Measurement error and power in twin-design extensions to Mendelian Randomization

Luis FS Castro-de-Araujo[1,2,6] (corresponding author) p: +1 804 502-4074 P.O. Box 980126, Richmond, VA 23298-0126, USA. luis.araujo@vcuhealth.org . orcid:0000-0002-0952-5052

Madhurbain Singh[1,5]**.** singhm18@vcu.edu, orcid:0000-0002-9396-2860

Yi (Daniel) Zhou[1]**.** zhouy33@vcu.edu

Philip Vinh[1,5]**.** vinhpb@vcu.edu

Hermine HM Maes[1,5]**.** hermine.maes@vcuhealth.org , orcid: 0000-0001-7489-2214

Brad Verhulst[3]. verhulst@tamu.edu

Conor V Dolan[4]**.** c.v.dolan@vu.nl. orcid:0000-0002-2496-8492

Michael C Neale[1,4,5,6]. michael.neale@vcuhealth.org. orcid:0000-0003-4887-659X

Author contributions

All authors contributed to the study conception and design. Material preparation, and analysis were performed by **Luis Castro-de-Araujo**, **Madhurbain Singh, Yi (Daniel) Zhou, Philip Vinh, Hermine Maes, Brad Verhulst, Conor V Dolan,** and **Michael C Neale**. The first draft of the manuscript was written by **Luis Castro-de-Araujo**, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Abstract

Mendelian Randomization (MR) has become an important tool for causal inference in the health sciences. It takes advantage of the random segregation of alleles to control for background confounding factors. In brief, the method works by using genetic variants as

[1] Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, 1-156
[2] Dept of Psychiatry, Austin Health. The University of Melbourne, Victoria, Australia.
[3] Department of Psychiatry and Behavioral Sciences, Texas A&M University. 2900 E 29th Street Bryan Texas, 77802, USA.
[4] Department of Biological Psychology, Vrije Universiteit. Amsterdam, Transitorium 2B03, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands.
[5] Department of Human and Molecular Genetics, Virginia Commonwealth University
[6] Department of Psychiatry, Virginia Commonwealth University

instrumental variables, but it depends on the assumption of exclusion restriction, i.e., that the variants affect the outcome exclusively via the exposure variable. Equivalently, the assumption states that there is no horizontal pleiotropy from the variant to the outcome. This assumption is unlikely to hold in nature, so several extensions to MR have been developed to increase its robustness against horizontal pleiotropy, though not eliminating the problem entirely (Sanderson et al. 2022). The Direction of Causation (DoC) model, which affords information from the cross-twin cross-trait correlations to estimate causal paths, was extended with polygenic scores to explicitly model horizontal pleiotropy and a causal path (MR-DoC, Minică et al 2018). MR-DoC was further extended to accommodate bidirectional causation (MR-DoC2 ; Castro-de-Araujo et al. 2023). In the present paper, we compared the performance of the DoC, MR-DoC, and MR-DoC2 models in regard to the effect of phenotypic measurement error and the effect of misspecification of unshared (individual-specific) environmental factors on the parameter estimates.

Keywords: causality, pleiotropy, twin design, Mendelian randomization

Reviewer comments

REVIEWER REPORTS

We want to start by noticing that since these models are in variance component style, we switched to refer too covA, covC, covE where there were ra, rc, re in the previous version.

Reviewer Comments:

Reviewer 1

Castro-de-Araujo et al present a simulation study examining the robustness of different flavours of twin Direction of causation model.

Major comments:

(1) As far as I can see, in MRDoC2, the addition of twin data provides no useful information on the direction, size and significance of the causal effect (g1 and g2) above what could be obtained by fitting a bidirectional model to unrelated individuals (i.e. essentially fitting the lower half of Figure 1c to unrelated individuals). Surely fitting this latter model would be superior compared to fitting in twins, simply because vastly more unrelated individuals could be analysed? What then is the point of fitting MR-DoC2 to twin data besides the rather more modest goal of estimating VA, VC and VE? To play devil's advocate, isn't MRDOC2 just "window dressing"?

Thank you for the comment. The model you allude to (a version of MR-DoC2 for unrelated persons, with no A, C, and E partitioning) is indeed identified. However, we conceptualized this paper as extending the tradition of the direction of causation models, as Minica (2018) did with MR-DoC. We see this as an option for a Mendelian randomization approach when \ data are available. A great advantage of the MR-DoC and MR-DoC2 models is that they form starting points for causal inference with data from relatives, when using multiple indicators, when the data are censored variables, and estimation of genotype-environment correlation (rAC). For the last two we have working models that we plan to make public soon. In other words, it is not window dressing when the available data come from relatives such as twin pairs. Many individuals were genotyped as part of family-based registries of twins, and to use these data in the most informative way seems expedient. It is also useful to consider the possible model extensions that cannot be evaluated in Mendelian randomization studies of unrelated individuals. Furthermore, Minica et al. (2020) reports, using simulations, that MR-DoC is robust to dynastic effects and assortative mating. Therefore, twin approaches provide causal inference robust to some important assumptions that are made when analyzing data from unrelated individuals. Undetected failures of these assumptions could substantially bias results of bidirectional MR analyses, and such failures are often hard to detect and correct. It is also worth noting that given access to twin data, MR-DoC (Minica et al. 2018) is not affected by weak instrument bias and allows for the identification of the pleiotropic path (denoted b2 in the paper), which can be useful when one is interested in estimating the amount of pleiotropy. Finally, the incorporation of genetic instruments in the DoC model leads the way for a wide variety of extensions to the method.

These include a variety of options, including: longitudinal designs, moderation of causal paths, and sources of cov(AC), such as cultural transmission and sibling interaction.

Minică, C., Boomsma, D., Dolan, C. V., Geus, E. D. de, & Neale, M. (2020). Empirical comparisons of multiple Mendelian randomization approaches in the presence of assortative mating. International Journal of Epidemiology. https://doi.org/10.1093/ije/dyaa013

(2) The authors claim that MR-DoC2 accommodates two sources of horizontal pleiotropy and "Of note, this type of pleiotropy is effectively equivalent to direct horizontal pleiotropy". I disagree with this statement, and the underlying undercurrent at several places in the manuscript i.e. that MR-DoC2 is robust to horizontal pleiotropy. The rf*b1 and rf*b3 paths in Figure 1c are not the same as e.g the pleiotropic path indexed by the b2 path coefficient in Figure 1b. For a start, the parameter rf will be primarily determined by the covariance between PS1 and PS2. This, I would expect, "constrain" the possible values that b1 and b3 could attain in the fitted model. Presumably also, if PS1=PS2, and rf=1 then the model would be empirically unidentified- correct? I think it would be interesting to examine the degree to which MR-DOC2 truly is (or is not) robust to horizontal pleiotropy of the type in the diagram in Figure 1B. If there is indeed some robustness, then this would be an important addition to the literature and would make me revise my opinion of the manuscript. If not, then I think the contribution is at best marginal.

We apologize: the phrasing is incorrect in the parts mentioned, as the 2023 paper does not address horizontal pleiotropy. We changed the paper throughout to mention that MR-DoC2 does not address all forms of horizontal pleiotropy, but does include other sources of association, thus realigning the text to the 2023 paper.

(3) The Methods section is diabolically difficult to follow, in particular the paragraph beginning "The five simulations performed were based on three factorial designs" and I gave up trying to understand Table 1. This needs a rewrite for clarity.

We reorganised the methods section into subsections to clarify how the data were generated and what tests were performed with those data.

Minor comments

Title:

The authors talk about "pleiotropy robustness" in the title to this paper, but it seems to me most of the simulations involve robustness to measurement error and the r_e = 0 assumption rather than pleiotropy?

We removed pleiotropy robustness, as we are not explicitly testing it.

Abstract:

- "It takes advantage of the random segregation…". It also takes advantage of independent assortment to protect from genetic confounding.

Adjusted to mention independent assortment.

- “that the variants affect the outcome exclusively via the exposure variable”. The variants don’t have to “affect” the exposure or the outcome, merely that they associate with differences in their means. Suggest a rewording.

Changed to “associates with” instead of ”affects”. There is some ambivalence with this change, since if the variants do affect the outcome, and other sources of association between variant and outcome have been excluded, then the causal language is reasonable.

Introduction:

- “This is particularly problematic in psychiatry, for example, when working with samples of children, which hinders intervention evaluation”. I’m not quite sure what the authors mean here. Perhaps they could reword?

We decided to remove it, the sentences above it already clearly pointed to limitations on RCTs applications.

- “…the association between the genetic variant(s) and the exposure must be sufficient…”. Maybe use another word other than “sufficient” which commonly has a special meaning in statistics for something quite different.

Edited accordingly in this excerpt. It now reads “First, the strength of the association between the genetic variant(s) and the exposure must be strong, which is known as the *relevance assumption*. “

- “..instruments effect on the outcome…”. Again, is doesn’t have to be a causal effect, it can just be an association. Suggest a reword.

Edited accordingly.

- Only two of the three instrumental variables assumptions are described in the introduction. I’d like to the see the exchangeability/no confounding assumption also mentioned. Also, MR assumes some kind of “gene-environment equivalence” (or “consistency” as it is confusingly referred to in the econometrics literature) whereby “intervening” on an exposure genetically produces the same effect on an outcome as intervening by environmental means. This could also be mentioned.

Added both exchangeability and the gene-environment equivalence assumptions, with proper citations.

- “Restriction assumption”- should be “restriction exclusion assumption”

Edited accordingly.

-The authors should also directly reference other MR sensitivity methods designed to address the issue of horizonal pleiotropy e.g. Weighted median methods (Bowden et al) and the modal estimator (Hartwig et al).

Thanks, mention of the mode method was missing, and we edited the introduction to better emphasize these methods.

-“two have to be constrained to zero”. Does the value have to be zero? I thought that the model would be identified if they were just constrained to any value (-1 < r_e < 1)?

Thank you. The model is identified if constrained to any valid value, edited to reflect this.

-“First, the model specifications will be presented…”- except the authors don’t really do this, rather they only reference the model specifications in the methods.

Thank you. In the revised version we explain the details of the specification instead of only linking to subsequent sections.

Discussion

“To identify the b2 parameter it is necessary to assume that specific environmental confounding (re) is zero.” Just constrained to some value not necessarily zero, correct? Also, you could identify the model by constraining r_c to some value like zero also, correct?

Thank you for the comment. This is correct, we changed the text to reflect that to identify b2 another constraint would be required such as fixing covC, covE, or others.

-“Causal” not “casual”

Corrected.

-“Feedback loops are frequent in nature, and most current MR methods cannot evaluate this type of relationship.” Methods exist for bidirectional MR and these should be cited. Whilst it is true that feedback loops are not directly modelled by traditional MR methods, their existence does not affect the ability of MR methods to obtain asymptotically consistent parameter estimates of the bidirectional effect regardless.

Thank you, yes. We rephrased to be a less strong statement about MR ability to evaluate feedback loops. We also included references to bidirectional non-twin modeling.

Figure 1: It would be great to make explicit the path coefficients from the latent A, C and E variables to the observed traits.

Thank you, this was an error. Corrected.

Table 1 Legend: Parameters x and y are not listed. Do you mean sigma_x and sigma_y?

Thank you. Corrected.

Reviewer 2

This paper gives a nice comparison of the DoC, MR-DoC and MR-DoC2 models and discusses some of their behaviours, especially how DoC and MR-DoC are sensitive to confounding in the E component (which seems a realistic possibility) while MR-DoC2 is robust to this and apparently also in measurement error in the exposure and/or outcome.

This adds to the literature on causal inference in twin models, which remains somewhat niche but could gain importance in light of these results. I learned a lot by reading it, and just raise a few points of clarification.

1. It's not clear what the MR-DoC model provides that the DoC model doesn't. It seems to introduce two new covariances to estimate two more parameters, so doesn't improve on the identifying assumptions. Perhaps this is explained in Minica's paper, but it would help readers of this paper to have it here too. Along these lines the results in figs 2-5 seem identical for DoC and MR-DoC, and this is worth discussing.

MR-DoC provides a way to estimate the horizontal pleiotropy path (with path coefficient denoted b2), which is biased in the presence of covE (correlation between the unshared environmental variables) in the data. In all tests performed, it behaves similarly to the DoC model. This is indeed an important finding, and we added a paragraph to discuss this.

2. There's a degree of forward referencing in the Methods section towards the figures and tables in Results. Better to present the results when discussing them.

Thank you, also following Reviewer 1 remarks we rewrote the methods section to state early what the purpose of the simulation was, and better organised the groups of simulations. This helped to eliminate the forward references to the results section.

3. I was a bit confused by "MR-DoC2 as the data generating process". It wasn't entirely clear that the first results were a combination of the three simulation models – this is mentioned in the figure captions but not the text. So you are first presenting all three models combined, then just the last one. Why not each model separately, or 1+2 together and 3 separately?

Thank you. We did exactly this (1+2 and 3) in this version.

4. Also, all three models are essentially the same structure except that when fitting the DoC model the PRS effects are subsumed into the A and C components. The only difference is the bidirectional effect in model 3. So comparing DoC in figs 2 and 4 and in figs 3 and 5 is essentially comparing bias under different A and C components – interesting to be sure, but not necessarily the result of having PRS effects present.

We agree, the PRS is capturing part of the heritability tagged by common variants, and the A variance is capturing the latent heritability implied by the MZ and DZ correlations. The PRS heritability is a part of the family-based estimate. Therefore the comparisons are not meant to check the effects of the presence of the PRSs on bias, but rather to explore the effect of the misspecified parameters on biases (having *covE* in data, when it is not present in the model) or the effect of unreliable parameters on biases.

5. Is there a reason why MR-DoC2 should do better under measurement error?

The presence of extra parameters like covE and rf in MR-DoC2 seems to help confine the error measurement to the ex and ey variances. The absence of that parameter in mrdoc on the other hand seems to be involved in the bias of cx, cy, g1, and b2. See figure 2.

6. References to "NCP variance" are inaccurate – NCP is a parameter that doesn't have a variance.

We made corrections to where we were referring to the parameter. Indeed the NCP is a parameter without a variance, however all other occurrences of the phrase “NCP variance” are referring to the regression run with the NCP as dependent variable (NCP ~ b1 + b3 + g1 + g2 + …), in this context we can refer to variation in the NCP that is explained by the other parameters in the regression model.

7. It’s also strange that power of MR-DoC is not affected by instrument strength. Again begs the question of what MR-DoC actually gives over DoC.

We further discussed this issue in a new paragraph at the discussion section. Indeed, in MR-DoC, the instrument strength is not relevant to the causal estimates or their precision (not shown in this paper). MR-DoC is a way to identify the horizontal pleiotropy path (b2) if one is interested in estimating this value, knowing that it will be biased if re is incorrectly modelled. So it is of interest in those cases where re is known, or when there are dynastic effects or assortative mating in the population.

8. Page 8, “power tests performed were under the hypothesis” – should be “of the hypothesis”

Corrected.

9. Page 9, “where the weaker the instrument, the greater the bias” – true, but this a different issue to the power to detect the causal effect, which is the surprising issue here.

Well observed, thank you. We decided to remove that statement, and further expand the issue in a separate paragraph later in the paper.

10. Accounting for direct horizontal pleiotropy is nice, but one should also be aware of correlated pleiotropy (path from PRS to C or E) which is likely more the norm in nature.

Yes, this is a very important point that we plan to address soon in a future paper. We can confirm the type of correlation rAC is identified under certain conditions in MR-DoC2. Tests to check rPRS-C are underway.

Introduction

A long-standing challenge in epidemiology has been to infer causality from correlational data in observational studies. Correlational studies are starting points for exploring the causal associations between variables. However, by themselves, correlations are insufficient to identify causality, due to the potential existence of background confounding and ambiguous direction of causality.

The primary alternative to observational studies is the randomised controlled trial (RCT), in which study participants are randomly allocated to treatment/experimental and control groups. This approach averages the effects of any confounders equally among the groups, so that any difference in the outcome can be attributed to the intervention. However, RCTs are not always feasible. Difficulties may arise due to ethical considerations, e.g., when the aim is to estimate the causal effects of risk factors on disease outcomes, as it is fundamental to not harm participants.

Mendelian randomization (MR) can be used to investigate causality in cases where RCTs are infeasible or unethical. It is based on Mendel's laws of inheritance of segregation and independent assortment. Specifically, it uses the randomization that happens in meiosis, when genetic information is shuffled between chromosomes (crossing-over) and these chromosomes then form the gametes (Madole and Harden 2022). Genetic variants associated with phenotypic exposures are identified in large scale genome-wide association studies (GWAS) and meta-analyses. These genetic variants, or weighted combinations thereof, are potentially useful instrumental variables: variables correlated with the predictor, but only indirectly associated with the outcome (Evans and Davey Smith 2015; Sanderson et al. 2022).

Several key assumptions are involved in causal inference based on MR (Sanderson et al. 2022). First, the strength of the association between the genetic variant(s) and the exposure must be strong, which is known as the *relevance assumption*. Second, the variant is not associated with a confounder in the relation between the exposure and outcome, or the exchangeability assumption. Third, the effect of a genetic liability change on the exposure variable is the same as an equivalently-sized environmental liability change, i.e., they both generate the same change in the outcome variable. This assumption is known as *environmental equivalence* (Howe et al. 2022). It highlights that, unlike in twin-designs, there is no partitioning of additive genetic and environmental variances in standard MR of unrelated individuals' data. Fourth, MR is based on the assumption that the instrument's association with the outcome is completely mediated by the exposure (known as the *exclusion restriction* assumption). In genetic studies, this assumption is known as *no horizontal pleiotropy*. This assumption may not hold, given that GWAS have shown that the same variant often influences multiple traits. Some variants may affect both exposure and outcome, i.e., if the exposure acts as a mediator between the variant and the outcome then it will *not* be a case of horizontal pleiotropy (Verbanck et al. 2018; Jordan et al. 2019). Horizontal pleiotropy, the **direct effect** of the instrument on both the exposure and the outcome, violates the assumption of exclusion restriction in MR.

Several solutions have been proposed to detect and/or accommodate horizontal pleiotropy in causal inference based on MR (Sanderson et al. 2022). One can use methods that relax this assumption, and only require the instrument strength to be independent of the direct effect of the exposure on the outcome (Bowden et al. 2015), or one can triangulate results from different MR methods to confirm that the strength and direction of the causal signal are consistent over tests (Burgess et al. 2020). Furthermore, estimators based on the mode (Hartwig et al. 2017) and on the weighted median (Bowden et al. 2015) were created to address the issue of horizontal pleiotropy. Alternative methods that integrate MR in the twin-design were proposed to address the exclusion restriction assumption (Hwang et al. 2021), notably MR-DoC (Minica et al. 2018), which combines MR with the Direction of Causation (DoC) twin design (Heath et al. 1993; Duffy and Martin 1994). The MR-DoC model includes a (horizontal) pleiotropic path, accommodating a direct relationship between the instrumental variable and the outcome, and thus allowing for a test of directional horizontal pleiotropy (path b2, Figure 1B). MR-DoC was extended to accommodate bidirectional causation in the presence of background confounding (Castro-de-Araujo et al. 2023; MR-DoC2, Figure 1) by adding a polygenic score that acts as an instrumental variable for the outcome. The model thus permits estimation of the effects of reverse causation (paths b1 and b3, Figure 1C). Henceforth, we refer to the Minica et al. (2018) model as MR-DoC, and the Castro-de-Araujo et al. (2023) model as MR-DoC2.

The DoC model uses cross-twin cross-trait correlations to extract information on possible causal paths between two phenotypes. However, it is known to have the following limitations. First, differences in reliability of the variables in the model may bias causal inference estimates (Heath et al. 1993; Gillespie et al. 2003). Specifically, the more reliable variable is more likely to be identified as the cause of the less reliable variable (Heath et al. 1993; Duffy and Martin 1994). Castro-de-Araujo et al. (2023) reported that this is not an issue in MR-DoC2, but it is unknown whether it is an issue in MR-DoC. Second, both the DoC and MR-DOC models require the assumption that the unshared environmental correlation(the parameter *re* in Figure 1) is zero in order to estimate the causal path between the exposure and the outcome. This constraint implies the assumption that unshared environmental influences are not a source of confounding. Violation of this assumption biases the causal estimates in DoC models (Rasmussen et al. 2019), but it is not a problem for MR-DoC2, which explicitly models this type of confounding. For MR-DoC it is unknown if $re \neq 0$ introduces biases to the estimates of interest, in particular the causal path (g1), or other paths (Figure 1, A).

The statistical power of the MR-DoC and MR-DoC2 models has been explored in Minica et al. (2018) and in Castro-de-Araujo et al. (2023), respectively. However, a comparison of the power profiles of the three models (i.e., DoC, MR-Doc, and MR-Doc2) is lacking. While all three models focus on causal inference, the models differ with respect to their assumptions. First, the DoC model can accommodate both unidirectional and bidirectional causation, provided that some other parameters are fixed. That is, in addition to the two causal paths, only one of three possible sources (A, C, E) of confounding can be accommodated. This implies that of the three correlations, covA, covC, and covE, two have to be constrained (to zero, usually, or to another value depending on the hypotheses). In general, a bivariate ACE model is identified with any three of the five possible path coefficients (parameters covA, covC, covE, g1 and g2; Figure 1), which connect the two phenotypes, freely estimated (Maes et al. 2021). Second, the MR-DoC model is usually specified with unidirectional causation (as bidirectional causation requires further constraints to the background ACE confounding), and assumes no unshared environmental confounding (*covE*=0, Figure 1). Third, MR-DoC2 is bidirectional and assumes no direct horizontal pleiotropy (b2 path in MR-DoC is fixed to zero in MR-DoC2). However, MR-DoC2 does accommodate two sources of indirect horizontal pleiotropy (*rf*b1* and *rf*b3* in Figure 1).

In this paper, we compared the statistical power profiles within and between the three models. We estimated the effect of phenotypic measurement error, and the effect of environmental confounding in DoC and MR-DoC (covE≠0). We finally identify situations in which each model performs optimally in terms of power. The outline of this paper is as follows. First, the model specifications will be presented; second, the simulation designs will be presented; third, bias due to measurement error will be tested by introducing unreliable phenotypes; fourth, results from simulations where exclusive environmental confounding (*re*, in Figure 1) is fixed to zero, when in fact it is present in the data generation process will be presented. Finally, the models' power profiles will be reported.

Methods

Model specification

Specifications of the three models were reported in the original papers: DoC (Heath et al. 1993; Neale & Cardon 1992), MR-DoC (Minică et al. 2018), and MR-DoC2 (Castro-de-Araujo et al. 2023). All three models are in the twin-design tradition of partitioning variance into additive genetic (A), shared environmental (C), and unique environmental (E) components (Figure 1). The models are bivariate: trait1 and trait2 are the phenotypes, and they are specified such that zero, one or two causal paths may exist between the two traits (g1 or g2, or both). The more classic model is in Panel A of Figure 1, known as the Direction of Causation model (Heath et al. 1993), which is nested in both the MR-DoC and MR-DoC2 models. However, MR-DoC is not nested in MR-DoC2 due to the presence of constraints needed for identification (dropped b2 and b4 in MR-DoC2). These models were specified in OpenMx code using matrix algebra and the variance component approach (Verhulst et al., 2019). Note that in previous papers the models were specified in path coefficients (Minică et al. 2018; Castro-de-Araujo et al. 2023). Notably, in the present specification the variances are estimated, and the paths are fixed, whereas in previous specifications the variances were fixed and paths estimated. In this version, variances are allowed to be negative, which is known to improve Type I error rates in the test of variance components (Verhulst et al. 2019). This approach is also faster to run, which helps in automation in larger analyses, and is consistent with a recent paper published by our group (Maes et al. 2022). The code for each model is publicly available as a function in the *umx* R package (Bates et al. 2019).

Simulation designs

We conducted simulation studies to compare the DoC, MR-DoC and MR-DoC2 models, and to investigate known limitations of the DoC and MR-DoC twin models. We addressed three issues: (A) the effect of unmodeled phenotypic measurement error on parameter estimates; (B) the effect of misspecification of the models with respect to the parameter *covE (incorrectly constrained covE=0)*; and (C) power, by assessing the associations between the models' parameters and the non-centrality parameter in tests of null-hypotheses in each model.

We used exact data simulation to generate raw data given a population covariance matrix. Using this method, we ensure that the covariance matrix of the generated raw data exactly equals the population covariance matrix. This equality means that when fitting the true model (given that it is identified) the parameter estimates match the population values exactly. Consequently, a hypothesis test based on the likelihood ratio (e.g., fixing a parameter to zero), will produce a non-centrality parameter, which we can use in power calculations (van der Sluis et al. 2008). The method comprises the following five steps. 1) Choose values for the parameters in the model of interest. 2) Simulate multivariate normally distributed raw data based on the expected model covariance matrices and means of the monozygotic (MZ) and dizygotic (DZ) twins. To this end, we used the function *mvrnorm()* in the R library MASS with MZ and DZ sample sizes at 1000 pairs (Venables et al. 2002). The option empirical=TRUE was used to remove sampling variation in the simulated samples' means, variances and covariances. 3) Fit the true model using maximum likelihood estimation, thus recovering the true parameter values. 4) Fit a false model by imposing the constraint(s) of interest (fixing non-zero parameter to zero), and refitting the model. (5) Extract the non-centrality parameter (NCP), which equals the difference in minus twice the log-likelihood of the models fitted in steps 3 and 4, and use it to calculate the power to reject the parameter of interest given a Type I error rate of .05.

In what follows, we present a total of five simulations, with a combination of factorial designs and data generation processes. The designs are listed in Table 1, in which parameters are featured as design factors, and their values as the factor levels. The values chosen were comparable to those used in a previous publication (Castro-de-Araujo et al. 2023), which attempted to model phenotypes with broad type heritability estimates from 0.32-0.5 and strong instruments (b1, and b3) between 0.16-0.22. All simulation designs included multivariate models with A, C, and E variances. In the interest of conciseness, factor levels (values of parameters) are equal across designs. However, note that the number of parameters in each design depends on the number of parameters in each of the models (DoC, MR-DoC, and MR-DoC2), thus Design 1 has fewer parameters (g1, covA, covC, covE, ax, cx, ex, ay, cy, and ey) than Design 2, which in turn has fewer parameters (b1, b2, g1, covA, covC, covE, ax, cx, ex, ay, cy, and ey) than Design 3 (b1, b3, g1,g2, covA, covC, covE,rf, ax, cx, ex, ay, cy and ey).

The two first simulations were designed to evaluate the effects of unreliable phenotypes and the effect of unshared environmental confounding on bias for all three models (Table 1, Designs 1, 2 and 3). The last three simulations fit each model to data generated exactly according to the MR-DoC2 covariance matrix generated by the parameter values in Table 1, Design 3. The reasoning behind generating data with the more complex model, was to facilitate comparisons between the three models using the least restrictive model to generate the data.

Simulations 1 and 2 (unreliable phenotypes and environmental confounding)

We first assessed the effect of unreliability of the phenotypic measurements on the parameter estimated. To do so, we added measurement error to the exposure and to the outcome. The reliabilities of the phenotypes were set to reflect the known shortcoming of the DoC model, in which the phenotype is more likely to be identified as the cause of the less reliable phenotype, thus reliability for the exposure was set to .90 and the reliability of the outcome was .70 (Table 1). Bias stemming from the unreliability was calculated as the mean difference between the true parameter values and the parameter estimates averaged across the exact data simulations (Figure 2).

In the second simulation, we assessed the effect of unshared environmental confounding on the estimates to evaluate how the unshared environmental correlation (*covE* in Figure 1) affects the parameter estimates, given that it is fixed to zero to identify both DoC and MR-DoC. In this simulation study, we added the parameter *covE* as a factor, with levels *covE*=(0.3 and -0.3) (Designs 1, 2, and 3), and then fitted models with *covE* fixed to equal zero, thus allowing us to address the consequence of violating the assumption *covE=0*, when in truth *covE*=0.3 or *covE*=-0.3. This *covE* level was chosen to be sufficiently large to produce observable bias effects, without being unrealistically large considering that E variance includes measurement error, which does not normally contribute to *covE.* Bias was calculated and plotted in Figure 3.

MR-DoC2 as the data generating process (simulations 3, 4, and 5)

In what follows, we generated data with the exact covariance matrix of the MR-DoC2 model, and we fitted the DoC, MR-DoC, and MR-DoC2 models to the data. The factorial design 3, Table 1, was used. This procedure aimed to evaluate potential biases, given: i) the constraints added to the model (unreliability or *covE*≠0), and ii) the absence of some parameters in DoC and MR-DoC in relation to MR-DoC.

First, we assessed the consequences of unreliability using this approach, elaborating on the results of the first simulation. Data were generated with the MR-DoC2 model, but the reliability of exposure and outcome were set to 0.90 and 0.70, respectively. DoC, MR-DoC, and MR-DoC2 were then fitted to these data. Whether this resulted in any bias in the parameter estimates of DoC or MR-DoC, was assessed visually in Figure 4, where differences between parameters and estimates were plotted.

Next, we examined the effect of the violation of the assumption *covE* = 0 in DoC and MR-DoC models. An extra factor *covE* with two levels *covE* = (-0.3 and +0.3) was added to design 3 (Table 1). Exact data generation was used to simulate data under the MR-DoC2 model, and then all three models were fitted to these data. In Figure 5, bias is plotted for when unmodeled *covE* is either -0.3 or +0.3 for all three models. Notice that MR-DoC2 includes *covE*, as a freely estimated parameter, whereas it is fixed in DoC and MR-DoC. The difference in covariance when *covE* is set to -0.3 in MR-DoC is shown in Table 2 as an example of the magnitude of this bias in the covariance structure.

The final simulation study aimed at evaluating the contributions of the parameters (a_x, c_x, a_y, c_y, b_1, b_2, b_3, g_1, g_2, *covA, covC, covE, rf*) in each model on the power to reject g1=0, i.e. the causal parameter in the regression of trait 2 on trait 1 (see Figure 1). This was performed for each model (DoC, MR-DoC, and MR-DoC2). We regressed the calculated NCPs (step 5, previous section) on the parameters' true values. The resulting R^2 statistics from these regressions represent the proportions of variance in the NCP explained by all the predictors, and the coefficients, the contributions of the individual predictors to the NCP. This allows us to gauge the effect of the parameters' values on the model's power. Note that all models were fitted to data generated with MR-DoC2, which includes a second causal path (*g2*), the second instrument path (*b3*), as well as environmental background confounding (*covE*) and the correlation of the instruments (*rf*). The coefficients of these regressions were calculated and plotted as stacked bar plots in Figure 6. The hypothesis tested was of rejecting g1=0. Finally, we test the effect of distinct inheritance patterns of the two phenotypes in the power to detect the causal path (either g1 or g2) in all models (Figure 7).

All analyses were performed in R version 4.1.3 (R Core Team 2021) running on a Linux OS (Solus OS distribution version 4.3). Modelling was performed using OpenMx version 2.20.7 (Neale et al. 2016).

Results

Bias due to measurement error

Phenotypic measurement error has been shown to bias the causal parameter estimates of the DoC model (Heath et al. 1993). To assess the impact of phenotypic measurement error, we introduced unreliability to both the exposure (10%; i.e., reliability .9) and the outcome (30%; i.e., reliability .7). The estimates obtained from the simulations are shown in Figures 2 and 4. In MR-DoC the causal path (*g1*) was underestimated given unreliable phenotypic measurement (Figure 2). Measurement error did not affect the causal path estimates (*g1, g2*) in MR-DoC2 (Figure 2). Therefore, MR-DoC2 was more robust than MR-DoC when the proportion of measurement error differs between the phenotypes.

MR-DoC2 includes parameters that are not present in the other models, specifically *g2*, *covE*, *rf*, and *b3*. When the data were generated using the MR-DoC2 model (Figure 4), we found more severe bias in DoC and MR-DoC. The absence of the paths from MR-DoC2 in MR-DoC (*covE, rf, b3*) and DoC (*b1, b2*), resulted in greater bias in the models' estimates than error measurement.

Bias due to environmental confounding

It has been previously noted that the causal estimates (*g1*), as obtained in the DoC model are biased if *re* is incorrectly fixed to zero (Rasmussen et al. 2019). The effect of the misspecification with respect to *re* in MR-DoC has not been previously explored. We set up simulations two and four to examine this, i.e., how the violation of the assumption that *re* = 0 affects the parameter estimates (Figure 3, and 5). If *re* is truly positive (+0.3), the specification *re*=0 resulted in an overestimation of the causal parameter *g1* in both DoC and MR-DoC. Conversely, *g1* is underestimated when *re*=-0.3, Figure 3. However, when models were fitted to data generated with MR-DoC2, the bias in the parameter estimate was larger and more pervasive for DoC than for MR-DoC (Figure 5). As expected, the causal (*g1, g2*) and pleiotropic (*b1, b3*) paths remain unbiased in the MR-DoC2 simulation, as this is the data generating model.

Power

We compared the statistical power profiles of each model. For this step, we generated data using the MR-DoC2 model (Table 1, Design 3), and then fitted each model to these data. MR-DoC2 is a bidirectional model, but here we focus on the power to reject the hypothesis that *g1*=0, at an alpha level of 0.05 with samples of 1000 MZ and 1000 DZ twin pairs. The NCPs from this power test were regressed on the parameter values, and the coefficients for each regression were plotted as a stacked bar plot in Figure 6. The total R^2 , the proportion of NCP variance explained by the parameters equalled 0.60 in the DoC model, 0.60 in the MR-DoC model, and 0.95 in the MR-DoC2 model. The longer the bars of a given parameter, the more NCP variance the parameter explains. We found that *g1* had the largest effect on DoC and MR-DoC power; and *g1*, and *b1* had large effects on the MR-DoC2 power. Also, *covA* and *covC* had small, but noticeable, effects on the power in the DoC and MR-DoC models. Note that, in MR-DoC, the path from the instrument to the exposure (*b1*) did not contribute to the power of the model. This means that the variance explained by the instrument does not affect the power to reject the hypothesis that *g1*=0 at the .05 significance level in MR-DoC. Furthermore, all power tests performed were of the hypothesis *g1*=0 in order to make the results more comparable between models. However, it is also possible to perform a 2 df power test in MR-DoC2 dropping both *g1* and *g2*. When this was done, the parameters *g2, b3* and *rf* also showed important influence in predicting the NCP variance (not shown).

It is known that the DoC models require different proportions of A, C and E variance components between phenotypes to enable estimation of causal paths between them (Heath et al. 1993). This is explicitly addressed here with a simulation set with parameter values b1 = $\sqrt{0.05}$, b2 = $\sqrt{0.05}$, b3= $\sqrt{0.05}$, g1 = $\sqrt{0.04}$, g2 = $\sqrt{0.04}$ (remainder values according to Table 1). The ax variance was increased in a .10 step within the range (0 to .80) while ay, cx, and cy were fixed to .10 (ex = 1-(ax+cx); ey = 1-(ay+cy)). Both phenotypes have A, C and E components. We report the NCP (Figure 7) associated with the LRT with the hypothesis of the rejection of g1=0 (1 df) or g1=0, g2=0 (2 df, bottom right, yellow). We found that all three models present better power with different ACE components between phenotypes, however this characteristic is less pronounced in MR-DoC2.

Discussion

We presented a series of simulations that address issues regarding: 1) measurement error, 2) misspecification of non-shared environmental confounding, and 3) the statistical power of three models: DoC, MR-DoC, and MR-DoC2. We found that the models differ in how they are affected by issues 1 and 2, and in the role played by the instrumental variable(s) between MR-DoC and MR-DoC2. The estimates of the causal path (*g1*) were biased in the DoC and MR-DoC models when there was measurement error of the phenotypes, or when *re* was misspecified as equal to zero. We also found that the power profiles differed between the models. For MR-DoC2, *b1* and *g1* were the parameters that had the largest effects on power to reject the hypothesis that *g1*=0, whereas in the DoC and MR-DoC the key parameters were *g1*, *ra* and *rc* (Figure 6). We color-coded Figure 1 to represent these results; the paths marked in blue have biassed estimates in the case of misspecification of *re,* the ones in red contribute relatively greatly to the NCP variance (power), and the ones in orange are important to power and are biassed.

When evaluating the power profile of MR-DoC, we found that the instrument strength does not explain any variance on the NCP to reject the false hypothesis of no causation. In other words, there is no requirement of a strong instrument in MR-DoC's case. MR-DoC explicitly includes horizontal pleiotropy in the parameter *b2*. It therefore is a model that addresses this problem directly, allowing causal inference adjusted for the presence of horizontal pleiotropy. To identify the *b2* parameter it is necessary to constrain other parameters, like *rc* or *re.* It should be noted, however, that *b2* was slightly biased by unmodeled non-zero *re* (Figure 2, green bar).

These MR-DoC characteristics (no effect of the instrument in explaining the NCP variance, and the biased *b2* in the presence of *re*) raises the question of what MR-DoC provides above the classic DoC causal estimation. In Figures 2-5 it is made explicit that MR-DoC behaves very similarly to DoC in terms of biases and power. The extra *b1* and *b2* paths also do not improve *g1* estimation precision (standard errors, not shown). Thus, MR-DoC is suitable in cases where *re* correlation has a previously established accurate estimate and can be fixed, allowing for the estimation of the horizontal pleiotropy path (b2). One can use a sensitivity analysis, fixing *re* to a range of values and reporting the obtained b2 for the range. Furthermore, MR-DoC is also suitable in cases where dynastic effects and assortative mating may be present, as this model is robust to these effects (Minică et al. 2020).

Parameter estimates from the MR-DoC2 model were consistently the least biased in all tests performed. The absence of *re* and partly pleiotropic pathways like *rf* * *b1* or *rf* * *b3* in MR-DoC and in DoC biased the causal path estimate (*g1*). Another strength of MR-DoC2 is its feedback loop structure, thereby allowing inference regarding bidirectional causation. Feedback loops are frequent in nature, and most current MR methods (often based on linear regression) can only evaluate this type of relationship by running the test twice, changing the instrument in each direction (Timpson et al. 2011).

It was found that with increasing difference in proportions of A variance between the two traits, power to reject the hypothesis that the causal path g1=0 increased for all models (Figure 7). In a scenario where the E variance is the complement of A + C and both Cs are kept at 0.1 and the difference in A proportion increases in each simulation run results in progressively higher NCP. DoC and MR-DoC behaved similarly, and MR-DoC2 was slightly less affected by it. This shows that MR-DoC2 is also be more versatile in types of phenotypes that can be tested.

The tests presented also revealed an important aspect of model comparisons in SEM. Due to the non-independence of model parameters, a change to one of them will usually result in changes to other paths (Figures 4 and 5). We also converted all three models to the variance component style in order to maintain coherence with our recent publications (Verhulst et al. 2019; Maes et al. 2022). The variance component style (as opposed to the more traditional RAM specification) does not inflate the Type I error (Verhulst et al. 2019). A practical advantage of this approach is that it is faster than the RAM specification.

This study should be interpreted in the light of the following limitations. The bias analysis and the results of the power analyses of these SEM models serve only as an aid to understanding how biases may arise and the power in specific scenarios considered. Changes to a single parameter in these models leads to changes in most other paths, making comparisons and interpretation not straightforward. For example, as noted above, setting *g2=0* in the power test revealed that *b3*, *rf*, and *g2* were influential to the NCP variance.

The models presented here overcome some limitations inherent in classical MR. MR-DoC does not require a strong instrument, and bidirectional causal inference is possible in the cross-sectional case. MR-DoC includes direct horizontal pleiotropy, as does MR-DoC2 which includes indirect horizontal pleiotropy. Furthermore, the models can be extended to relatives of any type (such as siblings, for example) and when such data are available, these models offer interesting new possibilities like true bidirectional causal inference or being able to test for causality while controlling for horizontal pleiotropy.

Tables and Figures

Table 1. Parameter values in the three factorial designs, with respective total number of cells for each design simulation. See Figure 1 for the model specification. Also, e_x was specified as $1 - a_x - c_x$ and e_2 as $1 - a_y - c_y$. Parameters σ_x and σ_y are not listed, as they remained unchanged across the designs.

θ	Design 1 (DoC)	Design 2 (MR-DoC)	Design 3 (MR-DoC2)
b_1		$\sqrt{0.025}, \sqrt{0.05}$	$\sqrt{0.025}, \sqrt{0.05}$
b_2		$\sqrt{0.025}, \sqrt{0.05}$	
b_3			$\sqrt{0.025}, \sqrt{0.05}$
g_1	$\sqrt{0.02}, \sqrt{0.04}$	$\sqrt{0.02}, \sqrt{0.04}$	$\sqrt{0.02}, \sqrt{0.04}$
g_2			$\sqrt{0.02}, \sqrt{0.04}$
covA	.0,.2	.0,.2	.0,.2
covC	.0,.2	.0,.2	.0,.2
covE			.2
rf			.2
a_x	.10	.10	.10
a_y	.10	.10	.10
c_x	.10	.10	.10
c_y	.10	.10	.10

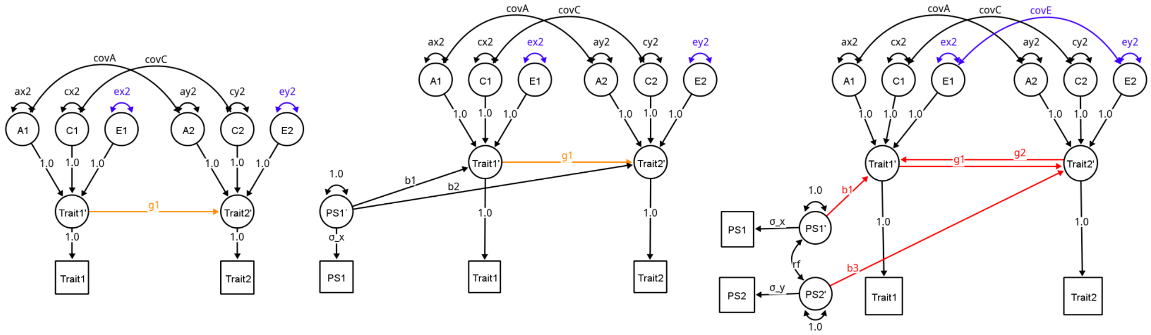
Total cells	2^3=8	2^5=32	2^6=64

Table 2:

Table 2. Change in covariance when re = 0.3 (in contrast with re = 0) for the MR-DoC model.

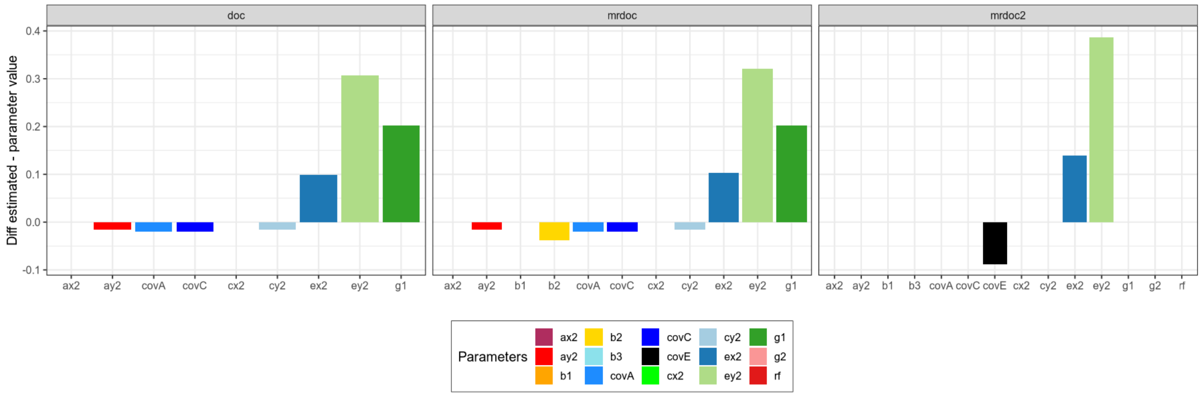
	X1	Y1	PSx1	X2	Y2
X1	0.000	0.300	0.000	0.000	0.000
Y1	0.300	0.085	0.000	0.000	0.000
PSx1	0.000	0.000	0.000	0.000	0.000
X2	0.000	0.000	0.000	0.000	0.300
Y2	0.000	0.000	0.000	0.300	0.085

Figure 1:



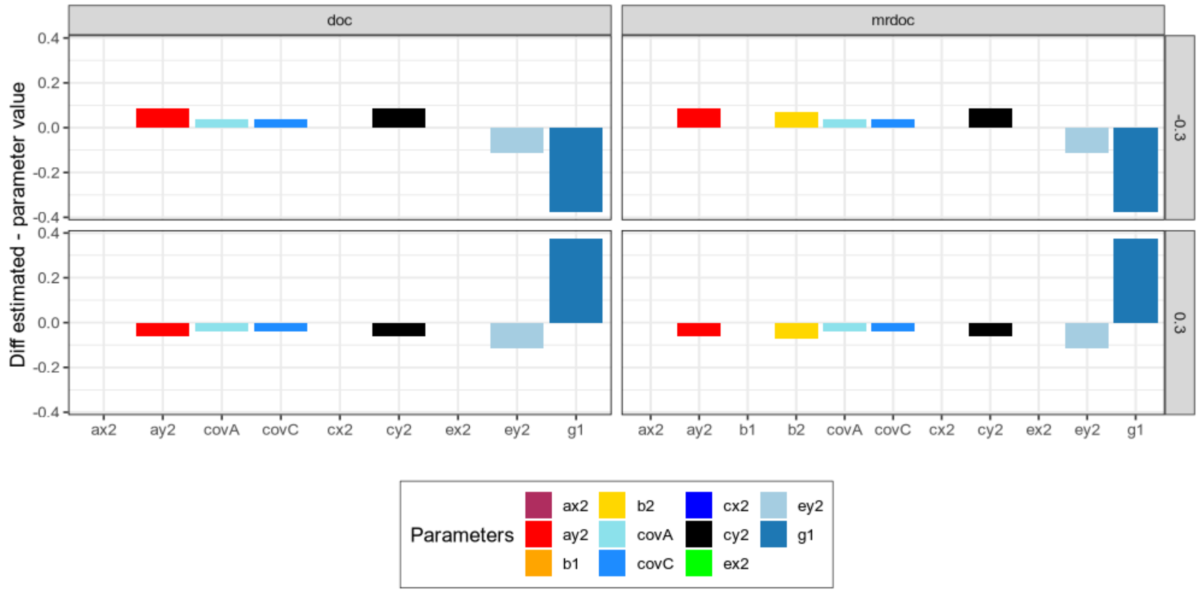
DoC (A), MR-DoC (B), and MR-DoC2 (C) model specifications for a single member of a twin pair. The genetic cross-twin correlations are 1 for MZs and 0.5 for DZs and the shared environmental variance cross-twin correlations are 1 MZs and DZs (not shown). They include the effects of additive genetic (A), common environment (C) and specific environment (E) factors for both Trait 1 and Trait 2, and their effects may correlate (parameters *covA*, *covC,* and *covE*). Path labels in red are important to the model's power, those susceptible to measurement error in blue, and in orange are those that are both susceptible to measurement error and are important to the model’s power. The latent variables Trait1' and Trait2' are not required for identification, but are kept to be coherent to Castro-de-Araujo et. al. (2023) paper, to emphasize the scaling solution of PS1' and PS2', and to point that one possible extension of these models is the use of multiple indicators.

Figure 2.



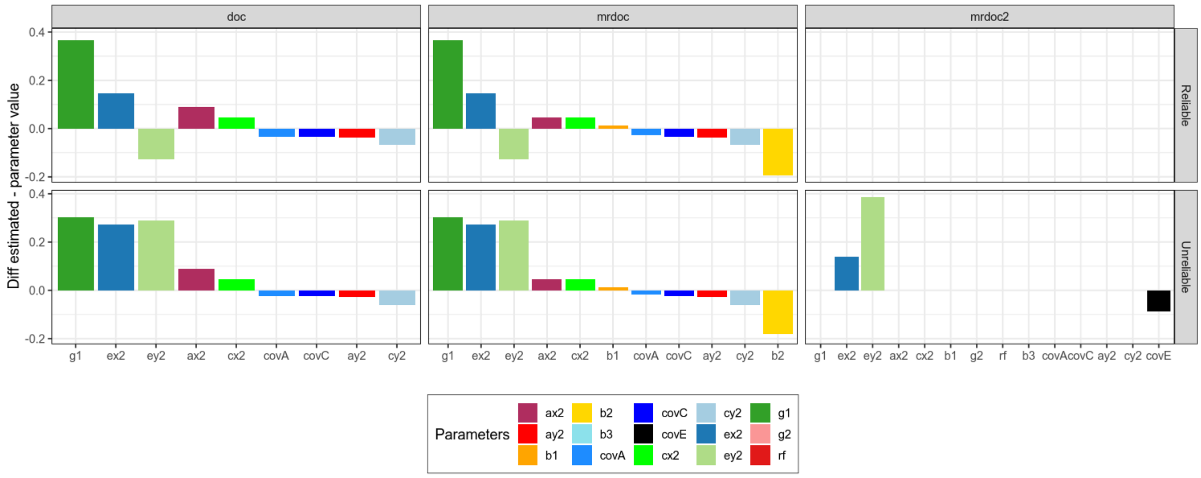
Robustness to measurement error in DoC, MR-DoC, and MR-DoC2. Error in measurement is an important source of bias for the classical DoC model and for MR-DoC. Reliability was set at 90% in the exposure (x) and 70% in the outcome (y) in an exact data simulation (Designs 1, 2, and 3; Table 1). Although the bias is not severe in either case there is overestimation of the causal path from Trait1 to Trait2 (g1) for DoC and MR-DoC models, which does not happen in the MR-DoC2 model.

Figure 3.



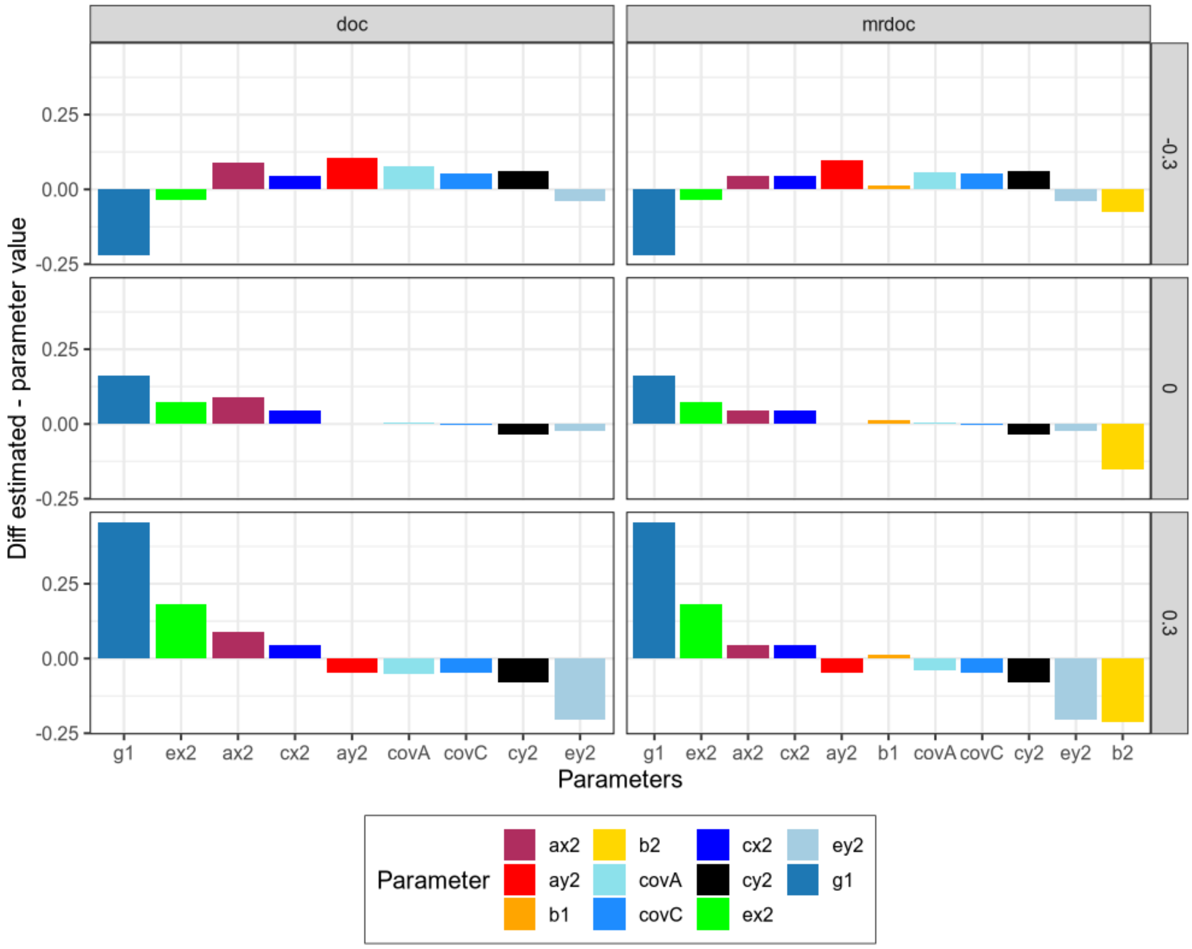
Bias due to *re* misspecification. This was based on designs 1, 2, and 3 (Table 1), and *covE* = [+0.3, -0.3] was added to the data generating process and then DoC and MR-DoC models were fitted with *covE* = 0 (and not free) . The presence of *re* introduces bias to DoC and MR-DoC notably in *g1*. No bias occurs with MR-DoC2.

Figure 4.



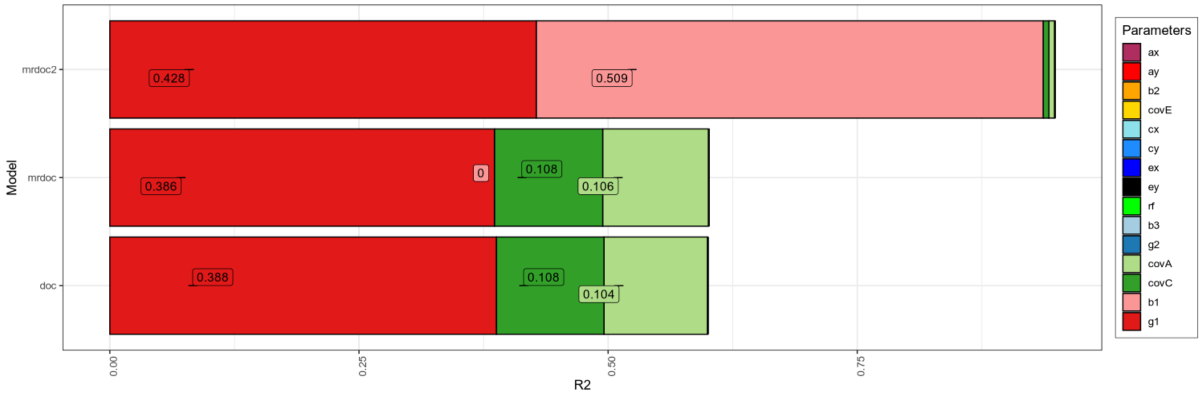
Bias due to unmodeled unreliability. This graph was based on exact data simulation with a reliability of 90% for the exposure and 70% for the outcome, and a panel with the simulation of the reliable phenotype measurement for comparison. MR-DoC2 was used as the data generating process (Design 3, Table 1) and all three models were fitted to the generated data. There are widespread biases in the DoC and in the MR-DoC estimates, with notable overestimation of g1. Contrast this with Figure 2, in which the independent exact data simulation revealed underestimation of g1 in both DoC and MR-DoC.

Figure 5.



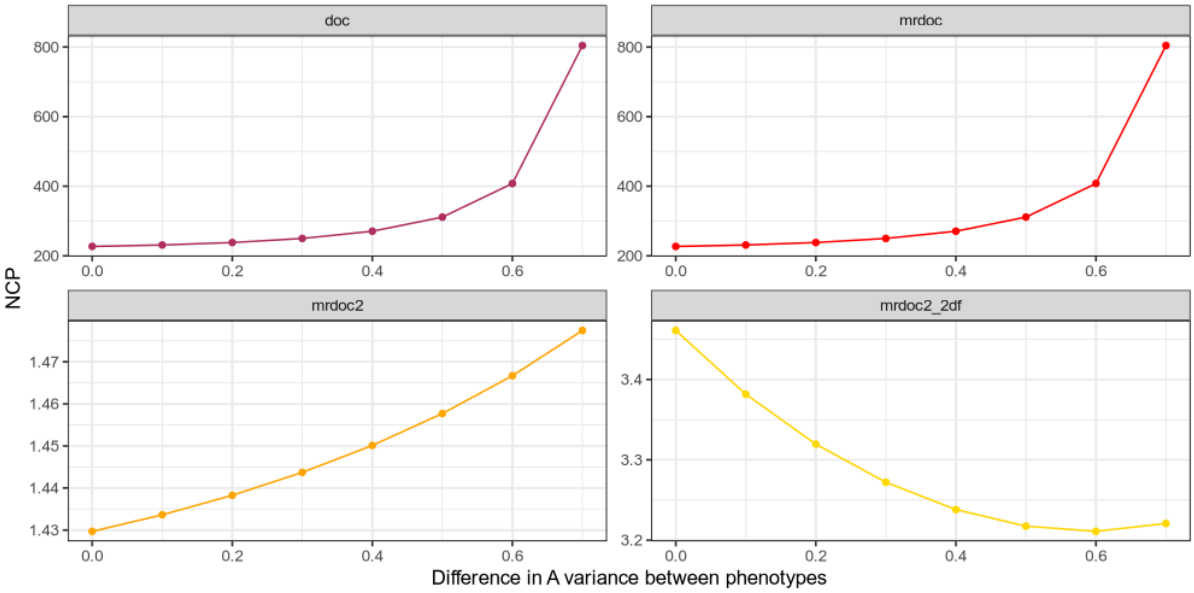
Bias due to the incorrect assumption that *covE* = 0. This was based on an exact data simulation, where MR-DoC2 was used as the data generating process (Design 3, Table 1) and each of the three models was then fit to the generated data. In the exact data simulation, an extra factor with two levels [*covE* = -0.3, +0.3] was added in the data generating process and then models were fitted with *covE* fixed at zero. The presence of *re* introduces bias to DoC and MR-DoC. No bias occurs with MR-DoC2.

Figure 6.



Variation in statistical power (non-centrality parameter; NCP), as a function of model parameter values. Using exact data simulation, the NCP was linearly regressed onto the parameter values to quantify their relative importance.MR-DoC2 was used as the data generating process (Design 3, Table 1) and all three models were then fitted to the generated data. The power test here is for the rejection of the hypothesis that *g1*=0. These are stacked bar plots with the coefficient of each predictor. The total R^2 for the DoC model was 0.60, 0.60 for MR-DoC, and 0.95 for MR-DoC2. Longer bars mean that *g1* has the largest effect on DoC and MR-DoC power, and that *g1* and *b1* have the largest effects on the MR-DoC2 power to reject g1=0.

Figure 7.



Variation in statistical power (non-centrality parameter; NCP), as a function of the difference in heritability (A variance). A simulation was set with parameter values b1 = $\sqrt{0.05}$, b2 = $\sqrt{0.05}$, b3= $\sqrt{0.05}$, g1 = $\sqrt{0.04}$, g2 = $\sqrt{0.04}$ (remainder values according to Table 1). The ax variance was increased in a .10 step within the range (0 to .80) while ay, cx, and cy was fixed to .10 (ex = 1-(ax+cx); ey = 1-(ay+cy)). The NCP plotted in the y-axis is associated with a LRT with the hypothesis of the rejection of g1=0 (1 df) or g1=0, g2=0 (2 df, bottom right, yellow).

Declarations

Funding

LFSCA is funded by NIH grant no 5T32MH-020030. MCN was funded by NIH grant DA-049867.

Conflicts of interest

Authors report no conflicts of interest

Ethics approval

Not applicable

Consent for publication

Not applicable

Availability of data and material

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Code availability

Code is available in a repository for replication.

References

Bates TC, Maes H, Neale MC (2019) umx: Twin and Path-Based Structural Equation Modeling in R. Twin Res Hum Genet 22:27–41. https://doi.org/10.1017/thg.2019.2

Bowden J, Davey Smith G, Burgess S (2015) Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 44:512–525. https://doi.org/10.1093/ije/dyv080

Burgess S, Smith GD, Davies NM, et al (2020) Guidelines for performing Mendelian randomization investigations. Wellcome Open Res. 186

Castro-de-Araujo LFS, Singh M, Zhou Y, et al (2023) MR-DoC2: Bidirectional Causal Modeling with Instrumental Variables and Data from Relatives. Behav Genet 53:63–73. https://doi.org/10.1007/s10519-022-10122-x

Duffy DL, Martin NG (1994) Inferring the direction of causation in cross-sectional twin data: theoretical and empirical considerations. Genet Epidemiol 11:483–502. https://doi.org/10.1002/gepi.1370110606

Evans DM, Davey Smith G (2015) Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality. Annu Rev Genomics Hum Genet 16:327–350. https://doi.org/10.1146/annurev-genom-090314-050016

Gillespie NA, Zhu G, Neale MC, et al (2003) Direction of Causation Modeling Between Cross-Sectional Measures of Parenting and Psychological Distress in Female Twins. Behav Genet 33:14

Hartwig FP, Davey Smith G, Bowden J (2017) Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol 46:1985–1998. https://doi.org/10.1093/ije/dyx102

Heath AC, Kessler RC, Neale MC, et al (1993) Testing hypotheses about direction of causation using cross-sectional family data. Behav Genet 23:29–50. https://doi.org/10.1007/BF01067552

Howe LJ, Tudball M, Davey Smith G, Davies NM (2022) Interpreting Mendelian-randomization estimates of the effects of categorical exposures such as disease status and educational attainment. Int J Epidemiol 51:948–957. https://doi.org/10.1093/ije/dyab208

Hwang L-D, Davies NM, Warrington NM, Evans DM (2021) Integrating Family-Based and Mendelian Randomization Designs. Cold Spring Harb Perspect Med 11:a039503. https://doi.org/10.1101/cshperspect.a039503

Jordan DM, Verbanck M, Do R (2019) HOPS: a quantitative score reveals pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases. Genome Biol 20:222. https://doi.org/10.1186/s13059-019-1844-7

Madole JW, Harden KP (2022) Building Causal Knowledge in Behavior Genetics. Behav Brain Sci 1–76. https://doi.org/10.1017/S0140525X22000681

Maes HH, Neale MC, Kirkpatrick RM, Kendler KS (2021) Using Multimodel Inference/Model Averaging to Model Causes of Covariation Between Variables in Twins. Behav Genet 51:82–96. https://doi.org/10.1007/s10519-020-10026-8

Maes HHM, Lapato DM, Schmitt JE, et al (2022) Genetic and Environmental Variation in Continuous Phenotypes in the ABCD Study®. Behav Genet. https://doi.org/10.1007/s10519-022-10123-w

Minică C, Boomsma D, Dolan CV, et al (2020) Empirical comparisons of multiple Mendelian randomization approaches in the presence of assortative mating. Int J Epidemiol. https://doi.org/10.1093/ije/dyaa013

Minică CC, Dolan CV, Boomsma DI, et al (2018) Extending Causality Tests with Genetic Instruments: An Integration of Mendelian Randomization with the Classical Twin Design. Behav Genet 48:337–349. https://doi.org/10.1007/s10519-018-9904-4

Neale MC, Hunter MD, Pritikin JN, et al (2016) OpenMx 2.0: Extended Structural Equation and Statistical Modeling. Psychometrika 81:535–549. https://doi.org/10.1007/s11336-014-9435-8

R Core Team (2021) R: A language and environment for statistical computing

Rasmussen SHR, Ludeke S, Hjelmborg JVB (2019) A Major Limitation of the Direction of Causation Model: Non-Shared Environmental Confounding. Twin Res Hum Genet Off J Int Soc Twin Stud 22:14–26. https://doi.org/10.1017/thg.2018.67

Sanderson E, Glymour MM, Holmes MV, et al (2022) Mendelian randomization | Nature Reviews Methods Primers. Nat Rev Methods Primer 2:6. https://doi.org/10.1038/s43586-021-00092-5

Timpson NJ, Nordestgaard BG, Harbord RM, et al (2011) C-reactive protein levels and body mass index: Elucidating direction of causation through reciprocal Mendelian randomization. Int J Obes 2005 35:300–308. https://doi.org/10.1038/ijo.2010.137

van der Sluis S, Dolan CV, Neale MC, Posthuma D (2008) Power Calculations Using Exact Data Simulation: A Useful Tool for Genetic Study Designs. Behav Genet 38:202–211. https://doi.org/10.1007/s10519-007-9184-x

Venables WN, Ripley BD, Venables WN (2002) Modern applied statistics with S, 4th ed.

Springer, New York

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. Nat Genet 50:693–698. https://doi.org/10.1038/s41588-018-0099-7

Verhulst B, Clark SL, Chen J, et al (2021) Clarifying the Genetic Influences on Nicotine Dependence and Quantity of Use in Cigarette Smokers. Behav Genet 51:375–384. https://doi.org/10.1007/s10519-021-10056-w

Verhulst B, Prom-Wormley E, Keller M, et al (2019) Type I Error Rates and Parameter Bias in Multivariate Behavioral Genetic Models. Behav Genet 49:99. https://doi.org/10.1007/s10519-018-9942-y