Supplementary materials: Orienting the causal relationship between imprecisely measured traits using GWAS summary data

10 October 2017

## Supplementary text 1

We assume the following model

where is the exposure on the outcome , is an instrument that has a direct effect on , is the measured quantity of , where measurement error is incurred from linear transformation in and and imprecision from , and is the measured quantity of , where measurement error is incurred from linear transformation in and and imprecision from . Our objective is to estimate the expected magnitude of association between and after conditioning on . Under the CIT, this is expected to be when causes , where is the predicted value of using the measured value of .

We can split into two parts, and .

**Part 1**

**Part 2**

Simpifying further

which can be substituted back to give

Finally

thus if the measurement imprecision in is . However if there is any imprecision then the condition will not hold.

## Supplementary text 2

Assuming that either or , the causal direction can be inferred by evaluating which of and is larger in magnitude. The Steiger test is a hypothesis test that provides a p-value for observing the difference in these correlations under the null hypothesis that they are equal.

Assuming the causal direction is , two stage MR is formulated using the following regression models:

for the first stage and

where . Writing in scale free terms, denotes the correlation between and the exposure variable , and it is expected that because , where is the causal association between and (which is likely to be less than 1).

In the presence of measurement error in and , however, the empirical inference of the causal direction will instead be based on evaluating , which can be simplified:

In order to assess how reliable the inference of the causal direction is in the presence of measurement imprecision, we can evaluate the range of potential values of measurement error in and over which the empirical difference in and would return the wrong causal direction.

For different values of , and . For different values of , and .

Call the true difference in the variance explained by the genetic variant in and . If then we infer that . There will be some values of and that do not alter whether . To evaluate the reliability, , of the inference of the causal direction with regards to measurement error, the objective is to compare the proportion of the parameter space that agrees with the inferred direction against the proportion which does not:

If then the direction of causality is equally probable across the range of possible measurement error values. If then times as much of the parameter space favours the inferred direction of causality. , the total volume of the function (Figure 4), can be obtained analytically by solving:

, the proportion of the volume that lies above the plane, can also be obtained analytically. The region of this volume is bound by the values of and where , which can be expanded to . Hence,

Thus .

## Supplementary text 3

We have assumed no unmeasured confounding in these simulations. Unmeasured confounding will however have potentially large influences on mediation-based methods for inferring causal directions, and can also adversely influence the estimate of the causal direction for the Steiger test.

### Unmeasured confounding in mediation

Including an unmeasured confounder, , after ignoring intercept terms the exposure and outcome variables can be modelled as

The observational estimate of the causal effect of on , is obtained from

From this it is clear that and will differ when both and are non-zero. Relating to mediation, where we attempt to test if associates with after adjusting for , such that

and

should any amount of unmeasured confounding exist, therefore, there will remain an association between and , which will introduce errors in inferring causal directions.

### Unmeasured confounding in the MR Steiger test

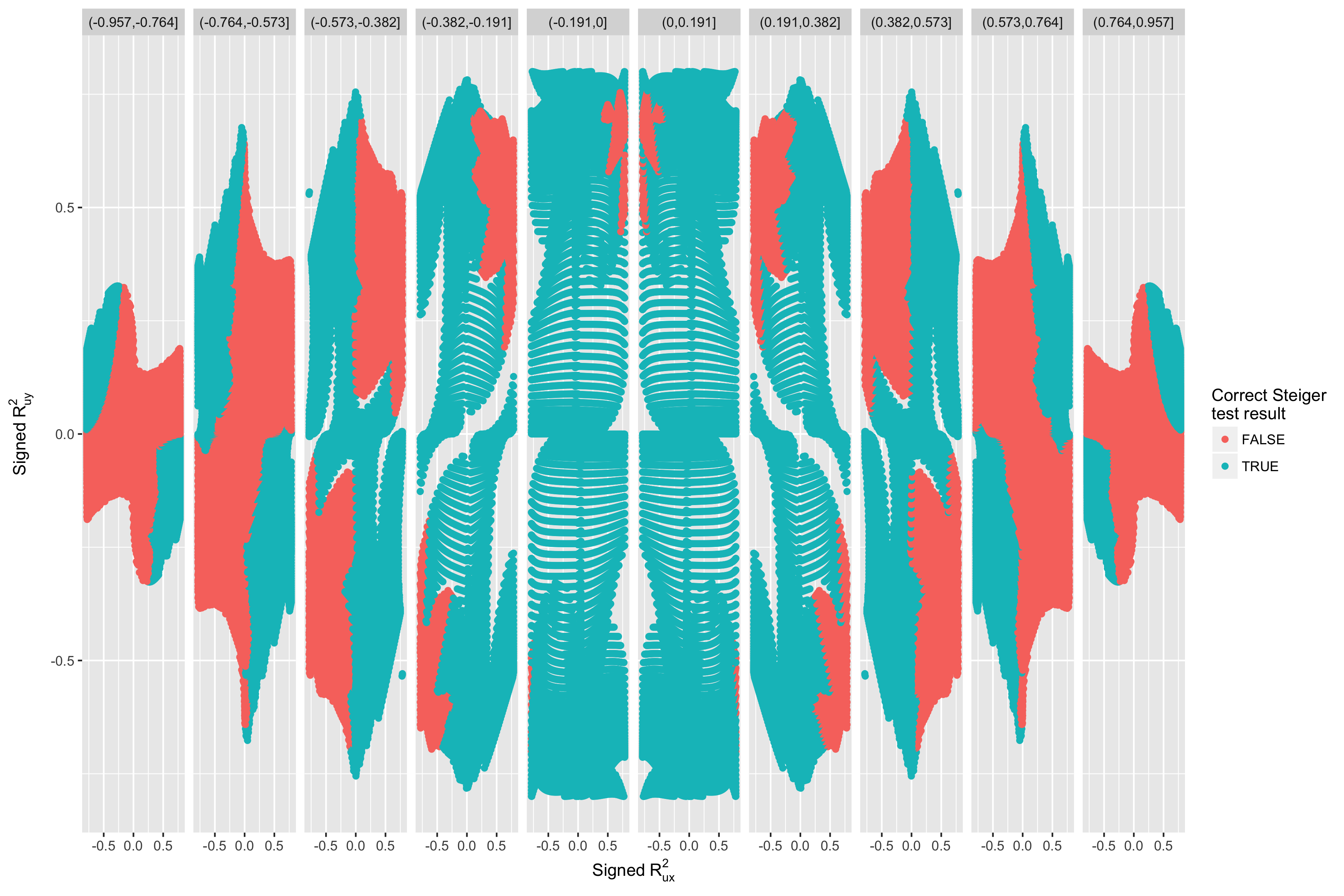
Similarly, we can investigate the extent to which unmeasured confounding will lead to the wrong causal direction between and using the MR Steiger test, evaluating the liability for the inequality being incorrect. After some algebra

and

Supplementary figure 2 shows the relationship between the magnitude of the correlations between , and the confounder for a range of , and a range of confounder effects. The pattern of results were similar for different values of . We note that in most cases for the parameter values explored, where the observational absolute is less than 0.2, unmeasured confounding will not incur the wrong causal direction in the MR Steiger test.



Supplementary figure 1: Illustrative simulations () showing the results from CIT analysis under a model of confounding. Here, the phenotypes and are not causally related, but there is a genetic effect and a confounder both influencing each phenotype. Each point represents a single simulation. Where power is high (when the absolute values of the and axes are large) the CIT returns a significant result () when testing the causal effect of on , and when testing the causal effect of on .



Supplementary figure 2: Graph representing the unmeasured confounding parameters that will lead to the MR Steiger test returning the wrong causal direction. Columns of boxes represent different signed values of the observational variance explained between and ().