Roboto

Mendelian randomization and MR-DoC

International Statistical Genetics Workshop

Luis Castro-de-Araujo

3/8/2024

Virginia Institute for Psychiatric and Behavioral Genetics &
The University of Melbourne &
Center of Data and Knowledge Integration for Health (CIDACS)

Outline

Recap, Mendelian randomization assumptions

Causal inference using twin-design

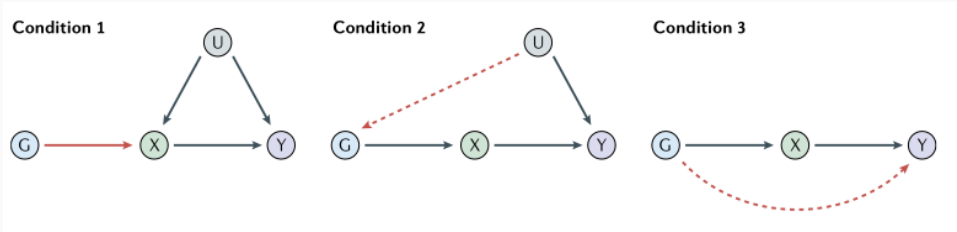
Practical

Summary and take home message

Recap, Mendelian randomization assumptions

MENDELIAN RANDOMIZATION

- Uses genetic variants as instrumental variables[1]
- Helps understand causation, but has strong assumptions[2]
  1. G (instrument) is robustly associated with X (“relevance”);
  2. G does not share common causes (C) with Y (Outcome) (“independence” or “exchangeability”); and
  3. G affects Y exclusively through its effect on X (“exclusion restriction”).



[1] Richmond et al., (2022), “Mendelian Randomization,” *Cold spring harb perspect med*.

[2] Sanderson et al., (2022), “Mendelian Randomization — Nature Reviews Methods Primers,” *Nature reviews methods primers*.

Causal inference using twin-design

Pleiotropy is pervasive

- Genetic variant influences more than one trait
- Horizontal vs Vertical pleiotropy
  - It has a central role in the genetic architecture[a]
  - Pleiotropy in over 48% of significant MR,[b] with large distortions on MR estimates.

[a] Jordan et al., (2019), "HOPS," *Genome biology*.

[b] Verbanck et al., (2018), "Detection of Widespread Horizontal Pleiotropy in Causal Relationships Inferred from Mendelian Randomization between Complex Traits and Diseases," *Nat genet*.

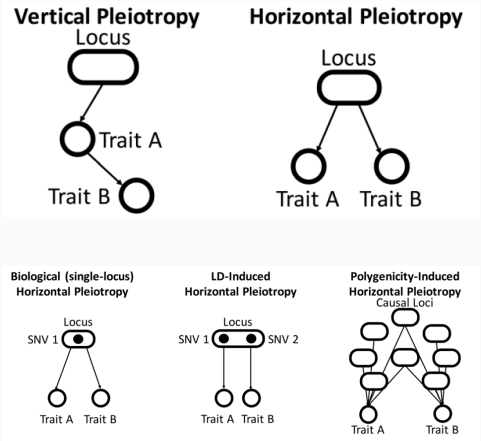


Figure 1: (LD) and polygenicity are expected to contribute to horizontal pleiotropy

Why twins?

- Access to twin data
- Triangulation of findings, confirmation of a causal relationship
- Interest in the variance components or in the background confounding elements

Structural equation modeling - equivalence to 2SLS

- SEM solutions have recovered exact estimates as 2sls, with caveats[a]
  - less convergence in weak instruments
  - slightly worse performance in ML-SEM

[a] Maydeu-Olivares et al., (2019), "Instrumental Variables Two-Stage Least Squares (2SLS) Vs. Maximum Likelihood Structural Equation Modeling of Causal Effects in Linear Regression Models," *Structural equation modeling: A multidisciplinary journal*.

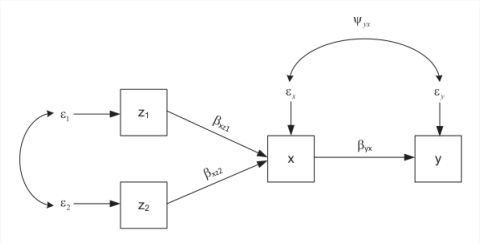


Figure 2: Instrumental Variables Regression (IVR) model that enables drawing causal inferences on the target regression mode

Direction of Causation model[3]

Model specification

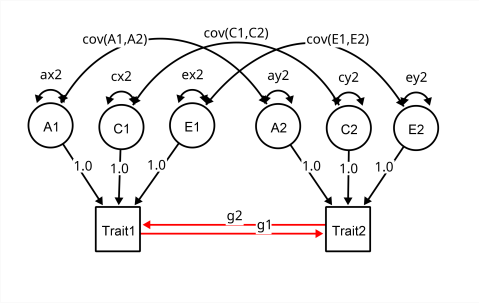


Figure 3: Path diagram representing a Bidirectional DoC for one twin.
`cov(A1,A2) = ra * sqrt(ax2 * ay2); cov(C1,C2) = rc * sqrt(cx2 * cy2); cov(E1,E2) = re * sqrt(ex2 * ey2)`

- Causal paths are estimated including information from the cross-twin cross-trait correlations.
- Cross-twin covariance between additive genetic effects is 0.5 (not shown) for DZ twins, as DZs are expected to share 50% of the genetic effects.
- Standard SEM symbology is used.

[3] Neale et al., *Methodology for Genetic Studies of Twins and Families* (New York, NY, US 1992).

Direction of Causation

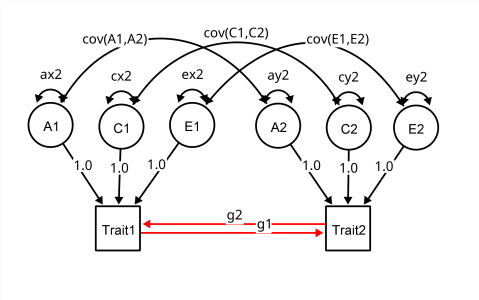


Figure 4: Path diagram representing a Bidirectional DoC for one twin.
`cov(A1,A2) = ra * sqrt(ax2 * ay2); cov(C1,C2) = rc * sqrt(cx2 * cy2); cov(E1,E2) = re * sqrt(ex2 * ey2)`

- Bias at the phenotypic level[a]
- Bias due to lack of unmodelled E confounding[b]
- Better detection of causal paths with different variance component proportions for each phenotype[c]

[a] Gillespie et al., (2003), "Direction of Causation Modeling Between Cross-Sectional Measures of Parenting and Psychological Distress in Female Twins," *Behav genet*.

[b] Rasmussen et al., (2019), "A Major Limitation of the Direction of Causation Model," *Twin res hum genet*.

[c] Maes et al., (2021), "Using Multimodel Inference/Model Averaging to Model Causes of Covariation Between Variables in Twins," *Behav genet*.

MR-DoC model[4]

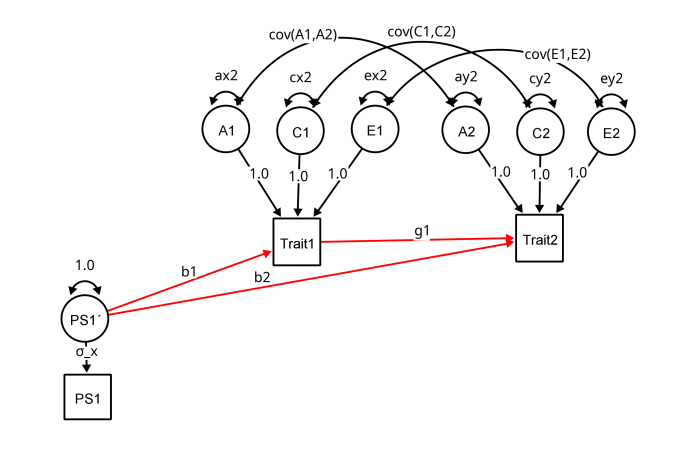


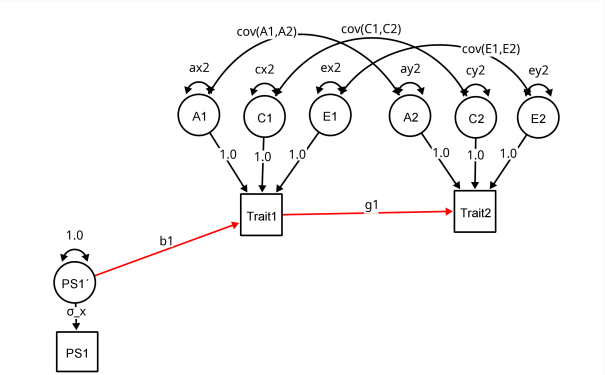
Figure 5: Path diagram for one twin. `cov(A1,A2) = ra * sqrt(ax2 * ay2); cov(C1,C2) = rc * sqrt(cx2 * cy2); cov(E1,E2) = re * sqrt(ex2 * ey2)`

[4] Minică et al., (2018), "Extending Causality Tests with Genetic Instruments," *Behav genet*.

x	aX	cX	eX	aY	cY	eY	ra	rc	re	b1	b2	g1	Id
fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	No
fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	fr	Yes
fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	Yes
fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	fr	fr	Yes
fr	fr	0	fr	fr	0	fr	fr	0	fr	fr	fr	fr	No
fr	fr	fr	fr	fr	0	fr	fr	0	fr	fr	fr	fr	Yes
fr	fr	0	fr	fr	fr	fr	fr	0	fr	fr	fr	fr	No

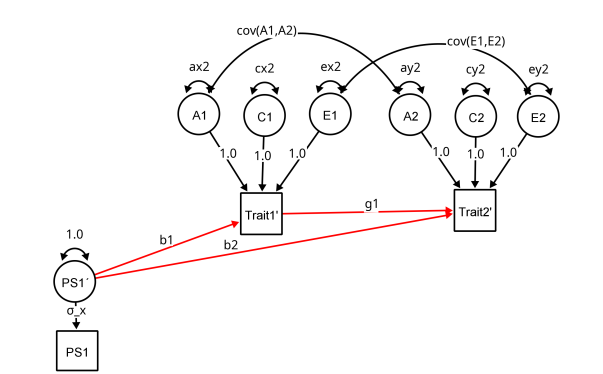
MR-DoC - IDENTIFIED CASES

x	aX	cX	eX	aY	cY	eY	ra	rc	re	b1	b2	g1	Id
fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	No
fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	fr	Yes
fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	Yes
fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	fr	fr	Yes
fr	fr	0	fr	fr	0	fr	fr	0	fr	fr	fr	fr	No
fr	fr	fr	fr	fr	0	fr	fr	0	fr	fr	fr	fr	Yes
fr	fr	0	fr	fr	fr	fr	fr	0	fr	fr	fr	fr	No



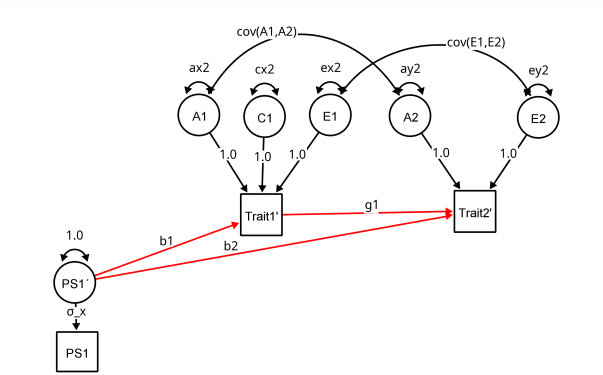
MR-DoC - IDENTIFIED CASES

x	aX	cX	eX	aY	cY	eY	ra	rc	re	b1	b2	g1	Id
fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	No
fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	fr	Yes
fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	Yes
fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	fr	fr	Yes
fr	fr	0	fr	fr	0	fr	fr	0	fr	fr	fr	fr	No
fr	fr	fr	fr	fr	0	fr	fr	0	fr	fr	fr	fr	Yes
fr	fr	0	fr	fr	fr	fr	fr	0	fr	fr	fr	fr	No



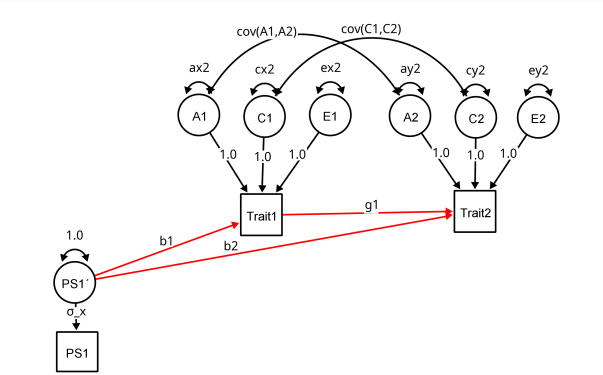
MR-DoC - IDENTIFIED CASES

x	aX	cX	eX	aY	cY	eY	ra	rc	re	b1	b2	g1	Id
fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	No
fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	fr	Yes
fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	Yes
fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	fr	fr	Yes
fr	fr	0	fr	fr	0	fr	fr	0	fr	fr	fr	fr	No
fr	fr	fr	fr	fr	0	fr	fr	0	fr	fr	fr	fr	Yes
fr	fr	0	fr	fr	fr	fr	fr	0	fr	fr	fr	fr	No

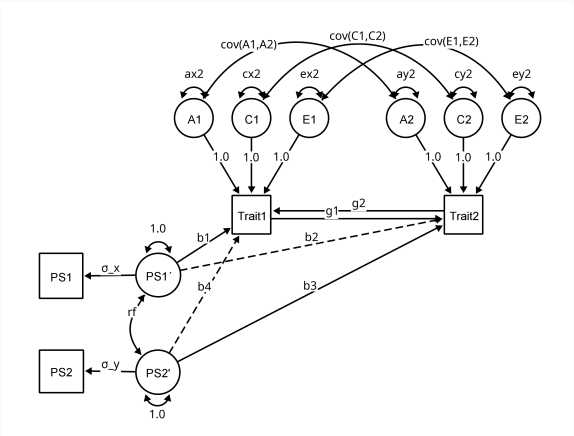


MR-DoC - IDENTIFIED CASES

x	aX	cX	eX	aY	cY	eY	ra	rc	re	b1	b2	g1	Id
fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	No
fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	fr	Yes
fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	Yes
fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	fr	fr	Yes
fr	fr	0	fr	fr	0	fr	fr	0	fr	fr	fr	fr	No
fr	fr	fr	fr	fr	0	fr	fr	0	fr	fr	fr	fr	Yes
fr	fr	0	fr	fr	fr	fr	fr	0	fr	fr	fr	fr	No



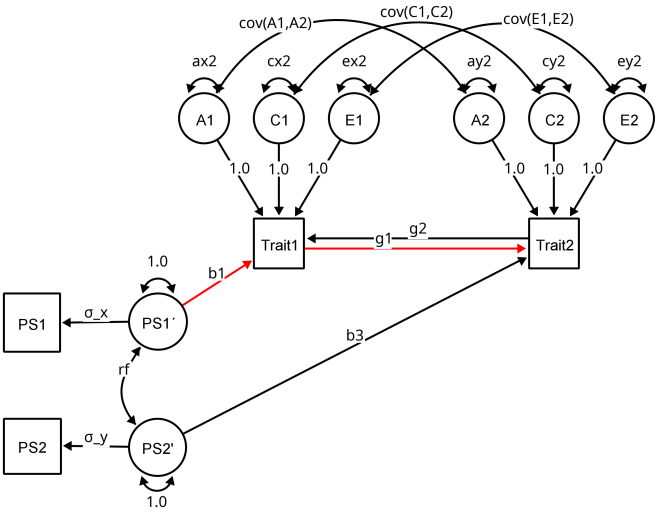
Model specification



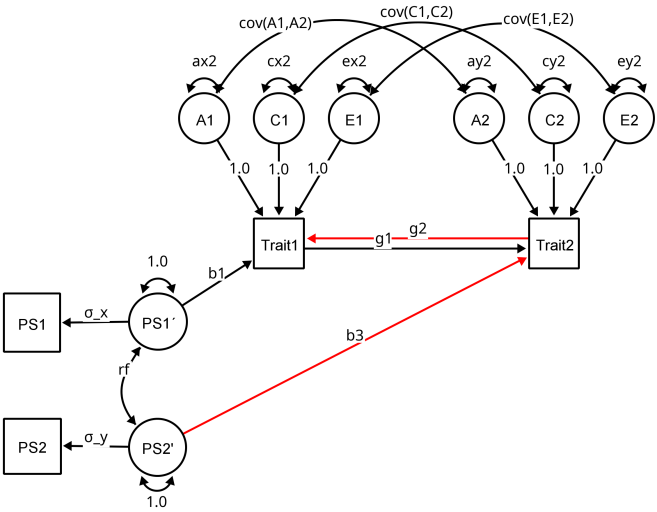
- Path diagram of the MR-DoC2 model for an individual.[a]
- The model includes the effects of additive genetic (A), common environment (C) and unique environment (E) factors for both X and Y, and their effects may correlate.

[a] Castro-de-Araujo et al., (2023), "MR-DoC2," *Behav genet*.

MR-DoC2 IN MORE DETAIL



MR-DoC2 IN MORE DETAIL



Comments on limitations and strengths[5]

- MR-DoC g1 is biased if modeled re = 0 (or cov(E1, E2) = 0), when re is not equal 0 in the data.
- MR-DoC2 requires very large sample sizes to detect g1, g2
- MR-DoC2 g1, g2 unbiased in the presence of measurement error

[5] Castro-de-Araujo et al., "Power, Measurement Error, and Pleiotropy Robustness in Twin-Design Extensions to Mendelian Randomization," Available at `https://www.researchsquare.com` (accessed November 6, 2023).

Practical

Copying files

```
# Make sure you are in your home folder by typing:
pwd

# Create a directory to hold today's work by typing:
mkdir mrdoc

# Change your directory to the new folder:
cd mrdoc

# Copy files, don't forget the dot at the end
cp /faculty/luis/2024/mrdoc/* .
```

Beta matrix

```
BE <- mxMatrix(name = "BE", type = "Full",nrow=3,  ncol = 3, byrow = TRUE,
          labels = c(NA,   "g2", "b1",
                     "g1", NA,   "b2",
                     NA,   NA,   NA),
          free = c(FALSE, FALSE, TRUE,
                   TRUE,  FALSE, TRUE,
                   FALSE, FALSE, FALSE),
          dimnames = list(c("X", "Y", "iX"),
                          c("X", "Y", "iX")))
```

A, C, E variances

```
# ACE decomposition
A <-  mxMatrix(name = 'A', type='Symm', nrow=3,ncol = 3,byrow = TRUE,
           labels=c("ax2",  "covA", NA,
                    "covA", "ay2",  NA,
                    NA,     NA,     "sigma_x"),
           free=c(TRUE, TRUE, FALSE,
                  TRUE, TRUE, FALSE,
                  FALSE,FALSE,TRUE),
           dimnames = list(c("X", "Y", "iX"),
                           c("X", "Y", "iX")))
```

A filter matrix is required to trop PRSs from twin 2 in MZs

```
# The filter matrix is used to remove the PRS from MZs,
# if we kept it the matrix would become redundant resulting
# in Hessian not positive
filter <- mxMatrix(name = 'filter', type='Full', nrow=5, ncol=6, free=FALSE,
          byrow = TRUE,
          values=c(1,0,0,0,0,0,
                   0,1,0,0,0,0,
                   0,0,1,0,0,0,
                   0,0,0,1,0,0,
                   0,0,0,0,1,0))
```

Practical - Implementation notes

- In the script you will find the pipe operator from R base `|>`, this is done to achieve concise blocks of code while still being expressive.
- The idea is that the resulting object from one instruction is passed on to the next instruction. Example:

```
unrel4 <- unrel3 |>
  omxSetParameters(name = "no_causal",  label="g1", free = FALSE,
                           values = 0) |>
  mxRun()
```

- In the case above, `omxSetParameters()` returns a model with modified parameters and this model is then passed to `mxRun()`, which returns the same model after estimation to the object `unrel4` using the assignment operator `<-`.

Copying files

```
# Make sure you are in your home folder by typing:
pwd

# Create a directory to hold today's work by typing:
mkdir mrdoc

# Change your directory to the TwinFacMod folder:
cd mrdoc

# Copy files, don't forget the dot at the end
cp /faculty/luis/2024/mrdoc/* .
```

Aims:

1. How to setup a simulation in OpenMx using mxGenerateData
2. How to specify a two-stage least squares test in OpenMx
3. Check that MR-DoC recovers the same estimates from 2sls test
4. Check that in the presence of horizontal pleiotropy estimates are biased in the 2sls test

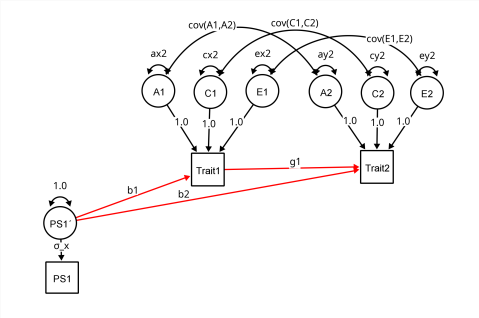
The script will exemplify four scenarios:

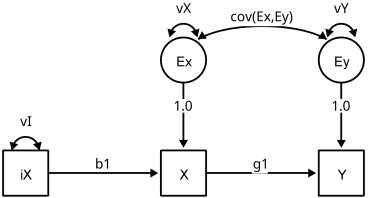
- no pleiotropy between the instrument and the outcome (Scenario 1),
- where there is pleiotropy (Scenario 2),
- presence of a bidirectional causal effect (Scenario 3),
- absence of a causal effect (Scenario 4)

Copying files

```
# Make sure you are in your home folder by typing:
pwd

# Create a directory to hold today's work by typing:
mkdir mrdoc

# Change your directory to the TwinFacMod folder:
cd mrdoc

# Copy files, don't forget the dot at the end
cp /faculty/luis/2024/mrdoc/* .
```

PRACTICAL


cov(A1,A2)
cov(C1,C2)
cov(E1,E2)
ax2
cx2
ex2
ay2
cy2
ey2
A1
C1
E1
A2
C2
E2
1.0
Trait1
Trait2
g1
b1
b2
PS1
o_x
PS1


vX
cov(Ex,Ey)
vY
Ex
Ey
1.0
vI
iX
b1
X
g1
Y

SUMMARY AND TAKE HOME MESSAGE

Aims:

1. How to setup a simulation in OpenMx using mxGenerateData
2. How to specify a two-stage least squares test in OpenMx
3. Check that MR-DoC recovers the same estimates from 2sls test
4. Check that in the presence of horizontal pleiotropy estimates are biased in the 2sls test

The script will exemplify four scenarios:

- no pleiotropy between the instrument and the outcome (Scenario 1),
- where there is pleiotropy (Scenario 2),
- presence of a bidirectional causal effect (Scenario 3),
- absence of a causal effect (Scenario 4)