Handling Survey Mode Effects in the UK Cohort Studies

Liam Wright, Georgia Tomova, Richard Silverwood

27 February 2025

CENTRE FOR LONGITUDINAL STUDIES

Economic and Social Research Council

1

Mixed Mode Survey Designs

CENTRE FOR LONGITUDINAL STUDIES

- Each new medium of communication begets a new way to conduct surveys
  - Face-to-face interview, postal questionnaire, telephone interview, web questionnaire, video interview.
- There are many ways to mix modes
  - Precise combination of modes
  - *Within*-sweep or *between*-sweep
  - How participants are assigned to modes

2

Benefits and Costs of Mixing Modes

CENTRE FOR LONGITUDINAL STUDIES

- Benefits
  - Cheaper to run (more N for your £).
  - (Potentially) higher response rates
  - More diverse (representative?) samples
- But …
  - Modes differ in how items are measured
  - This can cause bias

3

An Example Survey Item

CENTRE FOR LONGITUDINAL STUDIES

Please think back over the last 12 months about how your health has been.

Compared to people of your own age, would you say that your health has on the whole been…

1. …Excellent
2. Good
3. Fair
4. Poor, or
5. Very Poor?

4

Bias Can Affect All Types of Analysis

CENTRE FOR LONGITUDINAL STUDIES

- Descriptive
- Longitudinal
- Correlational

5

About CLS

CENTRE FOR LONGITUDINAL STUDIES

- The Centre for Longitudinal Studies is home to several national longitudinal cohort studies, which follow the lives of tens of thousands of people
- By collecting information from the same people over time, as they live their lives, our studies are powerful resources for answering important research questions
- Multi-purpose and multi-disciplinary studies collecting detailed information on different aspects of cohort members lives (economic, social and health)

6

IOE, Faculty of Education and Society

UCL

CENTRE FOR LONGITUDINAL STUDIES

Handling Mode Effects in the CLS Cohort Studies

User Guide

November 2024

CENTRE FOR LONGITUDINAL STUDIES

UKRI Economic and Social Research Council

7

Guidance Content

CENTRE FOR LONGITUDINAL STUDIES

1. An exhaustive list of mixed mode elements within CLS' studies
2. Review of item characteristics that generate mode effects
3. DAGs showing (many) situations in which DAGs cause bias
4. Some empirical estimates of mode effects and mode selection
5. Discussion of methods for handling mode effects (with worked examples in R and Stata)
6. Recommendations for accounting for mode effects

8

Outline of Webinar

CENTRE FOR LONGITUDINAL STUDIES

- 12:30 – 12:50: Background on Mode Effects (LW)
- 12:50 – 13:15: Conceptualising Mode Effects with DAGs (GT)
- 13:15 – 13:40: Methods for Handling Mode Effects (RJS)
- 13:40 – 13:45: Our Recommendations (LW)
- 13:45 – 14:00: Q & A

- Ask questions at any point in the chat box!

9

Background

CENTRE FOR LONGITUDINAL STUDIES

Economic and Social Research Council

10

Mixed Mode Elements in CLS' Cohorts

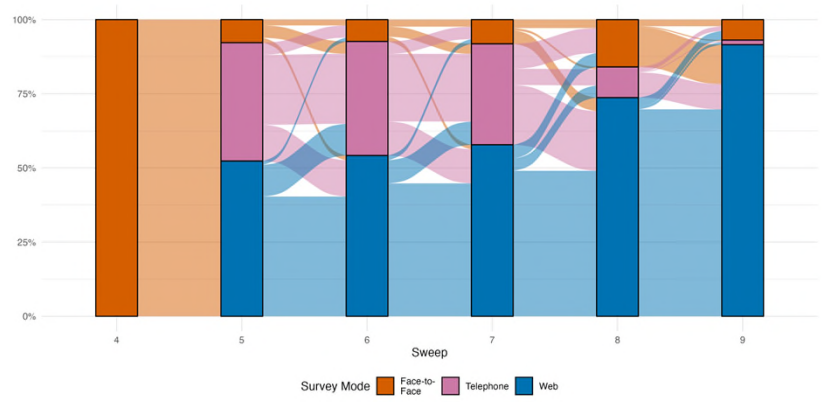
CENTRE FOR LONGITUDINAL STUDIES

1. Each study has mixed mode elements
2. Almost always, selection into mode is non-random
3. In mainstage and sub-studies
4. *Between-* and *within-person* mixing of modes
   - *Between-study* mixed of modes, too.

11

Mode Switching in Next Steps

CENTRE FOR LONGITUDINAL STUDIES



12

What Generates a Mode Effect?

CENTRE FOR LONGITUDINAL STUDIES

- d'Ardenne et al. (2017) highlight three (overlapping) factors:
  1. Interviewer Effects
     - Social desirability
  2. Satisficing (Doing 'Enough')
     - Motivation
     - Provision of explanatory information
  3. Question and Answer Presentation
     - Recency and primacy
     - Repeated responses and scales with mid-points.
- Note, mixed mode surveys are often designed to reduce mode effects.

13

Some Considerations

CENTRE FOR LONGITUDINAL STUDIES

Mode effects can be…

1. Contextual, rather than an inherent property of the item
2. On the *distribution* of responses
3. On the probability of answering an item
4. Difficult to predict direction and/or size *a priori*
5. Present or not, depending on the item

14

An Example Survey Item

CENTRE FOR LONGITUDINAL STUDIES

Please think back over the last 12 months about how your health has been.

Compared to people of your own age, would you say that your health has on the whole been…

1. …Excellent
2. Good
3. Fair
4. Poor, or
5. Very Poor?

15

Another Example Survey Item

CENTRE FOR LONGITUDINAL STUDIES

Do you have any children?

Please include any adopted children, step-children or foster children of whom you consider yourself to be a parent in addition to your own biological children. Please also include children who do not currently live with you.

1. Yes
2. No

16

Empirical Evidence on Mode Effects

CENTRE FOR LONGITUDINAL STUDIES

- Several studies have estimated mode effects experimentally
- Reviewing a handful of major surveys, we concluded:
  - Mode effects are not a feature of every survey item (negative controls!)
  - Social desirability appears to be the main motivator.
  - Even so, mode effects on sample means are usually small (~ 0.3 SD or so maximum)
  - Some evidence of recency / primary effects too.
- Larger systematic review underway

17

Evidence of Mode Selection

CENTRE FOR LONGITUDINAL STUDIES

- We review factors related to selection into mode in Next Steps (Sweeps 5-8) and NCDS sequential mixed mode designs.
- Broadly, factors related to responding to later offered modes are similar to those you'd expect for non-response in general
  - Lower cognitive ability
  - More disadvantaged socioeconomic status
  - Poorer health
- But controlling for these factors does not seem to fully account for mode selection.

18

Some More Considerations

CENTRE FOR LONGITUDINAL STUDIES

- Selection depends on the particulars of the survey design.
  - Read the technical reports!
- Mode effects potentially heterogeneous according to population
- Mode effects also depend on the survey design particulars
- Bias depends on the type of analysis and the strength of mode effects and selection

19

UCL

Conceptualising Mode Effects using DAGs

CENTRE FOR LONGITUDINAL STUDIES

UKRI Economic and Social Research Council

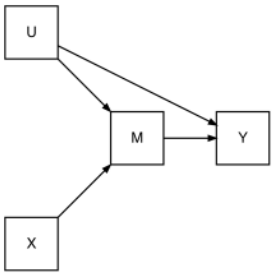
20

Directed acyclic graphs (DAGs)

CENTRE FOR LONGITUDINAL STUDIES

DAGs are **graphical representations** of **causal relationships** between variables

They encode our assumptions about the data generating process



variables == nodes
relationships == arcs/edges

e.g. we can state:
- U and M are causes of Y;
- If we intervene on U, then the probability/expectation of Y will change

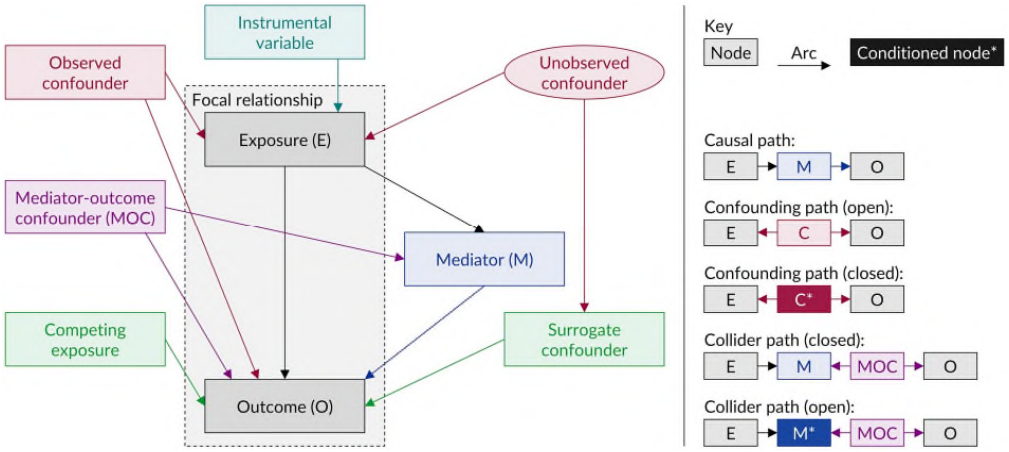
Key features of DAGs:

- directed
- acyclic
- non-parametric

21

Directed acyclic graphs (DAGs): variable roles and paths

CENTRE FOR LONGITUDINAL STUDIES



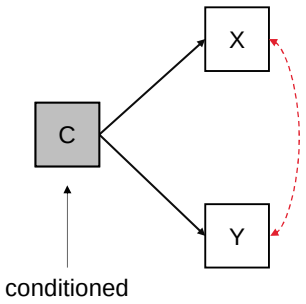
Tennant et al. 2021. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol*

22

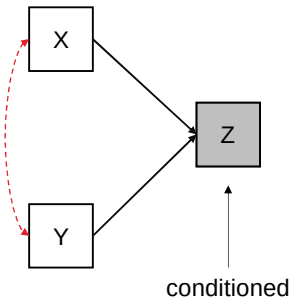
Directed acyclic graphs (DAGs): key types of bias

CENTRE FOR LONGITUDINAL STUDIES

Confounding bias



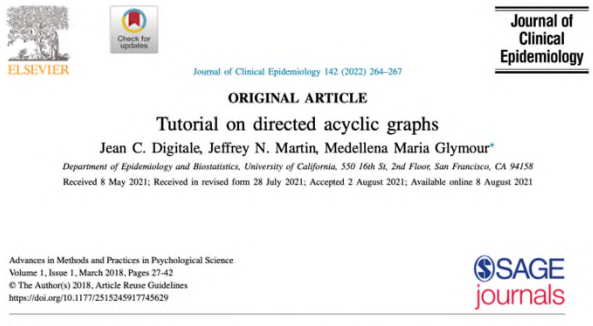
Collider bias



23

Directed acyclic graphs (DAGs): resources

CENTRE FOR LONGITUDINAL STUDIES


ELSEVIER

Journal of Clinical Epidemiology 142 (2022) 264–267

Journal of Clinical Epidemiology

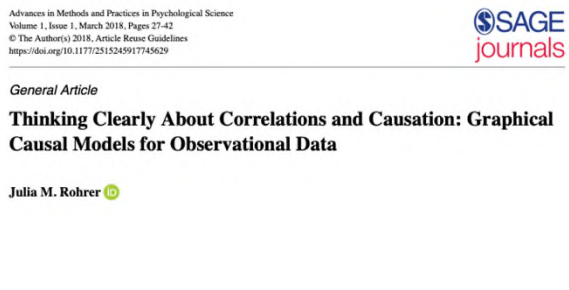
ORIGINAL ARTICLE

Tutorial on directed acyclic graphs

Jean C. Digitale, Jeffrey N. Martin, Medellena Maria Glymour*

Department of Epidemiology and Biostatistics, University of California, 550 16th St, 2nd Floor, San Francisco, CA 94158

Received 8 May 2021; Received in revised form 28 July 2021; Accepted 2 August 2021

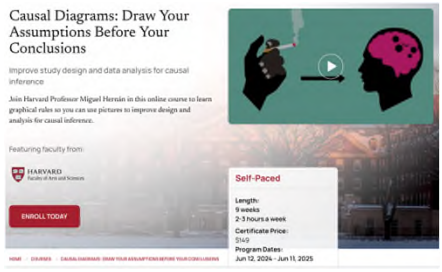

Advances in Methods and Practices in Psychological Science
Volume 1, Issue 1, March 2018, Pages 27-42
© The Author(s) 2018, Article Reuse Guidelines
https://doi.org/10.1177/2515245917745629

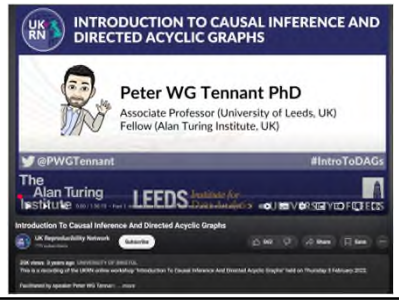
SAGE journals

General Article

Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data

Julia M. Rohrer





24

DAGs for methodological challenges

CENTRE FOR LONGITUDINAL STUDIES

- DAGs are often considered in the context of applied research
- However, they are extremely useful for understanding and explaining methodological challenges

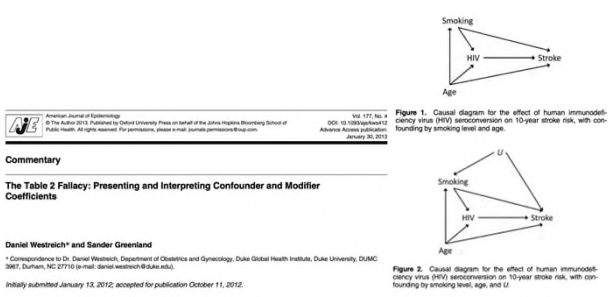
Practice of Epidemiology

The Birth Weight "Paradox" Uncovered?

Sonia Hernández-Díaz[1,2], Enrique F. Schisterman[3], and Miguel A. Hernán[1]

[1] Department of Epidemiology, Harvard School of Public Health, Boston, MA.
[2] Slone Epidemiology Center, Boston University, Boston, MA.
[3] Epidemiology Branch, National Institute of Child Health and Human Development, Bethesda, MD.

Received for publication February 7, 2005; accepted for publication January 23, 2006.



Commentary

The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients

Daniel Westreich* and Sander Greenland

* Correspondence to Dr. Daniel Westreich, Department of Obstetrics and Gynecology, Duke Global Health Institute, Duke University, DUMC 3967, Durham, NC 27710 (e-mail: daniel.westreich@duke.edu).

Initially submitted January 13, 2012; accepted for publication October 11, 2012.

Figure 1. Causal diagram for the effect of human immunodeficiency virus (HIV) seroconversion on 10-year stroke risk, with confounding by smoking level and age.

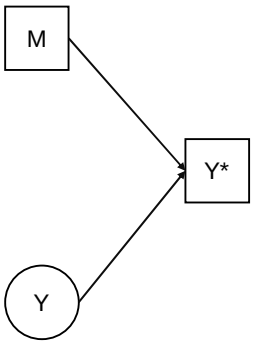
Figure 2. Causal diagram for the effect of human immunodeficiency virus (HIV) seroconversion on 10-year stroke risk, with confounding by smoking level, age, and U.

25

Depicting mode effects in DAGs: variable measurement

CENTRE FOR LONGITUDINAL STUDIES

- M represents survey mode
  typically binary, e.g. web and telephone
- Y is a variable of interest, the true value of which is unknown (i.e. latent)
  e.g. alcohol intake
- Y* is the measure of Y obtained via a survey



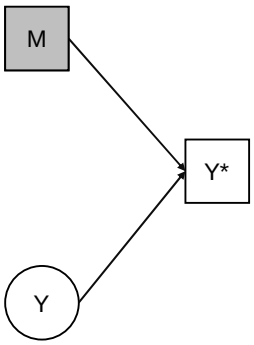
26

Depicting mode effects in DAGs: variable measurement

CENTRE FOR LONGITUDINAL STUDIES

- M is a source of variation (or error) in Y*
- Conditioning on M will remove the variation in Y* that is due to M
- This might make conditioning on M seem like an obvious generic solution to mode effects…

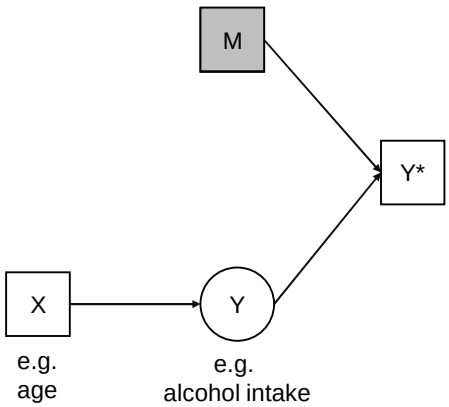
  But the reality is a lot more complex.

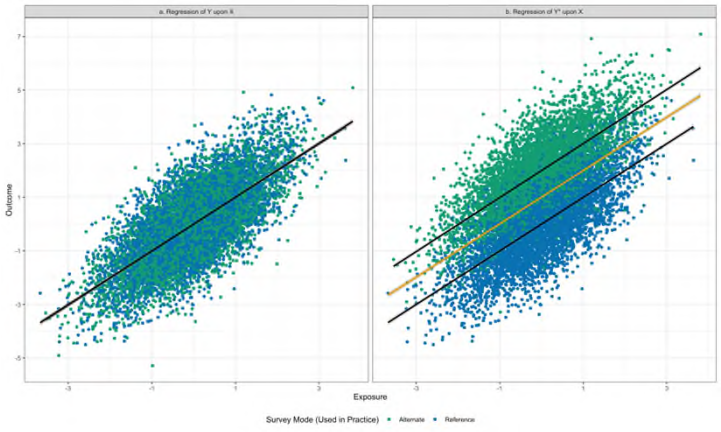


27

Depicting mode effects in DAGs: outcome mode effect

CENTRE FOR LONGITUDINAL STUDIES

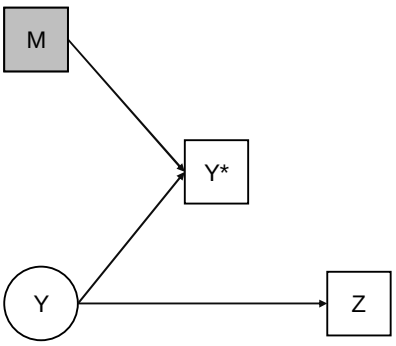




28

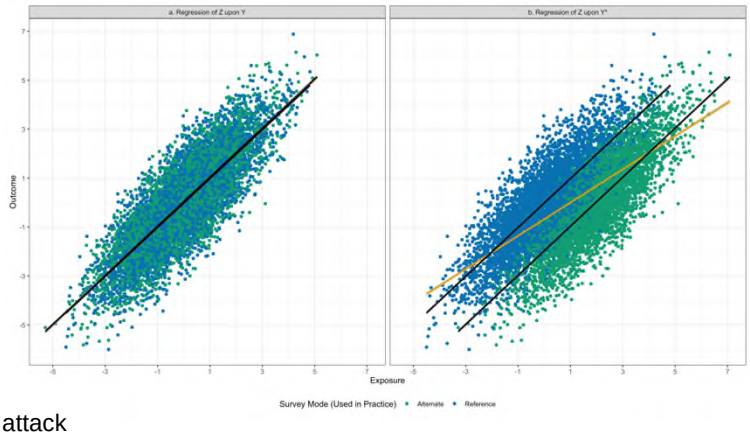
Depicting mode effects in DAGs: exposure mode effect

CENTRE FOR LONGITUDINAL STUDIES



e.g.
alcohol intake

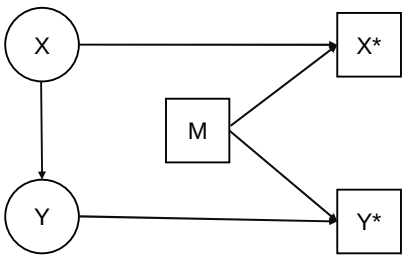
e.g.
history of heart attack



29

Depicting mode effects in DAGs: exposure and outcome mode effects

CENTRE FOR LONGITUDINAL STUDIES



Same concepts apply. Adjustment for mode will:

- reduce error introduced by mode effect in the outcome
- reduce bias introduced by mode effect in the exposure

However…

30

Depicting mode effects in DAGs: selection into mode

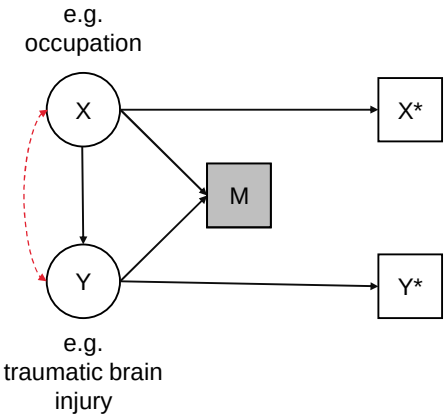
CENTRE FOR LONGITUDINAL STUDIES

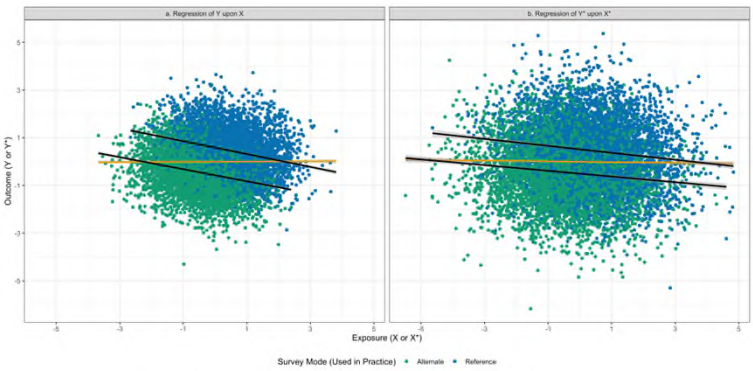
- So far, we have only considered mode as an exogenous source of variation in a variable.
- In reality, this is unlikely to be the case.
- It may be more plausible that:
  - availability/offer of mode may affect selection into the study;
  - relevant variables of interest (e.g. exposure, outcome) may affect selection into a mode.

31

Depicting mode effects in DAGs: selection into mode

CENTRE FOR LONGITUDINAL STUDIES

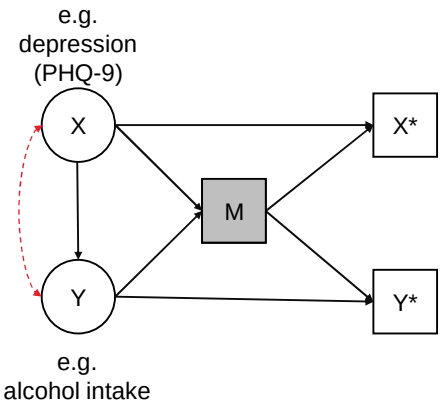


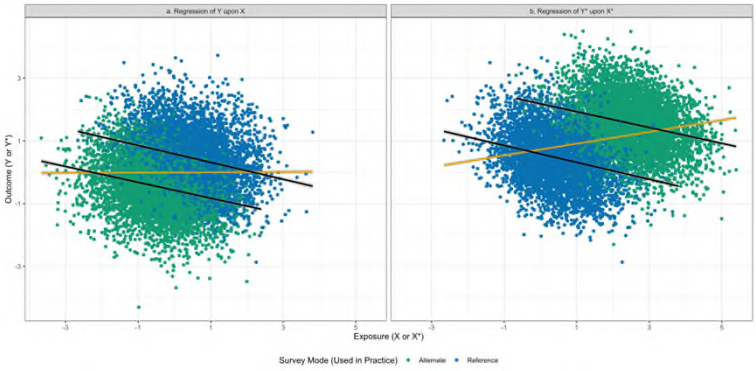


32

CENTRE FOR LONGITUDINAL STUDIES

Depicting mode effects in DAGs: mode effects and selection into mode

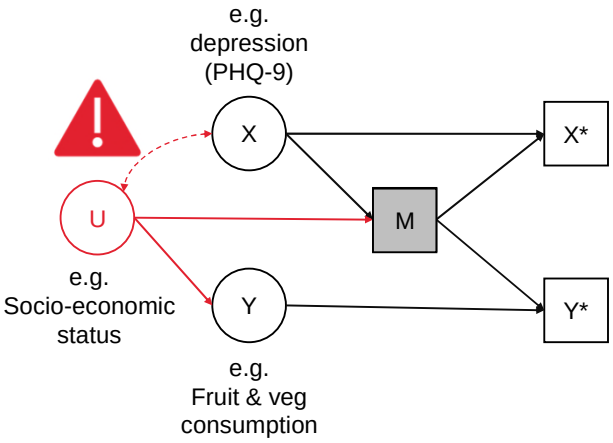




33

CENTRE FOR LONGITUDINAL STUDIES

Depicting mode effects in DAGs: mode effects and selection into mode



34

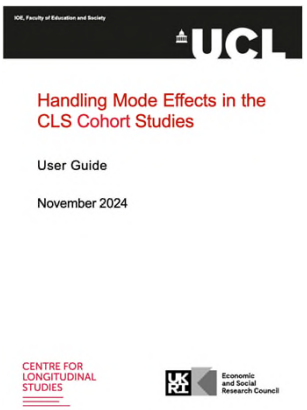
CENTRE FOR LONGITUDINAL STUDIES

Depicting mode effects in DAGs

These principles extend to other scenarios and examples that may exist

e.g. longitudinal analyses, mediation, moderation, etc.

User guide contains more comprehensive examples and explanations



35

CENTRE FOR LONGITUDINAL STUDIES

Caveats

- Whether there are mode effects AND/OR selection into mode will be **context-specific**
  - There may exist some literature to help guide you towards this
  - DAGs can help you illustrate any assumptions you make
- *How much* bias may be introduced is also context-specific – both mode effects and selection can be weak or strong

36

CENTRE FOR LONGITUDINAL STUDIES

Key messages

- Both the presence and strength of any mode effects or mode selection will be entirely context-dependent
- To best guide your analyses, it is important to understand the underlying theory
- Where selection into/according to mode occurs, conditioning on mode risks introducing spurious associations from collider bias, that may be larger than any benefit gained.
- Where mode effects occur without selection into mode, it may be appropriate/safe to simply condition on mode.
  - But this is a very strong assumption to make! When in doubt, assume that selection is at play and consider alternative methods…

37

UCL

Methods for Handling Mode Effects

CENTRE FOR LONGITUDINAL STUDIES

UKRI Economic and Social Research Council

38

(Proposed) Methods for Handling Mode Effects

CENTRE FOR LONGITUDINAL STUDIES

1. Statistical Control
2. Multiple Imputation
3. Sensitivity Analysis
4. Worked Example

39

Statistical Control – Overview

CENTRE FOR LONGITUDINAL STUDIES

- Where the aim is to obtain an estimate of association, it may be possible to account for mode effects using control variables.
- Where there is no relevant mode selection, this could simply involve adding an indicator variable for mode to the substantive model (or, alternatively, stratifying by mode).
- Can instead estimate the mode effect and using this to predict counterfactuals for those observed in the alternate mode.

40

Statistical Control – Overview

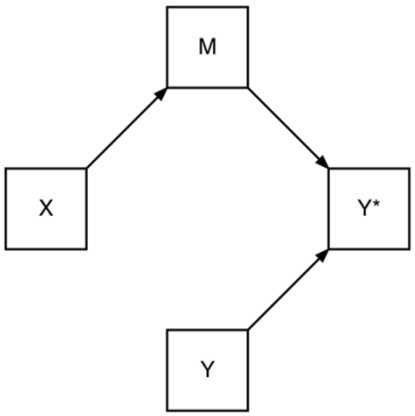
CENTRE FOR LONGITUDINAL STUDIES

- However, where there is mode selection according to a relevant variable, accounting for mode alone may not be sufficient and could even increase bias.
- In this case, a larger set of control variables may be required – it is possible that this set of control variables does not or could not exist in the data (e.g. if Y is both subject to mode effects and a source of mode selection).

41

Statistical Control – Examples

CENTRE FOR LONGITUDINAL STUDIES

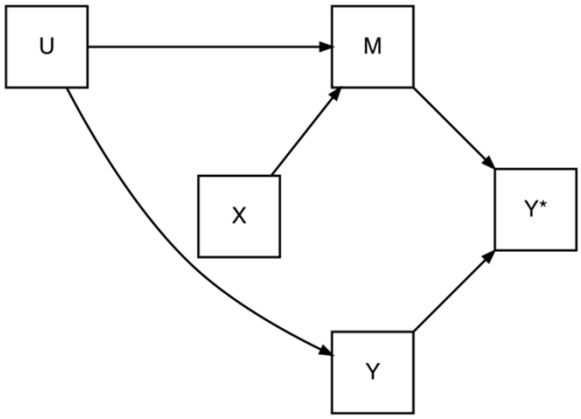


Regress Y* on X **controlling for M**.

42

Statistical Control – Examples

CENTRE FOR LONGITUDINAL STUDIES

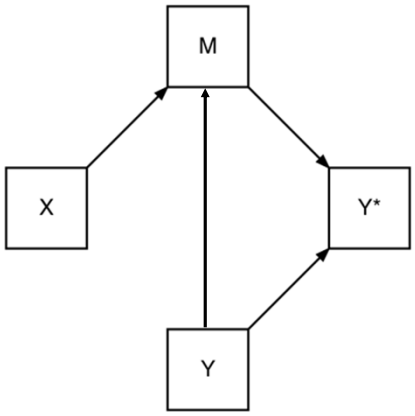


Regress Y* on X controlling for M **and also U** (if observed).

43

Statistical Control – Examples

CENTRE FOR LONGITUDINAL STUDIES



Regression of Y* on X biased whether controlling for M or not.

44

Statistical Control

CENTRE FOR LONGITUDINAL STUDIES

Advantages	Disadvantages
• Straightforward method, easily understood and implemented. • Given the richness of variables captured in long-running cohort studies, the required set of control variables (or something sufficiently approximating it) may be available.	• Strong assumption that mode selection correctly accounted for. • Required set of control variables may be unknown, unmeasured, or poorly measured, meaning bias persists. • Adjusting for causes of mode selection may change the interpretation of the estimate being produced.

45

Multiple Imputation – Overview

CENTRE FOR LONGITUDINAL STUDIES

- Values of variables hypothesized to exhibit mode effects are artificially set to missing for individuals in the alternate survey mode(s).
- Predictive models are developed based on data from those in the reference survey mode.
- Predictive models applied to data for those in the alternate survey mode to generate counterfactuals.
- 'Completed' dataset then used to provide descriptive statistics or analysed in substantive regression models.

46

Multiple Imputation – Overview

CENTRE FOR LONGITUDINAL STUDIES

- Multiple imputed datasets generated by this procedure.
- Each imputed dataset analysed using the substantive model then estimates pooled to obtain standard errors that account for uncertainty inherent in the imputation process.

47

Multiple Imputation – Example

CENTRE FOR LONGITUDINAL STUDIES

- [illegible] web (which was anonymous).

id	mode	sum	gender	sc	health	ed	cog
1	tel	[illegible]	[illegible]	[illegible]	2	[illegible]	[illegible]
2	tel	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3	web	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	web	[illegible]	[illegible]	[illegible]	4	[illegible]	[illegible]
5	tel	[illegible]	1	[illegible]	[illegible]	3	[illegible]
…	…	…	…	…	…	…	…

48

Multiple Imputation

CENTRE FOR LONGITUDINAL STUDIES

Advantages	Disadvantages
• Increasingly commonly used so may already be familiar to researchers. • Easy-to-use functionality in major statistical programming languages. • Straightforward to implement for a wide variety of variable types. • Can combine with MI for missing data handling.	• Does not use information from the observed values in the alternate mode(s) – potentially very wasteful. • Strong assumption that data are 'missing at random' (MAR): 'missingness' is independent of the (counterfactual) value of Y, conditional on the covariates used to generate the imputed values.

49

Sensitivity Analysis – Overview

CENTRE FOR LONGITUDINAL STUDIES

- Broad area of 'quantitative bias analysis'.
- Can be approached in different ways – here focus on record-level sensitivity analysis.
- The size of the mode effect is assumed and is used to simulate a counterfactual response for those in the alternate survey mode(s).
- Substantive models are then run using these simulated data and compared against substantive models using observed values to examine whether, and to what extent, results change.

50

Sensitivity Analysis – Overview

CENTRE FOR LONGITUDINAL STUDIES

How to choose mode effects for sensitivity analysis?

1. Choose single plausible mode effect to obtain 'best estimate'.
2. Fine net of mode effects can be tried and sensitivity of findings explored (usually graphically).
3. Intentionally choose implausible mode effect.

51

Sensitivity Analysis – Example

CENTRE FOR LONGITUDINAL STUDIES

- [illegible] (no mode effect) and 2 units (large mode effect) to explore if/when the substantive conclusions change.

id	mode	sum	sum5	sum01	…	sum19	sum20
1	tel	12	13	10	.2.	93	11
2	tel	15	13	14	.4.	17	17
3	web	13	13	13	.2.	13	14
4	web	22	25	25	.4.	22	22
5	tel	10	13	9	.2.	83	8
…	…	…	…	…	…	…	…

52

Sensitivity Analysis

CENTRE FOR LONGITUDINAL STUDIES

Advantages	Disadvantages
• Mode effects are assumed rather than estimated, so detailed understanding of mode selection not required. • Use all available information, unlike MI. • Extremely flexible approach, e.g. heterogeneity in mode effects, multiple variables subject to mode effects, mixing modes between sweeps.	• Plausible mode effects for a given situation may not be known. • Estimation of mode effects from non-experimental data (if necessary) requires appropriate modelling of mode selection. • No out-of-the-box functionality for performing general sensitivity analysis.

53

Worked Example: NCDS Sweep 9

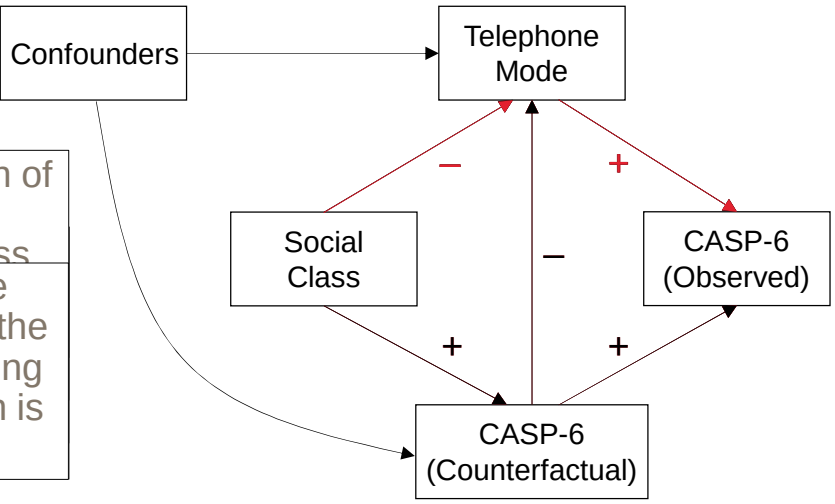
CENTRE FOR LONGITUDINAL STUDIES

- Evidence that individuals in more advantaged socioeconomic positions have better health on average.
- Interested in the (cross-sectional) association between social class (professional vs. not) and quality of life (QoL; CASP-6 score) in NCDS Sweep 9 (age 55 years).

54

Worked Example: NCDS Sweep 9

CENTRE FOR LONGITUDINAL STUDIES



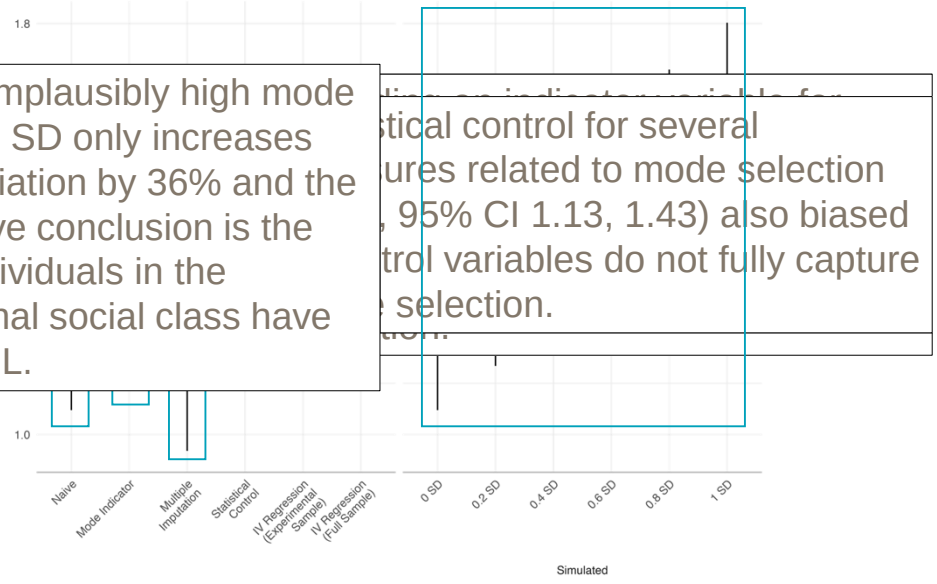
A simple regression of observed CASP-6 score on social class …

… and the negative association due to the mode effect, meaning that the association is understated.

55

Worked Example: NCDS Sweep 9

CENTRE FOR LONGITUDINAL STUDIES



Even an implausibly high mode effect of 1 SD only increases the association by 36% and the substantive conclusion is the same: individuals in the professional social class have higher QoL.

…stical control for several …ures related to mode selection …, 95% CI 1.13, 1.43) also biased …trol variables do not fully capture … selection.

56

Recommendations

CENTRE FOR LONGITUDINAL STUDIES

Economic and Social Research Council

57

First Steps

CENTRE FOR LONGITUDINAL STUDIES

1. Investigate where a variable is likely to suffer mode effects, *a priori*
2. Determine the likely size of the mode effect based on previous literature
3. Draw out assumptions on mode effects and selection unsing DAGs

58

Analysis

CENTRE FOR LONGITUDINAL STUDIES

4. Report descriptive statistics on survey mode
5. Run a naïve analysis not accounting for mode
6. Run substantive analysis accounting for mode (if possible)
7. Run a sensitivity analysis positing values for mode effect

59

Reporting

CENTRE FOR LONGITUDINAL STUDIES

8. Report results of sensitivity and other analyses
9. Discuss mode effects in strengths and limitations sections

60

Watch Out For…

CENTRE FOR LONGITUDINAL STUDIES

1. Systematic review of experimental mode effects estimates (with searchable database)
2. Tutorial on sensitivity analysis for handling mode effects
3. Empirical analysis examining the efficacy of statistical control
4. Empirical analysis of whether mode effects materially bias existing studies

61

UCL

Thank you!

Further Questions? liam.wright@ucl.ac.uk

CENTRE FOR LONGITUDINAL STUDIES

UKRI Economic and Social Research Council

62