# Causal inference test can be misleading in the presence of measurement error

## Abstract

With our ability to characterize the human phenome ever improving it is becoming increasingly common to use statistical tests to infer the causal relationships between correlated variables. A simple method for inferring if an exposure is on the causal pathway to an outcome is through mediation analysis, where the effect of an instrument on the outcome is tested before and after adjusting for the exposure. We show that in the face of measurement error this can lead to erroneous results, and that increasing sample size, rather than reducing bias, can often have the alarming effect of increasing certainty in the wrong answer. We argue that because measurement error is ubiquitous in phenotypic data, that mediation-based causal inference should be treated with caution. Finally, we demonstrate that Mendelian randomization is a method for causal inference that is robust to measurement error.

## Introduction

Mediation is a commonly used statistical method for making causal inference in observational data1–4. It exists in many forms, from simple regression based systems to structural equation modeling. There has been a number of recent publications that employ mediation based approaches to make causal inference, often in large scale ‘omics datasets5–7. The principle behind mediation analysis can be explained as follows. Supposing an exposure has a causal effect on an outcome, then an ‘instrument’ that causes the exposure (a SNP, for example) should also influence the outcome. Therefore the influence of the instrument on the outcome conditional on the exposure should be zero. This forms the basis of a number of tests, such as the causal inference test (CIT).

While mediation based approaches are simple and convenient to implement, it is important to note where they may lead to unreliable results. One such mechanism is measurement error8–11. Measurement (or observational) error is the difference between a measured value of a quantity and its true value. Such variability can arise through a whole plethora of mechanisms, which are often unique to the study design and difficult to avoid12,13.

Array technology is now widely used to obtain high throughput phenotyping at low cost, but they come with the problem of having imperfect sensitivity, so for example methylation levels as measured by the Illumina450k chip are prone to have some amount of noise around the true value14,15. Sensitivity is also an issue, for example if the unit of measurement of biological interest is the methylation level in a T cell, then measurement error of this value can be introduced by using methylation levels from whole blood samples because the measured value will be an assay of many cell types.

Measurement error can indeed arise in more low-tech data too, for example when measuring body mass index one is typically interested in using this as a proxy for fat mass, but it is clear that the correlation between fat mass and obesity is not perfect. A similar problem of biological misspecification is unavoidable in disease diagnosis. Measurement error can also be introduced after the data has been collected, for example the transformation of non-normal data for the purpose of statistical analysis will lead to a new variable that will typically have both bias and imprecision compared to the original variable. The sources of measurement error are not limited to this list 13, and its impact has been explored in the context of mediation analysis in the epidemiological literature extensively 8,11. When being objective one must assume that measurement error is ubiquitous, and any measured variable is only an imperfect proxy of the biological quantity that the researcher intended to obtain. Here we show using theory and simulations how measurement error can lead to unreliable causal inference in the mediation-based CIT method.

### Methods and results

Here we consider a simple causal model where an exposure, *x*, has an effect on an outcome of interest, *y*, and that there is a known SNP, *g*, which has a direct effect on the exposure. Measurement error of an exposure can be modeled as some transformation of the true value that leads to the observed value, . For example, we can define , where and influence the bias in the measurement of , and represents the imprecision in the measurement of . Here the true value of the exposure is partially explained by the genetic instrument, *g*, such that

where is the effect of the SNP on the exposure, and is the residual value of *x*; and the outcome is partially explained by the exposure

where is the true effect of the exposure on the outcome. In the causal inference test (CIT), an omnibus p-value is generated from four hypothesis tests: 1) is associated with ; 2) is associated with ; 3) is associated with ; and 4) is independent of . The 4th condition is necessary for causal inference, and can be expressed as , where . When accounting for the possibility of measurement error we can show through simple algebra that

where

Thus an observational study will find when the true model is causal only when D = 1. Therefore, if there is any measurement error that incurs imprecision (i.e. ) then there will remain an association between and , which is in violation of the the 4th condition of the CIT. Note that measurement bias alone is insufficient to lead to a violation of the test statistic assumptions.

An alternative method to infer causality that is in principle robust to measurement error in the exposure variable, is Mendelian randomization (MR). The origins and applications of MR have been explained in detail, but the principle in which this is achieved is straightforward. Whereas mediation makes causal inference by contrasting the association of the instrument on the outcome before and after adjusting for the exposure, MR uses the instrument as a proxy for the exposure that is assumed to be devoid of measurement error and independent of potential confounders. In this setting, the problem is reformulated: instead of the researcher being agnostic about whether the instrument, *g*, has a direct influence on *x* or *y*, the researcher must select an instrument which is known to have a direct effect on *x*. Thus if *g* serves as a proxy for *x*, one can reason that the influence of *g* on *y* is the proportion of the effect of *x* on *y* that is explained by the effect of *g* on *x*. Hence the effect estimate of *x* on *y* is estimated as

which has the attractive feature that

thus we obtain an estimate of the effect of on that is unbiased by measurement error in .

The impact of measurement error on interpretation of the CIT can be shown through simulation. We generated a simple system where had a causal effect on , and is instrumented by , such that

and measurement error in was simulated such that where values of and were chosen to obtain a range of values of . We then used the CIT method to obtain a test statistic for two causal models, the true model of causing , and the false model of causing . Figure 1 shows the test statistics from each model against varying degrees of measurement error, demonstrating a number of issues. Firstly, increasing measurement increases the test statistic for the false model. Secondly, increasing measurement error decreases the test statistic for the true model, such that at a moderate level of measurement error the comparison of p-values between the two orientations will fail to differentiate the true from the false model, and as measurement error continues to increase the false model eventually has a bigger test statistic than the true model. And thirdly, increasing sample size will not solve the problem; on the contrary, it increases certainty about the erroneous causal inference.

The alternative framework that we discussed above, MR, can also be applied to these simulations. Note, here we

### Discussion

Researchers are often confronted with the problem of making causal inferences using a statistical framework on observational data, and unfortunately there are no perfect solutions. In the face of measurement error it is evident that causal inference drawn through mediation analysis will be difficult to interpret, and this is a serious problem considering that it is often impossible to estimate the extent of measurement error for most experimental designs. Mediation analysis does have its advantages though. For example, it is not necessary to have biological

In omics it’s not such a big stretch to know the direct effect

An alternative method that is robust to measurement error in inferring causal direction is Mendelian randomization16,17. Here, the influence of the exposure on the outcome is not estimated as a function of the association of the measured exposure on the measured outcome, rather it is estimated as a function of the exposure’s instrument on the measured outcome. Figure 2 demonstrates that measurement error does not lead to bias in the inference of causality when the conditions of Mendelian randomization are satisfied.

## References

1. Millstein, J., Zhang, B., Zhu, J. & Schadt, E. E. Disentangling molecular relationships with a causal inference test. *BMC Genet.* **10,** 23 (2009).

2. Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nat. Genet.* **37,** 710–717 (2005).

3. Chen, L. S., Emmert-Streib, F. & Storey, J. D. Harnessing naturally randomized transcription to infer regulatory relationships among genes. *Genome Biol.* **8,** R219 (2007).

4. Sritara, C. *et al.* Causal relationship between the AHSG gene and BMD through fetuin-A and BMI: multiple mediation analysis. *Osteoporos. Int.* **25,** 1555–62 (2014).

5. Koestler, D. C. *et al.* Integrative genomic analysis identifies epigenetic marks that mediate genetic risk for epithelial ovarian cancer. *BMC Med. Genomics* **7,** 8 (2014).

6. Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* **31,** 142–7 (2013).

7. Waszak, S. M. *et al.* Variation and genetic control of chromatin architecture in humans. *Cell* **162,** 1039–1050 (2015).

8. le Cessie, S., Debeij, J., Rosendaal, F. R., Cannegieter, S. C. & Vandenbroucke, J. P. Quantification of bias in direct effects estimates due to different types of measurement error in the mediator. *Epidemiology* **23,** 551–60 (2012).

9. Nagarajan, R. & Scutari, M. Impact of noise on molecular network inference. *PLoS One* **8,** e80735 (2013).

10. Shpitser, I., Vanderweele, T. & Robins, J. M. On the Validity of Covariate Adjustment for Estimating Causal Effects.

11. Blakely, T., McKenzie, S. & Carter, K. Misclassification of the mediator matters when estimating indirect effects. *J. Epidemiol. Community Health* **67,** 458–66 (2013).

12. Houle, D., Pélabon, C., Wagner, G. & Hansen, T. Measurement and meaning in biology. *Q. Rev. Biol.* **86,** 3–34 (2011).

13. Hernán, M. a & Cole, S. R. Invited Commentary: Causal diagrams and measurement bias. *Am. J. Epidemiol.* **170,** 959–62; discussion 963–4 (2009).

14. Harper, K. N., Peters, B. a & Gamble, M. V. Batch effects and pathway analysis: two potential perils in cancer studies involving DNA methylation array analysis. *Cancer Epidemiol. Biomarkers Prev.* **22,** 1052–60 (2013).

15. Chen, Y. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8,** 203–9 (2013).

16. Smith, G. D. & Ebrahim, S. Mendelian randomization : prospects , potentials , and limitations. **33,** 30–42 (2004).

17. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23,** 89–98 (2014).

Figure : Influence of measurement error on causal inference made by CIT

The CIT was performed on simulated variables where the exposure influenced the outcome and the exposure was instrumented by a SNP. The test statistic from CIT when testing if the exposure caused the outcome (the true model) is in red, and the test for the outcome causing the exposure (false model) is in green. Simulations were performed using sample sizes of 100, 1000 and 10000 (rows of plots). As measurement error increases (decreasing values on x-axis) the test statistic for the incorrect model gets stronger and the test statistic for the correct model gets weaker.

Macintosh HD:Users:gb13047:repo:MethylationIV:cit:images:proxy_cit_n.pdf

Figure : Influence of measurement error on Mendelian randomization

Simulations used in Figure 1 were also tested using MR. The test was parameterized to return a result of correct, incorrect, or inconclusive inference of causality. The y-axis shows the proportion of times each outcome was inferred. Measurement error does not incur bias in MR.

Macintosh HD:Users:gb13047:repo:MethylationIV:cit:images:proxy_bdmr.pdf