estimation

To empirically evaluate the variation of MR-BOIL, we employ the bootstrap sampling procedure (Efron & Tibshirani, 1986). The number of bootstrap sampling is 200. Let $\hat{\beta}_{1i}$ be the estimate of causal effect with MR-BOIL method based on the ith boostrap sample,

$i = 1, ..., 200$, and $\bar{\hat{\beta}}_1$ be corresponding the average of these 200 estimates. The standard error of MR-BOIL method by bootstrap sampling is calculated as

$$\text{Standard Error (SE)} = \sqrt{\frac{1}{199}\sum_{i=1}^{200}(\hat{\beta}_{1i} - \bar{\hat{\beta}}_1)^2}.$$

For application, we evaluate the variation of our MR-BOIL method using this bootstrap method for estimating SBP-hypertension causal effect. In Table A1, we show point estimates, p-values and standard error of SBP-hypertension causal effect with two-stage method, adjusted-IV method and MR-BOIL method.

Other combinations of exposure and outcome

In addition to the validation examples, we add three additional exposure-outcome combinations in real data analysis. For ARIC data set, we study the causal relationship of SBP-CHD (coronary heart disease), BMI (body mass index in kg/(m × m))-CHD, BMI-IDI (incident diabetes indicator). Both SBP and BMI variables are continuous and both CHD and IDI variables are binary. The data preprocess and MR-BOIL method implementation flow is similar to that in SBP-hypertension study.

For exposure as SBP and BMI, 8 and 15 SNPs are selected respectively. The estimates and p values of causal effect are shown in Tables A2–A4. In these results, no significant causal effects are identified for all methods.

Runtime and memory usage

We provide runtime and memory usage for one scenario of parameter setting in simulation and one SNP in real data example with direct method, two-stage method, adjusted-IV method and MR-BOIL method. The computing environment is 100 CPU cores in one computer node.

In the simulation, we record runtime and memory usage for one scenario of parameter setting of maximum likelihood estimation in Table A5. The parameters are $\beta_0 = 2, \beta_1 = 1, \gamma = 1, \sigma_1 = 1, \sigma_2 = 1, \rho = 0.8$ and sample

TABLE A1 The point estimates, p values and standard error (SE) of the SBP-hypertension causal effect with two-stage method, adjusted-IV method and MR-BOIL method.

	Two-stage			Adjusted-IV			MR-BOIL		
SNP location	Estimate	p Value	SE	Estimate	p Value	SE	Estimate	p Value	SE
rs11695862	−0.239	0.974	4.237	0.802	0.926	5.071	0.724	2.72E−06	0.060
rs11884411	−1.330	0.802	32.673	−4.900	0.436	39.109	3.271	0.004244	0.512
rs16854522	5.392	0.302	1.739	6.508	0.299	2.073	1.098	2.90E−06	0.344
rs7356321	2.456	0.004	1.709	3.278	0.001	2.043	2.662	6.33E−08	0.110
rs4140431	2.015	0.091	1.873	2.548	0.073	2.245	2.497	1.48E−06	0.700
rs10237149	3.486	0.345	1.953	4.859	0.268	2.330	1.102	1.02E−06	0.055
rs7100920	−0.975	0.822	1.331	−0.362	0.944	1.591	3.439	0.027484	0.643
rs1896341	6.874	0.625	2.012	4.413	0.792	2.398	1.824	3.77E−06	0.316

Note: The SE of MR-BOIL is estimated by the bootstrap method.

TABLE A2 The point estimates and p values of the SBP-CHD causal effect for two-stage method, adjusted-IV method and MR-BOIL method.

	Two-stage		Adjusted-IV		MR-BOIL	
SNP location	Estimate	p Value	Estimate	p Value	Estimate	p Value
rs11695862	19.580	0.118	19.866	0.115	−0.515	0.761
rs11884411	3.225	0.262	3.093	0.284	0.000	0.998
rs16854522	−1.140	0.715	−1.281	0.682	−0.515	0.678
rs7356321	−0.333	0.954	−0.260	0.964	0.000	0.997
rs4140431	−17.817	0.111	−18.158	0.106	−2.423	0.383
rs10237149	3.375	0.140	3.329	0.148	0.318	0.691
rs7100920	5.719	0.168	5.695	0.171	−1.788	0.772
rs1896341	3.687	0.077	3.630	0.082	0.323	0.626

TABLE A3 The point estimates and *p* values of the BMI-CHD causal effect for two-stage method, adjusted-IV method and MR-BOIL method.

	Two-stage		Adjusted-IV		MR-BOIL	
SNP location	Estimate	*p* Value	Estimate	*p* Value	Estimate	*p* Value
rs6603803	−0.472	0.754	−0.487	0.747	−0.001	0.897
rs9424977	9.659	0.690	9.541	0.695	0.405	0.845
rs12406019	4.991	0.672	5.011	0.672	−0.645	0.849
rs6658131	2.261	0.805	2.211	0.810	0.000	0.997
rs713586	−0.860	0.832	−0.880	0.829	0.392	0.923
rs12986742	1.622	0.516	1.632	0.514	0.476	0.706
rs1530559	2.836	0.115	2.761	0.126	2.105	0.166
rs4958361	13.100	0.639	13.039	0.642	−0.614	0.896
rs6976031	0.026	0.992	−0.123	0.963	0.099	0.927
rs913783	−8.685	0.063	−8.900	0.058	−0.409	0.547
rs10733682	−1.015	0.466	−1.120	0.424	−0.405	0.564
rs4929942	4.844	0.158	4.720	0.170	0.000	0.997
rs4898534	1.657	0.276	1.613	0.291	0.000	0.997
rs7141420	−1.181	0.795	−1.271	0.780	0.000	0.916
rs9922047	−7.656	0.135	−7.774	0.130	−2.000	0.297

TABLE A4 The point estimates and *p* values of the BMI-INI causal effect for two-stage method, adjusted-IV method and MR-BOIL method.

	Two-stage		Adjusted-IV		MR-BOIL	
SNP location	Estimate	*p* Value	Estimate	*p* Value	Estimate	*p* Value
rs6603803	−4.384	0.617	−4.507	0.620	−1.696	0.784
rs9424977	−3.044	0.416	−3.778	0.325	0.287	0.400
rs12406019	−1.132	0.744	−0.842	0.813	−2.661	0.382
rs6658131	2.668	0.468	1.904	0.616	0.299	0.512
rs713586	3.299	0.423	3.352	0.430	0.308	0.751
rs12986742	0.530	0.902	0.814	0.854	0.103	0.699
rs1530559	−3.857	0.476	−5.357	0.339	0.000	0.996
rs4958361	−5.964	0.166	−6.592	0.139	−2.497	0.540
rs6976031	5.356	0.600	4.020	0.706	0.000	0.997
rs913783	−6.237	0.772	−7.019	0.753	0.295	0.404
rs10733682	−3.884	0.355	−4.527	0.293	0.721	0.836
rs4929942	3.047	0.951	0.601	0.991	0.304	0.294
rs4898534	3.534	0.494	3.611	0.494	−1.697	0.674
rs7141420	−23.084	0.115	−24.648	0.101	−1.999	0.212
rs9922047	10.593	0.051	10.544	0.062	0.295	0.726

size is $n = 1000$. 1000 replications are performed and means of runtime and memory usage are recorded.

In real data analysis, we record runtime and memory usage for estimation of SBP-hypertension causal effect with SNP rs11695862 as the IV in Table A6. Both tables show that to obtain a higher accuracy, our MR-BOIL method requires a higher computational and storage cost compared with other methods.

TABLE A5 Runtime and memory usage for estimating causal effect with direct method, two-stage method, adjusted-IV method and MR-BOIL method.

Direct		Two-stage		Adjusted-IV		MR-BOIL	
Runtime(s)	Memory usage (kB)	Runtime (s)	Memory usage (kB)	Runtime (s)	Memory usage (kB)	Runtime (s)	Memory usage (kB)
0.661	3.063	0.627	2.144	0.813	11.456	2712.976	4789.976

Note: The parameters are $\beta_0 = 2$, $\beta_1 = 1$, $\gamma = 1$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\rho = 0.8$ and sample size is $n = 1000$. 1000 replications are performed and means of runtime and memory usage are recorded. The computing environment is 100 CPU cores in one computer node.

TABLE A6 Runtime and memory usage for estimating SBP-hypertension causal effect with SNP rs11695862 as the IV.

Two-stage		Adjusted-IV		MR-BOIL	
Runtime (s)	Memory usage (kB)	Runtime (s)	Memory usage (kB)	Runtime (s)	Memory usage (kB)
2.184	2.304	3.286	72.848	5193.631	2038.960

Note: The computing environment is 100 CPU cores in one computer node.