# The statistical power of hypothesis-free causal inference

Power - sample size, instrument effect sizes, trait correlation

Multiple testing

Pleiotropy between instruments

Measurement error of traits

Measurement error of instruments

## Abstract

The emergence of high throughput ‘omics assays is fast making the characterisation of the human phenome a reality. But as correlations between different phenotypic measures are uncovered, there remains the pervasive issue of making robust inference about their causal relationships. Here we demonstrate that implementing bi-directional Mendelian randomisation can be used to make logical inference about the causal direction between traits, or if the correlation is being caused by an unobserved confounder. We show that the power of such analyses is sufficient over a range of sample sizes when analysing low level phenotypes with large genetic effects, or high level disease traits with large-scale meta-GWAS data. We show that pleiotropy between correlated traits can lead to erroneous inferences about the causal direction between correlated traits, and we demonstrate that in most cases this can be overcome by orthogonalising the instrumental variables. Finally we show that the frequently used Causal Inference Test (CIT) is susceptible to making erroneous inferences when there is measurement error in the traits, and that BDMR is robust to this issue. In summary, BDMR can be carried out in an hypothesis-free framework to survey potentially thousands of correlations for their causal effects, but caution is advised under certain circumstances.

## Introduction

causal inference is not the business of statistics.

the objective of omics data is to be able to disentangle molecular networks from the path of genes to high level phenotypes.

statistical methods are required to assign causal directionality for the intermediate phenotypic relationships. causal inference tests are commonly used, where inference is made by

instrumental variable analysis is used to make unbiased estimates in regression equations. but it can be shown that an ordinary 2sls will not generate statistical support for the assumed direction of causality.

here we demonstrate that bi-directional 2sls can be used to make inference on direction of causality, and we discuss how it can be used in practice. included in these analyses are the accuracy of causal inference, impact of multiple testing and how to overcome problems of pleiotropy.

## Results

impact of r2

impact of zA / zB

impact of n

multiple testing

pleiotropy

hazards with causal inference test

## Methods

### Statistical power of BDMR

The objective of bi-directional Mendelian randomisation (BDMR) is to make inference on the direction of causality between two correlated traits (A and B). This method is a simple extension of standard Mendelian randomisation (MR) in that it uses an instrument (genetic predictor) for trait A and for trait B, and the null hypothesis is that the correlation between A and B is due to a confounder. In standard MR inferring the causal relationship between trait A and B is made using the following logical process:

1. If trait A causes trait B then the instrument causing trait A should also cause trait B
2. If a sufficiently powered instrument for trait A is significantly associated with trait B then this is evidence towards A causing B
3. If a sufficiently powered instrument for trait A does not significantly predict trait B then this is evidence for B causing A

There are several pitfalls to these inferences, for example point (2) does not logically distinguish between the event of A causing B and B causing A, and point (3) makes inference based upon acceptance of the null hypothesis of a statistical test, which relies on being able to distinguish between a true negative and a false negative .

If instruments are available for both traits A and B then many of these problems can be circumvented and the strength of evidence for causal inference can be improved. In BDMR causal inference is made using the following logical process:

1. If trait A causes trait B then the instrument for A should also be associated with trait B.
2. If trait A causes trait B then the instrument for B should not be associated with trait A.
3. The reverse implies that B causes A
4. Absence of significant associations of B's instrument on A or A's instrument on B means we cannot reject the null hypothesis that correlation between trait A and B is caused by an unobserved confounder.

Assuming independent instruments for traits A and B, the statistical power for BDMR is a function of the correlation between traits A and B (), the variance of *A* explained by its instrument (), the variance of B explained by its instrument (), the sample size (*n*), and the significance threshold (). Power of a statistical test is calculated as

where is the type-II error rate, is the value from a central distribution with 1 degree of freedom (df) at type-I error rate , and is the random variable from the non-central distribution with 1 df and non-centrality parameter NCP. The statistical power of a two-stage least squares (2SLS) analysis is then obtained by calculating

where is the variance explained in trait A by instrument . BDMR is performed as a series of four 2SLS tests and from these test statistics causal inference is made, as shown in Table 1.

Table : Table of equations and expectations outlining the procedure for calculating

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Stage 1 | Stage 2 | A causes B1 | B causes A1 | Confounder1 |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  |  |
| 3 |  |  |  |  |  |
| 4 |  |  |  |  |  |
|  |  | **Power** |  |  |  |

1. Expected values for each of the 2SLS coefficients under various causal models

### Measurement error simulations

In order to assess the effect of phenotypic measurement error on causal inference in BDMR or CIT, we simulated causal networks such that trait A causes trait B, but an imperfect proxy for trait *A* () was used for testing causal inference.

where is the bias of the measurement and is the imprecision of the measurement.

### Simulations for pleiotropy and genome-wide profile scores

- pleiotropy

- imperfect allele scores