Interpretation of two-sample Mendelian randomization for binary exposures and outcome

Zixuan Wu[1] and Jingshu Wang[1,*]

Abstract

Mendelian randomization (MR) employs genetic variations to infer causal effects of modifiable exposures on various outcomes, with two-sample MR using GWAS summary statistics from separate studies for exposure and outcome traits. However, challenges arise when either the exposure or outcome trait, or both, are binary. Conventional two-sample MR methods assume a linear and homogeneous causal effect of the exposure on the outcome, which is unrealistic for binary traits. To address this, we propose a model that treats binary traits as dichotomized continuous traits. We show that summary statistics for a binary trait, derived from either logistic or linear regressions, maintain nearly proportional relationships to those from linear regressions on the underlying continuous trait, provided the associations between the continuous trait and a single SNP are limited. The ratio is also easily calculable given the prevalence of the binary trait. Consequently, MR frameworks on the underlying continuous traits can be resolved using summary statistics from binary traits. Our analytical approximations and simulations indicate that no modifications to two-sample MR methods are needed for binary traits, only an adjustment in the interpretation of results. This novel interpretation also extends to complex MR frameworks, enhancing the applicability of MR to binary and ordered categorical traits.

[1]Department of Statistics, The University of Chicago, Chicago, IL 60615, USA

[*]Corresponding author, jingshuw@uchicago.edu

1 Main

Mendelian randomization (MR) uses genetic variations to infer the causal effect of genetically modifiable exposures on health, developmental or social outcomes [11]. With the increasing availability of Genome-Wide Association Study (GWAS) summary statistics, two-sample Mendelian Randomization (MR), using only GWAS summary statistics of the exposure and outcome traits from separate studies, has become the most popular approach to performing MR [19]. The linear relationship of marginal associations between a genetic instrument and the outcome and exposure traits serves as a fundamental model in two-sample MR [2, 5]. Specifically, in the absence of horizontal pleiotropy, denoting Γ_j and γ_j as the true marginal associations between Single Nucleotide Polymorphism (SNP) j and the outcome and exposure traits respectively, it is assumed that $\Gamma_j = \beta\gamma_j$, where β represents the true causal effect of interest.

However, when either the exposure or outcome traits are binary, justification of the linear relationship $\Gamma_j = \beta\gamma_j$ is unclear even under no horizontal pleiotropy. Besides assuming the common three assumptions in MR that guarantee SNP j to be a valid instrument [16], deriving $\Gamma_j = \beta\gamma_j$ typically relies on an additional assumption that the causal relationship between the exposure and outcome is both linear and homogeneous [15, 20]. Yet, a linear structural equation becomes implausible when the outcome trait is binary, and the homogeneity assumption, assuming that the causal effect is the same for every individual, is less tenable if the exposure trait is binary [7, 26]. In cases where the exposure trait is binary, β might also be interpreted as the complier average causal effect under the monotonicity assumption [7, 30]. However, the SNP-specific complier groups, defined as individuals who get the exposure if and only if the SNP is mutated, have limited practical interpretation [7] and the consistency of the average causal effects across all the SNP-specific complier groups still implies an underlying homogeneity assumption. In addition, when traits are binary, summary statistics are often obtained not through linear regressions, but via logistic regressions, especially in case-control studies. Logistic regressions provide estimated odds ratios between the SNP and the binary trait, rather than the marginal associations estimated in linear regression models. These challenges have raised significant concerns regarding the application of two-sample MR to binary traits [7, 15, 26].

Here, we provide an interpretation of the model $\Gamma_j = \beta\gamma_j$ where Γ_j and γ_j can be estimated from GWAS summary statistics when either the exposure or outcome phenotype, or both, are binary. We justify this model by treating a binary trait as a dichotomization of a continuous trait [7, 22]. Many binary traits represent diagnoses of common diseases, often determined by whether a continuous test measurement exceeds a specific threshold. For example, diabetes is diagnosed if blood glucose levels surpass a certain threshold, and chronic kidney disease is diagnosed if the estimated glomerular filtration rate (eGFR) falls below a specified level. This viewpoint is known as the liability model of diseases, where liability is the underlying continuous trait that reflects the genetic, environmental, and stochastic components of the disease [12, 13, 17]. Recent research has also expanded this liability model to include ordered categorical exposure traits such as education attainment [22].

Figure 1 illustrates our assumption regarding the causal relationship between the observed exposure X and outcome Y, and their respective underlying continuous traits, $X^\star$ and $Y^\star$, under the liability model of binary traits. If the observed exposure (or outcome trait) is continuous, then $X = X^\star$ (or $Y = Y^\star$). However, if the observed exposure trait is binary, we assume $X = 1$ if $X^\star > c_X$, where c_X represents a threshold. Similarly, if the observed outcome trait is binary, then $Y = 1$ if $Y^\star > c_Y$. Our

main assumption is that, if the observed traits are binary, they are treated as "noisy" representations of the underlying continuous measurements. Under this setup, the causal relationship is always between the continuous exposure and outcome traits, and we only require that the usual MR assumptions, such as linearity and assumptions on horizontal pleiotropic effects, hold for these continuous traits.

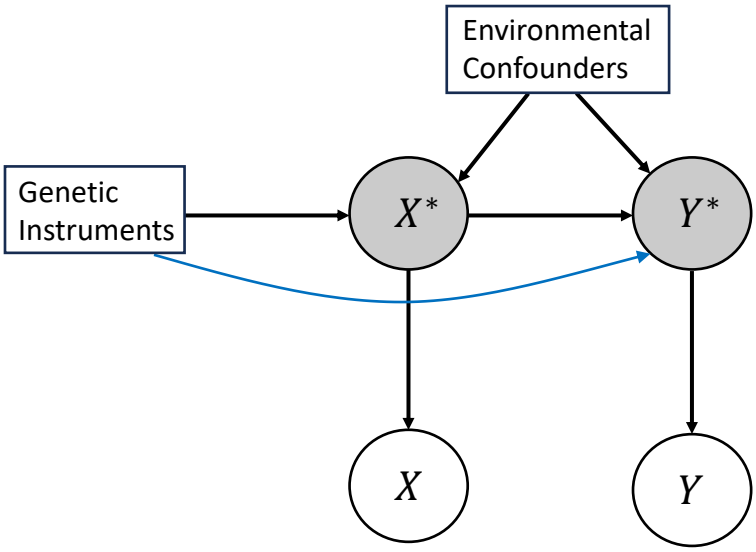

Figure 1: Z denotes the genetic variants; U denotes the unmeasured environmental confounders; X^*, Y^* denote the unobserved continuous exposures and outcomes; X, Y denote the observed potentially binary exposures and outcomes. The blue arrow indicates possible horizontal pleiotropy.

Our argument is based on the fact that, if a single SNP always has a small SNP-trait association, the expectations of summary statistics on continuous traits exhibit a near-proportional relationship with those on their dichotomized binary counterparts, regardless of the regression model applied to the binary trait to obtain the summary statistics. This result is supported both by our analytical approximations and by simulations. Specifically, let $\Gamma_j^\star = \beta\gamma_j^\star + \alpha_j^\star$ where $\Gamma_j^\star$, $\gamma_j^\star$ and $\alpha_j^\star$ denote the marginal associations and pleiotropic effects for the underlying continuous exposure and outcome traits. Then the true log odds ratio or marginal association between SNP j and the observed binary traits satisfies $\Gamma_j \approx s_Y \Gamma_j^\star$ and $\gamma_j \approx s_X \gamma_j^\star$. The scaling factors $s_X > 0$ and $s_Y > 0$ are constants across genetic variants and can be computed as functions of the prevalence of the binary traits, the types of regression models used to get the summary statistics, and the distributions of the underlying continuous traits. As a consequence, if we perform MR using summary statistics from the binary exposure or/and outcome trait, then we can estimate a scaled version of the causal effect of the underlying continuous exposure on the continuous outcome under common assumptions on horizontal pleiotropy of the continuous traits. We can further calculate and correct for the scaling factor given knowledge of the prevalence of the binary traits and if we assume that the underlying continuous traits follow standard Gaussian distributions (or any other known distributions after standardization).

Our findings challenge the conventional belief that working with binary traits poses a concern in two-sample MR analyses. Based on our interpretation, we believe that no modifications to MR methods are necessary when either the exposure or outcome trait, or both, are binary. The modification is just in the interpretation of the MR results. Furthermore, our novel interpretation not only applies to standard univariate MR but also facilitates the interpretation of more complex MR frameworks, such as multivariate MR [6, 9], within-family MR [4, 21], and life-course MR [25, 28], on binary or ordered

categorical traits.

2 Analytical derivations

Summary statistics between SNP j and a continuous trait

If we observe a continuous trait $X^\star$ (or $Y^\star$), by convention, the summary statistics are always obtained by linear regressions. Specifically, the marginal association between SNP Z_j and trait $X^\star$ is the projection of $X^\star$ on SNP j satisfying

$$X^\star = \mu_j + \gamma_j^\star Z_j + \epsilon_j$$

where $\boldsymbol{E}(\epsilon_j) = 0$, and ϵ_j and Z_j are independent. However, when we only observe a binary trait X, the continuous trait $X^\star$ is not observed so we can not directly obtain the summary statistics estimating $\gamma_j^\star$ by regressing trait $X^\star$ on SNP j.

Latent variable thresholding regression model on binary trait X

Now we follow the latent variable thresholding model framework to build a binary generalized linear regression model (GLM) on the binary trait X based on the distribution of $X^\star$ [23].

Assume that the distribution of $-X^\star$ follows some distribution with cumulative distribution function (CDF) $F(x)$. We assume that for a single SNP j, its association with the trait is relatively small ($\gamma_j^\star = o(1)$), then $\mu_j + \epsilon_j \approx X^\star$. Thus, following our assumption that the binary trait X is a dichotomization of X^*, we have

$$\begin{aligned}\mathbb{P}(X = 1) &= \mathbb{P}(X^\star \geq c_X) \\ &= \mathbb{P}\left(-\mu_j - \epsilon_j \leq \gamma_j^\star Z_j - c_X\right) \\ &\approx F\left(\gamma_j^\star Z_j - c_X\right).\end{aligned}$$

As a consequence, if we run a binary GLM model on X with the inverse link function $F(x)$, then the regression coefficient γ_j^F is approximately $\gamma_j^\star$. For instance, if the continuous trait is standardized following $X^\star \sim N(0,1)$, then

$$\mathbb{P}(X = 1) \approx \Phi\left(-c_X + \gamma_j^\star Z_j\right) \tag{1}$$

where Φ is the CDF of a standard Gaussian distribution. This indicates that if we run a probit regression of the binary X on a single SNP j, then the true coefficient γ_j^Φ of SNP j will approximately also be $\gamma_j^\star$. In terms of the accuracy of the approximation in (1), in the supplement (SI text S1.1), we show that the approximation error $|\gamma_j^\Phi - \gamma_j^\star|/\gamma_j^\star$ where γ_j^Φ is the coefficient from probit regression is only about 5% if Z_j explains 10% of the variance of $X^\star$, and the approximation error is less than 3% if the variance explained by the SNP is 5%. Though we assume $\gamma_j^\star = o(1)$ in our derivation, these results show that the approximation in this step is accurate as long as the SNP-trait association is moderately small.

Summary statistics from logistic or linear regression on binary trait

Next, we show that if we use the logistic regression for the binary trait X on a single SNP j to get summary statistics, the regression coefficients will be approximately proportional to $\gamma_j^\star$. This approximation

is also based on the assumption that the SNP-trait association $\gamma_j^\star$ is small.

Denote the standard logistic function as $H(x) = e^x/(1+e^x)$. The logistic regression fits the model $H^{-1}(\mathbb{P}(X=1)) = \gamma_{0j} + \gamma_j^H Z_j$. Since $\gamma_j^\star = o(1)$, using Taylor expansion at 0, we have

$$H^{-1}(\mathbb{P}(X=1)) \approx H^{-1}F\left(\gamma_j^\star Z_j - c_X\right) \tag{2}$$

$$\approx (H^{-1}F)(-c_X) + (H^{-1}F)'(-c_X)\gamma_j^\star Z_j \tag{3}$$

Thus, the regression coefficient γ_j^H is approximately proportional to the marginal association $\gamma_j^\star$ for the underlying continuous trait:

$$\gamma_j^H \approx s_X \gamma_j^\star$$

where the constant ratio $s_X = (H^{-1}F)'(-c_X)$ does not change across SNPs. By definition $\mathbb{P}(X=1) = \mathbb{P}(X^\star > c_X) = F(-c_X)$, thus if we denote the prevalence of the binary trait as $p_X = \mathbb{P}(X=1)$, then the ratio has the form

$$s_X = (H^{-1}F)'\left(F^{-1}(p_X)\right) \tag{4}$$

which only depends on the distribution F of the continuous trait and the prevalence p_X of the binary trait. When $X^\star \sim N(0,1)$, then $s_X = (H^{-1}\Phi)'\left(\Phi^{-1}(p_X)\right)$, as illustrated in Figure 2a. Figure 2a shows that when $X^\star$ is standardized Gaussian and the prevalence $P(X=1)$ is not too extreme (between $15\% - 85\%$), the ratio s_X is very close to 1.7, which is the ratio between the logistic and probit regression coefficients indicated by the well-known approximation $H(x) \approx \Phi(x/1.7)$ [27]. However, if $P(X=1)$ is too low or too high, then s_X can be much larger than 1.7, and changes rapidly with the prevalence.

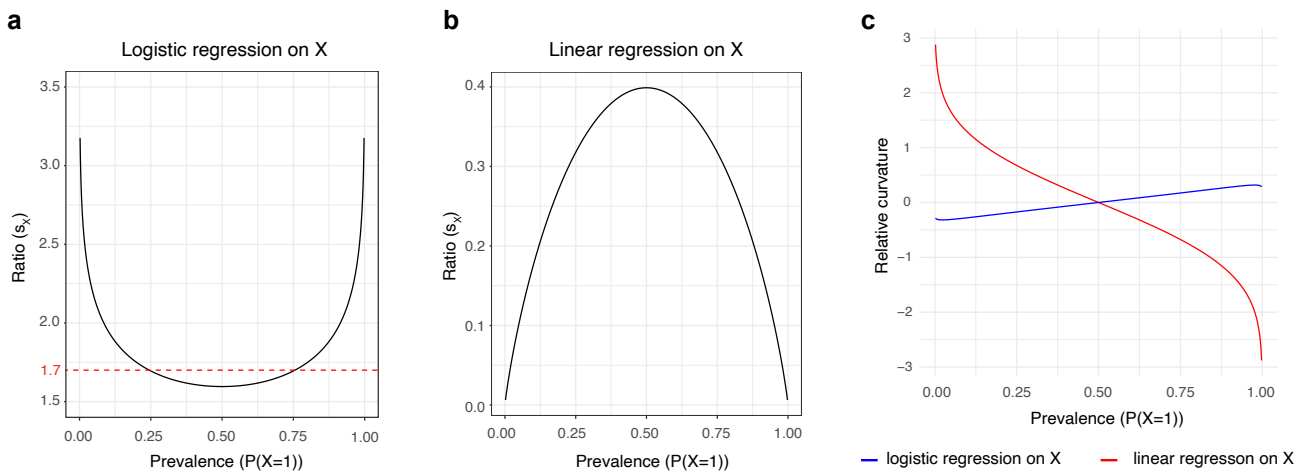

Figure 2: a) and b): Ratio of the expected summary statistics between using logistic regression (a) or linear regression (b) on the binary trait, and using linear regression on the underlying continuous trait, as a function of the prevalence $c_X = P(X=1)$. The dotted red line indicates $y = 1.7$. c) Relative curvatures in approximating the logistic and linear regression coefficients on the binary trait X.

In some GWAS studies, particularly those involving biobank data [ukb], linear regression is employed to obtain summary statistics on binary traits. Using a similar approach, we can demonstrate that the true regression coefficient from linear regression on the binary trait for SNP j is also approximately proportional to $\gamma_j^\star$, although the precision of this approximation may be lower.

The linear regression on X fits the model $\mathbb{P}(X=1) = \mu_{0j} + \gamma_j^L Z_j$. Using Taylor expansion at 0, we

have

$$\mathbb{P}(X=1) \approx F\left(\gamma_j^\star Z_j - c_X\right) \tag{5}$$

$$\approx F(-c_X) + F'(-c_X)\gamma_j^\star Z_j. \tag{6}$$

Thus, similar to the logistic regression, we also have

$$\gamma_j^L \approx s_X \gamma_j^\star, \quad s_X = F'\left(F^{-1}\left(p_X\right)\right). \tag{7}$$

When $X^\star \sim N(\mu, 1)$, the ratio $s_X = \Phi'\left(\Phi^{-1}\left(p_X\right)\right)$ as a function of p_X is shown in Figure 2b.

Approximation accuracy in logistic and linear regression coefficients

The accuracy of the true logistic or linear regression coefficients γ_j^H and γ_j^L on the binary trait for approximating the marginal association γ_j^* with the continuous trait depends on the precision of the Taylor expansion approximation in equations (3) and (5). This precision is influenced by the magnitude of the omitted higher-order terms in the expansion. Specifically, the accuracy of the Taylor expansion is closely related to the relative curvature, defined as the ratio between the second and first order derivatives of the corresponding function.

Figure 2c illustrates the relative curvature for both logistic regression and linear regression when $X^\star \sim N(0, 1)$. For logistic regression, the relative curvature remains very close to zero, irrespective of the binary trait's prevalence. In contrast, the relative curvature for the linear regression approximation is significantly higher, particularly when $P(X = 1)$ is near 0 or 1. Consequently, the logistic regression coefficients γ_j^H can approximate γ_j^* much more accurately than the linear regression coefficients for binary traits.

Scaling of the causal effect in MR

Given the causal relationship between the continuous traits $X^\star$ and $Y^\star$ (Figure 1), two-sample MR often starts with the single-SNP linear models

$$\Gamma_j^\star = \beta\gamma_j^\star + \alpha_j^\star.$$

Here, β represents the causal effect of the exposure on the outcome. The terms $\Gamma_j^\star$ and $\gamma_j^\star$ are marginal associations between SNP j and the outcome and exposure traits, respectively. The term $\alpha_j^\star$ denotes the horizontal pleiotropic effect of SNP j. Additional assumptions on $\alpha_j^\star$, such as the InSIDE assumption [3] assuming that the pleiotropic effects are independent of genetic associations, are necessary in MR methods addressing horizontal pleiotropy to ensure the identifiability of β [8, 24, 29, 31].

Based on our derivations, when the observed exposure or outcome trait, or both, are binary, the mean regression coefficient on the observed traits γ_j and Γ_j approximately follow $\gamma_j \approx s_X\gamma_j^\star$ and $\Gamma_j \approx s_Y\Gamma_j^\star$. Thus, we have

$$\Gamma_j \approx \beta s_Y / s_X \gamma_j + \alpha_j^\star$$

with the same pleiotropic assumptions as the scenario when underlying continuous traits are observed. Therefore, two-sample MR on the binary traits provides an estimate of the scaled causal effect $\beta s_Y / s_X$. Given our previous derivations, the size of the scaling factors s_Y and s_X depends on whether the trait is

binary, the regression model used to compute summary statistics from the binary trait, the prevalence of the binary trait, and the distribution of the underlying continuous trait. The first three factors are either observed or can be easily calculated. The distribution of the underlying continuous trait after standardization can be reasonably assumed to follow a standard Gaussian distribution.

GWAS summary statistics after adjusting for additional covariates

In practice, many GWAS summary statistics are obtained after adjusting for additional environmental covariates U, such as age, gender, and the top principle components representing population structure. Under common MR assumptions, the genetic variants Z are unconfounded and therefore independent of the environmental covariates. Our argument that the true coefficients from the logistic or linear regression on binary traits are approximately proportional to the linear regression coefficients on the underlying continuous traits still holds under such unconfoundedness.

Under the assumption that $U \perp\!\!\!\perp Z$, the coefficients from linear regressions, either $\gamma_j^\star$ on the continuous trait $X^\star$ or γ_j^L on the binary trait X, remain unchanged regardless of whether U is adjusted for or not. However, the logistic regression coefficient γ_j^H will differ after adjusting for U, a phenomenon known as the non-collapsibility of odds ratios [18].

Consider the logistic regression model with the adjustment of U as $\mathbb{P}(X = 1 \mid Z_j, U) = H(\mu_{0j} + \gamma_{j,U}^H Z_j + U^T \beta_j)$. Since the effect of a single SNP is usually small, the coefficients of the environmental variables, β_j, should be similar across j. If $P(X = 1)$ is not too extreme and U follows an approximately normal distribution, given the approximation $H(x) \approx \Phi(x/1.7)$, following similar derivations as in [31], we will have

$$\mathbb{P}(X = 1 \mid Z_j) \approx H\left(\frac{\mu_{0j} + \gamma_{j,U}^H Z_j}{\sqrt{1 + \mathrm{Var}(U^T \beta_j)/1.7^2}}\right).$$

Given that β_j is similar across j, this approximation indicates that $\gamma_{j,U}^H$ and the unadjusted logistic regression coefficient γ_j^H are approximately proportional to each other, thus are also proportional to the regression coefficient on the underlying continuous trait. The ratio of the coefficients depends on the contribution of U on the binary trait X.

3 Empirical evaluations

We generate synthetic GWAS data for both continuous and binary traits to evaluate whether the mean summary statistics are indeed approximately proportional to each other. Specifically, we generate a latent continuous trait $X^\star$ using a linear regression model with 100 independent SNPs (for details, see SI text S2). The effective allele frequencies (EAF) of these SNPs are randomly sampled from the empirical EAFs in the UK Biobank data (`http://www.nealelab.is/uk-biobank/`), with effect sizes randomly sampled from a standard Gaussian distribution. We then introduce Gaussian noise to $X^\star$ such that these SNPs in total explain approximately 60% of its variability. Subsequently, we standardize the continuous trait and dichotomize it to generate the binary trait X where $X = 1$ if $X^\star > c$. The threshold c is chosen to match three prevalence values: 50%, 1% and 0.1%. As discussed earlier, larger effect sizes of the SNPs can lead to worse approximation accuracy. We compared the range of the per-SNP effect sizes on the continuous trait after standardization in our synthetic data with the estimated effect sizes of common continuous traits in GWAS summary statistics and observed that the empirical effect sizes are within the range of our synthetic data (Figure 3ab).

We perform 50 repeated simulations to generate summary statistics for these 100 SNPs and compute the mean regression coefficients as the average of these simulations (Figure 3b). Empirically, the regression coefficients on binary traits, whether from logistic or linear regression, are indeed proportional to the linear regression coefficients on the underlying continuous traits. Additionally, after scaling the original binary trait coefficients by $1/s_X$, defined as in (4) and (7), they align well with the continuous trait coefficients. As predicted in our empirical derivations, the accuracy of the approximations is determined by three factors, the prevalence $P(X = 1)$, the strength of the associations, and the type of regression model on the binary trait. Our argument generally works well for logistic regression on any binary traits or linear regression on binary traits whose prevalence is not too extreme in the population of interest.

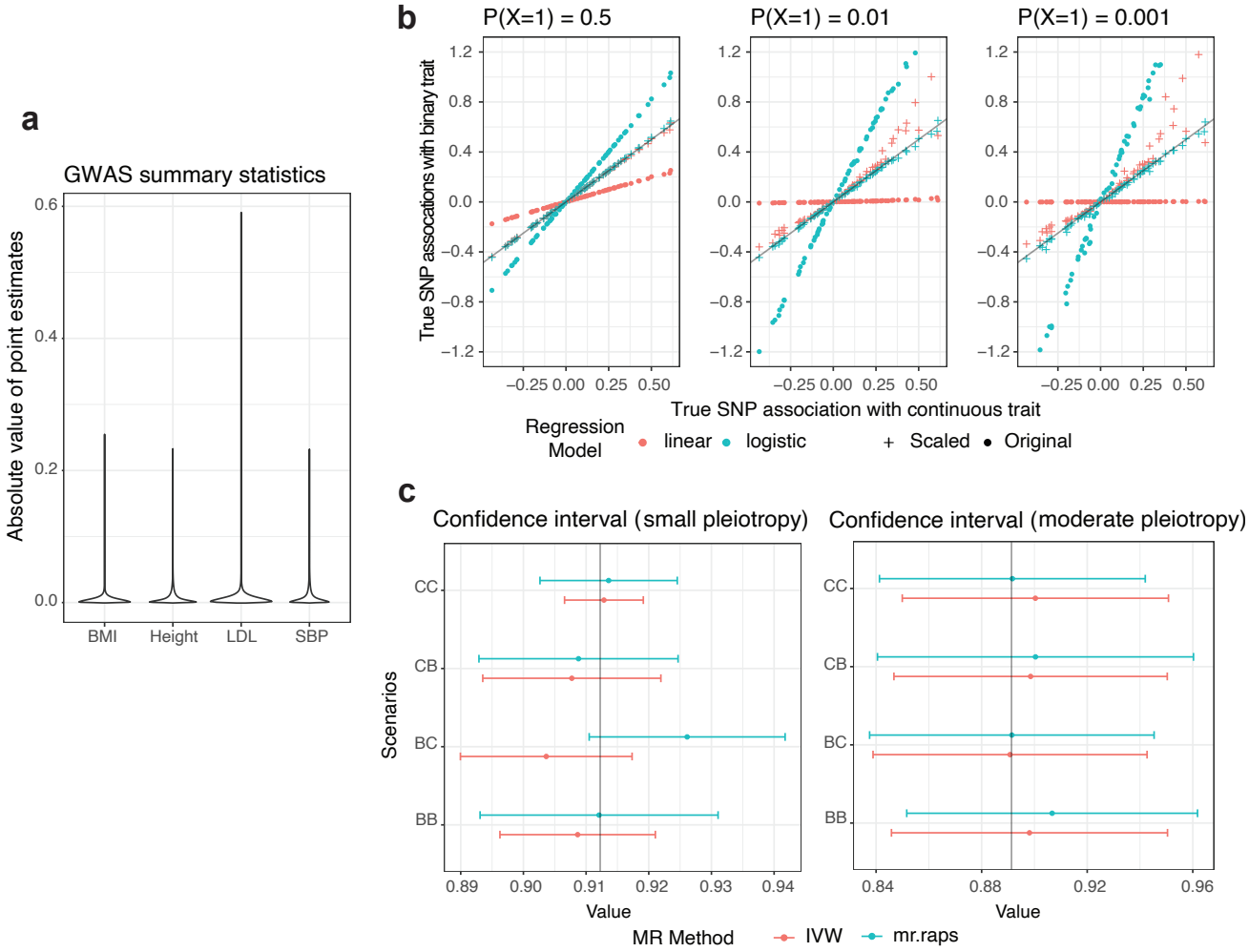

Figure 3: a) Range of observed summary statistics on common continuous traits. b) True per SNP associations on continuous traits v.s. true per SNP associations on the binary traits, for both linear and logistic regression, before and after scaling by s_X. The true associations are computed as the average summary statistics across 50 repeated simulations. The black line indicates $y = x$. c) MR confidence intervals after proper scaling under four scenarios: continuous exposure and outcome traits (CC), continuous exposure and binary outcome (CB), binary exposure and continuous outcome (BC), binary exposure and outcome (BB). The results shown are based on one random simulation experiment.

Finally, we compare MR results using binary traits with MR estimates derived directly from the underlying continuous traits. We consider four scenarios: both exposure and outcome traits are binary,

either one is binary, and both are continuous. For binary traits, MR estimates the scaled causal effect $\beta s_Y/s_X$. Therefore, we scale the estimates of each scenario by s_X/s_Y to facilitate comparison across different scenarios (Figure 3c). Mendelian Randomization is performed using two widely used methods: IVW [5] and mr.raps [31]. Figure 3c shows that using binary traits for either the exposure or the outcome yields MR estimates comparable to those obtained using continuous traits. However, directly observing the continuous traits tends to provide more precise estimates, as binarization intuitively reduces the information in the data and increases the errors in GWAS summary statistics. Surprisingly, the MR estimates across the four scenarios can have similar uncertainty levels when the horizontal pleiotropic effects are moderately large.

4 Discussion

We propose a new interpretation of MR for binary exposures and outcomes based on the liability model, treating binary traits as dichotomization of underlying continuous traits. Through mathematical derivations and empirical evidence, we demonstrate that existing two-sample MR methods remain applicable to GWAS summary statistics for binary traits. The results can be interpreted as a scaled causal effect of the underlying continuous exposure on the continuous outcome, with the scale easily computed given the prevalence of the binary traits. Since our interpretation relies solely on the observation that the summary statistics of the binary trait are proportional to those of the underlying continuous trait, our argument not only validates the use of univariate MR methods but also supports the application of standard MR methods in more complex frameworks, such as multivariate MR, life-course MR, and within-family MR, on binary traits.

While the idea of treating binary or ordinal categorical exposures as discretizations of latent continuous exposures has been proposed earlier in MR [7, 22], there have been no further justifications for using conventional two-sample MR methods under this interpretation. It is also believed in these works that MR results are not easily interpretable and more assumptions are required when the exposure is a binary trait. We clarify this issue and provide detailed discussions on the impact of the liability model of binary traits on conducting two-sample MR. Besides binary traits, we also extend our mathematical argument to ordered categorical traits such as education attainment (see SI text S1.2). The mathematical derivations used in this work, based on small SNP-trait associations, are closely related to our earlier interpretation of the MR model when the outcome trait is binary ([31], supplement A.1). Here, we provide a unified interpretation for both binary exposure and outcome, and offer more accurate approximations, especially when the prevalence of the binary traits is close to 0 or 1. Additionally, there are specialized methods developed for binary outcomes [10, 14], but they typically require individual-level data instead of summary statistics.

Our interpretation of the MR model can also be applied to GWAS summary statistics that are obtained from case-control studies. In case-control studies, where the GWAS data are collected retrospectively rather than prospectively, the logistic regression coefficients remain unchanged. However, the coefficients from linear regression on the binary trait will differ. Despite this, even if the GWAS summary statistics are derived from linear regression in case-control studies, the true linear coefficient γ_j^L should still be approximately proportional to γ_j^H in the case-control population, and thus approximately proportional to γ_j^*.

One limitation of our work is that we assume a simplified relationship between the liability contin-

uous trait and observed binary trait, allowing only a common threshold across all individuals in the population. In practice, the threshold may be individual-specific, depending on other features such as age. These features may also be confounding factors between the exposure and outcome, resulting in a more complicated causal relationship between the traits than illustrated in Figure 1. Additionally, when GWAS summary statistics are obtained after adjusting for other covariates, our interpretation does not apply if these covariates are confounding factors between the genotypes and the trait, as discussed earlier in the paper.

5 Declaration of Interests

The authors have declared that no competing interests exist.

6 Acknowledgments

J.W. is partly supported by the National Science Foundation under grants DMS-2113646 and DMS-2238656.

References

[ukb] Uk biobank summary statistics description. `https://www.nealelab.is/blog/2017/9/11/details-and-considerations-of-the-uk-biobank-gwas`. Accessed: 2024-05-16.

[2] Bowden, J., Davey Smith, G., Haycock, P. C., and Burgess, S. (2016). Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic epidemiology*, 40(4):304–314.

[3] Bowden, J., Del Greco M, F., Minelli, C., Davey Smith, G., Sheehan, N., and Thompson, J. (2017). A framework for the investigation of pleiotropy in two-sample summary data mendelian randomization. *Statistics in medicine*, 36(11):1783–1802.

[4] Brumpton, B., Sanderson, E., Heilbron, K., Hartwig, F. P., Harrison, S., Vie, G. Å., Cho, Y., Howe, L. D., Hughes, A., Boomsma, D. I., et al. (2020). Avoiding dynastic, assortative mating, and population stratification biases in mendelian randomization through within-family analyses. *Nature communications*, 11(1):1–13.

[5] Burgess, S., Butterworth, A., and Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic epidemiology*, 37(7):658–665.

[6] Burgess, S., Dudbridge, F., and Thompson, S. G. (2015). Re:"multivariable mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects". *American journal of epidemiology*, 181(4):290–291.

[7] Burgess, S. and Labrecque, J. A. (2018). Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *European journal of epidemiology*, 33:947–952.

[8] Burgess, S. and Thompson, S. G. (2017). Interpreting findings from mendelian randomization using the mr-egger method. *European journal of epidemiology*, 32:377–389.

[9] Carter, A. R., Sanderson, E., Hammerton, G., Richmond, R. C., Davey Smith, G., Heron, J., Taylor, A. E., Davies, N. M., and Howe, L. D. (2021). Mendelian randomisation for mediation analysis: current methods and challenges for implementation. *European journal of epidemiology*, 36(5):465–478.

[10] Clarke, P. S. and Windmeijer, F. (2010). Identification of causal effects on binary outcomes using structural mean models. *Biostatistics*, 11(4):756–770.

[11] Davey Smith, G. and Ebrahim, S. (2003). 'mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International journal of epidemiology*, 32(1):1–22.

[12] de Villemereuil, P. (2018). Quantitative genetic methods depending on the nature of the phenotypic trait. *Annals of the New York Academy of Sciences*, 1422(1):29–47.

[13] Dempster, E. R. and Lerner, I. M. (1950). Heritability of threshold characters. *Genetics*, 35(2):212.

[14] Deng, Y., Tu, D., O'Callaghan, C. J., Jonker, D. J., Karapetis, C. S., Shapiro, J., Liu, G., and Xu, W. (2023). A bayesian approach for two-stage multivariate mendelian randomization with mixed outcomes. *Statistics in Medicine*, 42(13):2241–2256.

[15] Didelez, V. and Sheehan, N. (2007). Mendelian randomization as an instrumental variable approach to causal inference. *Statistical methods in medical research*, 16(4):309–330.

[16] Emdin, C. A., Khera, A. V., and Kathiresan, S. (2017). Mendelian randomization. *Jama*, 318(19):1925–1926.

[17] Falconer, D. S. (1965). The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of human genetics*, 29(1):51–76.

[18] Greenland, S., Pearl, J., and Robins, J. M. (1999). Confounding and collapsibility in causal inference. *Statistical science*, 14(1):29–46.

[19] Hartwig, F. P., Davies, N. M., Hemani, G., and Davey Smith, G. (2016). Two-sample mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique.

[20] Hernán, M. A. and Robins, J. M. (2006). Instruments for causal inference: an epidemiologist's dream? *Epidemiology*, 17(4):360–372.

[21] Howe, L. J., Rasheed, H., Jones, P. R., Boomsma, D. I., Evans, D. M., Giannelis, A., Hayward, C., Hopper, J. L., Hughes, A., Lahtinen, H., et al. (2023). Educational attainment, health outcomes and mortality: a within-sibship mendelian randomization study. *International Journal of Epidemiology*, 52(5):1579–1591.

[22] Howe, L. J., Tudball, M., Davey Smith, G., and Davies, N. M. (2022). Interpreting mendelian-randomization estimates of the effects of categorical exposures such as disease status and educational attainment. *International journal of epidemiology*, 51(3):948–957.

[23] McCullagh, P. (2019). *Generalized linear models*. Routledge.

[24] Morrison, J., Knoblauch, N., Marcus, J. H., Stephens, M., and He, X. (2020). Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nature genetics*, 52(7):740–747.

[25] Power, G. M., Sanderson, E., Pagoni, P., Fraser, A., Morris, T., Prince, C., Frayling, T. M., Heron, J., Richardson, T. G., Richmond, R., et al. (2023). Methodological approaches, challenges, and opportunities in the application of mendelian randomisation to lifecourse epidemiology: A systematic literature review. *European Journal of Epidemiology*, pages 1–20.

[26] Richmond, R. C. and Smith, G. D. (2022). Mendelian randomization: concepts and scope. *Cold Spring Harbor perspectives in medicine*, 12(1):a040501.

[27] Roddam, A. W. (2008). Measurement Error in Nonlinear Models: A Modern Perspective. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 171(2):505–506.

[28] Sanderson, E., Richardson, T. G., Morris, T. T., Tilling, K., and Davey Smith, G. (2022). Estimation of causal effects of a time-varying exposure at multiple time points through multivariable mendelian randomization. *PLoS Genetics*, 18(7):e1010290.

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

[30] Yau, L. H. and Little, R. J. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association*, 96(456):1232–1244.

[31] Zhao, Q., Wang, J., Hemani, G., Bowden, J., and Small, D. S. (2020). Statistical inference in two-sample summary-data mendelian randomization using robust adjusted profile score.

Supplementary Information

S1 More mathematical details

S1.1 Approximation accuracy in the latent variable thresholding model

To proceed, we make the following assumptions

- Gaussian latent traits: both $X^\star$ and ϵ_j are Gaussian distributed with $\mathrm{Var}(X^\star) = \sigma^2_{X^\star}$ and $\mathrm{Var}(\epsilon_j) = \sigma_j^2$
- Ignorable variance explained by single SNP:

$$\gamma_j^2 \ll \mathrm{Var}(X^*)$$

- Independence: Z_j and ϵ_j's are independent and $\mathbb{E}(\epsilon_j) = 0$.

Then,

$$\begin{aligned}
\mathbb{P}(X = 1) &= \mathbb{P}(X^* > c_X) \\
&= \mathbb{P}(\mu_j + \gamma_j Z_j + \epsilon_j > c_X) \\
&= \mathbb{P}\left(\frac{-\epsilon_j}{\sigma_j} < \frac{\mu_j + \gamma_j Z_j - c_X}{\sigma_j}\right) \\
&= \Phi\left(\frac{\mu_j - c_X}{\sigma_j} + \frac{\gamma_j}{\sigma_j} Z_j\right)
\end{aligned}$$

Thus, probit regression of X on Z_j leads to the coefficient γ_j/σ_j. Now we compute the relative difference between this coefficient and the linear coefficient $\gamma_j/\sigma_{X^\star}$ on the standardized continuous trait. Note

$$\begin{aligned}
\left|\frac{\gamma_j}{\sigma_j} - \frac{\gamma_j}{\sigma_{X^\star}}\right| \Big/ \left|\frac{\gamma_j}{\sigma_{X^\star}}\right| &= \left|\frac{\sigma_X^\star - \sigma_j}{\sigma_j}\right| \\
&= \frac{\sigma_{X^\star}}{\sqrt{\sigma^2_{X^\star} - \gamma_j^2 \mathrm{Var}(Z_j)}} - 1 \\
&= \frac{1}{\sqrt{1 - \gamma_j^2 \mathrm{Var}(Z_j)/\sigma^2_{X^\star}}} - 1
\end{aligned}$$

Let $a = \gamma_j^2 \mathrm{Var}(Z_j)/\sigma^2_{X^\star}$, which is the proportion of the variance of $X^\star$ explained by the SNP Z_j. Then the relative error is

$$\left|\frac{\gamma_j}{\sigma_j} - \frac{\gamma_j}{\sigma_{X^\star}}\right| \Big/ \left|\frac{\gamma_j}{\sigma_{X^\star}}\right| = \frac{1}{\sqrt{1-a}} - 1.$$

The relative error is less than 0.0541 when $a \leq 0.1$ and less than 0.026 when $a \leq 0.05$.

S1.2 Summary statistics from linear regression for ordered categorical trait

Similar to the binary trait setting, for an ordered categorical trait, we assume that it is a discretization of the underlying continuous liability X^* [22]. Specifically, for an ordered categorical variable X with K levels, we assume

$$X = k \text{ if } X^* \in (c_k, c_{k+1}]$$

where $k \in \{0, 1, \ldots, K-1\}$ and $-\infty = c_0 < c_1 < \cdots < c_K = \infty$.

Denote the CDF of the continuous trait $X^\star$ as $F(\cdot)$. For its marginal association with a single SNP j, we have the following decomposition:

$$X^\star = \mu_j + \gamma_j Z_j + \epsilon_j$$

Then approximately we have

$$\begin{aligned}
\mathbb{P}(X \leq k) &= \mathbb{P}(\mu_j + \gamma_j Z_j + \epsilon_j \leq c_{k+1}) \\
&\approx F(c_{k+1} - \gamma_j Z_j) \\
&\approx F(c_{k+1}) - F'(c_{k+1})\gamma_j Z_j,
\end{aligned}$$

where the first approximation is due to ignorable γ_j and the second approximation is due to Taylor expansion.

If we perform linear regression on the categorical variable X to obtain GWAS summary statistics, then the linear regression on X fits the model $\mathbb{E}[X \mid Z_j] = \mu_{0j} + \gamma_j^L Z_j$. And we have

$$\begin{aligned}
\mathbb{E}[X \mid Z_j] &= \sum_{k=0}^{K-2} \mathbb{P}(X > k \mid Z) \\
&= (K-1) - \sum_{k=0}^{K-2} \mathbb{P}(X \leq k \mid Z) \\
&\approx (K-1) - \sum_{k=0}^{K-2} \big(F(c_{k+1}) - F'(c_{k+1})\gamma_j Z_j\big) \\
&= \left(K - 1 - \sum_{k=0}^{K-2} F(c_{k+1})\right) + \left(\sum_{k=0}^{K-2} F'(c_{k+1})\right) \gamma_j Z_j.
\end{aligned}$$

This indicates that

$$\gamma_j^L \approx s_X \gamma_j$$

where the ratio $s_X = \sum_{k=0}^{K-2} F'(c_{k+1})$ is a constant across all the SNPs.

S2 Simulation details

Here we provide more details of our simulation setup. First we generate the effect size γ_j^* of 100 independent SNPs j on the trait X^* from standard Gaussian distribution

$$\gamma_j \sim \mathcal{N}(0, 1), \quad j = 1, 2, \cdots, 100,$$

The EAFs $(p_1, \cdots, p_{100})$ of the SNPs are randomly sampled from the empirical EAFs in the UK Biobank.

Then we conduct experiments for 50 times with these fixed effect sizes and EAFs. At each replication, we generate a population of $n = 400K$ individuals. Specifically, the SNPs Z_{ij}, the continuous traits X_i^*

and the binary traits X_i are generated following:

$$
\begin{aligned}
Z_{ij} &\sim \text{Bernoulli}(p_j), \quad i = 1, 2, \cdots, 4 \cdot 10^5, \\
\epsilon_i &\sim \mathcal{N}\left(0, \text{Var}\left(\sum_{j=1}^{100} Z_{ij}\gamma_j\right) \cdot \frac{2}{3}\right), \\
\tilde{X}_i^\star &= \sum_{j=1}^{100} Z_{ij}\gamma_j + \epsilon_i, \\
X_i^\star &= \left(\tilde{X}_i^\star - \mathbb{E}[\tilde{X}_i^\star]\right) / \sqrt{\text{Var}(\tilde{X}_i^\star)}, \\
X_i &= \mathbf{1}\{X_i^\star \geq \text{quantile}(X_i^\star, c)\}, \quad c = 0.5, 0.99, 0.999.
\end{aligned}
$$

Here the variance of ϵ_i is designed in the way such that these 100 SNPs in total explain approximately 60% of the variability in X. The threshold c is chosen such that $\mathbb{P}(X_i = 1) = 0.5, 0.01, 0.001$, respectively. Here $\tilde{X}_i^\star$ represents the unscaled trait and $X_i^\star$ represents the standardized trait. And we let

$$\gamma_j^\star = \gamma_j / \sqrt{\text{Var}(\tilde{X}_i^\star)}$$

be the true effect sizes of Z_j on the standardized $X^\star$. In (Figure 3b) we plot the true SNP association with binary trait against true SNP association with continuous trait. Since the EAFs can be relatively low for some SNPs, quasi-complete separation could occur in these 50 replications for logistic regression estimates. Here we just remove the logistic regression estimates for such SNPs if such computational errors occur.

For the evaluation of MR results, we generate both the exposure and outcome traits in a similar approach as discussed above. Specifically, we generate the continuous exposure trait $X^\star$ as above and set the binary trait X to have a prevalence 0.1:

$$X_i = \mathbf{1}\{X_i^\star \geq \text{quantile}(X_i^\star, 0.9)\}$$

for individual i. Then we generate outcome traits Y and $Y^\star$ given the continuous exposure trait $X^\star$ and the pleiotropic effects α:

$$
\begin{aligned}
\alpha_j &\sim \mathcal{N}(0, \sigma^2), \quad \sigma = 0.008 \text{ or } 0.08, \quad j = 1, 2, \cdots, 100, \\
\epsilon_i^Y &\sim \mathcal{N}(0, 0.2), \\
\tilde{Y}_i^\star &= X_i^\star + \sum_{j=1}^{100} Z_{ij}\alpha_j + \epsilon_i^Y, \\
Y_i^\star &= \left(\tilde{Y}_i^\star - \mathbb{E}[\tilde{Y}_i^\star]\right) / \sqrt{\text{Var}(\tilde{Y}_i^\star)}, \\
Y_i &= \mathbf{1}\{Y_i^\star \geq \text{quantile}(Y_i^\star, 0.9)\}, \\
\beta &= 1 / \sqrt{\text{Var}(\tilde{Y}_i^\star)}.
\end{aligned}
$$

Here $\sigma = 0.08$ corresponds to the medium pleiotropy setting and $\sigma = 0.008$ corresponds to the weak pleiotropy setting. The parameter β can be understood as the causal effect of one unit increase of $X^\star$ on one unit of $Y^\star$. The IVW analysis was conducted using the package (`https://github.com/`

WSpiller/MVMR), and for mr.raps analysis we use *grappleRobustEst* function from (https://github.com/jingshuw/GRAPPLE).