ROYAL STATISTICAL SOCIETY

DATA | EVIDENCE | DECISIONS

Handling survey mode effects

12 November 2025

Outline

CENTRE FOR
LONGITUDINAL
STUDIES

Viewing survey mode effects through the lens of causal directed acyclic graphs	Richard Silverwood (University College London)
Methods for handling survey mode effects	Liam Wright (University College London)
Break	
A systematic review of the experimental literature on mode effects	Georgia Tomova (University College London)
Adaptive mixed-mode survey design	Barry Schouten (Statistics Netherlands)

Please use chat throughout for any questions.

Viewing mode effects through the lens of causal directed acyclic graphs

Liam Wright, Georgia Tomova, Richard Silverwood

CENTRE FOR LONGITUDINAL STUDIES

RSS Event: Handling survey mode effects
12 November 2025

Economic and Social Research Council

Outline

CENTRE FOR LONGITUDINAL STUDIES

1. Background
2. Quick introduction to DAGs
3. Representing mode effects in DAGs

CENTRE FOR LONGITUDINAL STUDIES

Background

Background

CENTRE FOR LONGITUDINAL STUDIES

Many ways to collect survey data (modes):

- Face-to-face interview
- Postal questionnaire
- Telephone interview
- Web questionnaire
- Video interview

These are frequently used in combination (mixed-mode design):

- Within-sweep
- Between-sweep

Background

CENTRE FOR LONGITUDINAL STUDIES

Benefits of mixed-mode:

- Cheaper to run
- (Potentially) higher response rates
- More diverse (representative?) samples

But…

- Modes necessarily differ in how items are measured
- This can influence responses and potentially cause bias

What generates a mode effect?

CENTRE FOR LONGITUDINAL STUDIES

Interviewer effects

- Social desirability

Satisficing (doing ‘enough’)

- Explanation/additional information
- Motivation

Question and answer presentations

- Primacy and recency
- Repeated responses

Background

CENTRE FOR LONGITUDINAL STUDIES

Mode (measurement) effects: Differences in responses between modes due to how items are measured.

Mode selection: Differences in who is being measured under each mode.

Background

We want…

- To understand whether, how and why our analysis may be biased
- To clarify any steps that could be followed to remove bias
- To communicate this efficiently to others

Try causal DAGs?

CENTRE FOR LONGITUDINAL STUDIES

Quick introduction to DAGs

What is a causal DAG?

CENTRE FOR LONGITUDINAL STUDIES

- Causal directed acyclic graphs (DAGs) are graphical representations of causal models.
- Variables are represented by nodes.
- Causal relationships are represented by directed arcs (arrows).

What is a causal DAG?

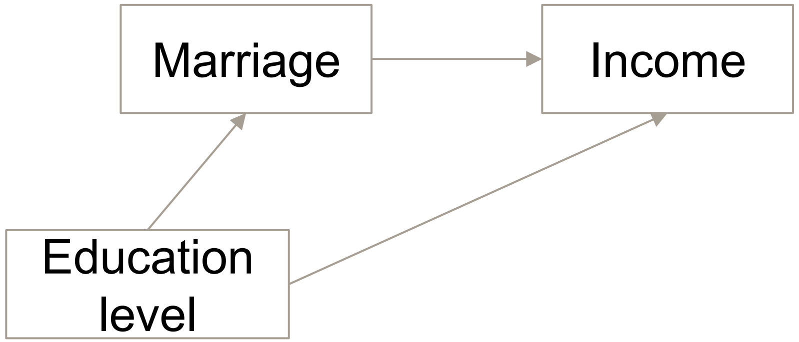
DAGs encode beliefs:

- X has a causal effect on Y.
- Y does not have a causal effect on X (acyclic).
- There are no variables that cause X and Y.

Paths

CENTRE FOR LONGITUDINAL STUDIES

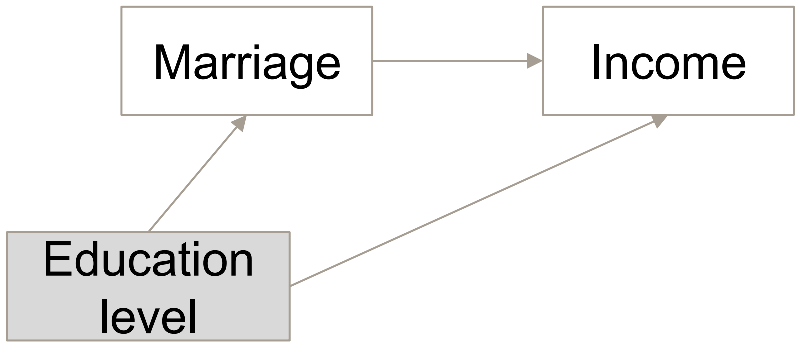
- A path exists between two variables if they are connected by one or more arrows (regardless of their directions).
- An open path between two variables implies a statistical association between the two.
- A closed path implies no statistical association (through that path).



Some conventions

CENTRE FOR LONGITUDINAL STUDIES

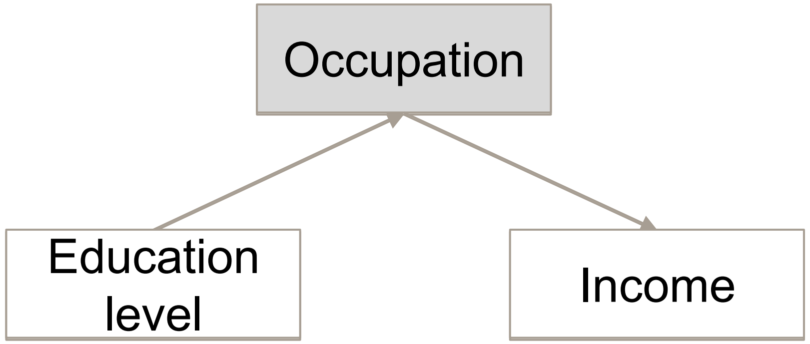
- Variables can be conditioned upon (e.g. by stratification, matching or statistical control).
- Unconditioned variables will be represented with a white box.
- Conditioned variables will be represented with a grey box.



Causal paths

CENTRE FOR LONGITUDINAL STUDIES

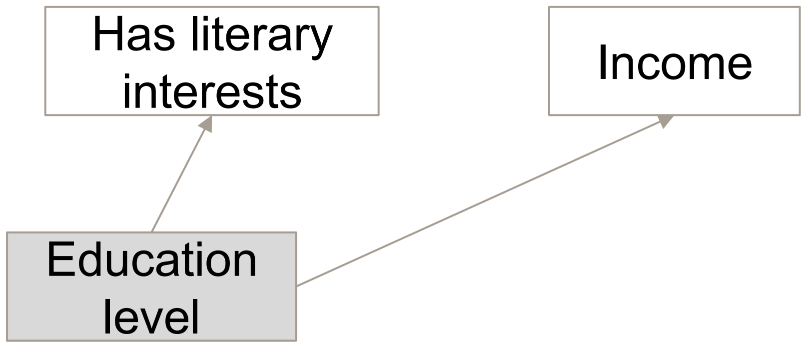
- A causal path is a path comprised of arrows running in the same direction (a directed path).
- Causal paths are open, unless a variable on the path is conditioned upon.



Backdoor paths

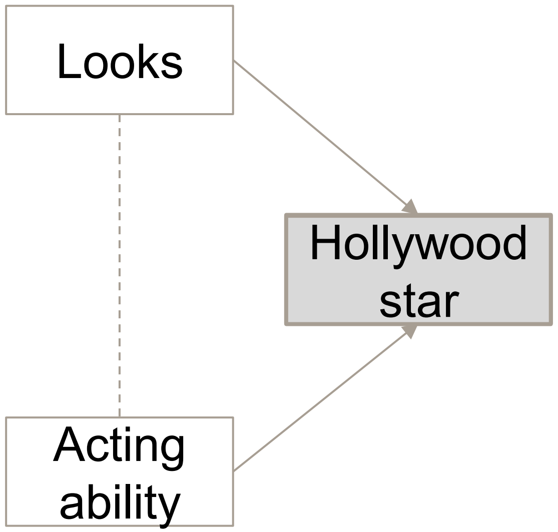
CENTRE FOR LONGITUDINAL STUDIES

- Open paths do not have to move in the direction of causality
- Backdoor paths transmit (non-causal) association.
- They can be closed by conditioning on one or more variables on the path.
- Open backdoor paths reflect confounding bias.



Collider paths

- Collider paths include a common consequence.
- They are closed unless the common consequence (or a descendant of this) is conditioned upon.
- Conditioning induces collider bias.



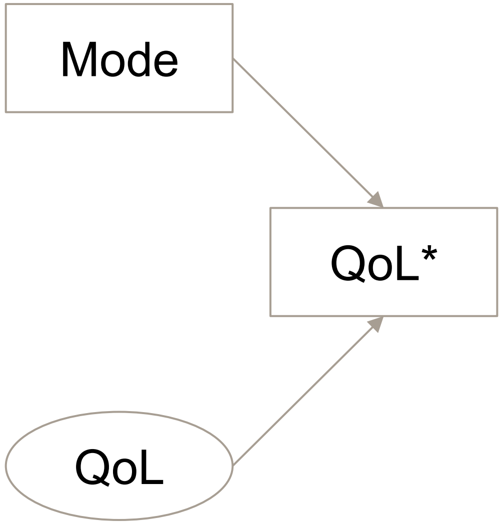
CENTRE FOR
LONGITUDINAL
STUDIES

Representing mode effects in DAGs

Representing a mode effects in DAGs

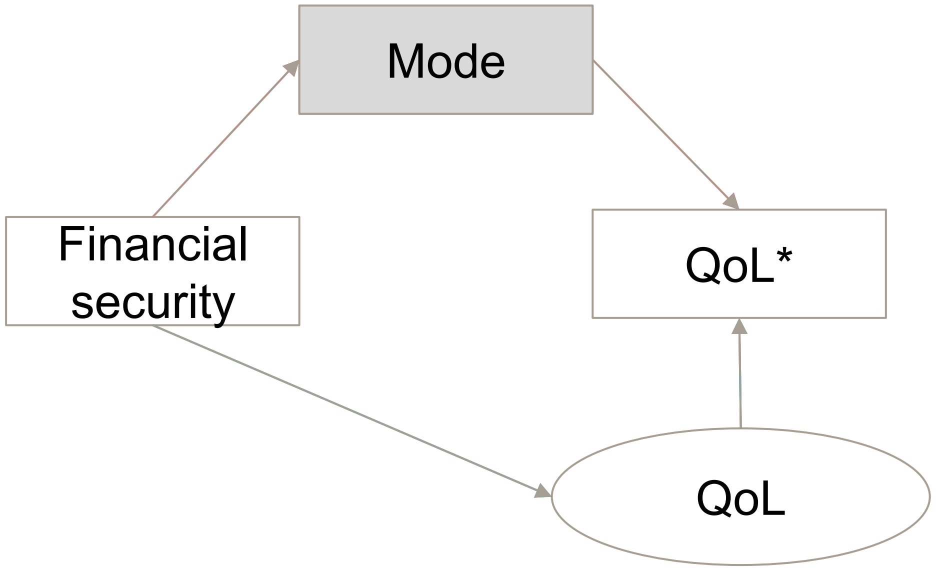
CENTRE FOR LONGITUDINAL STUDIES

- Mode effects are a form of systematic measurement error.
- Ellipses are used to denote latent variables.
- The measured variable (here, QoL*) is caused by the underlying latent (unobserved) variable and the mode of response.
- Motivating example: does financial security affect QoL?



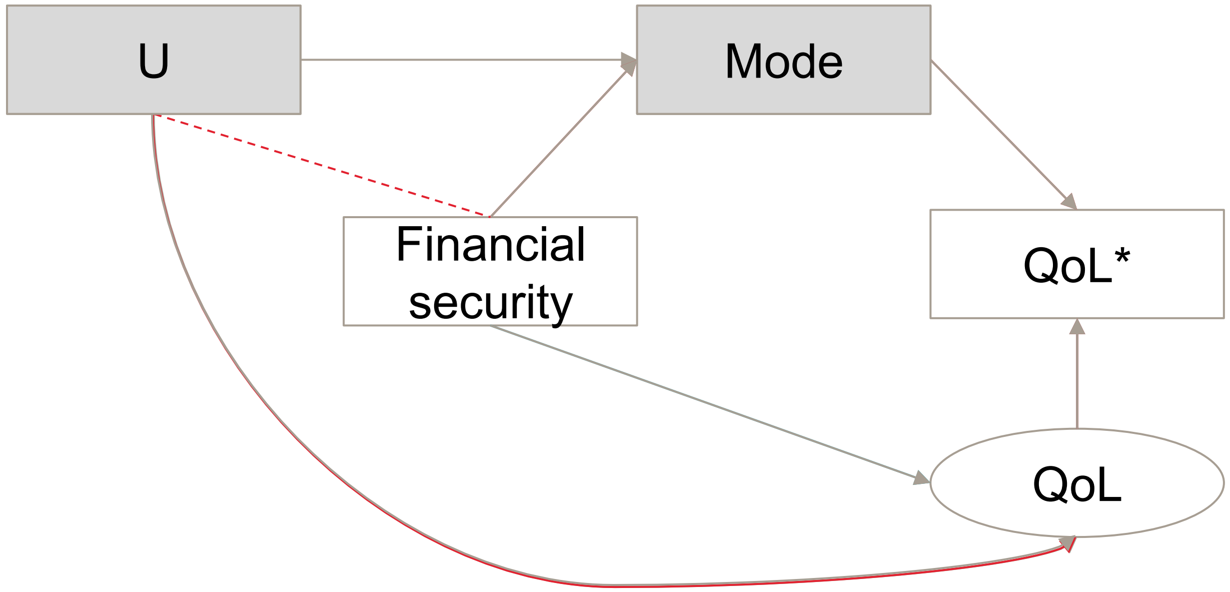
Outcome mode effects; mode selection on exposure

CENTRE FOR LONGITUDINAL STUDIES



Condition on mode.

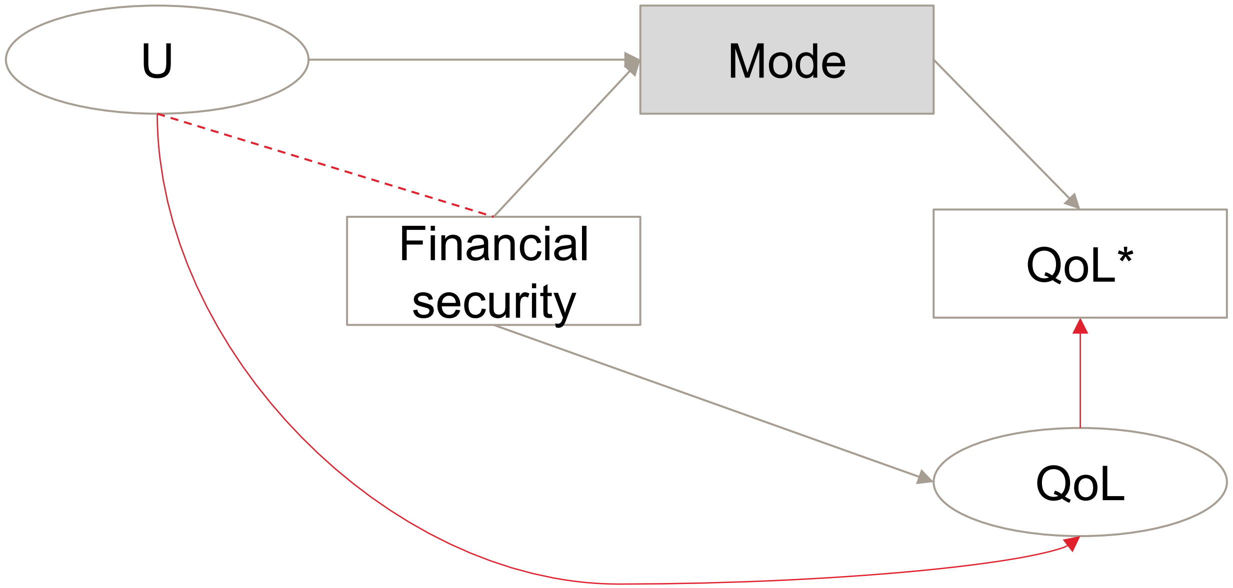
Outcome mode effects; mode selection on exposure and observed U



Condition on mode and observed U.

Outcome mode effects; mode selection on exposure and unobserved U

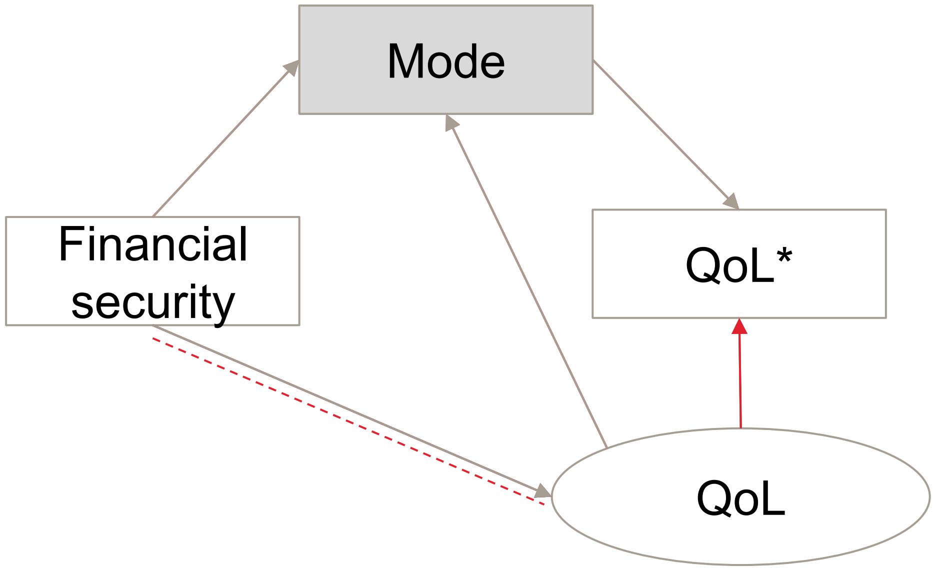
CENTRE FOR LONGITUDINAL STUDIES



Cannot condition on unobserved U → bias.

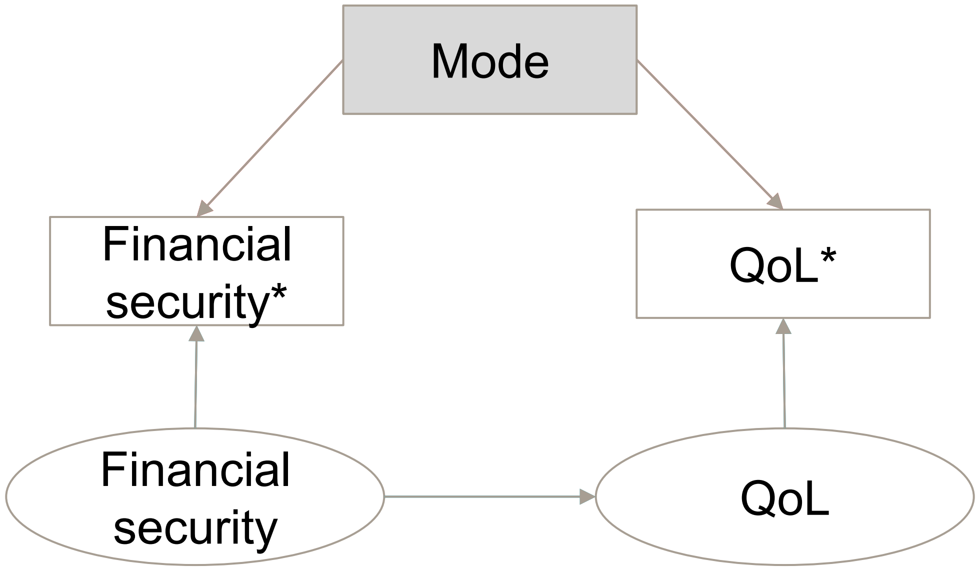
Outcome mode effects; mode selection on outcome

CENTRE FOR LONGITUDINAL STUDIES



Cannot condition on latent (unobserved) outcome → bias.

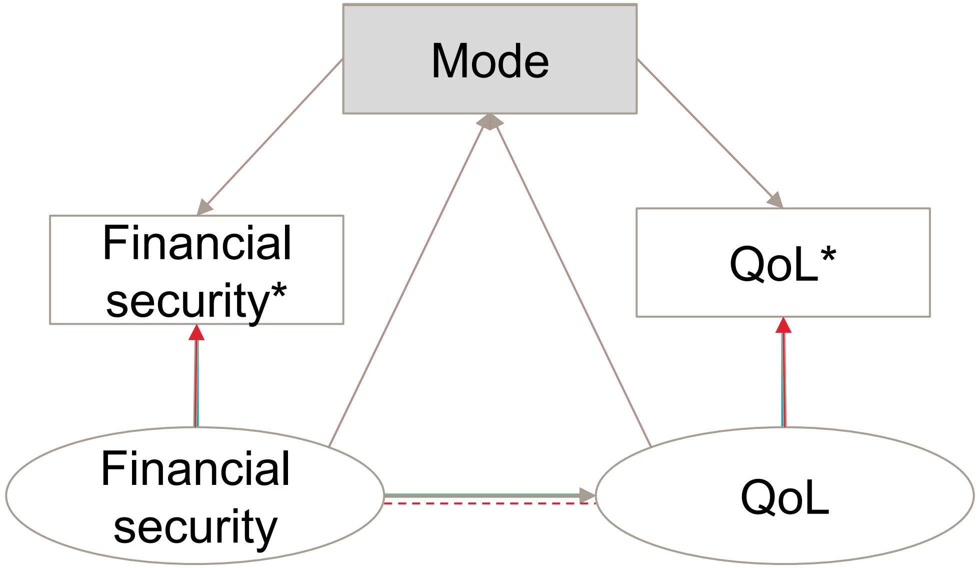
Exposure and outcome mode effects



Condition on mode.

Exposure and outcome mode effects; mode selection on exposure and outcome

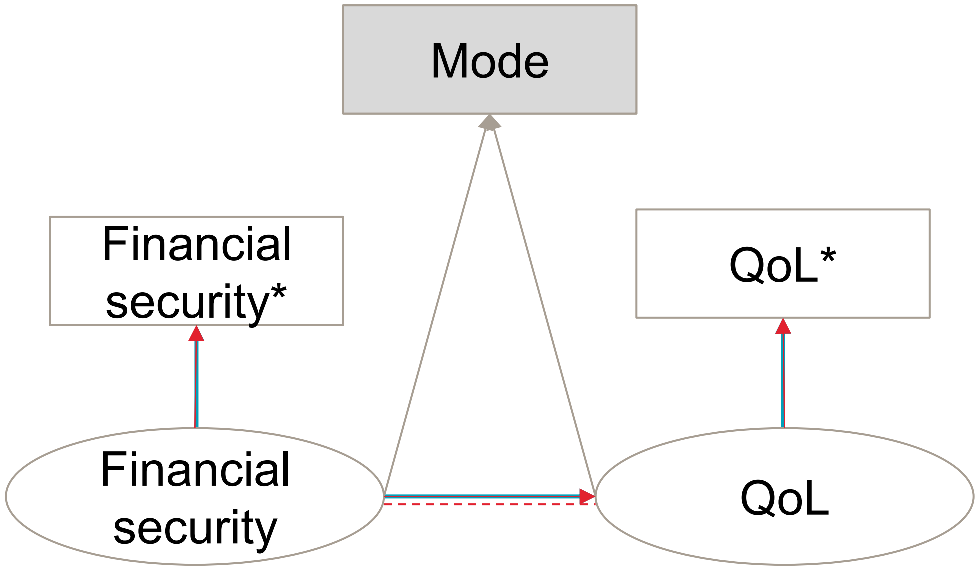
CENTRE FOR LONGITUDINAL STUDIES



Cannot condition on latent (unobserved) exposure and outcome → bias.

No mode effects; mode selection on exposure and outcome

CENTRE FOR
LONGITUDINAL
STUDIES



Conditioning on mode → bias. Don't condition on anything!

Summary

CENTRE FOR LONGITUDINAL STUDIES

- Where mode effects occur without selection into mode (strong assumption), it may be safe to simply condition on mode.
- Where selection into mode occurs, conditioning on mode risks introducing collider bias – this could be larger than the amount of confounding reduced(!).
- We therefore really need to know which variables affect mode selection and to have observed these.

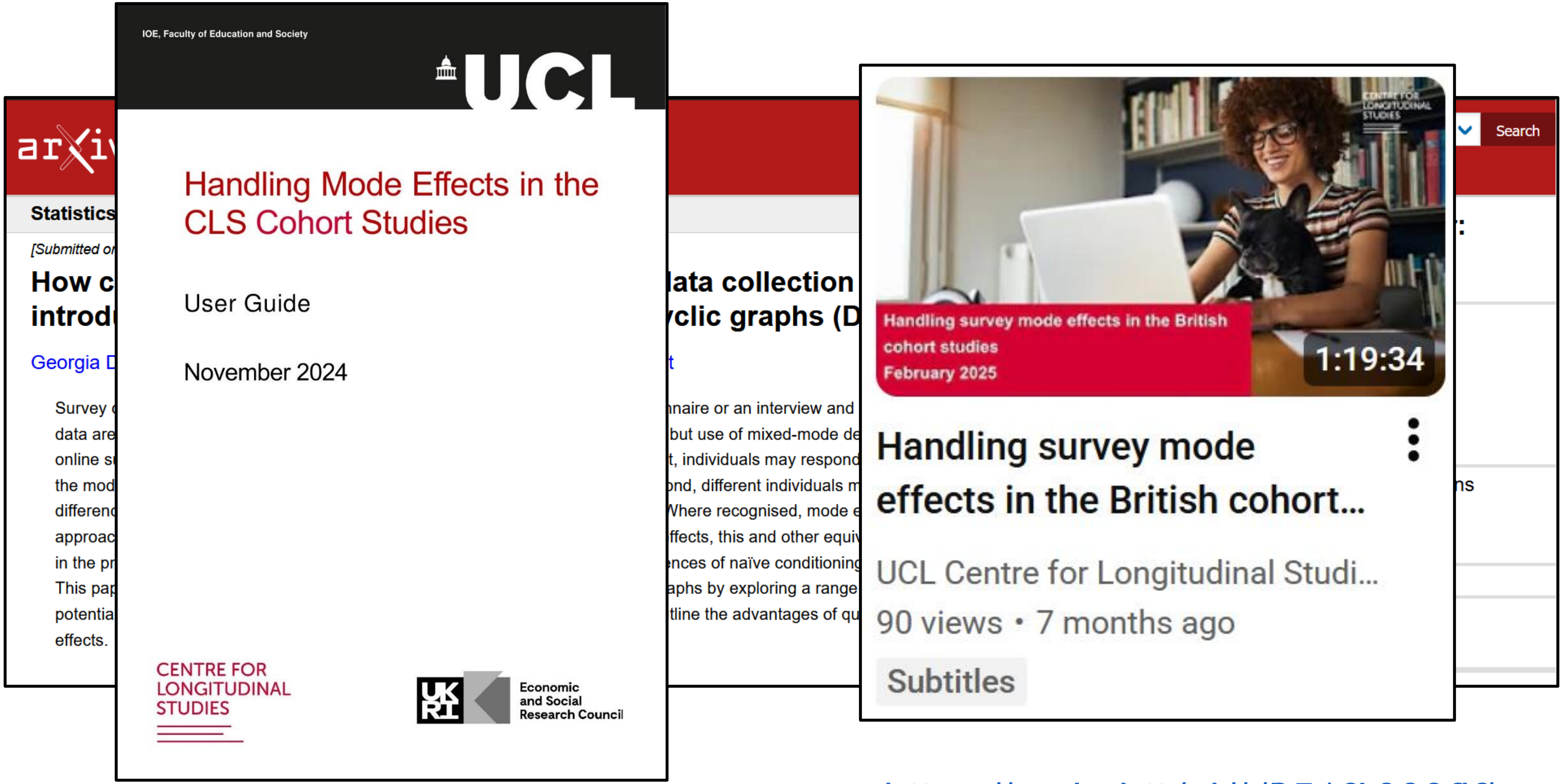
Summary

CENTRE FOR LONGITUDINAL STUDIES

- We have placed the problem of mode effects within the simple and intuitive causal DAG framework.
- Realistic settings may be much more complicated.
- But we can always encode our assumptions in a DAG and interrogate it to see whether, how and why our analysis will be biased, and what we might be able to do about it.
- Can also use the DAGs to determine whether/how the mode effect itself can be unbiasedly estimated.
- Not sure what to assume? Draw multiple DAGs.

Resources

CENTRE FOR
LONGITUDINAL
STUDIES



https://arxiv.org/abs/2510.00900
https://cls.ucl.ac.uk

Funding

- The Centre for Longitudinal Studies is supported by the Economic and Social Research Council (ES/W013142/1).
- This work was further supported by the ESRC-funded Survey Futures collaboration.

UKRI Economic and Social Research Council

SURVEY FUTURES
SURVEY DATA COLLECTION
METHODS COLLABORATION

Thank you.

CENTRE FOR LONGITUDINAL STUDIES

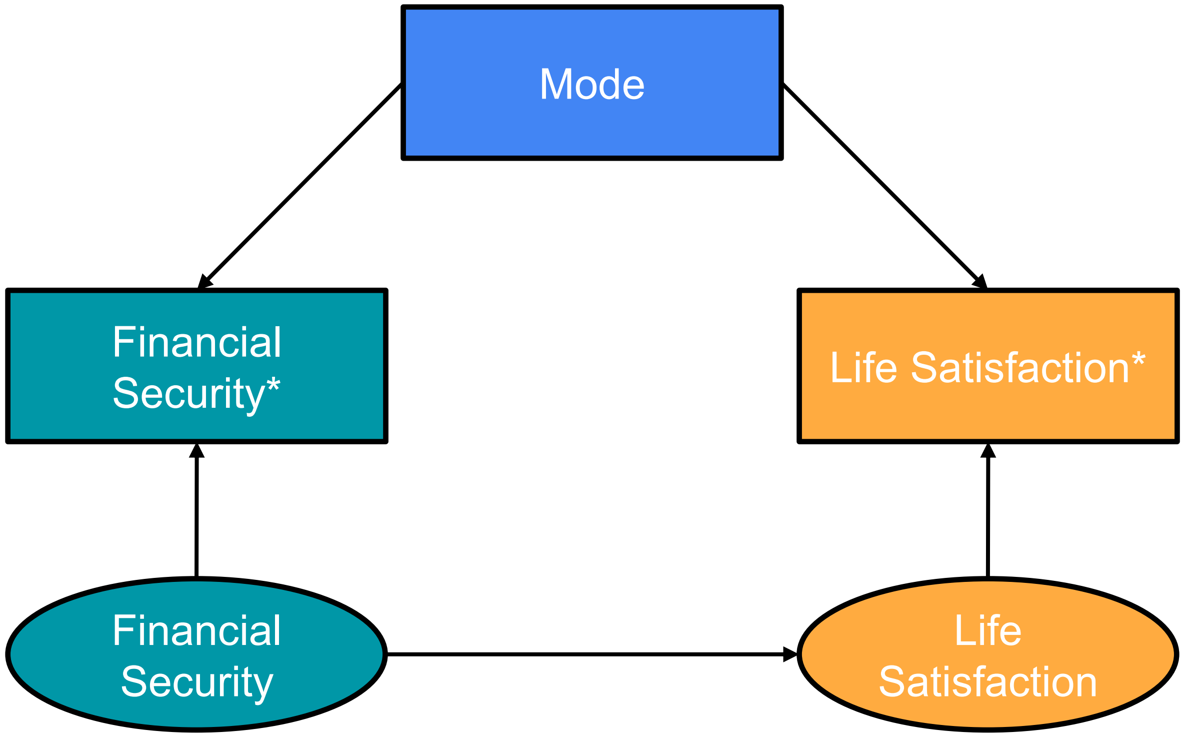
Economic and Social Research Council

Methods for Handling Mode Effects

Royal Statistical Society, November 2025

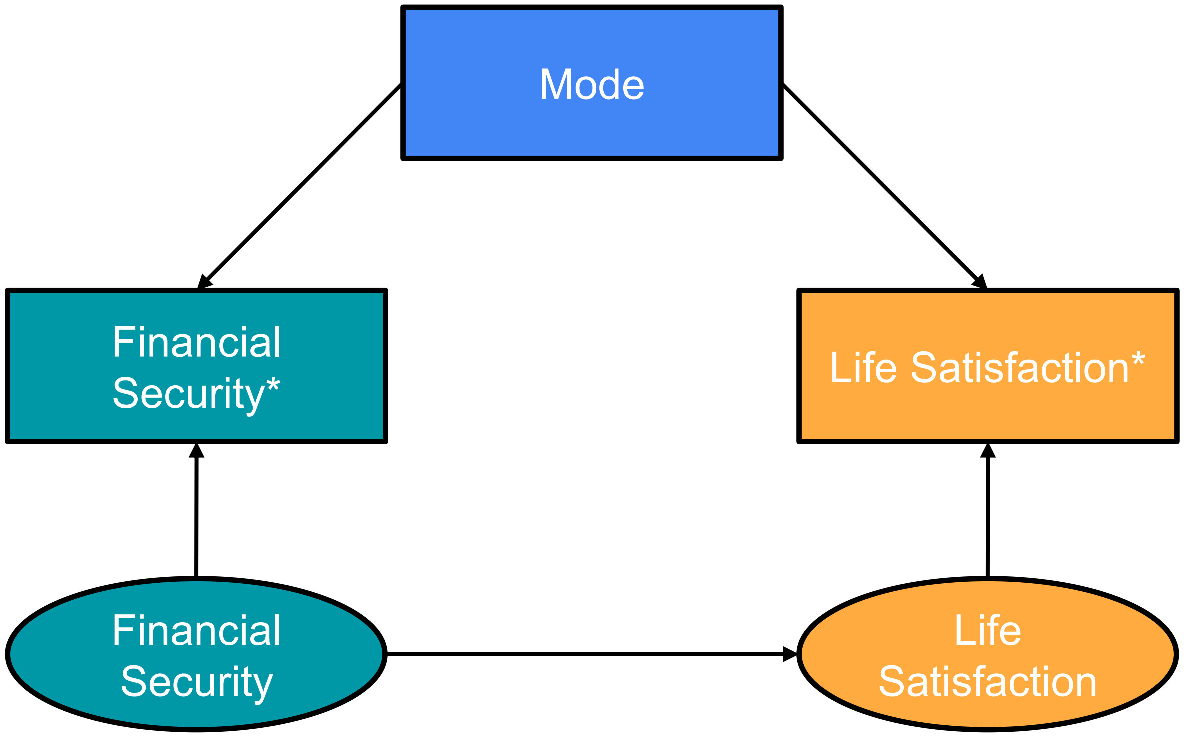
Introduction

- Unaccounted for mode effects may bias analyses of mixed-mode survey data.
- Surveys can be designed to reduce mode effects *ex ante*.
- But some differences are inherent. Methods are required to deal with, or quantify, this bias *post hoc*.



Introduction

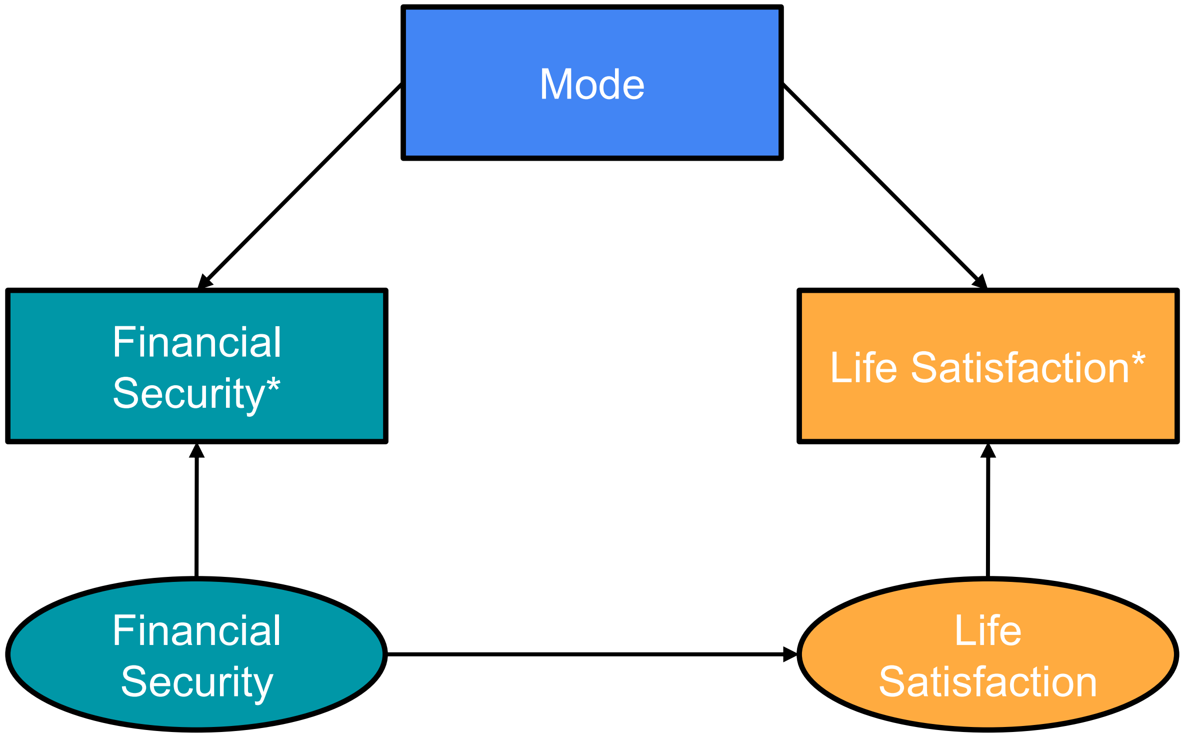
- Several methods has been proposed and used in the literature (see, e.g., Maslovskava et al., 2023).
- I will discuss:
  1. Statistical control
  2. Multiple imputation
  3. Quantitative Bias Analysis



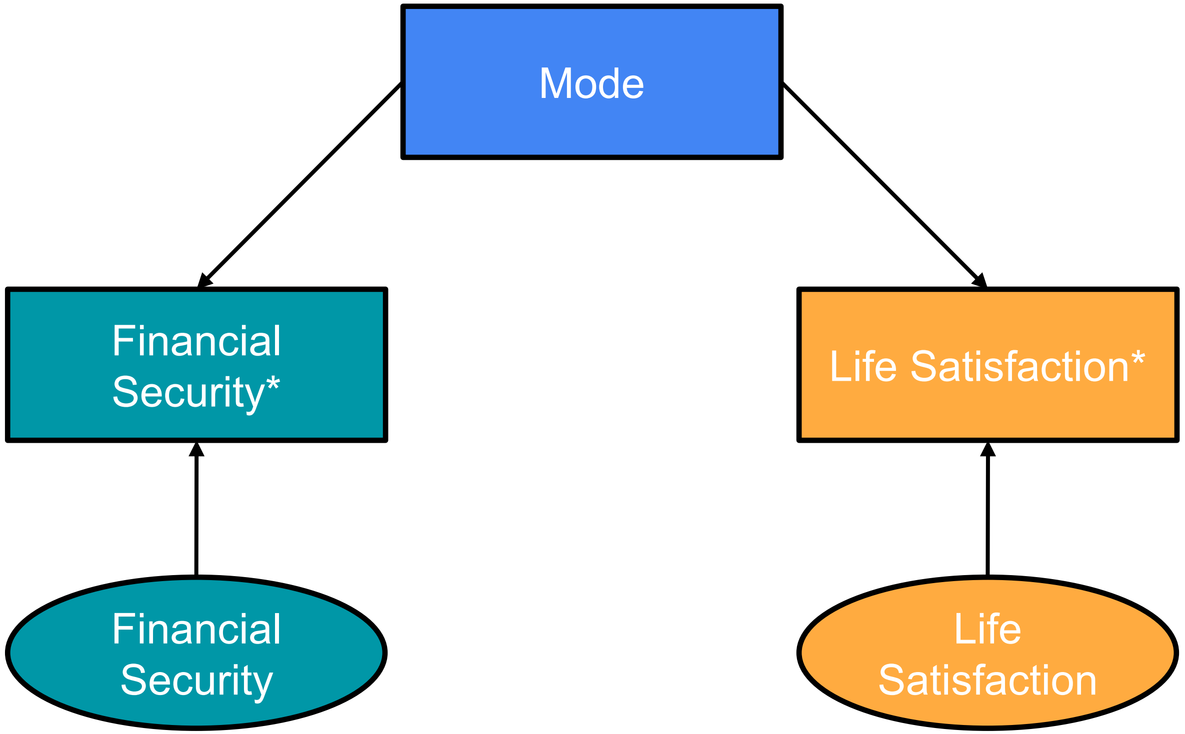
Statistical Control

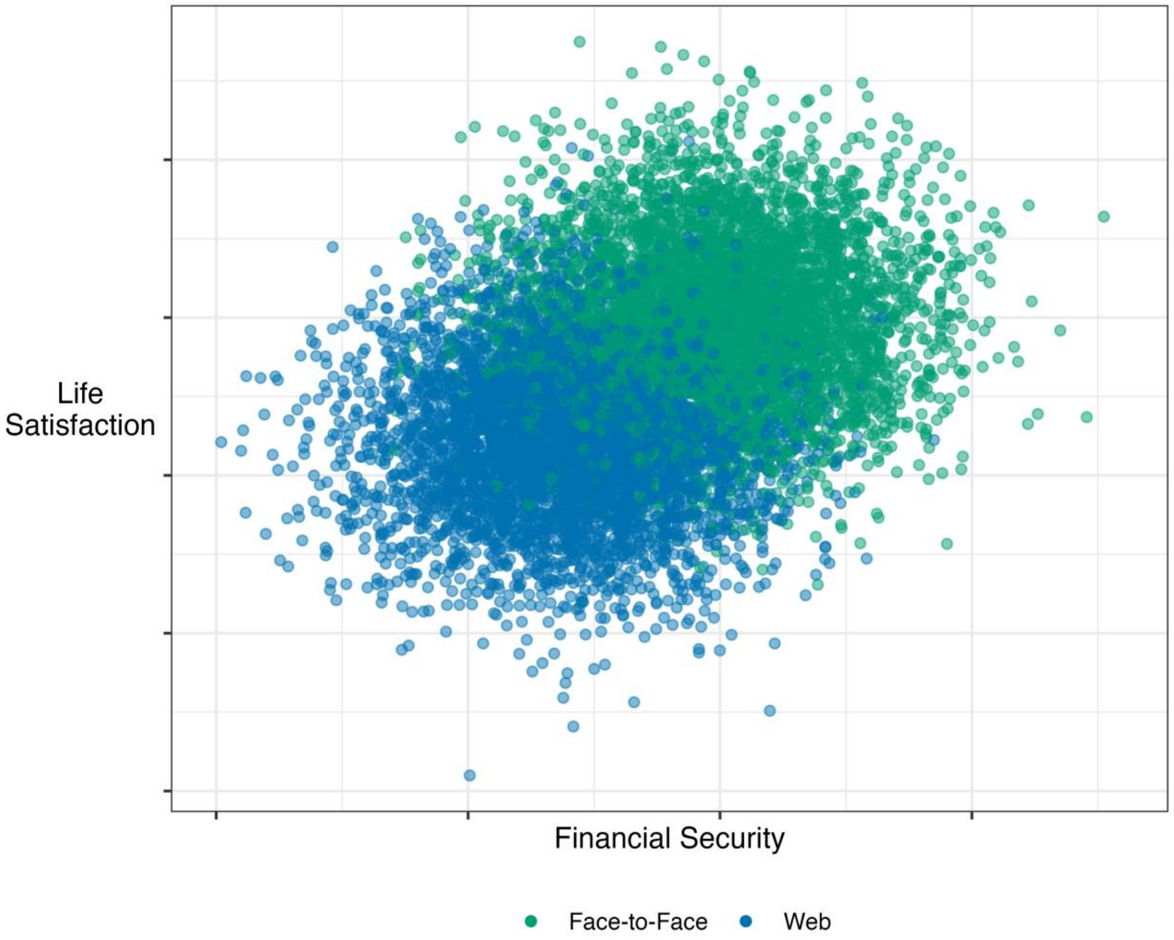
Statistical Control

- It may be possible to account for mode effects by conditioning upon *mode*.
- This could simply involve adding an indicator variable for mode into regressions or stratifying by mode.
- In the presence of mode selection, it may also require additional control variables.

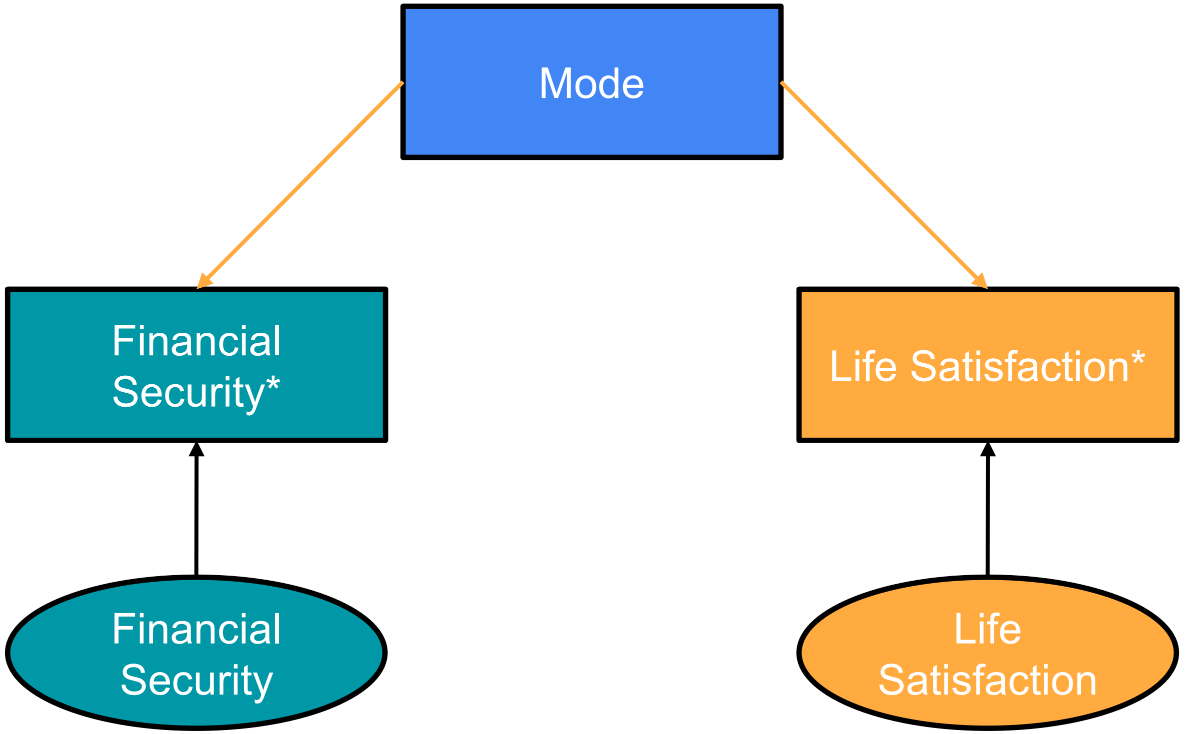


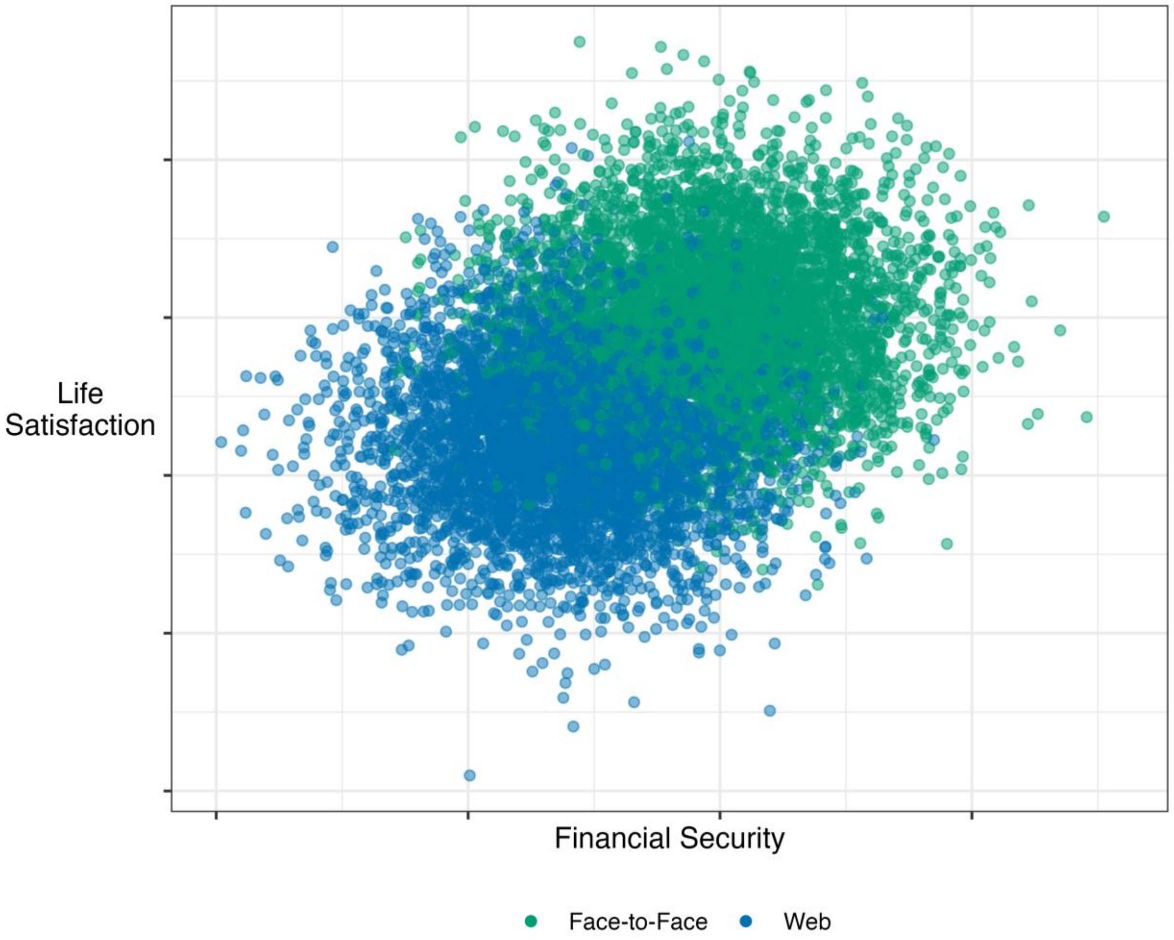
Joint Mode Effects



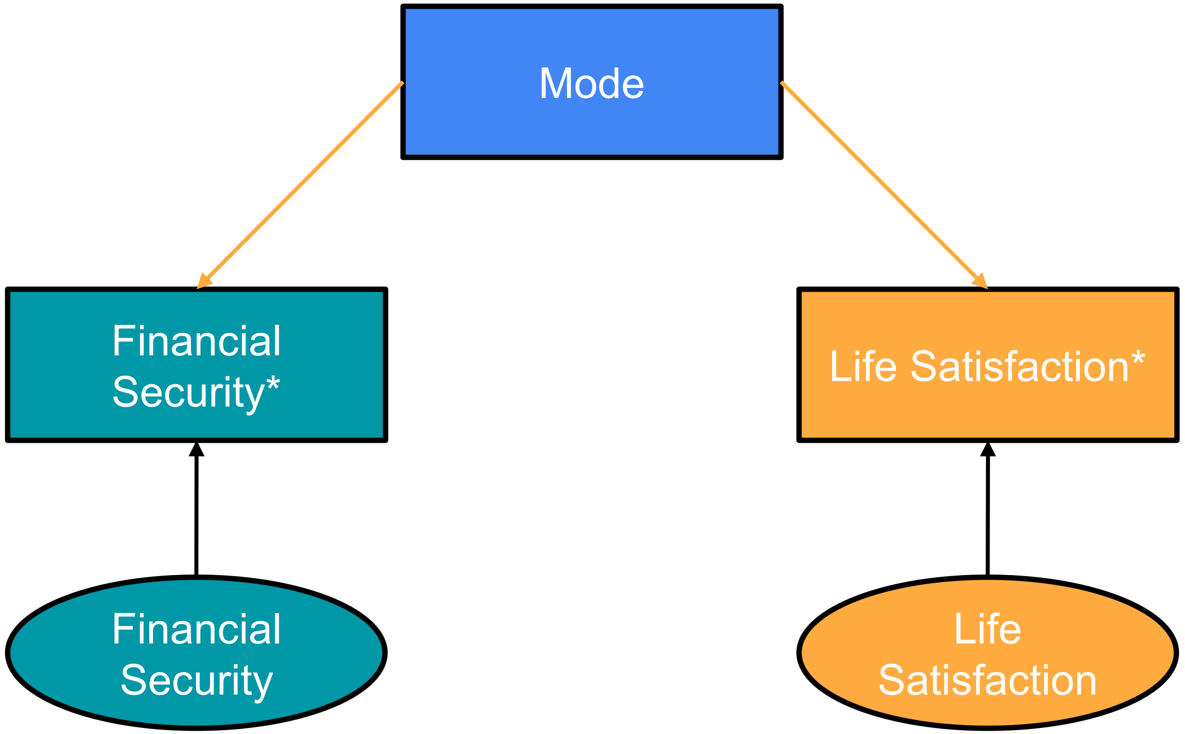


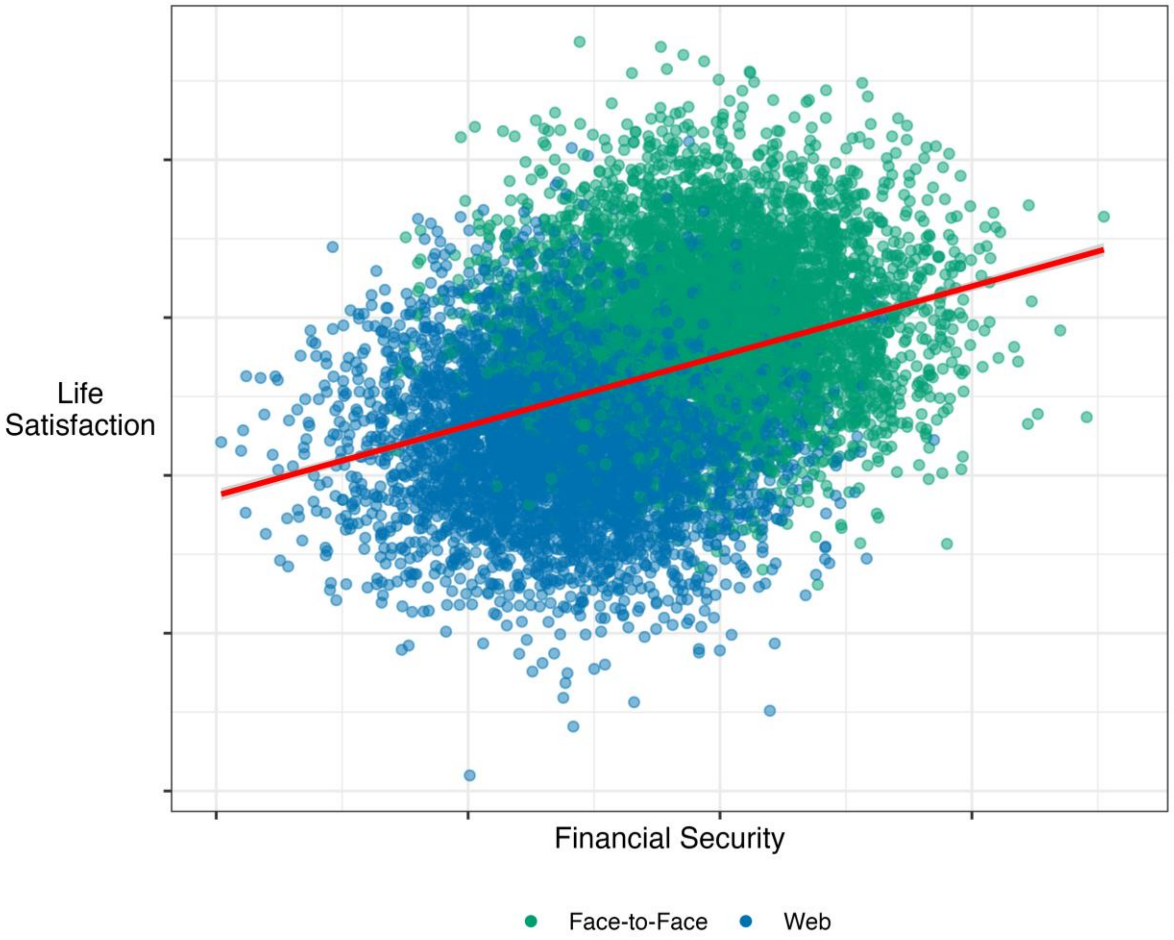
Joint Mode Effects



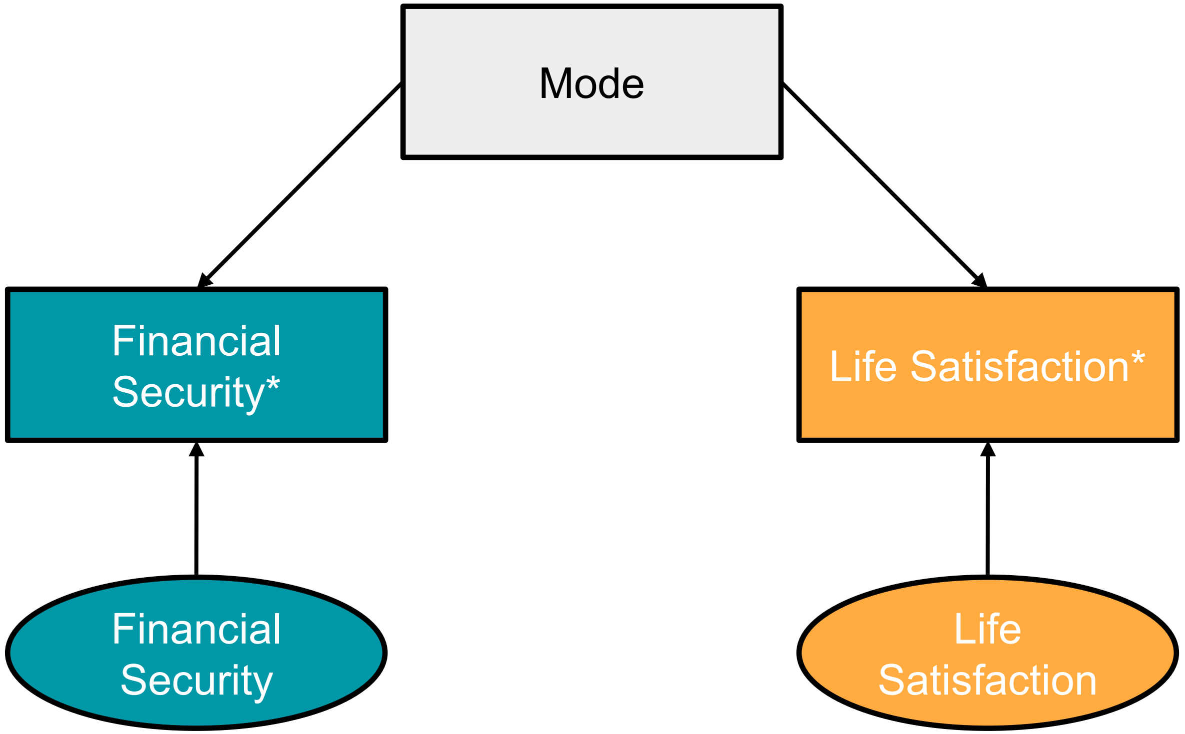


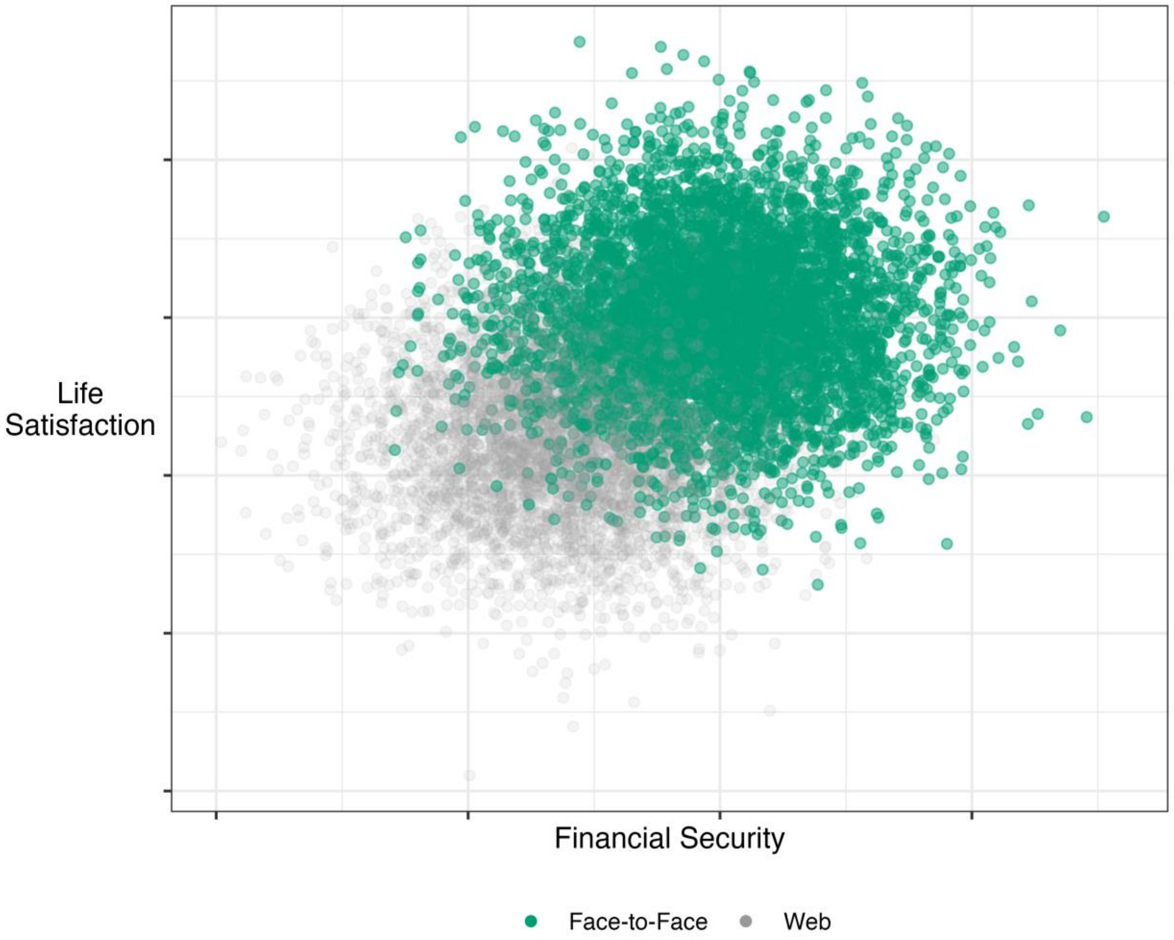
Joint Mode Effects



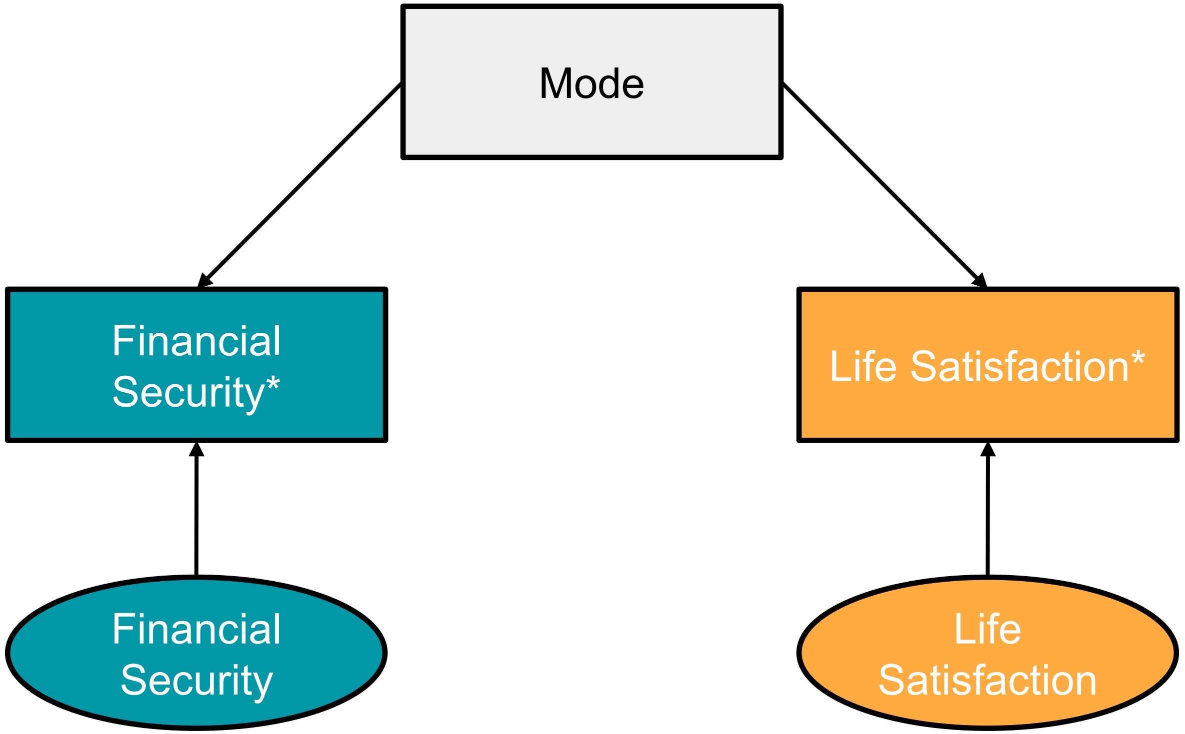


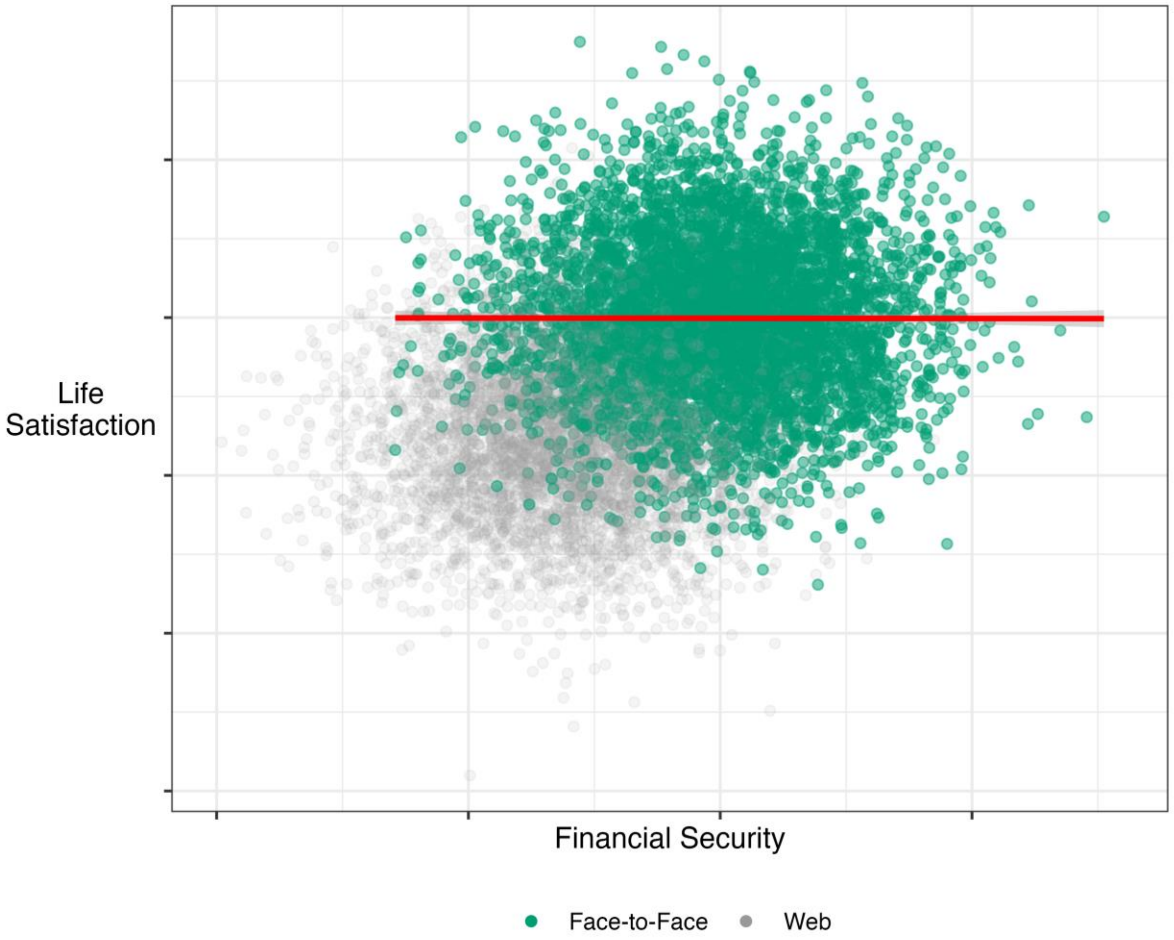
Joint Mode Effects



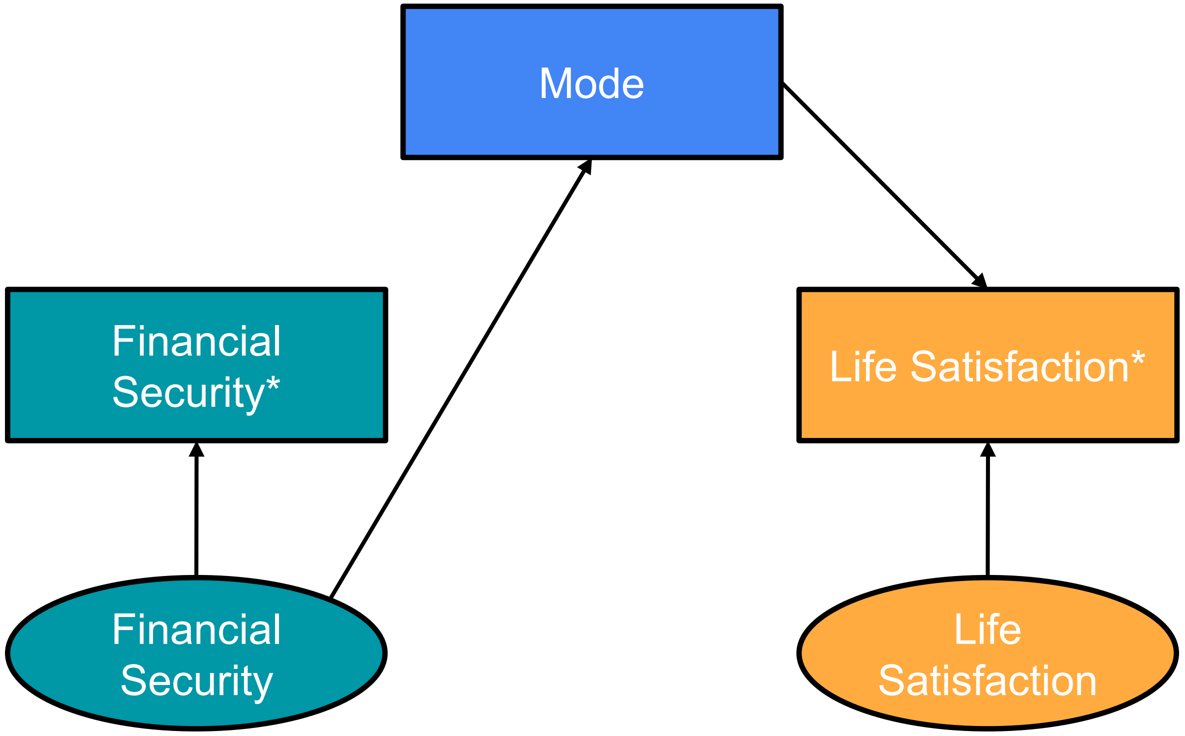


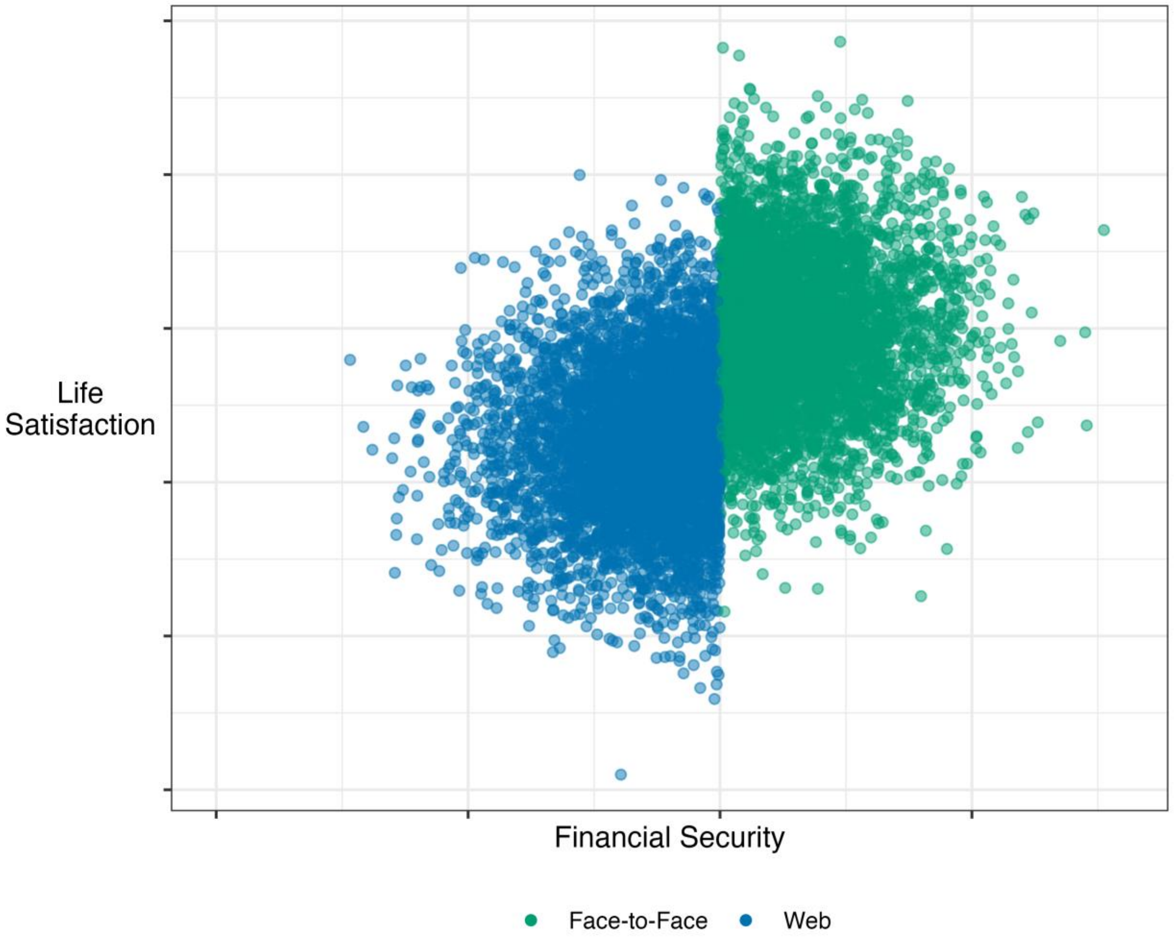
Joint Mode Effects



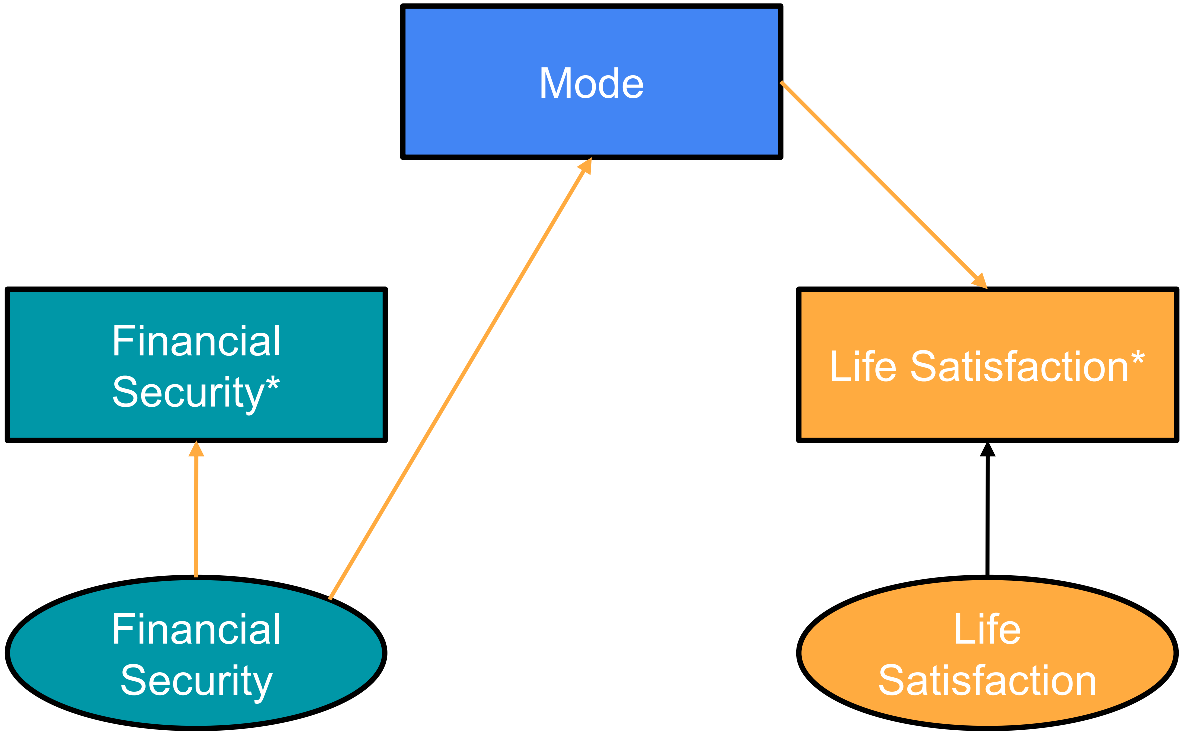


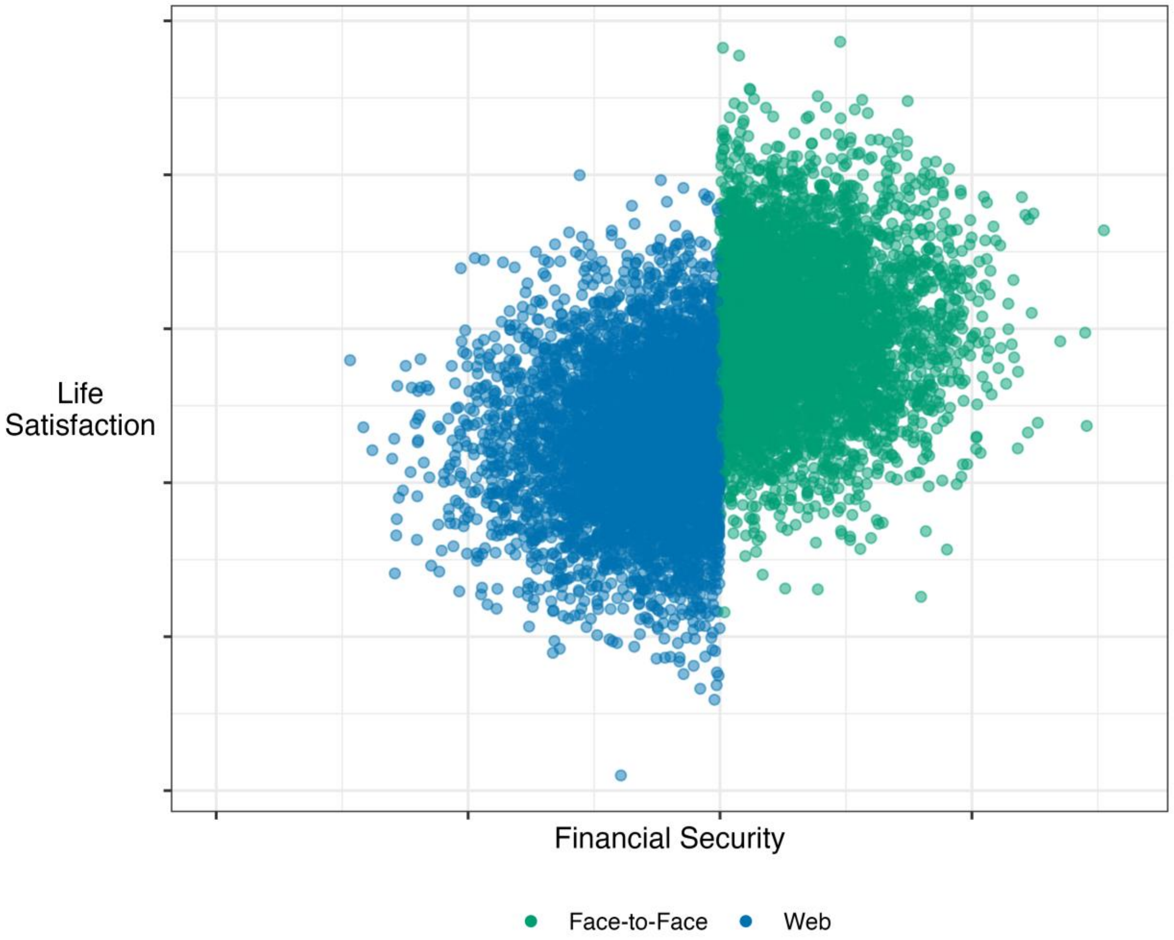
Mode Effects and Mode Selection



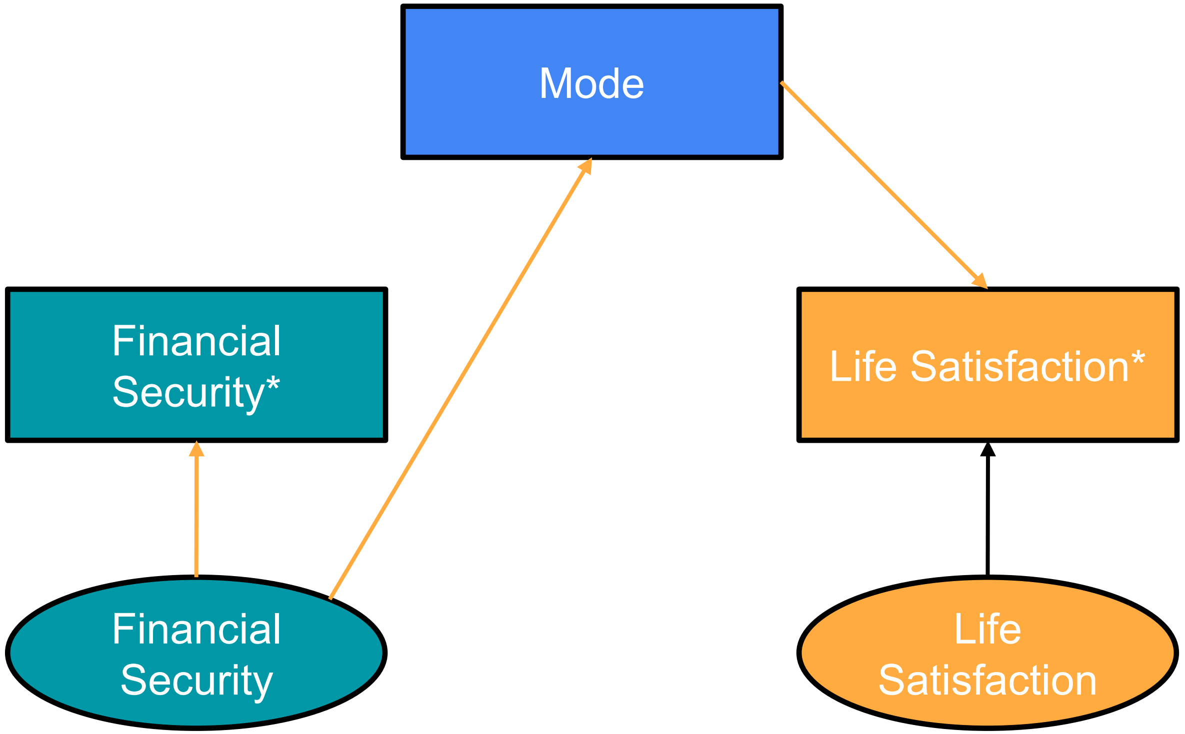


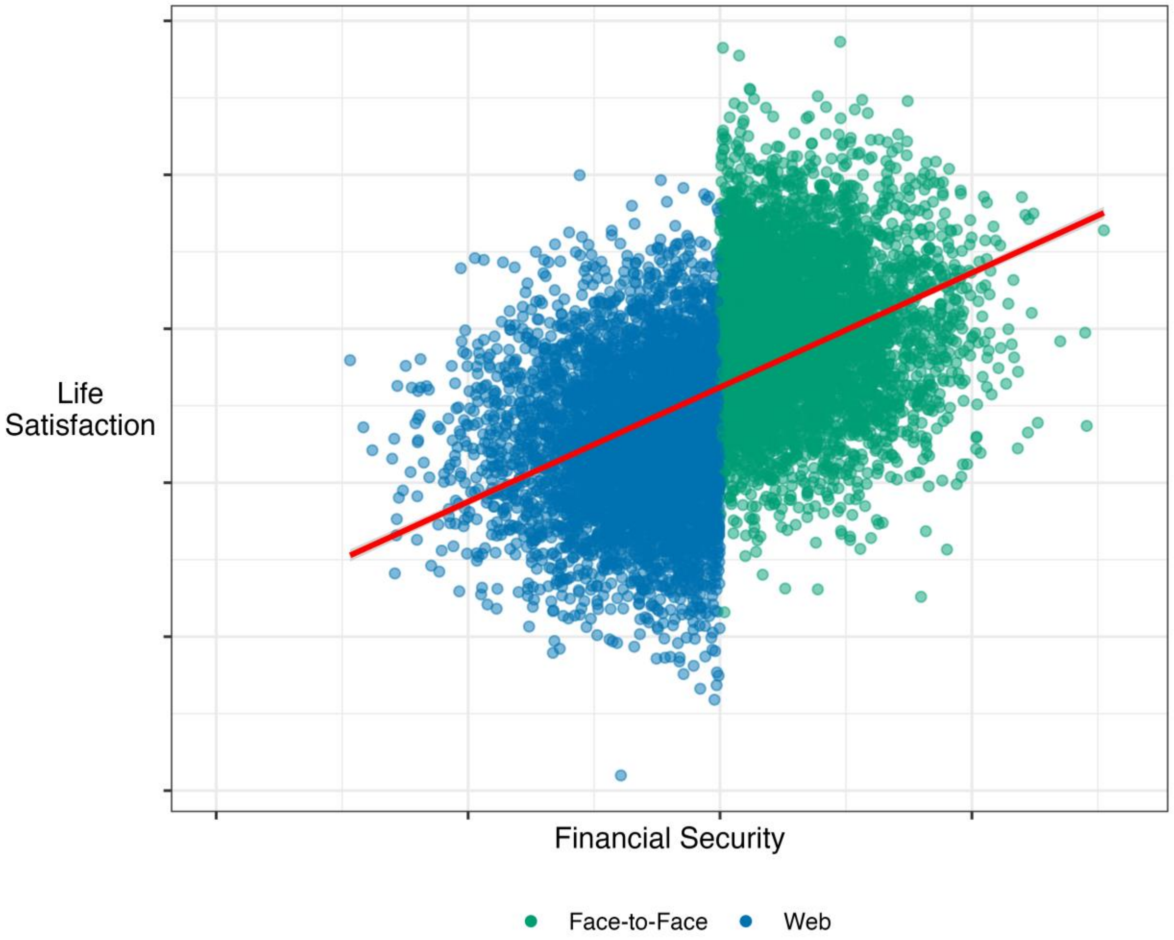
Mode Effects and Mode Selection



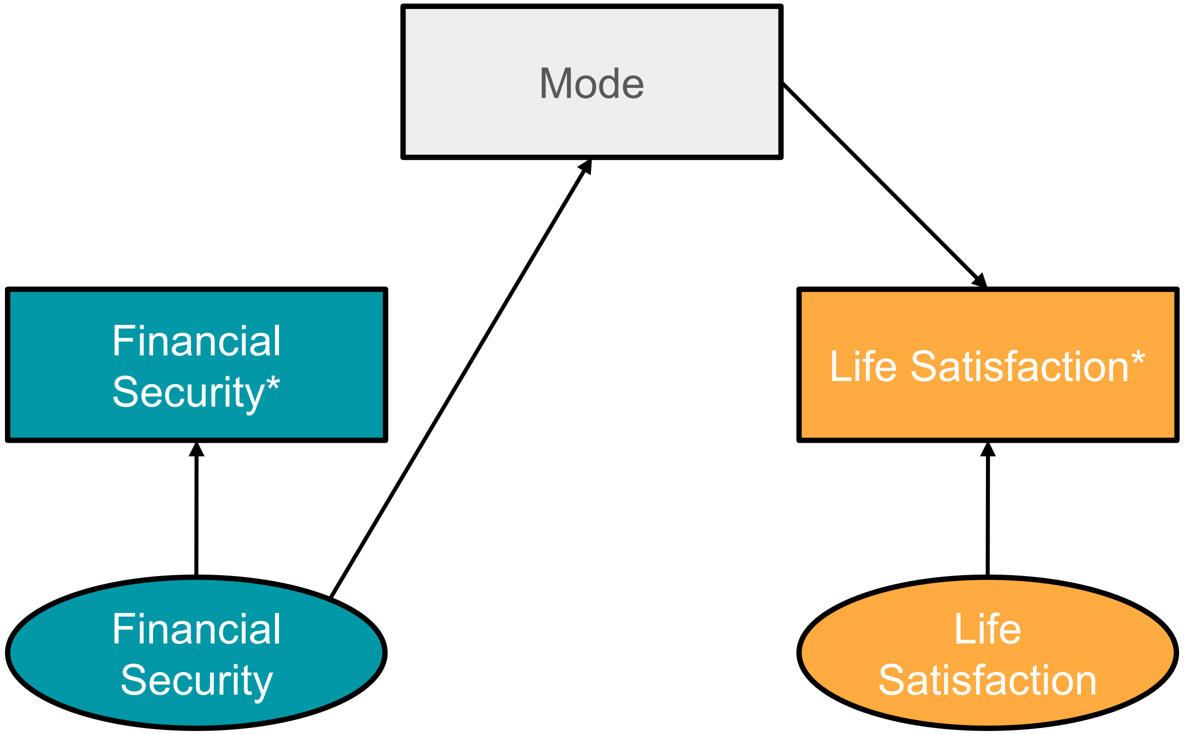


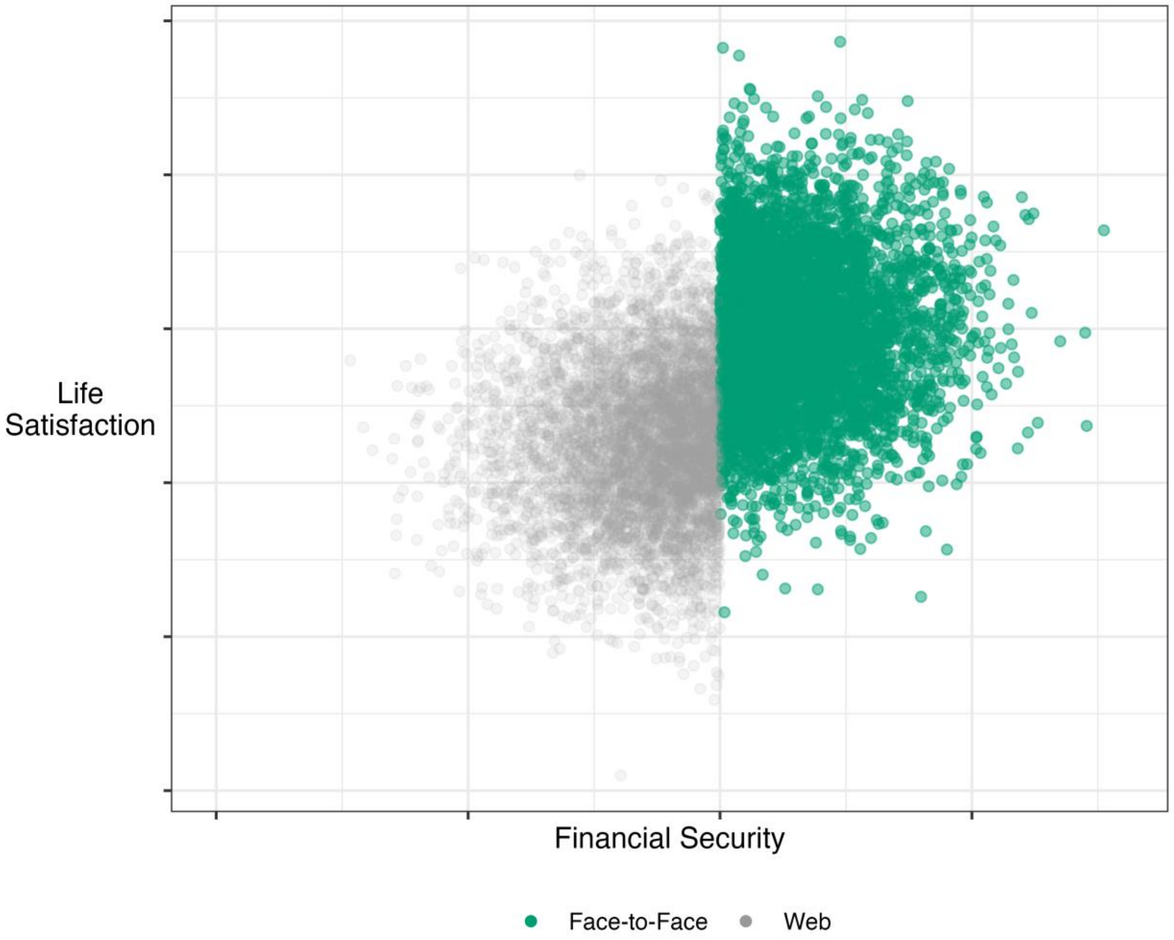
Mode Effects and Mode Selection



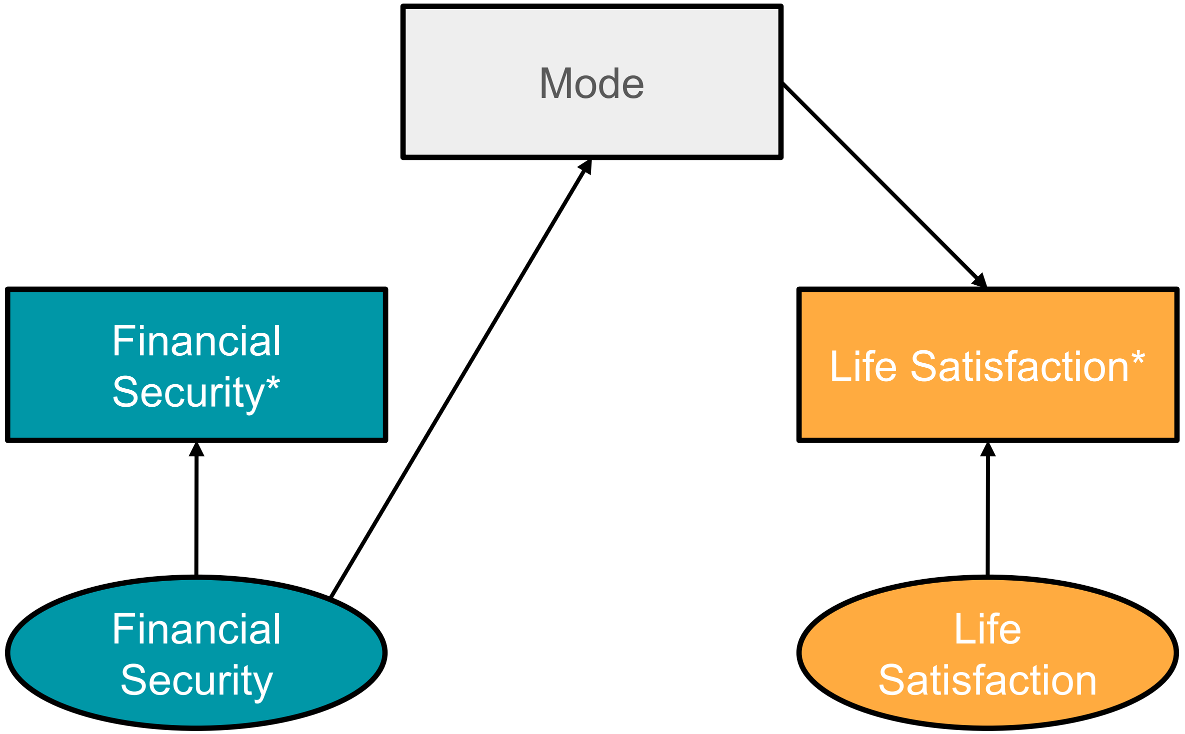


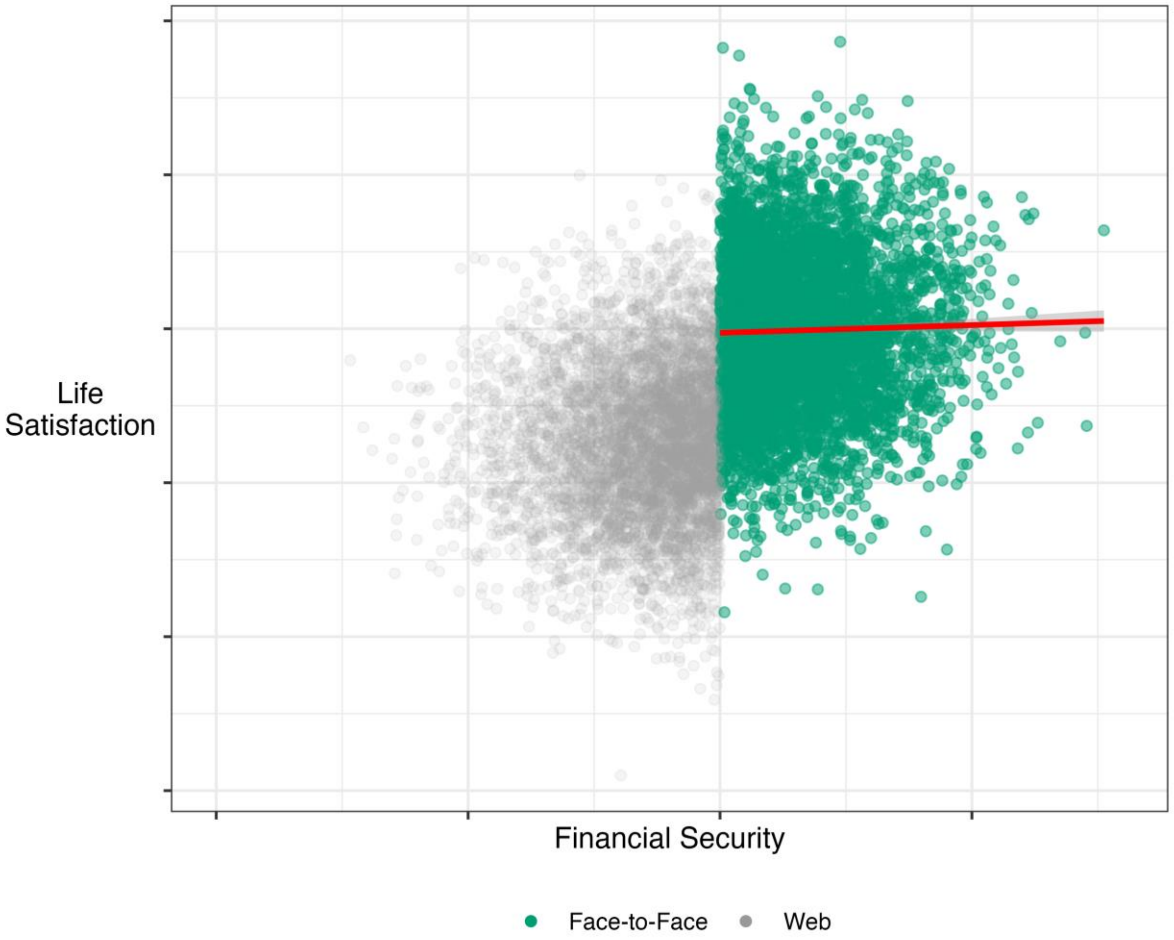
Mode Effects and Mode Selection



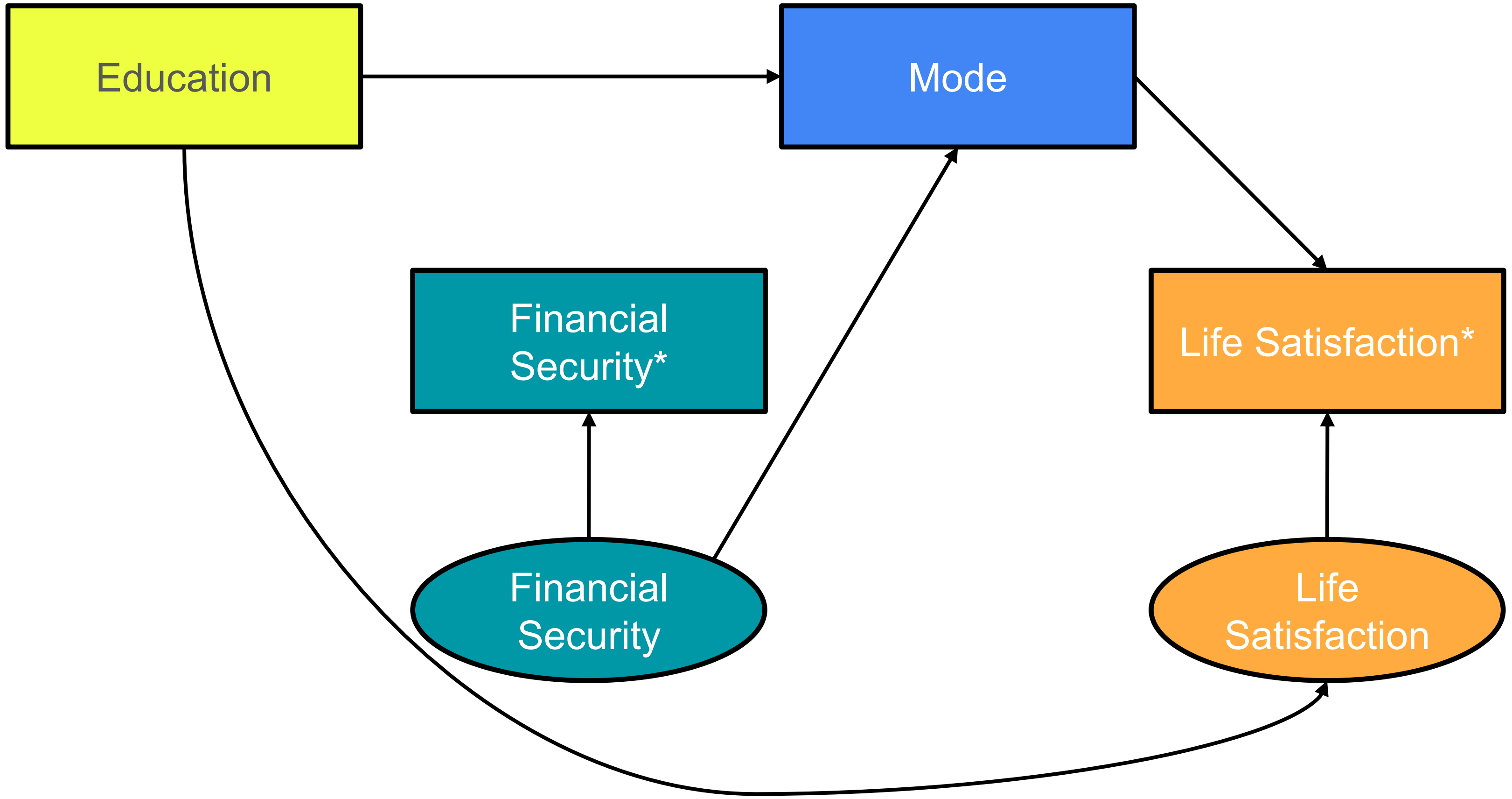


Mode Effects and Mode Selection

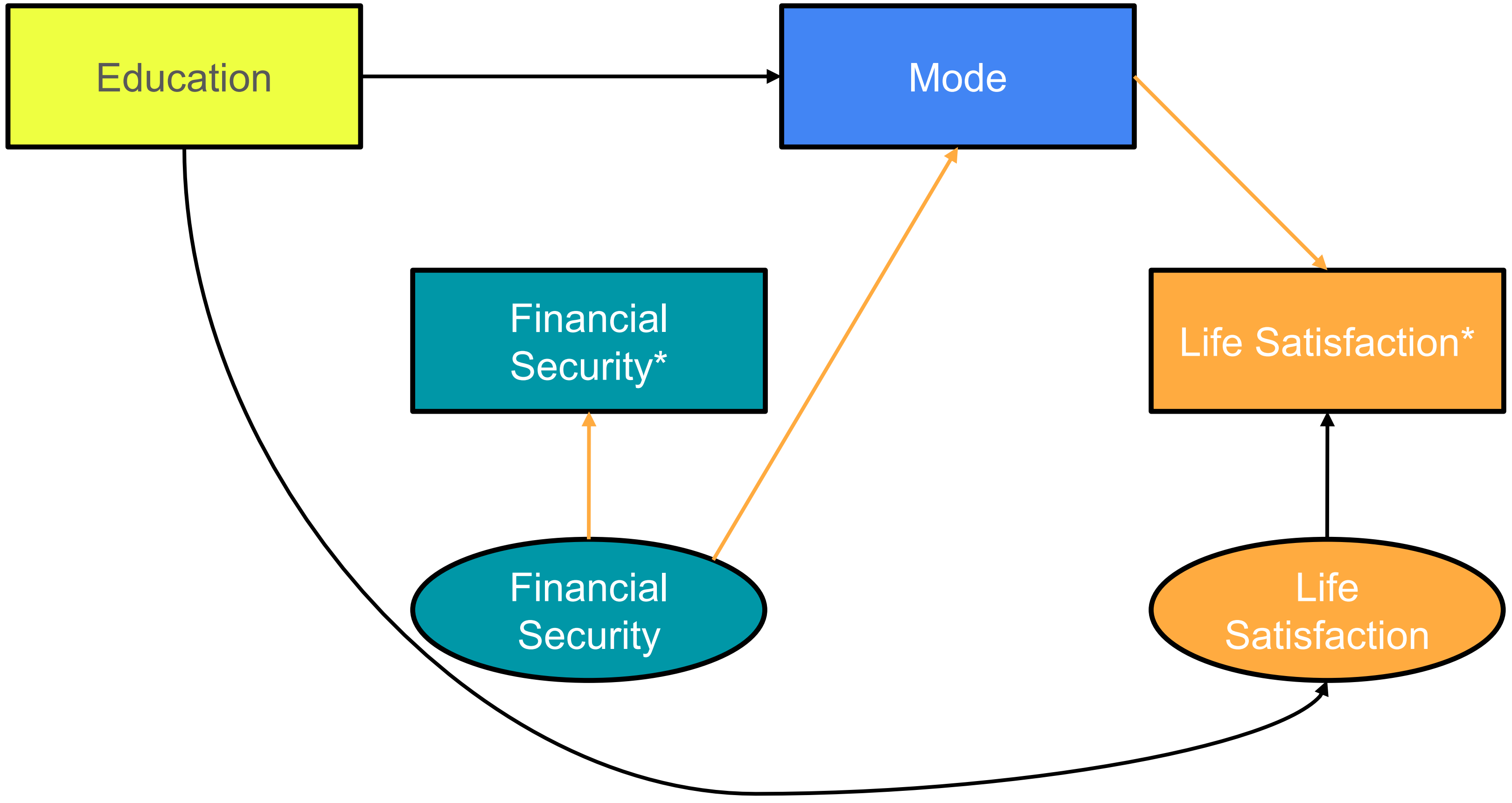




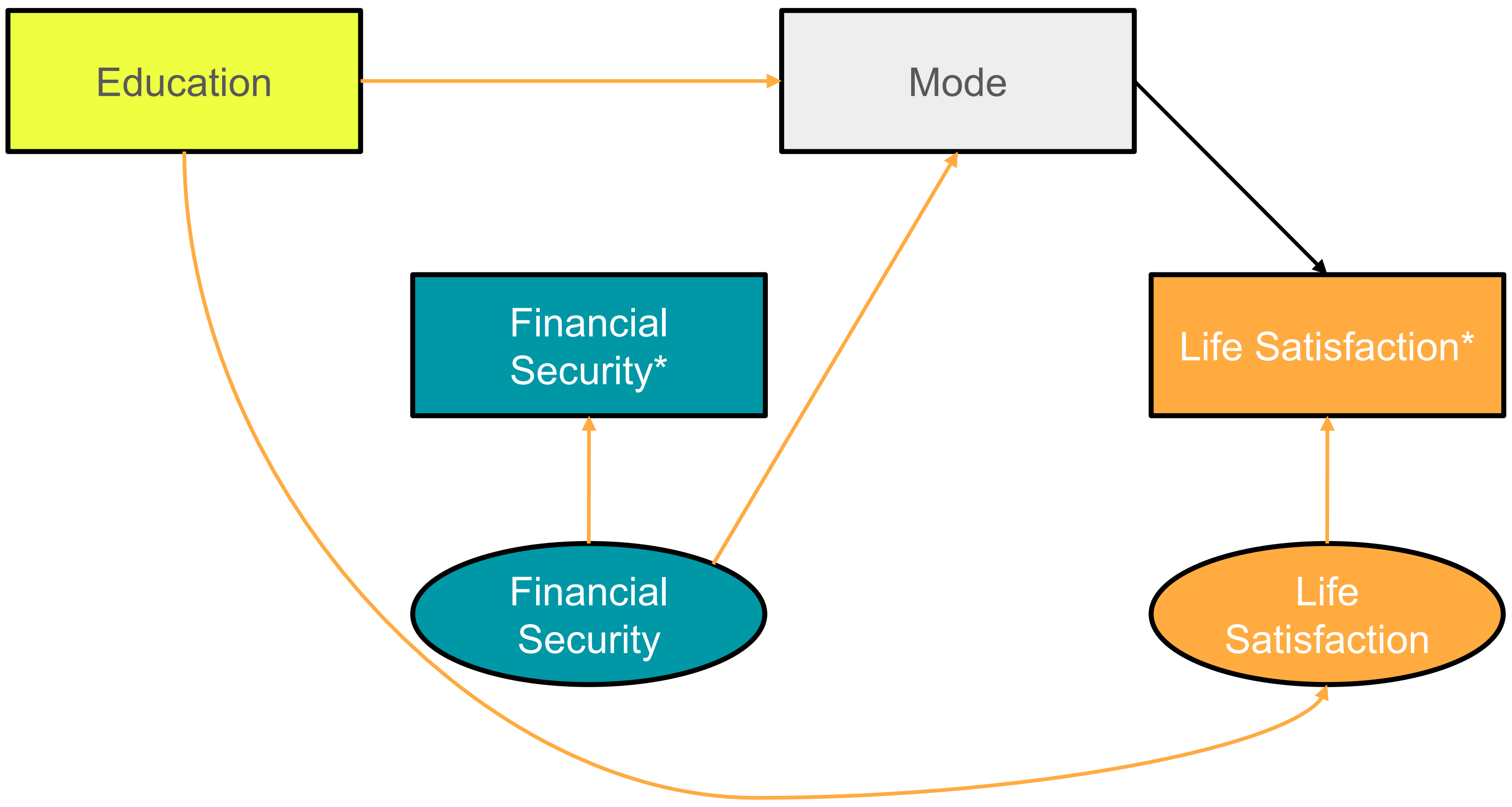
But…Multiple Sources of Mode Selection



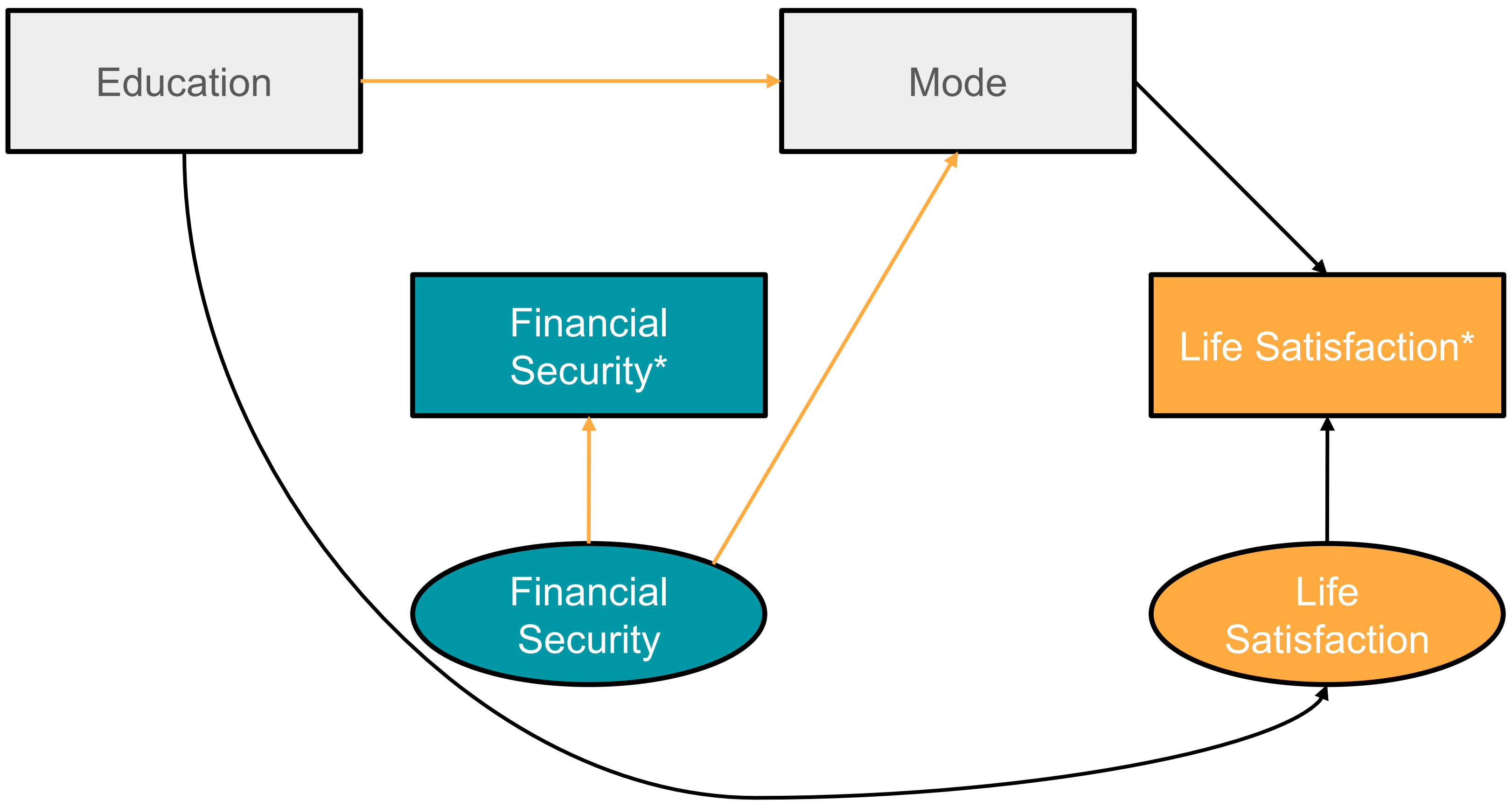
But…Multiple Sources of Mode Selection



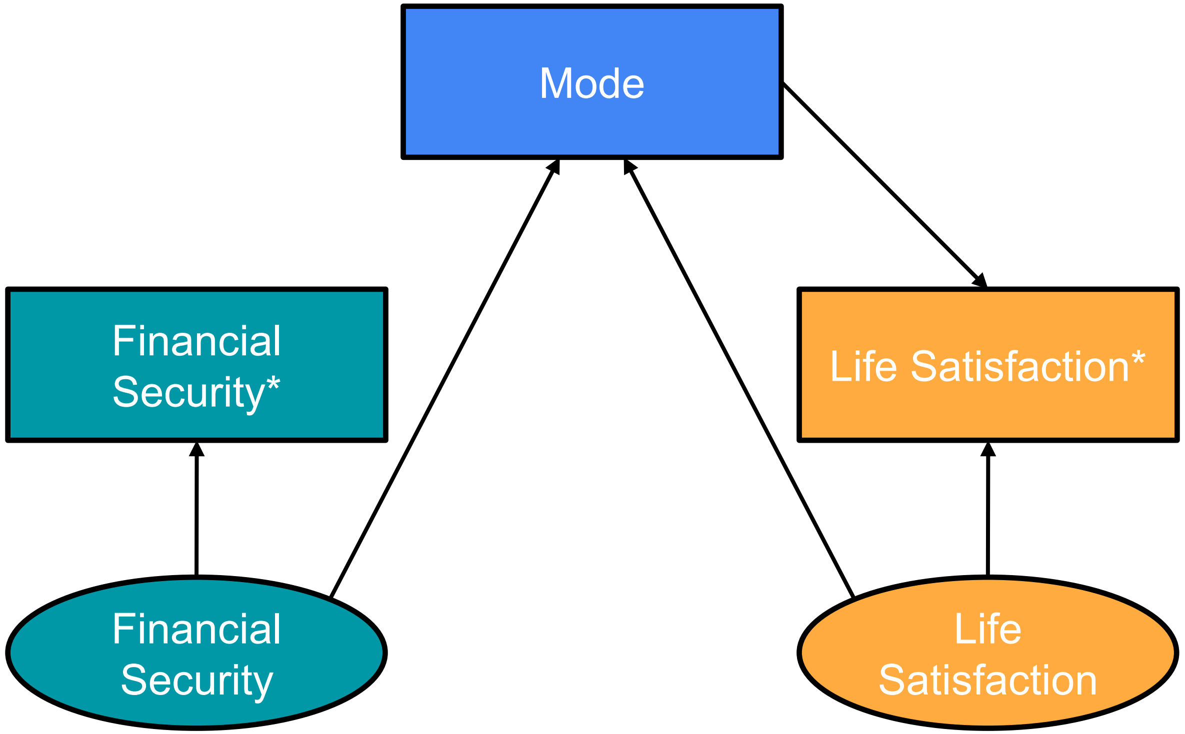
But…Multiple Sources of Mode Selection



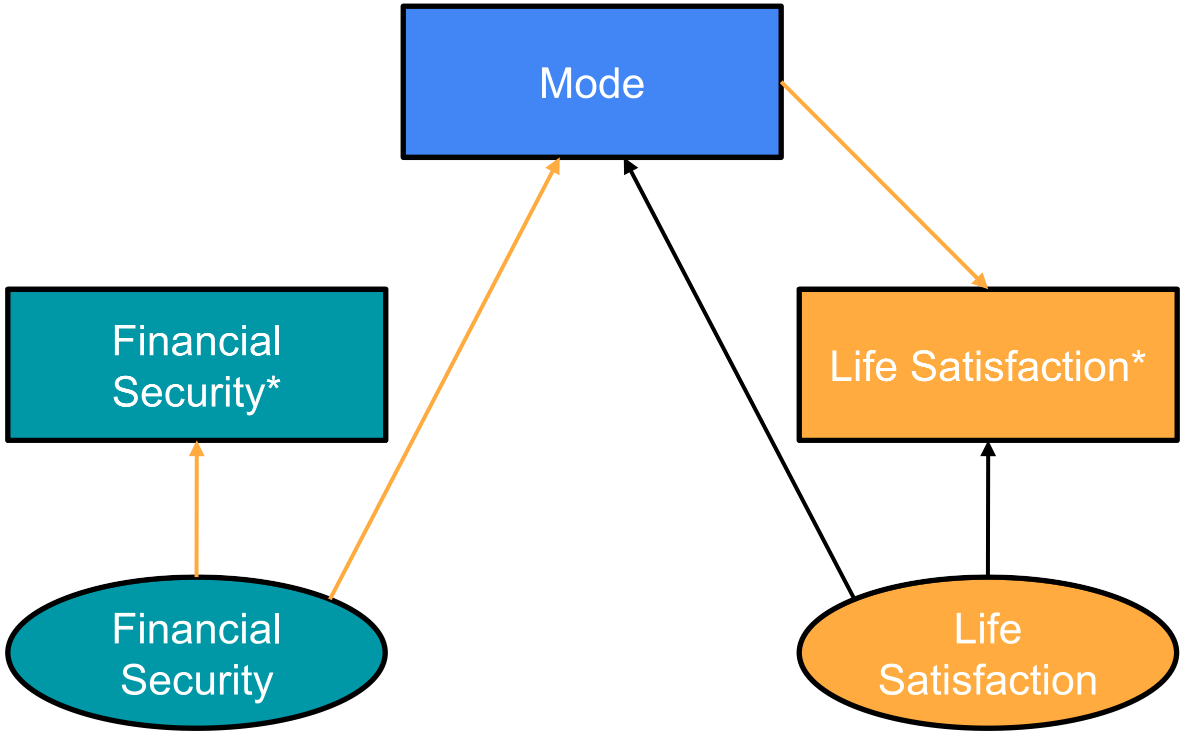
But…Multiple Sources of Mode Selection



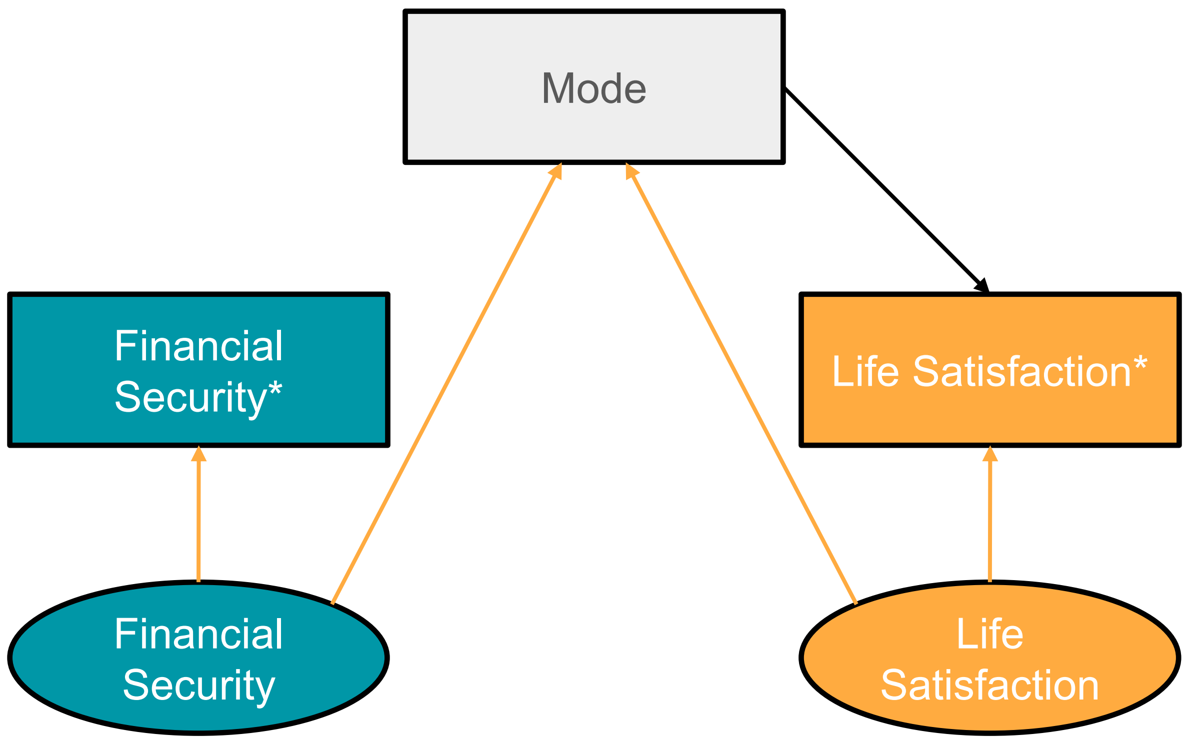
But…Joint Mode Selection



But…Joint Mode Selection

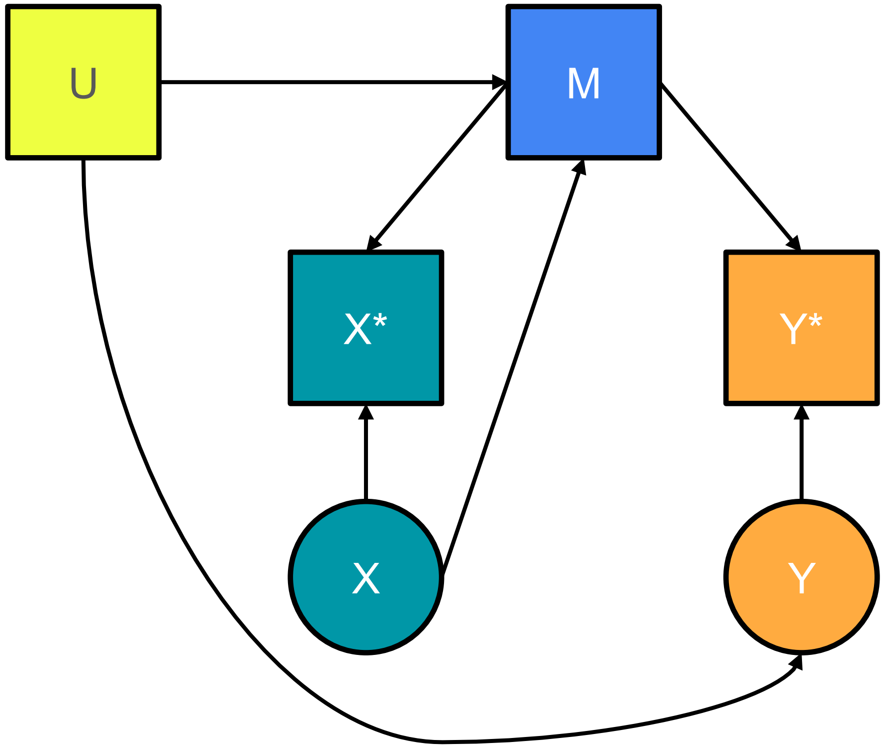


But…Joint Mode Selection



Alternative Approach to Statistical Control

- Can instead estimate the mode effect and use this to predict counterfactuals for those observed in the alternate mode.
    - e.g. Y* ~ M + U
- Then analyse as if observed.
- Bootstrapping the whole thing to appropriately incorporate uncertainty.



Statistical control

Advantages	Disadvantages
• Straightforward method, easily understood and implemented. • Given the richness of variables captured in large social surveys, 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.

Multiple Imputation

Multiple Imputation

- 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.

Multiple Imputation

- 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.

Multiple Imputation - Example

- A battery of mental health questions were asked by telephone (which used an interviewer) and web (which was anonymous).

id	mode	sum
1	tel	9
2	tel	17
3	web	14
4	web	22
5	tel	15
…	…	…

Multiple Imputation - Example

- Values of sum score artificially set to missing for telephone respondents.

id	mode	sum	gender	sc	health	ed	cog
1	tel	.	1	3	2	2	-0.5
2	tel	.	1	1	2	1	0.7
3	web	14	2	2	3	1	0.4
4	web	22	2	2	4	2	1.4
5	tel	.	1	4	5	3	-0.8
…	…	…	…	…	…	…	…

Multiple Imputation - Example

- Develop imputation model for sum score among web respondents.

sum ~ gender + social class + health + education + cognitive ability

id	mode	sum	gender	sc	health	ed	cog
1	tel	.	1	3	2	2	-0.5
2	tel	.	1	1	2	1	0.7
3	web	14	2	2	3	1	0.4
4	web	22	2	2	4	2	1.4
5	tel	.	1	4	5	3	-0.8
…	…	…	…	…	…	…	…

Multiple Imputation - Example

- Apply predictive model to data for telephone respondents, imputing counterfactual sum score.

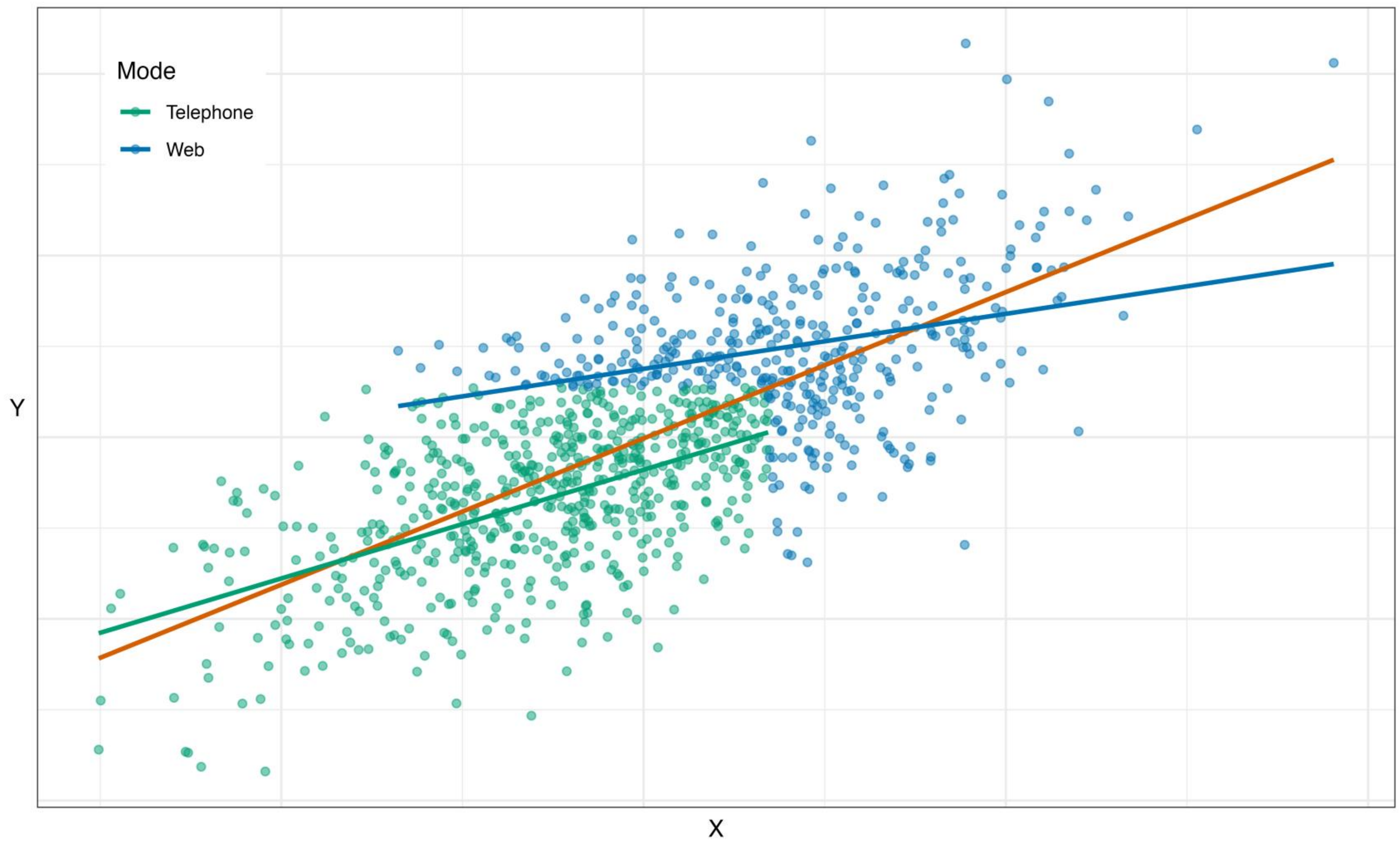
id	mode	sum	gender	sc	health	ed	cog
1	tel	11	1	3	2	2	-0.5
2	tel	18	1	1	2	1	0.7
3	web	14	2	2	3	1	0.4
4	web	22	2	2	4	2	1.4
5	tel	18	1	4	5	3	-0.8
…	…	…	…	…	…	…	…

Multiple Imputation - Example

- Repeat process to create multiple imputed datasets.

id	mode	sum	gender	sc	health	ed	cog
1	tel	8	1	3	2	2	-0.5
2	tel	16	1	1	2	1	0.7
3	web	14	2	2	3	1	0.4
4	web	22	2	2	4	2	1.4
5	tel	14	1	4	5	3	-0.8
…	…	…	…	…	…	…	…

Multiple Imputation Conditions on Mode



Multiple Imputation

Advantages	Disadvantages
• Increasingly commonly used so may already be familiar to researchers. • Easy-to-use functionality in major statistical software. • 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): conditional on the variables used, answering in the alternate mode is not informative about the value of the variables to be imputed – equivalent to requiring that mode selection is correctly accounted for.

Quantitative Bias Analysis for Mode Effects

Quantitative Bias Analysis (QBA)

- Simply knowing there is bias is unsatisfying. How biased? Is it sufficient to change substantive conclusions?
- We may have good evidence on the size of mode effects and mode selection which can inform these judgements.
- Enter QBA: Broad array on methods aimed at correcting for bias or assessing the impact of bias using quantitative information (Fox et al., 2021).
- QBA methods vary in the what the judgements they are allowing people to make, and how quantitative information is inputted or outputted.

QBA Methods

- Here we will focus on two broad approaches:

  1. Simple Sensitivity Analysis: Given an observed association, how strong would mode effects and mode selection have to be to fully explain the association?
  2. Counterfactual Simulation: e.g., Given a hypothesised mode effect and a mixed mode survey, what would a counterfactual single mode survey have produced?

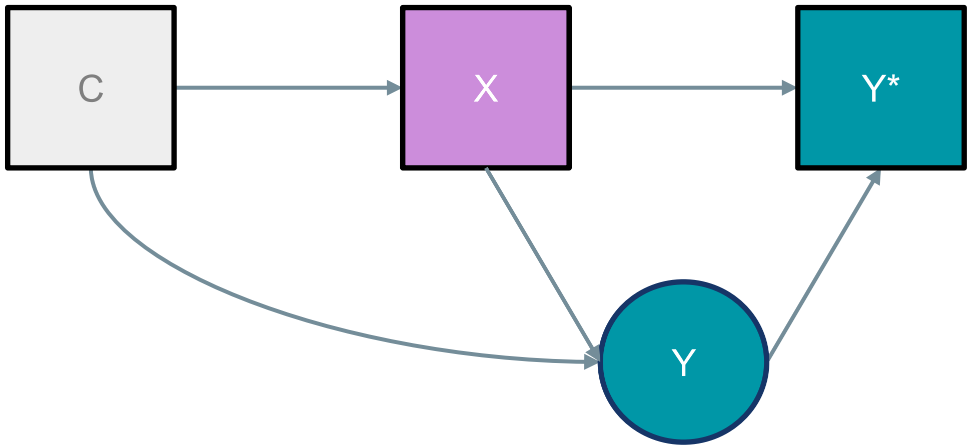
Simple Sensitivity Analysis: Cornfield Conditions

view. 2) There is a quantitative question. Cigarette smokers have a ninefold greater risk of developing lung cancer than nonsmokers, while over-two-pack-a-day smokers have at least a 60-fold greater risk. Any characteristic proposed as a measure of the postulated cause common to both smoking status and lung-cancer risk must therefore be at least ninefold more prevalent among cigarette smokers than among nonsmokers and at least 60-fold more prevalent among two-pack-a-day smokers. No such characteristic has yet been produced despite diligent search.

Cornfield et al. (1959)

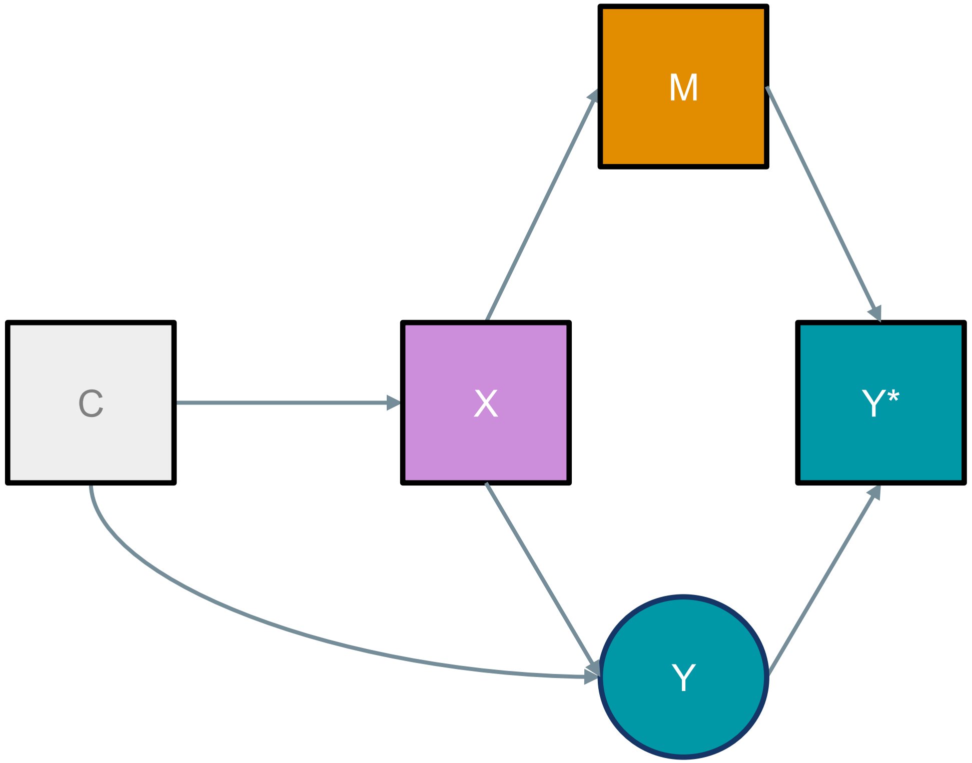
Simple Sensitivity Analysis

- Assuming a differential measurement error mechanism, the size of the mode effect can be bounded by the observed association.
- VanderWeele & Li (2019) derives bounds for cases where exposure and outcome are binary



Simple Sensitivity Analysis

- Assuming a differential measurement error mechanism, the size of the mode effect can be bounded by the observed association.
- VanderWeele & Li (2019) derives bounds for cases where exposure and outcome are binary



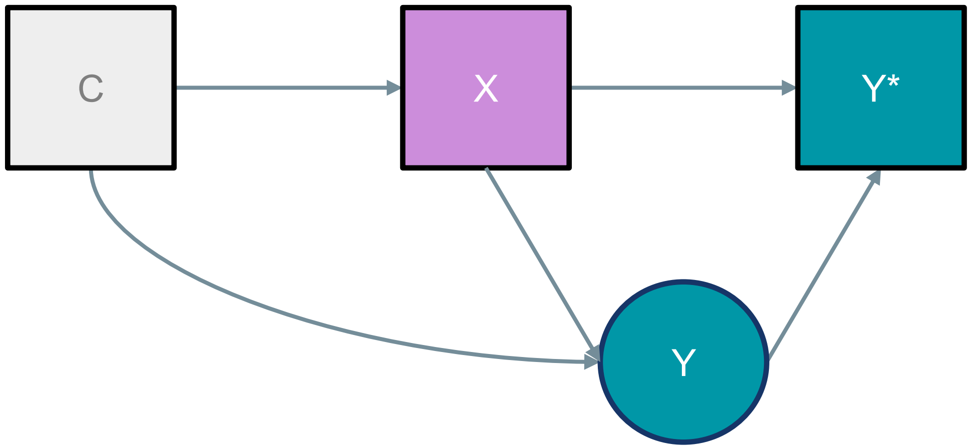
Simple Sensitivity Analysis: VanderWeele & Li (2019)

Where $RR_{XY^*} \geq 1$

- $RR_{XY^*} \leq$ Sensitivity(X = 1)/ Sensitivity(X = 0)

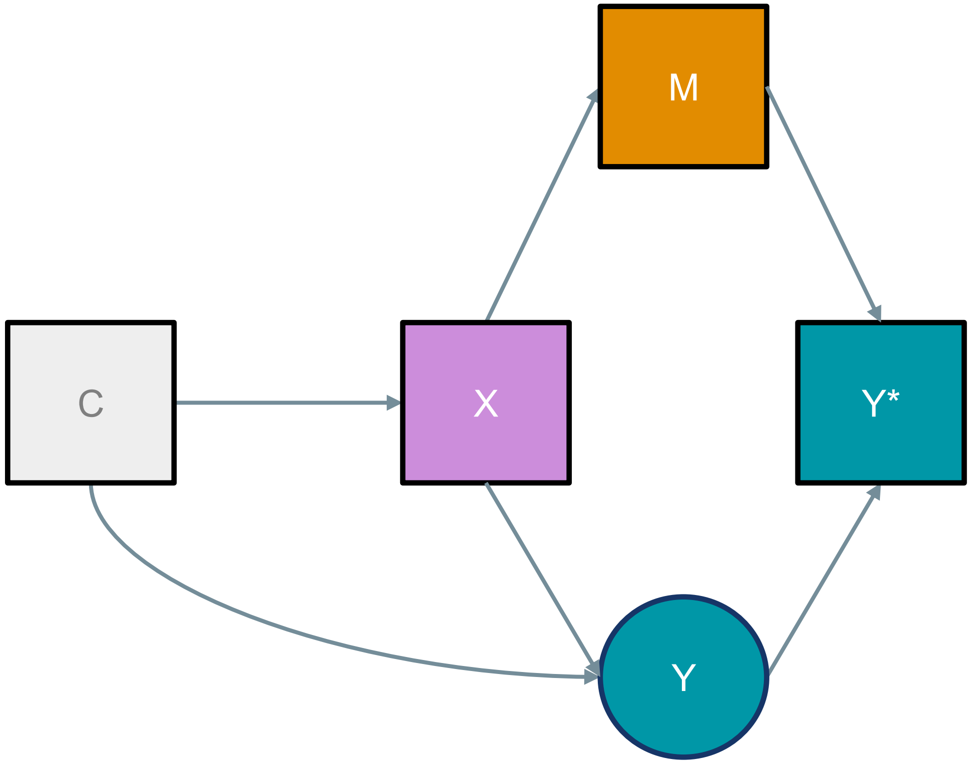
OR …

- $RR_{XY^*} \leq$ False Positive Probability(X = 1)/ False Positive Probability(X = 0)



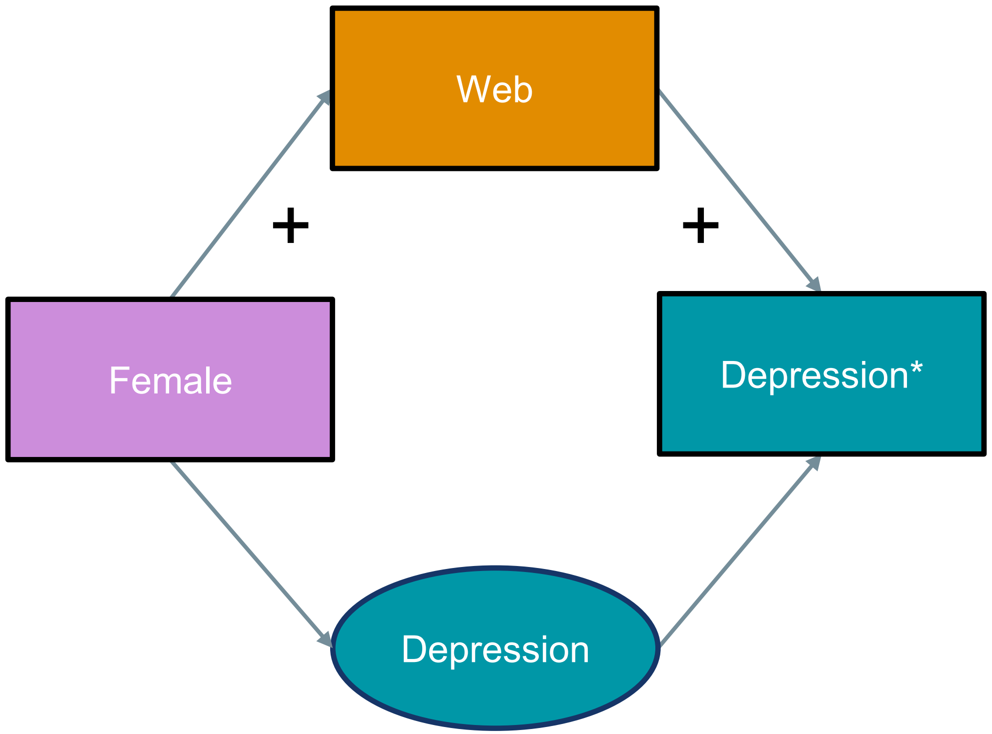
Simple Sensitivity Analysis

- Sensitivity(X = 1) = Sensitivity(M = 1)·Prob(M = 1 | X = 1) + Sensitivity(M = 0)·Prob(M = 0 | X = 1)
- Sensitivity(M = 0) ≤ Sensitivity(X = 1) ≤ Sensitivity(M = 1)
- So … the mode effect would have to be even stronger!

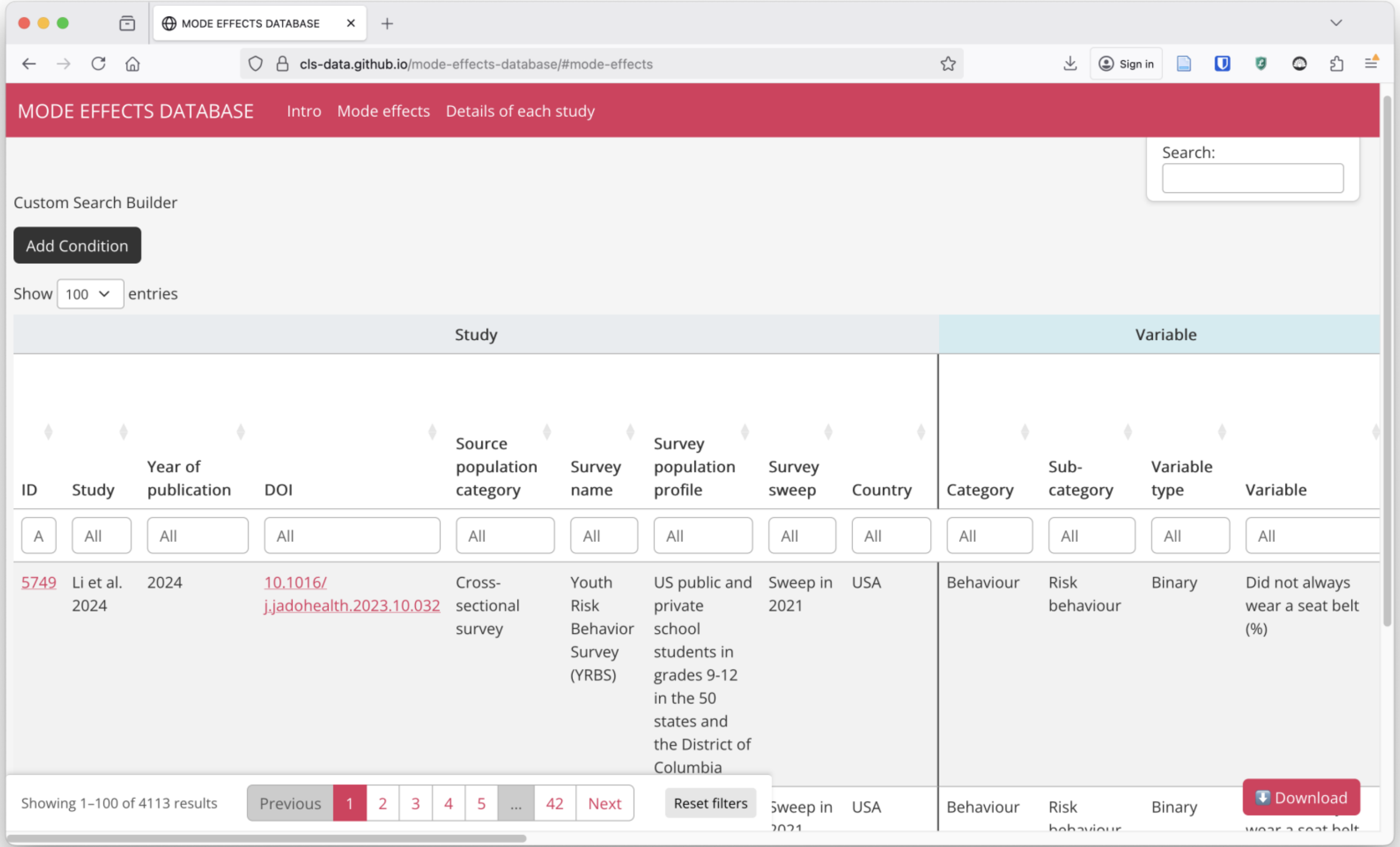


Worked Example: Next Steps Sweep 9

- RR(Depression* | Female) = 1.57 (95% CI = 1.45 – 1.71)
- P(Web | Female) = 0.904
- P(Web | Male) = 0.826
- Goodman et al. (2022): RR(Depression | Web) = 1.23
  - Sensitivity | Web = 1.00 (Assumption)
  - Sensitivity | Other = 0.813 (Corollary)
- Ratio of Sensitivities = 1.015
  - Sensitivity | Female = 1 * 0.904 + 0.813 * (1 – 0.904) = 0.982
  - Sensitivity | Male = 1 * 0.826 + 0.813 * (1 – 0.826) = 0.967



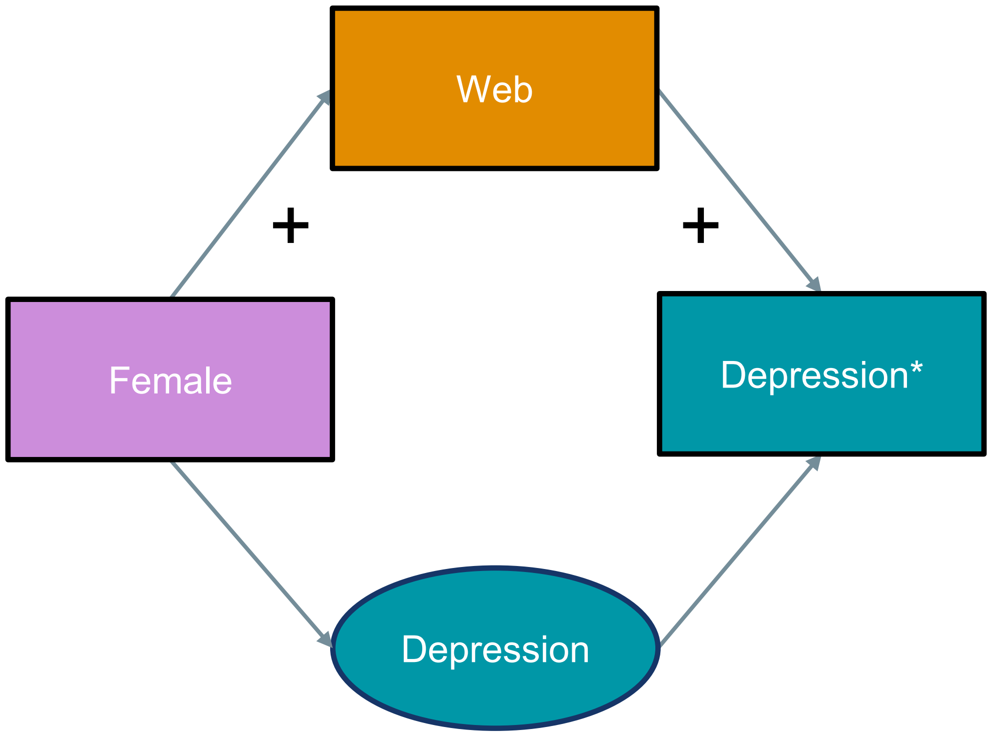
A Database of Mode Effect Estimates



Tomova (2025)

Worked Example: Next Steps Sweep 9

- RR(Depression* | Female) = 1.57 (95% CI = 1.45 – 1.71)
- P(Web | Female) = 0.904
- P(Web | Male) = 0.826
- Goodman et al. ([2022](#)): RR(Depression | Web) = 1.23
  - Sensitivity | Web = 1.00 (Assumption)
  - Sensitivity | Other = 0.813 (Corollary)
- Ratio of Sensitivities = 1.015
  - Sensitivity | Female = 1 * 0.904 + 0.813 * (1 – 0.904) = 0.982
  - Sensitivity | Male = 1 * 0.826 + 0.813 * (1 – 0.826) = 0.967



Counterfactual Simulation – Overview

- We may have some idea how large a mode effect will be
- We could use this to ‘correct’ the observed data
  - What would this person’s mental health score have been if they had responded in the reference mode?
  - Best Answer: True Mental Health = Observed Mental Health – Anticipated Mode Effect
- Substantive models can then be run using these simulated data to examine whether, and to what extent, results change.

Counterfactual Simulation – Process

id	mode	sum
1	tel	11
2	tel	19
3	web	14
4	web	22
5	tel	10
…	…	…

Counterfactual Simulation – Process

id	mode	sum	sum1
1	tel	11	10
2	tel	19	18
3	web	14	14
4	web	22	22
5	tel	10	9
…	…	…	…

Counterfactual Simulation – Process

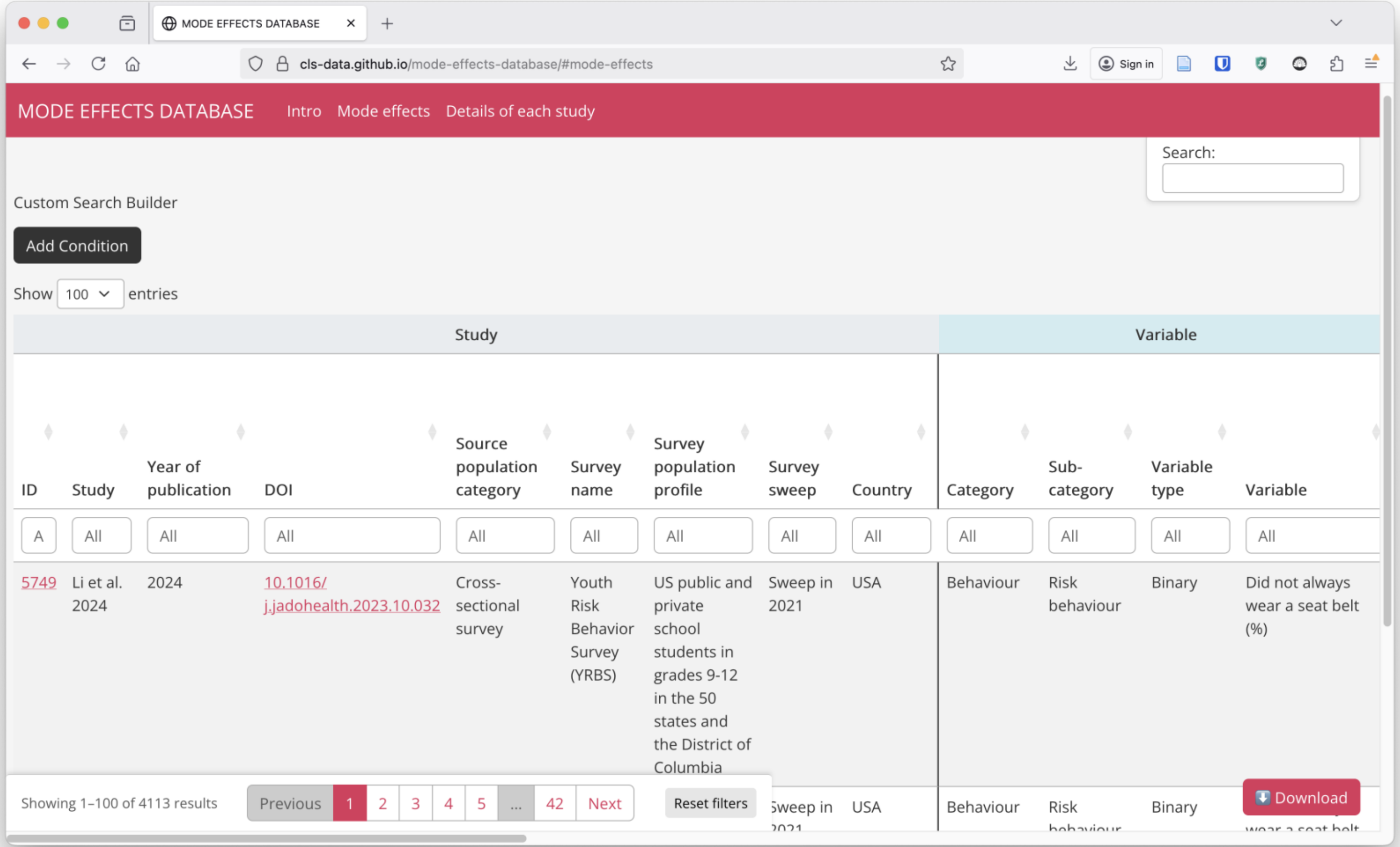
id	mode	sum	female	sum1f
1	tel	11	0	11
2	tel	19	1	18
3	web	14	1	14
4	web	22	0	22
5	tel	10	0	10
…	…	…	…	

Counterfactual Simulation – Overview

How to choose mode effects for sensitivity analysis?

1. Choose single most plausible mode effect to obtain ‘best estimate’.
   - Best estimates are just that – *estimates*. They come with uncertainty which should be propagated.
2. Use fine net of mode effects and examine robustness
   - Fox et al. (2021) provide advice on how to convert statements about uncertainty into distributions.
3. Intentionally choose implausible mode effect.
   - Mode effects > 1 SD are not often seen in the wild.

A Database of Mode Effect Estimates



Tomova (2025)

Counterfactual Simulation – Overview

How to choose mode effects for sensitivity analysis?

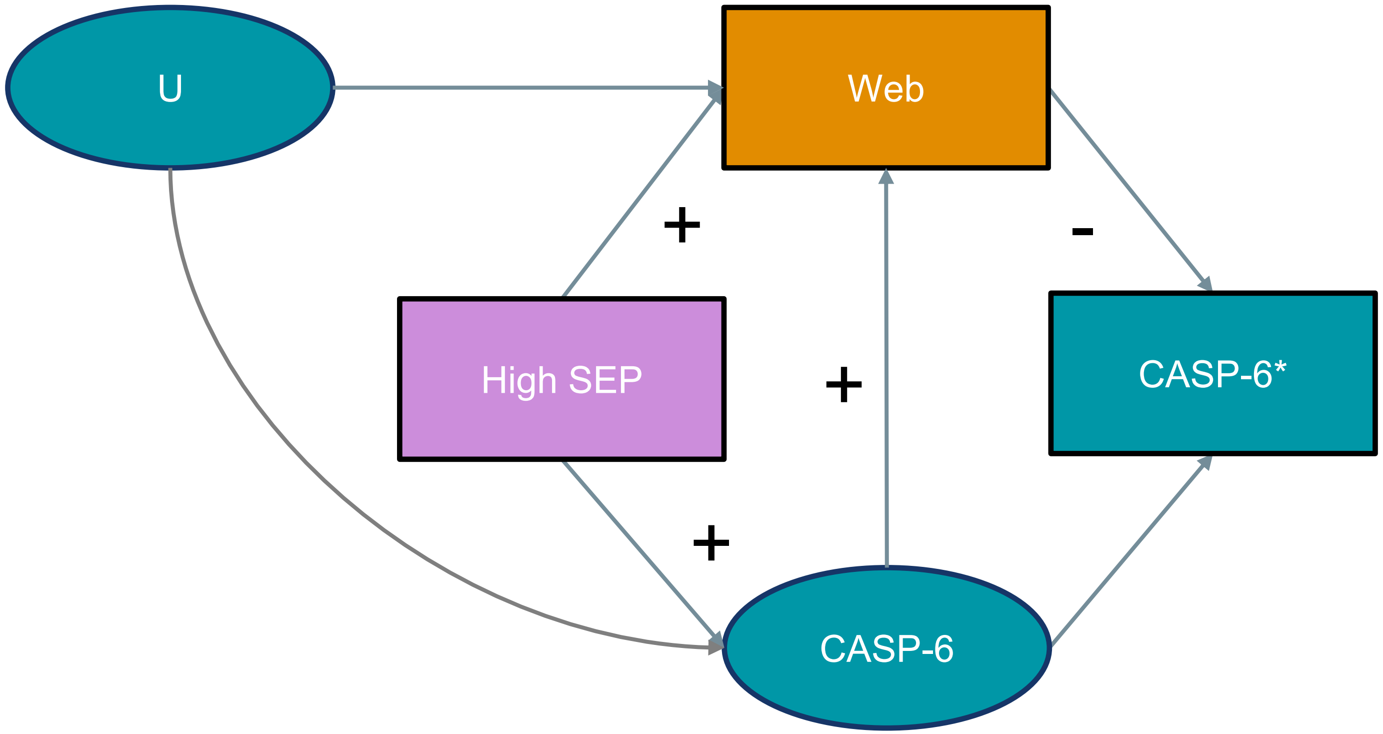
1. Choose single most plausible mode effect to obtain ‘best estimate’.
    - Best estimates are just that – *estimates*. They come with uncertainty which should be propagated.
2. Use fine net of mode effects and examine robustness
    - Fox et al. (2021) provide advice on how to convert statements about uncertainty into distributions.
3. Intentionally choose implausible mode effect.
    - Mode effects > 1 SD are not often seen in the wild.

Quantitative Bias Analysis

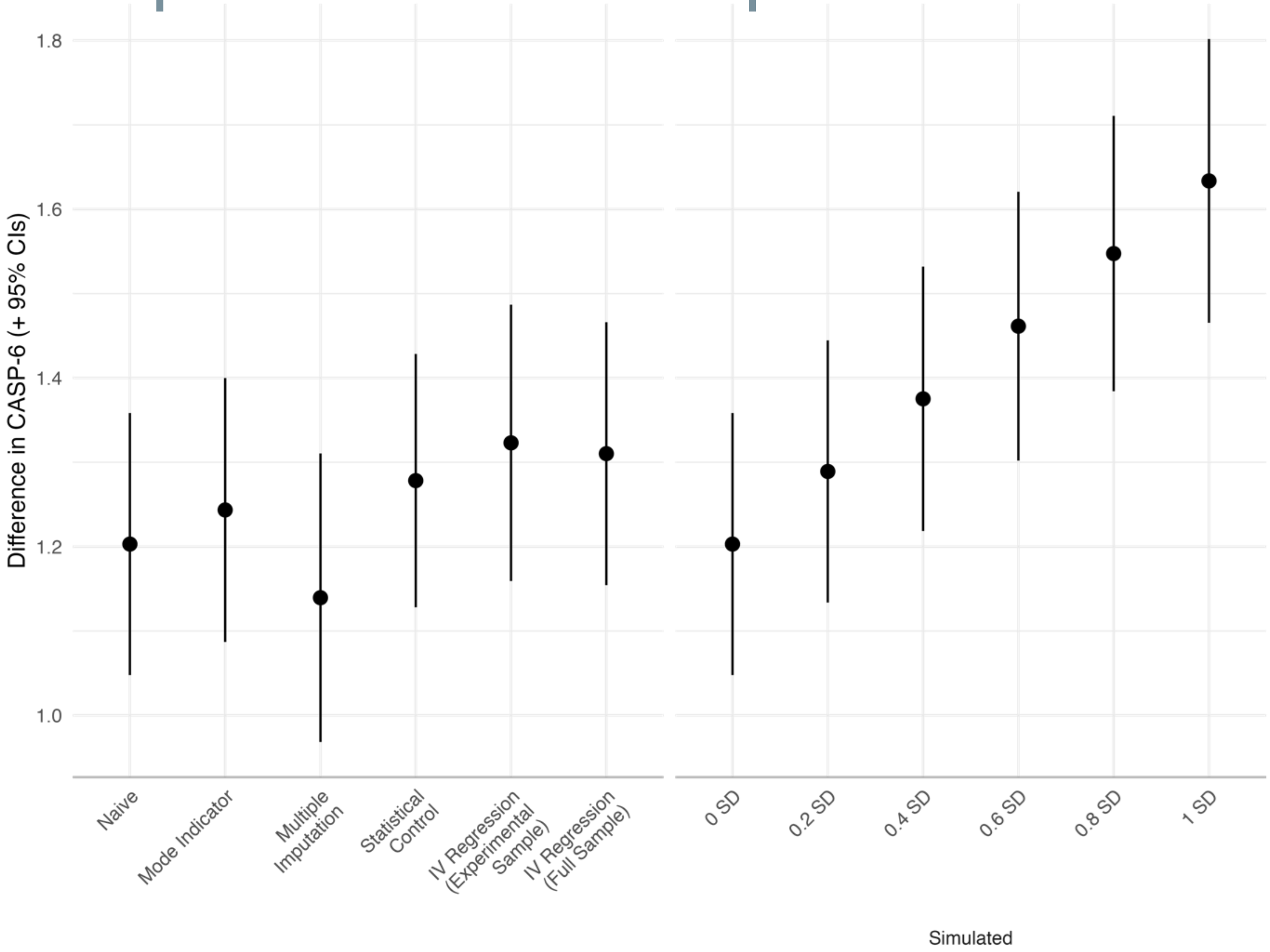
Advantages	Disadvantages
• Detailed understanding of mode selection may not be required. • Can sometimes use all available information, unlike MI. • Flexible approach, e.g., heterogeneity in mode effects, multiple variables subject to mode effects, mixing modes between sweeps, multiplicative errors • Simple sensitivity analysis can be performed *post hoc*	• Plausible mode effects for a given situation may not be known (though, implausible *might* be). • Estimation of mode effects from non-experimental data (if necessary) requires appropriate modelling of mode selection. • No simple out-of-the-box functionality for performing general sensitivity analysis. • For non-continuous variables, MC-SIMEX could be used.

Worked Example

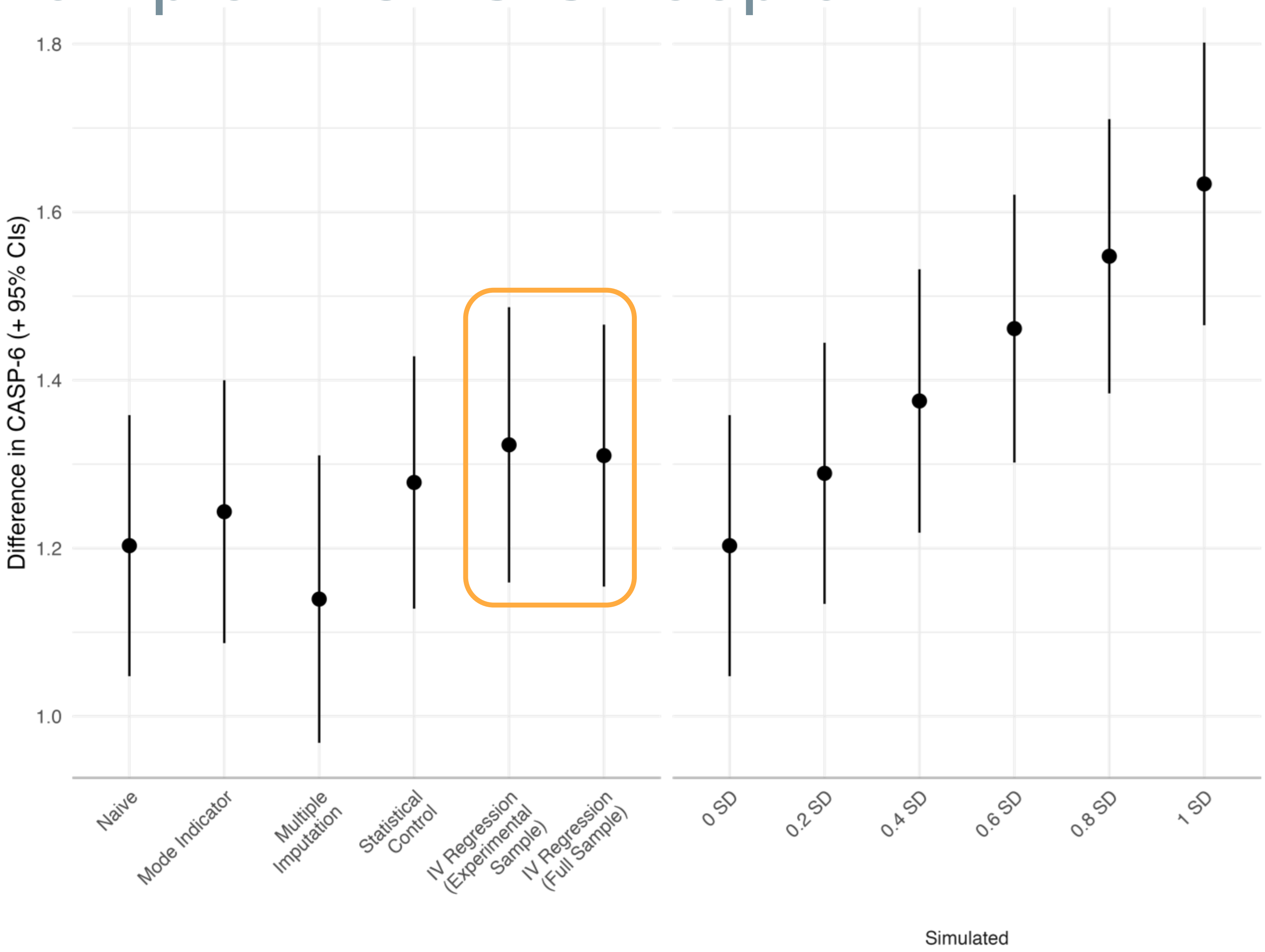
Worked Example: NCDS Sweep 9



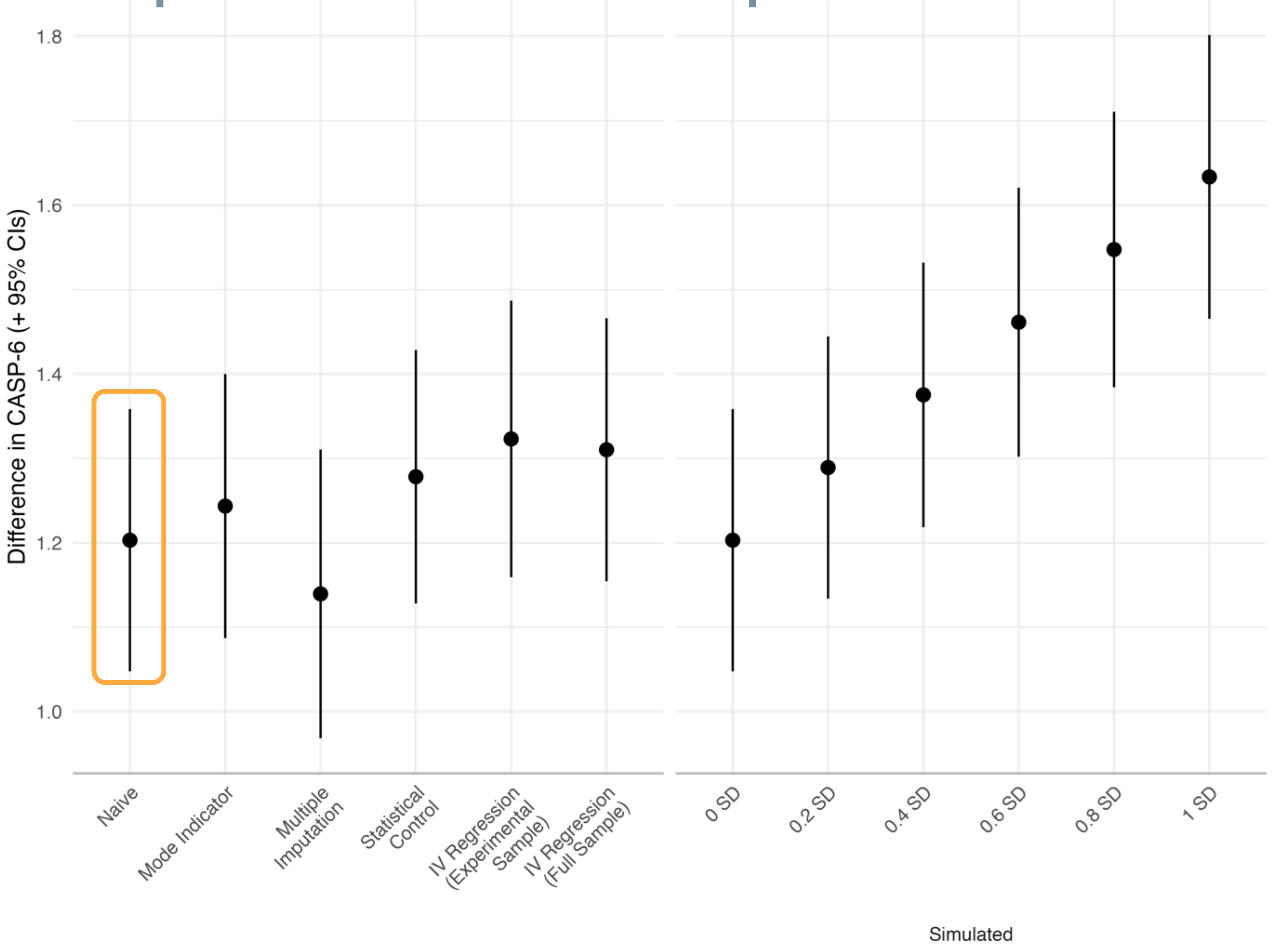
Worked Example: NCDS Sweep 9



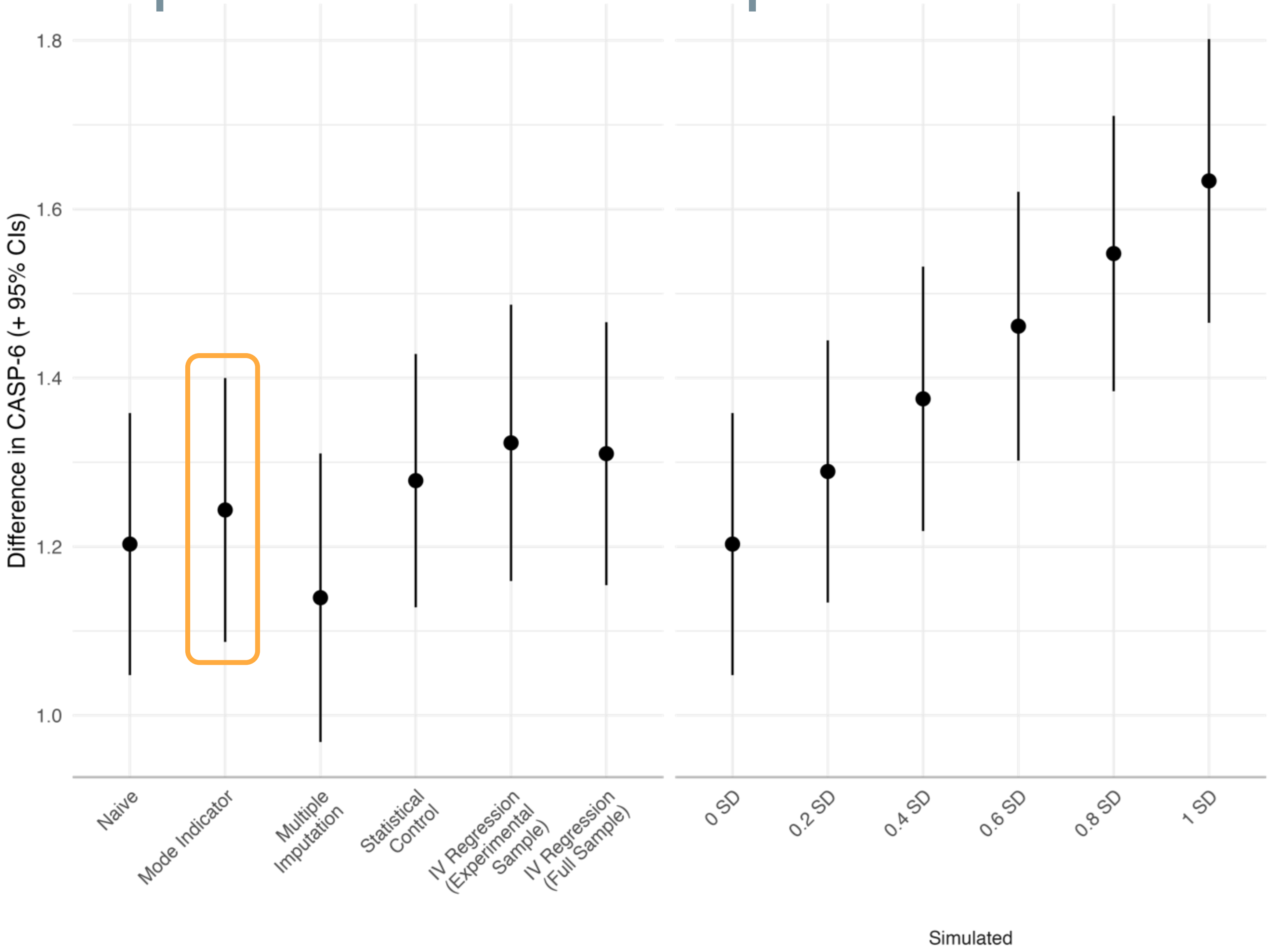
Worked Example: NCDS Sweep 9



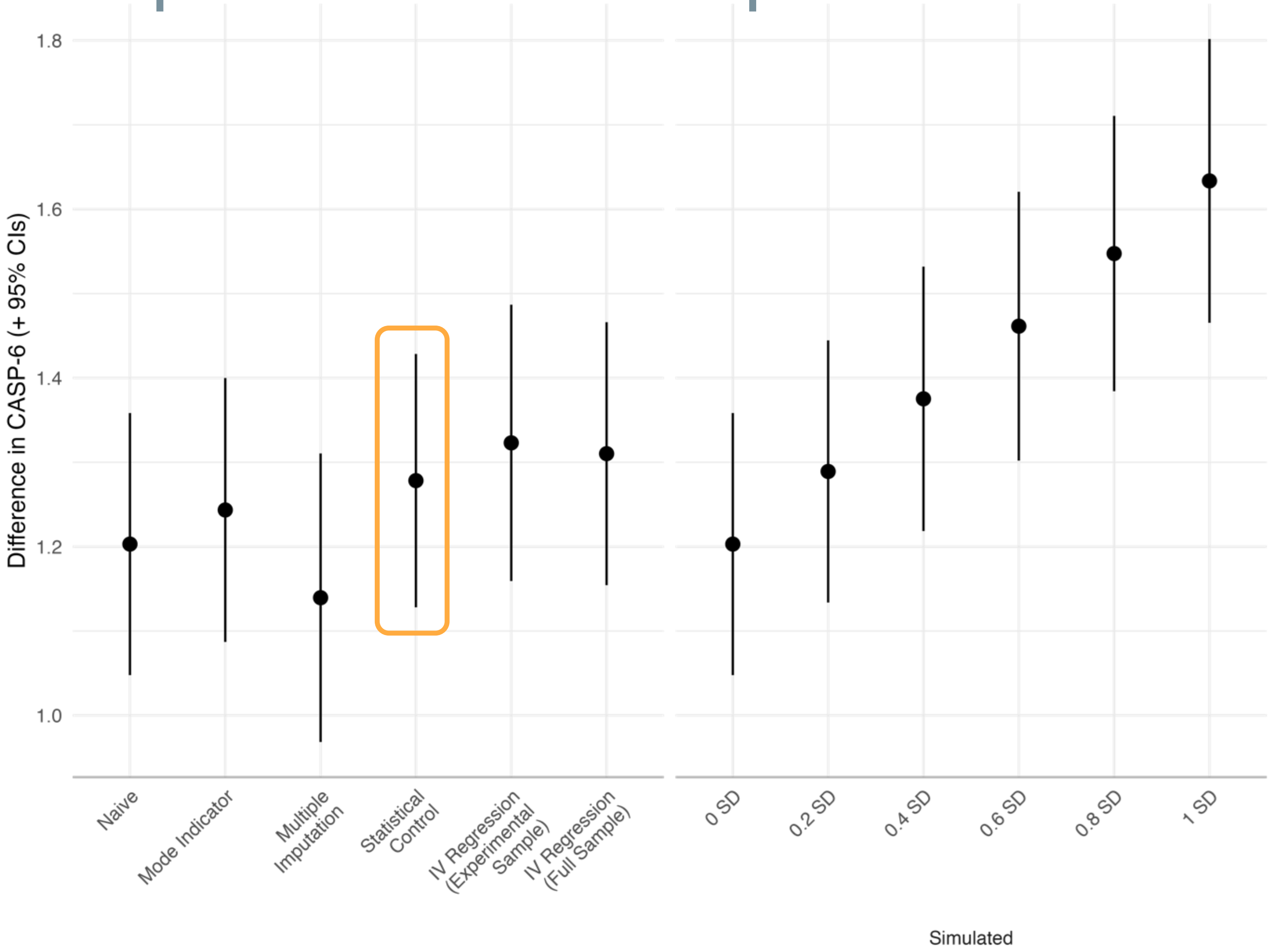
Worked Example: NCDS Sweep 9



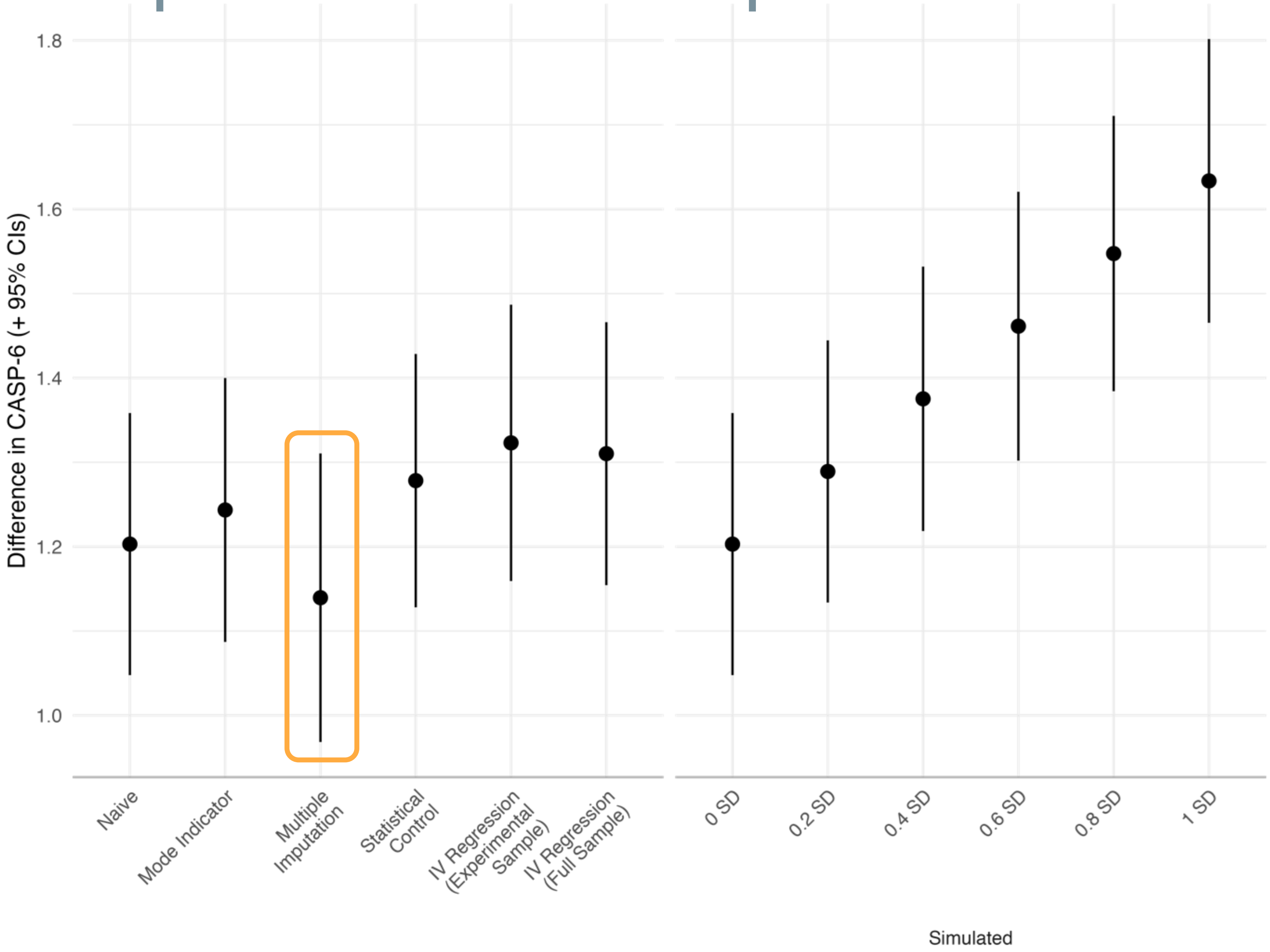
Worked Example: NCDS Sweep 9



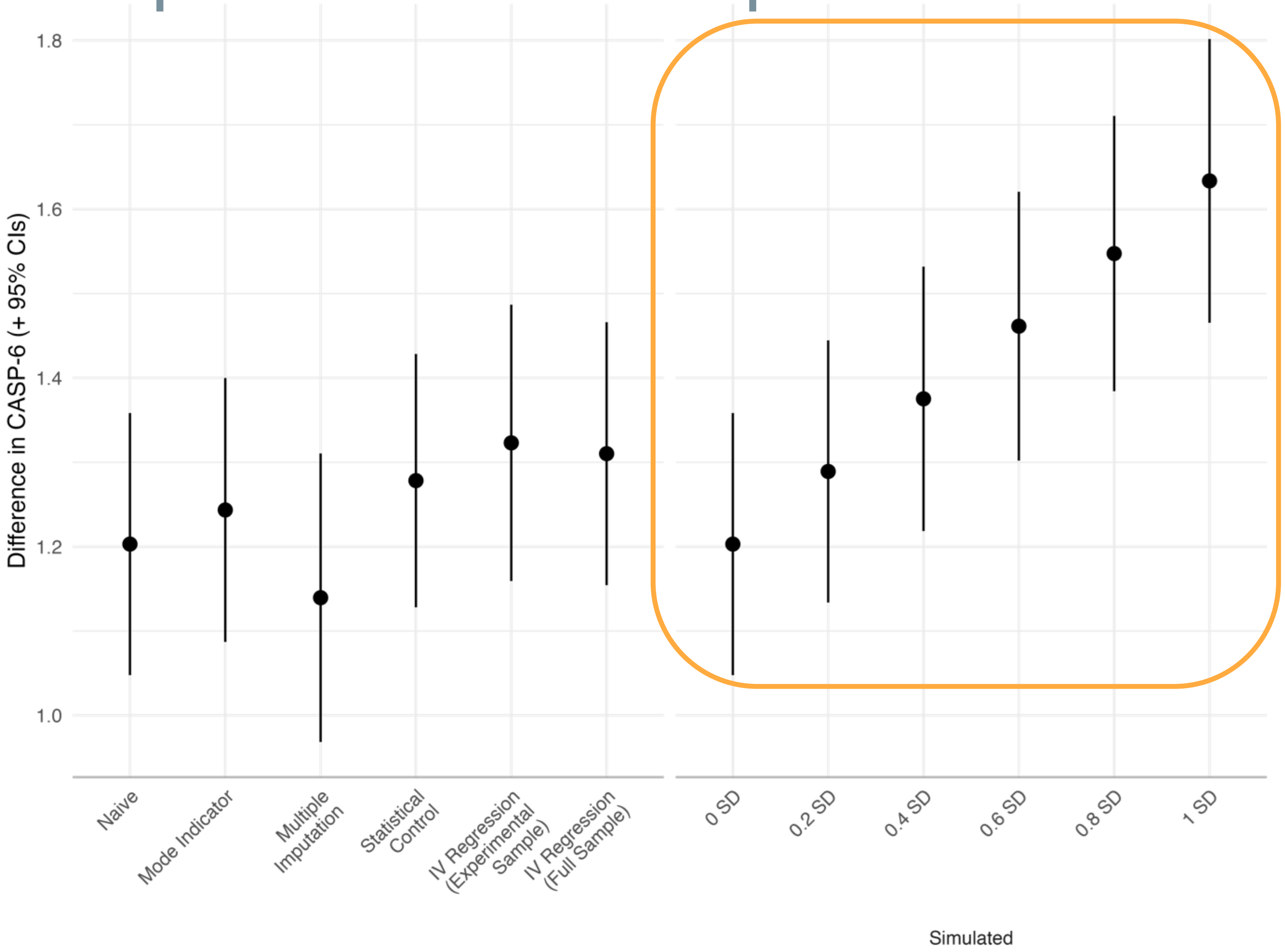
Worked Example: NCDS Sweep 9



Worked Example: NCDS Sweep 9



Worked Example: NCDS Sweep 9



Questions? liam.wright@ucl.ac.uk

CENTRE FOR LONGITUDINAL STUDIES

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

UKRI
Economic and Social Research Council

A systematic review of the experimental literature on mode effects

Georgia Tomova, Richard Silverwood, Liam Wright

RSS Handling Survey Mode Effects 2025

CENTRE FOR LONGITUDINAL STUDIES

Economic and Social Research Council

Background

- Several proposed approaches for handling mode effects exist
- Many require an unrealistic assumption that **mode selection** is either not present or can be fully controlled using available data
- **Quantitative bias analysis** offers an alternative that does not require this assumption
- However, it requires some existing information on what plausible mode effects might be

Background

- Although many mode effect experiments have been published, they may not be straightforward to locate and extract, potentially hindering utilization
- There is also **limited evidence synthesis** on mode effects
- Existing systematic reviews focus on specific causes of mode effects, specific variables, mode comparisons, or other outcomes

Aims

Aim 1. Systematically review the literature to identify survey mode effect estimates:

- obtained using experimental and quasi-experimental designs
- for survey items relevant to health and social science
- conducted within existing surveys, in the general population, or a sex-, age-, or region- specific stratum of the population

Aim 2. Synthesise the findings of the systematic review into a searchable **database of mode effect estimates**.

The study was pre-registered on the Open Science Framework (osf.io/bs5dw) after a pilot search, screen, and extraction.

Search

Stage 1. We searched the following **bibliographic databases**:

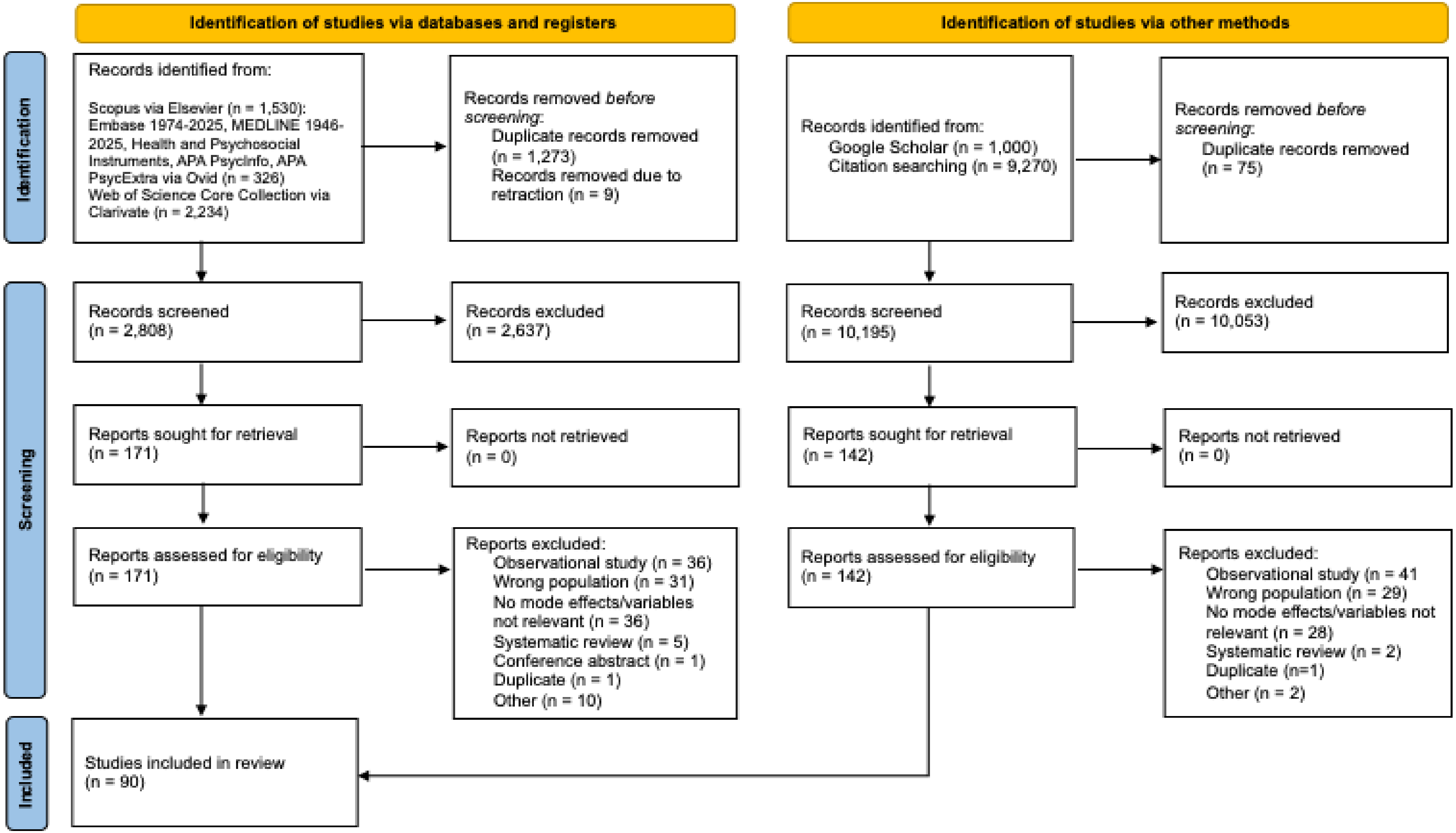
- Scopus (via Elsevier)
- Embase 1974-2025, MEDLINE 1946-2025, Health and Psychosocial Instruments, APA PsycInfo, APA PsycExtra (via Ovid)
- Web of Science Core Collection (via Clarivate)

Stage 2. And the following **alternative sources**:

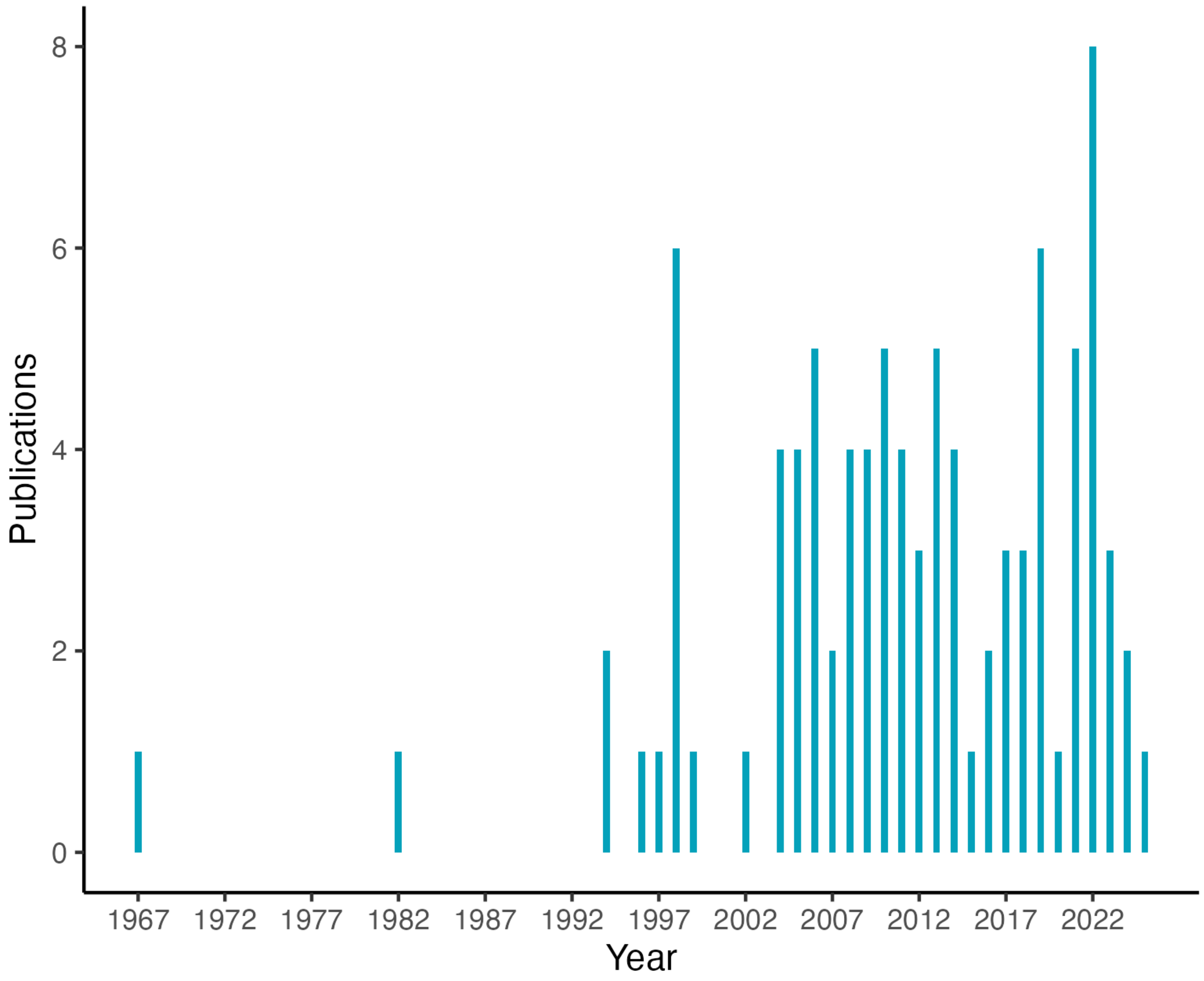
- Google Scholar (100 pages)
- Backwards and forwards citation screen of articles included in the first stage

Inclusion and exclusion

Inclusion criteria	Exclusion criteria
➢ report mode effect estimates (or mode-specific estimates) on survey item measurements ➢ experimental or quasi-experimental design ➢ health and social science domain ➢ sample from the general population or age-, sex-, or region-specific strata of the population ➢ in English ➢ published any time since database inception	➢ mode effect estimate reported for response rate, change over time, or association ➢ study is observational (unless adopting a quasi-experimental design) ➢ sample from a population defined by clinical, occupational or other characteristics, except those specified in the inclusion criteria ➢ no full text available


Identification of studies via databases and registers
Identification of studies via other methods
Identification
Screening
Included
Records identified from:
Scopus via Elsevier (n = 1,530):
Embase 1974-2025, MEDLINE 1946-2025, Health and Psychosocial Instruments, APA PsycInfo, APA PsycExtra via Ovid (n = 326)
Web of Science Core Collection via Clarivate (n = 2,234)
Records removed *before screening*:
Duplicate records removed (n = 1,273)
Records removed due to retraction (n = 9)
Records screened (n = 2,808)
Records excluded (n = 2,637)
Reports sought for retrieval (n = 171)
Reports not retrieved (n = 0)
Reports assessed for eligibility (n = 171)
Reports excluded:
Observational study (n = 36)
Wrong population (n = 31)
No mode effects/variables not relevant (n = 36)
Systematic review (n = 5)
Conference abstract (n = 1)
Duplicate (n = 1)
Other (n = 10)
Studies included in review (n = 90)
Records identified from:
Google Scholar (n = 1,000)
Citation searching (n = 9,270)
Records removed *before screening*:
Duplicate records removed (n = 75)
Records screened (n = 10,195)
Records excluded (n = 10,053)
Reports sought for retrieval (n = 142)
Reports not retrieved (n = 0)
Reports assessed for eligibility (n = 142)
Reports excluded:
Observational study (n = 41
Wrong population (n = 29)
No mode effects/variables not relevant (n = 28)
Systematic review (n = 2)
Duplicate (n=1)
Other (n = 2)

Publications timeline



Data extraction

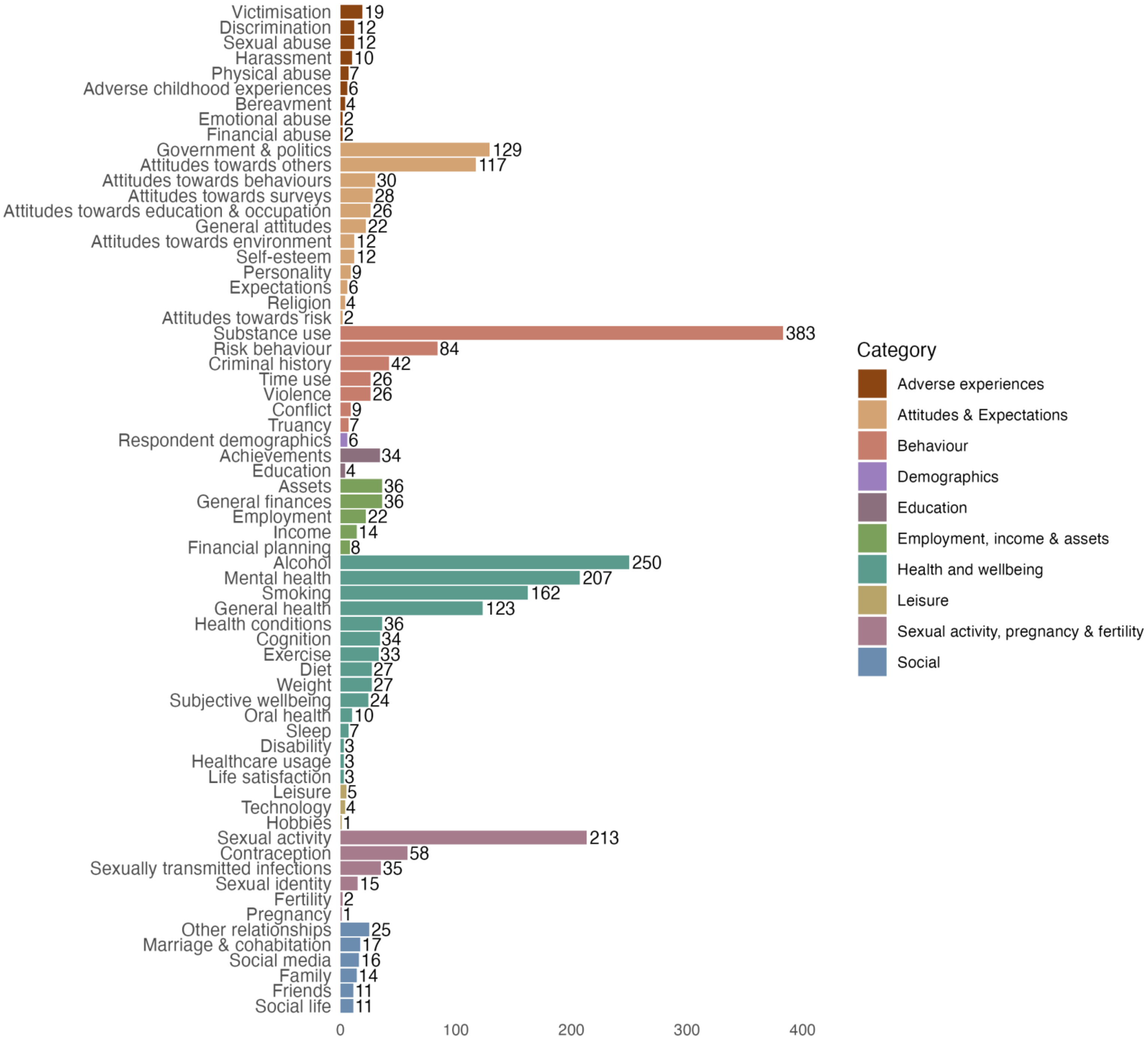
General	Study design	Mode effect estimates	Appraisal
• Year of publication • Authors • Title • Journal (or repository) • DOI	• Source population category • Survey population profile • Survey name • Survey sweep • Country • Sampling approach • Experimental study design • Modes • Response rate (overall, by mode) • Post-response exclusions • Compliance • Sample size	• Category/sub-category • Item name • Reference/alternate mode • Estimand • Effect measure • Mode effect estimate • Standard error and confidence interval • Standardised effect size • Standard error and confidence interval For each mode arm: • Item response rate • Item mean/median • Item SD	• Quality of reporting • Selection • Item non-response • General/other comments

Populations

Country	N studies
USA	41
UK	6
the Netherlands	6
Sweden	4
Germany	3
Italy	3
Spain	3
Switzerland	3
Belgium	2
Australia, Botswana, Canada, China, Denmark, Hungary, Japan, Kenya, Lithuania, Malawi, South Korea, Taiwan, Tanzania, Turkey, Vietnam, Zimbabwe	1 each
International	3

Source	N papers
Survey members	
Longitudinal survey	25
Cross-sectional survey	21
General population	
Adults	23
Adolescents	12
Adolescents and young adults	4
Children	2
Older adults	1
Other	
Large-scale educational assessment	2

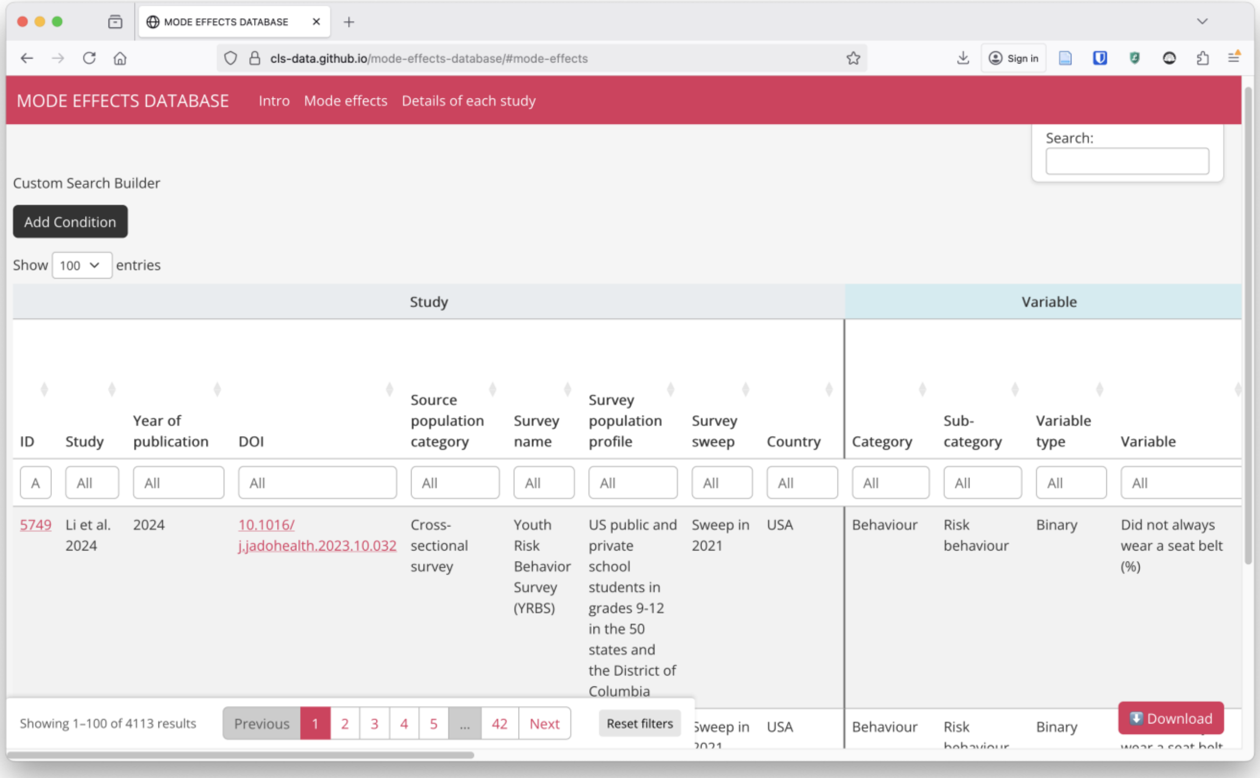
Items



Victimisation 19
Discrimination 12
Sexual abuse 12
Harassment 10
Physical abuse 7
Adverse childhood experiences 6
Bereavment 4
Emotional abuse 2
Financial abuse 2
Government & politics 129
Attitudes towards others 117
Attitudes towards behaviours 30
Attitudes towards surveys 28
Attitudes towards education & occupation 26
General attitudes 22
Attitudes towards environment 12
Self-esteem 12
Personality 9
Expectations 6
Religion 4
Attitudes towards risk 2
Substance use 383
Risk behaviour 84
Criminal history 42
Time use 26
Violence 26
Conflict 9
Truancy 7
Respondent demographics 6
Achievements 34
Education 4
Assets 36
General finances 36
Employment 22
Income 14
Financial planning 8
Alcohol 250
Mental health 207
Smoking 162
General health 123
Health conditions 36
Cognition 34
Exercise 33
Diet 27
Weight 27
Subjective wellbeing 24
Oral health 10
Sleep 7
Disability 3
Healthcare usage 3
Life satisfaction 3
Leisure 5
Technology 4
Hobbies 1
Sexual activity 213
Contraception 58
Sexually transmitted infections 35
Sexual identity 15
Fertility 2
Pregnancy 1
Other relationships 25
Marriage & cohabitation 17
Social media 16
Family 14
Friends 11
Social life 11
0
100
200
300
400
Category
Adverse experiences
Attitudes & Expectations
Behaviour
Demographics
Education
Employment, income & assets
Health and wellbeing
Leisure
Sexual activity, pregnancy & fertility
Social

Database

https://cls-data.github.io/mode-effects-database/



- **4,113** mode effect estimates in total
- **3,545** unique items
  (i.e. excluding items analysed using more than one effect measure or estimand)

Mode effects

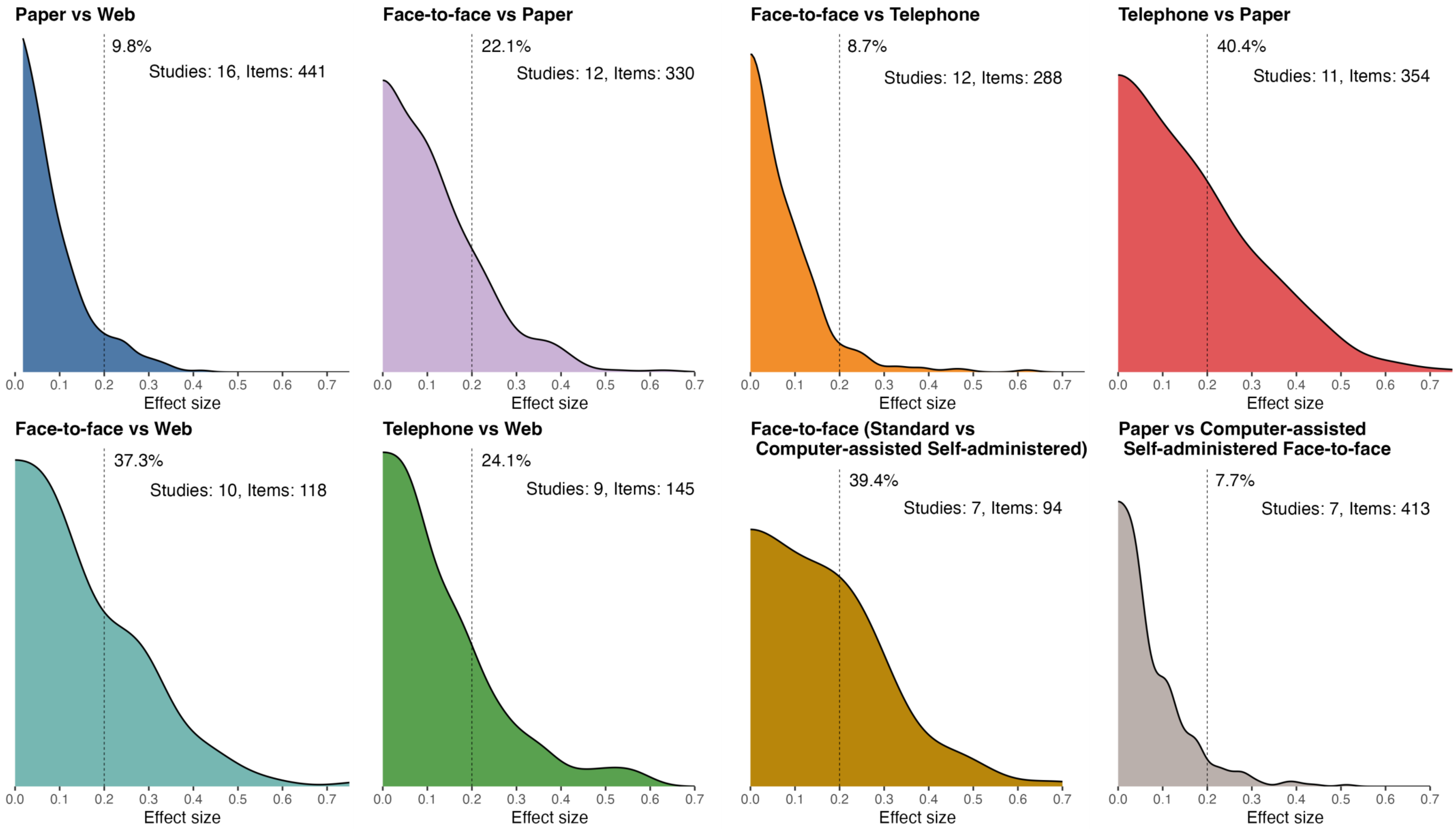
We aggregated and compared modes in two different ways:

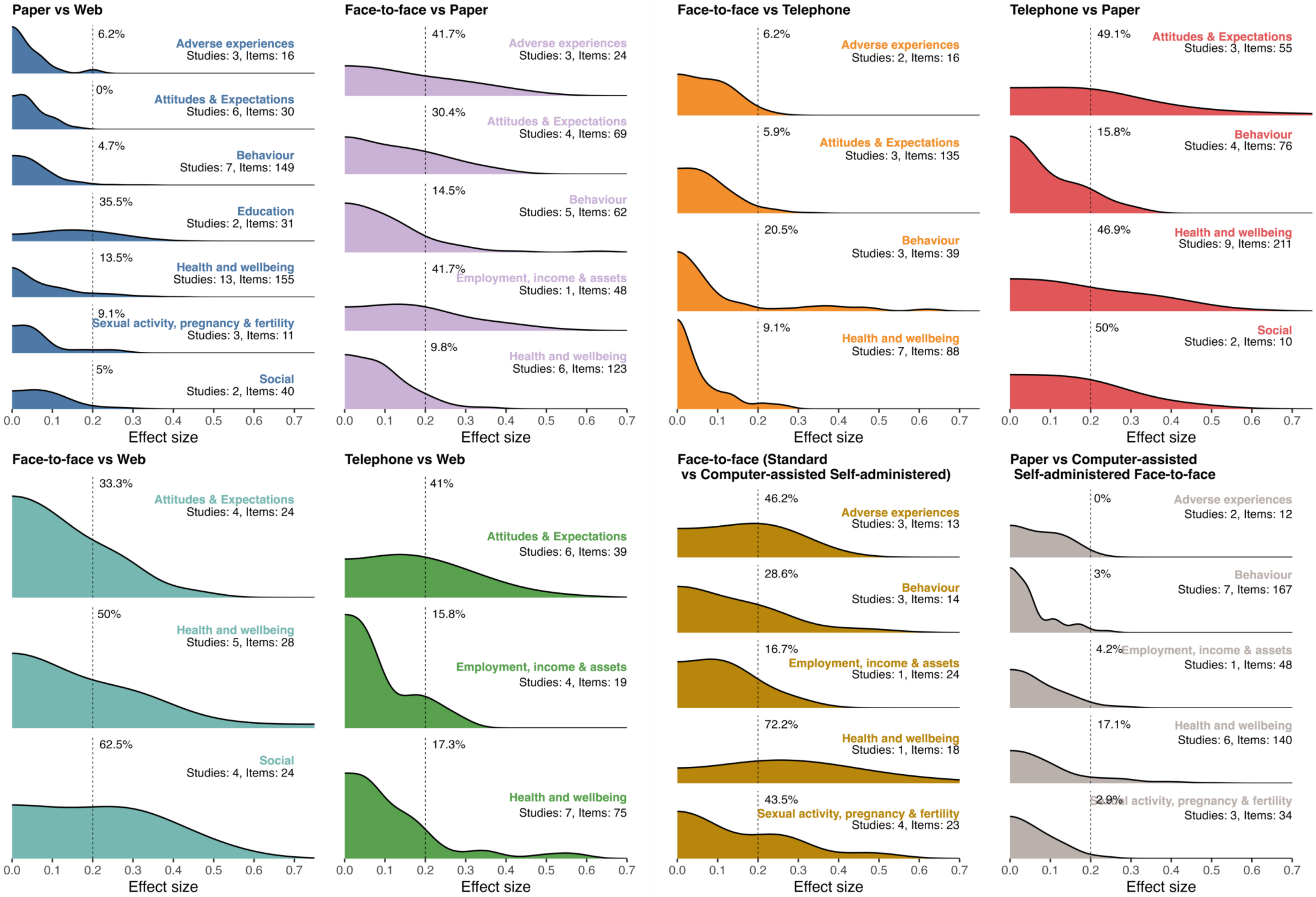
Mode groups	Mode characteristics
Face-to-face	Interviewer presence
Face-to-face – (A)CASI	Written vs aural questions
Paper	Computer-assisted vs not
Telephone	Reported to an interviewer vs not
Web	Interviewer: in-person or phone
Mobile	Self-reported: paper vs web
Other	

Mode comparisons

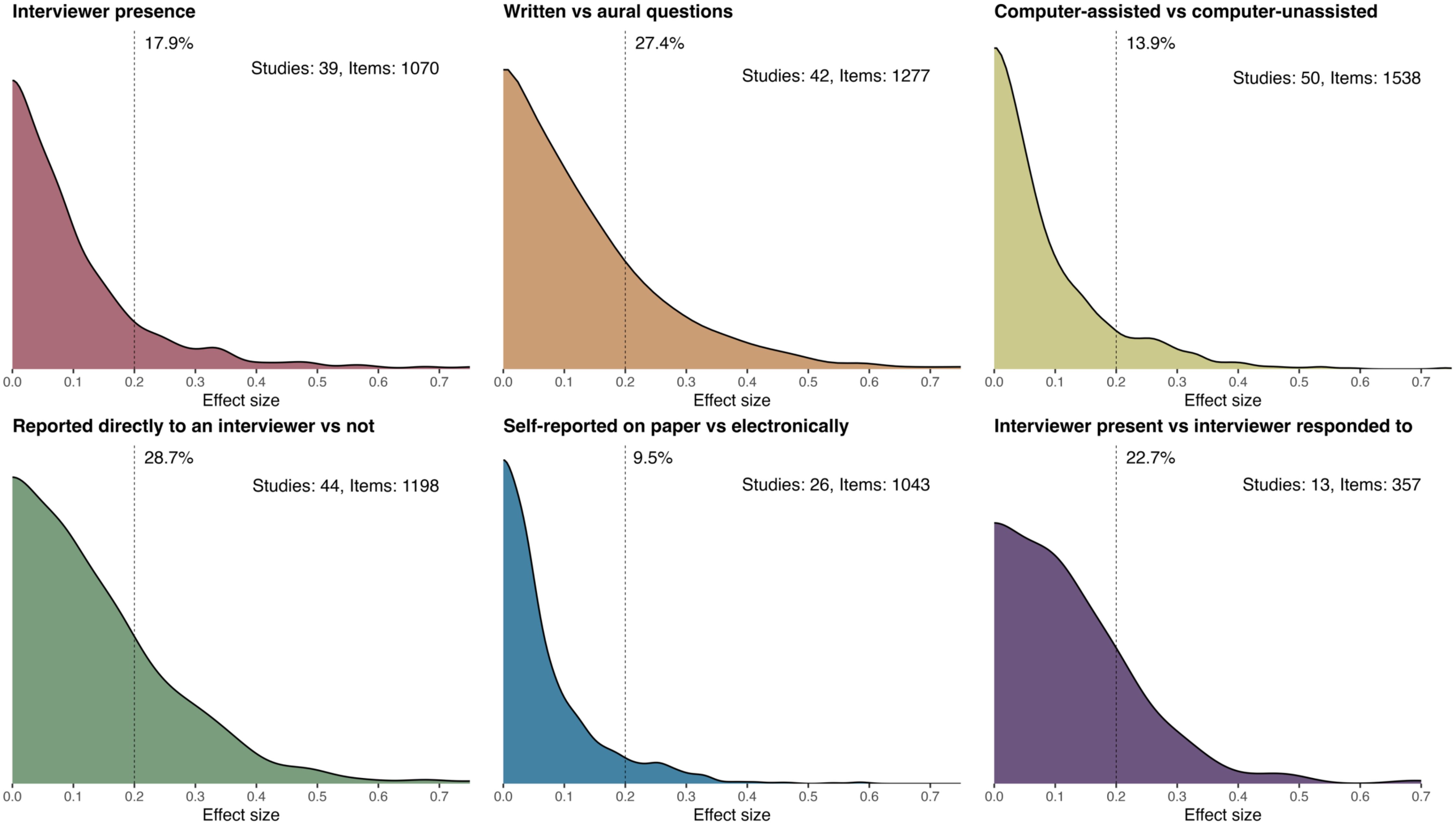
- Only included **standardised** mode effects (available for 84.2% of estimates)
- Only considered them in terms of **absolute** size
- Indicated proportion above 0.2 SD (“small” effect size)

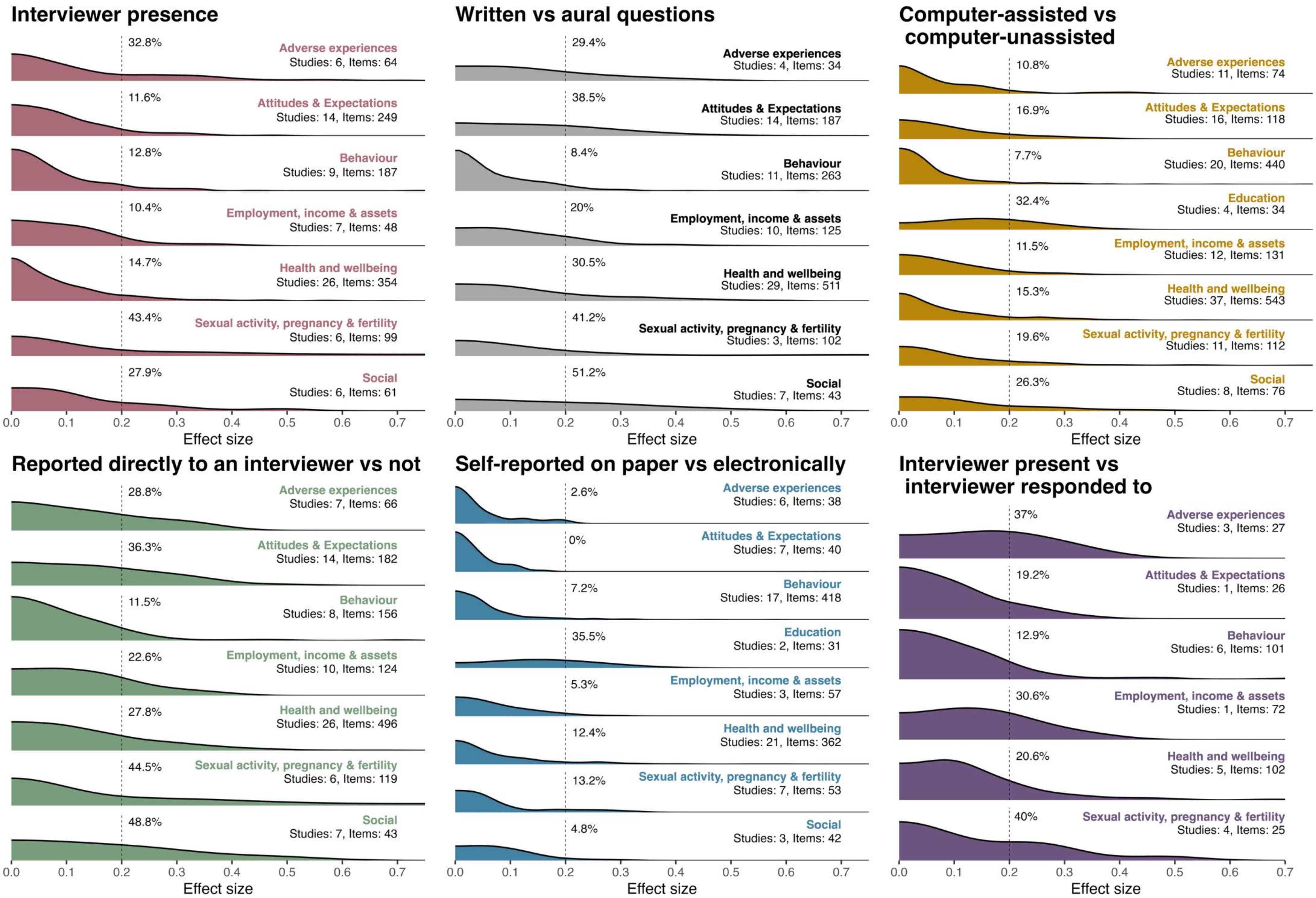
Mode effects



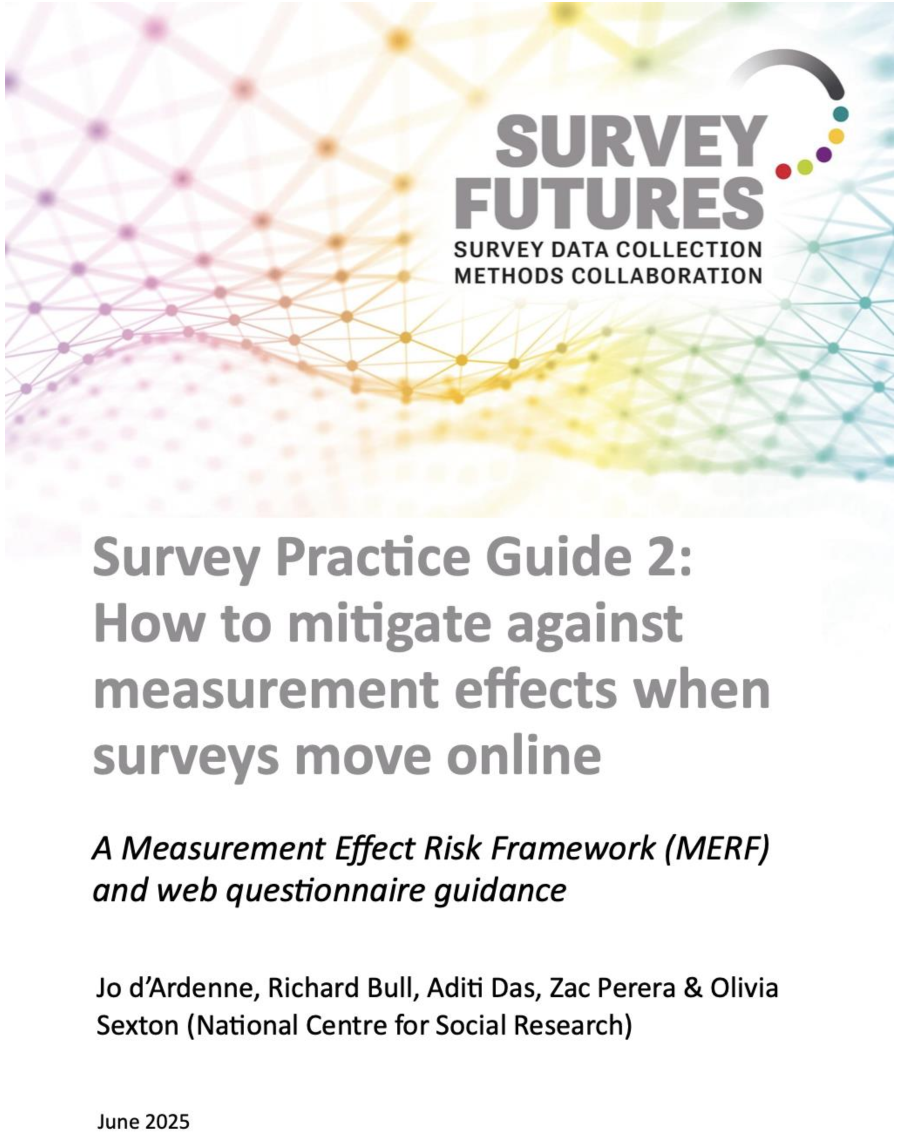


Paper vs Web
6.2%
Adverse experiences
Studies: 3, Items: 16
0%
Attitudes & Expectations
Studies: 6, Items: 30
4.7%
Behaviour
Studies: 7, Items: 149
35.5%
Education
Studies: 2, Items: 31
13.5%
Health and wellbeing
Studies: 13, Items: 155
9.1%
Sexual activity, pregnancy & fertility
Studies: 3, Items: 11
5%
Social
Studies: 2, Items: 40
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Effect size
Face-to-face vs Paper
41.7%
Adverse experiences
Studies: 3, Items: 24
30.4%
Attitudes & Expectations
Studies: 4, Items: 69
14.5%
Behaviour
Studies: 5, Items: 62
41.7%
Employment, income & assets
Studies: 1, Items: 48
9.8%
Health and wellbeing
Studies: 6, Items: 123
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Effect size
Face-to-face vs Telephone
6.2%
Adverse experiences
Studies: 2, Items: 16
5.9%
Attitudes & Expectations
Studies: 3, Items: 135
20.5%
Behaviour
Studies: 3, Items: 39
9.1%
Health and wellbeing
Studies: 7, Items: 88
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Effect size
Telephone vs Paper
49.1%
Attitudes & Expectations
Studies: 3, Items: 55
15.8%
Behaviour
Studies: 4, Items: 76
46.9%
Health and wellbeing
Studies: 9, Items: 211
50%
Social
Studies: 2, Items: 10
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Effect size
Face-to-face vs Web
33.3%
Attitudes & Expectations
Studies: 4, Items: 24
50%
Health and wellbeing
Studies: 5, Items: 28
62.5%
Social
Studies: 4, Items: 24
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Effect size
Telephone vs Web
41%
Attitudes & Expectations
Studies: 6, Items: 39
15.8%
Employment, income & assets
Studies: 4, Items: 19
17.3%
Health and wellbeing
Studies: 7, Items: 75
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Effect size
Face-to-face (Standard
vs Computer-assisted Self-administered)
46.2%
Adverse experiences
Studies: 3, Items: 13
28.6%
Behaviour
Studies: 3, Items: 14
16.7%
Employment, income & assets
Studies: 1, Items: 24
72.2%
Health and wellbeing
Studies: 1, Items: 18
43.5%
Sexual activity, pregnancy & fertility
Studies: 4, Items: 23
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Effect size
Paper vs Computer-assisted
Self-administered Face-to-face
0%
Adverse experiences
Studies: 2, Items: 12
3%
Behaviour
Studies: 7, Items: 167
4.2%
Employment, income & assets
Studies: 1, Items: 48
17.1%
Health and wellbeing
Studies: 6, Items: 140
2.9%
al activity, pregnancy & fertility
Studies: 3, Items: 34
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Effect size


Interviewer presence
17.9%
Studies: 39, Items: 1070
Written vs aural questions
27.4%
Studies: 42, Items: 1277
Computer-assisted vs computer-unassisted
13.9%
Studies: 50, Items: 1538
Reported directly to an interviewer vs not
28.7%
Studies: 44, Items: 1198
Self-reported on paper vs electronically
9.5%
Studies: 26, Items: 1043
Interviewer present vs interviewer responded to
22.7%
Studies: 13, Items: 357
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Effect size


Interviewer presence
32.8%
Adverse experiences
Studies: 6, Items: 64
11.6%
Attitudes & Expectations
Studies: 14, Items: 249
12.8%
Behaviour
Studies: 9, Items: 187
10.4%
Employment, income & assets
Studies: 7, Items: 48
14.7%
Health and wellbeing
Studies: 26, Items: 354
43.4%
Sexual activity, pregnancy & fertility
Studies: 6, Items: 99
27.9%
Social
Studies: 6, Items: 61
Effect size
Written vs aural questions
29.4%
Adverse experiences
Studies: 4, Items: 34
38.5%
Attitudes & Expectations
Studies: 14, Items: 187
8.4%
Behaviour
Studies: 11, Items: 263
20%
Employment, income & assets
Studies: 10, Items: 125
30.5%
Health and wellbeing
Studies: 29, Items: 511
41.2%
Sexual activity, pregnancy & fertility
Studies: 3, Items: 102
51.2%
Social
Studies: 7, Items: 43
Effect size
Computer-assisted vs computer-unassisted
10.8%
Adverse experiences
Studies: 11, Items: 74
16.9%
Attitudes & Expectations
Studies: 16, Items: 118
7.7%
Behaviour
Studies: 20, Items: 440
32.4%
Education
Studies: 4, Items: 34
11.5%
Employment, income & assets
Studies: 12, Items: 131
15.3%
Health and wellbeing
Studies: 37, Items: 543
19.6%
Sexual activity, pregnancy & fertility
Studies: 11, Items: 112
26.3%
Social
Studies: 8, Items: 76
Effect size
Reported directly to an interviewer vs not
28.8%
Adverse experiences
Studies: 7, Items: 66
36.3%
Attitudes & Expectations
Studies: 14, Items: 182
11.5%
Behaviour
Studies: 8, Items: 156
22.6%
Employment, income & assets
Studies: 10, Items: 124
27.8%
Health and wellbeing
Studies: 26, Items: 496
44.5%
Sexual activity, pregnancy & fertility
Studies: 6, Items: 119
48.8%
Social
Studies: 7, Items: 43
Effect size
Self-reported on paper vs electronically
2.6%
Adverse experiences
Studies: 6, Items: 38
0%
Attitudes & Expectations
Studies: 7, Items: 40
7.2%
Behaviour
Studies: 17, Items: 418
35.5%
Education
Studies: 2, Items: 31
5.3%
Employment, income & assets
Studies: 3, Items: 57
12.4%
Health and wellbeing
Studies: 21, Items: 362
13.2%
Sexual activity, pregnancy & fertility
Studies: 7, Items: 53
4.8%
Social
Studies: 3, Items: 42
Effect size
Interviewer present vs interviewer responded to
37%
Adverse experiences
Studies: 3, Items: 27
19.2%
Attitudes & Expectations
Studies: 1, Items: 26
12.9%
Behaviour
Studies: 6, Items: 101
30.6%
Employment, income & assets
Studies: 1, Items: 72
20.6%
Health and wellbeing
Studies: 5, Items: 102
40%
Sexual activity, pregnancy & fertility
Studies: 4, Items: 25
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Effect size

Comparison to existing theory



SURVEY FUTURES

SURVEY DATA COLLECTION METHODS COLLABORATION

Survey Practice Guide 2: How to mitigate against measurement effects when surveys move online

A Measurement Effect Risk Framework (MERF) and web questionnaire guidance

Jo d'Ardenne, Richard Bull, Aditi Das, Zac Perera & Olivia Sexton (National Centre for Social Research)

June 2025

Interviewer effects:

We found that mode effects were indeed most common in:

- Face-to-face vs web
- Telephone vs paper
- Reported directly to an interviewer or not

Socially desirable/sensitive items:

We found that mode effects were often (but no always) common in:

- Sexual activity, behaviour, pregnancy
- Social life
- Health and wellbeing

Satisficing/presentation effects were difficult to assess in our data.

Reporting

There were substantial inconsistencies in reporting across studies, in terms of:

- study design and sampling
- response rates
- sample size
- item non-response
- selection of reported estimates
- modes

This makes it challenging to both summarise and utilise the evidence.

Reporting: sample size

Out of 90 studies

32 did not report sample size appropriately:

- **8** did not report the sample sizes of the comparison groups
- **24** provided incomplete or inconsistent information

Reporting: post-randomisation issues

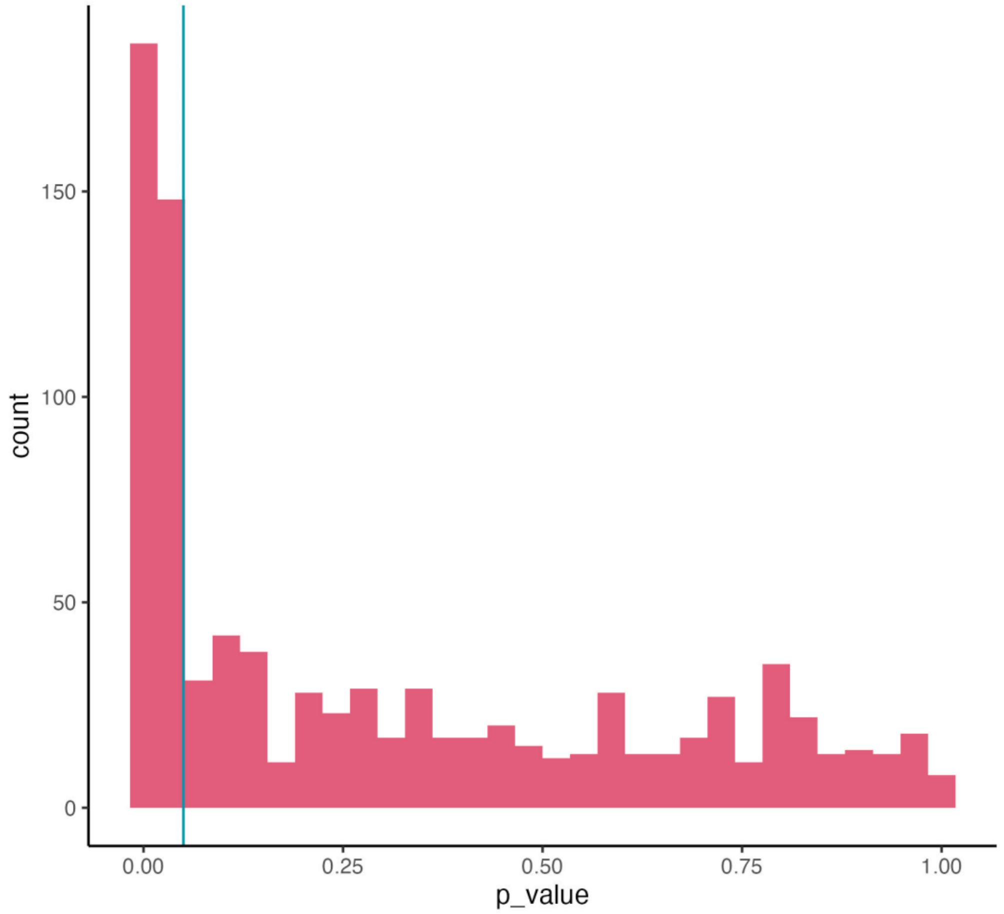
Out of 90 studies

53 did not report the extent of compliance to the randomly allocated mode

40 did not discuss **any** post-randomisation potential issues such as e.g. non-compliance, differential non-response

Publication bias: statistical significance

- 86.2% comparisons had no associated p-values reported



- majority of estimates did not have p-values
- including only significant p–values is common
- ➢ **publication bias difficult to assess**

Examples of selective reporting:

“The results are only shown for the X variables with a significant mode effect.”

”Estimates for selected indicators [variables]…”

Publication bias: pre-existing beliefs

- Studies are more likely to examine mode comparisons and items for which they **anticipate mode effects will exist**
  - evidenced by the number of studies examining substance use, mental health, alcohol, sexual activity
- This may skew results and suggest that mode effects are **more common than they are**
  - Most were already under 0.2 SD (“small”)
- However:
  - These thresholds are arbitrary
  - The degree to which a mode effect of any size would impact substantive findings depends on many factors

Recommendations

We encourage researchers to:

- ✓ clearly report the sampling process and study design
- ✓ report the sample size, non-response, randomisation compliance, and any other exclusions
- ✓ report all conducted analyses, not just those with significant results
- ✓ report confidence intervals for all estimates, not just indicators of significance

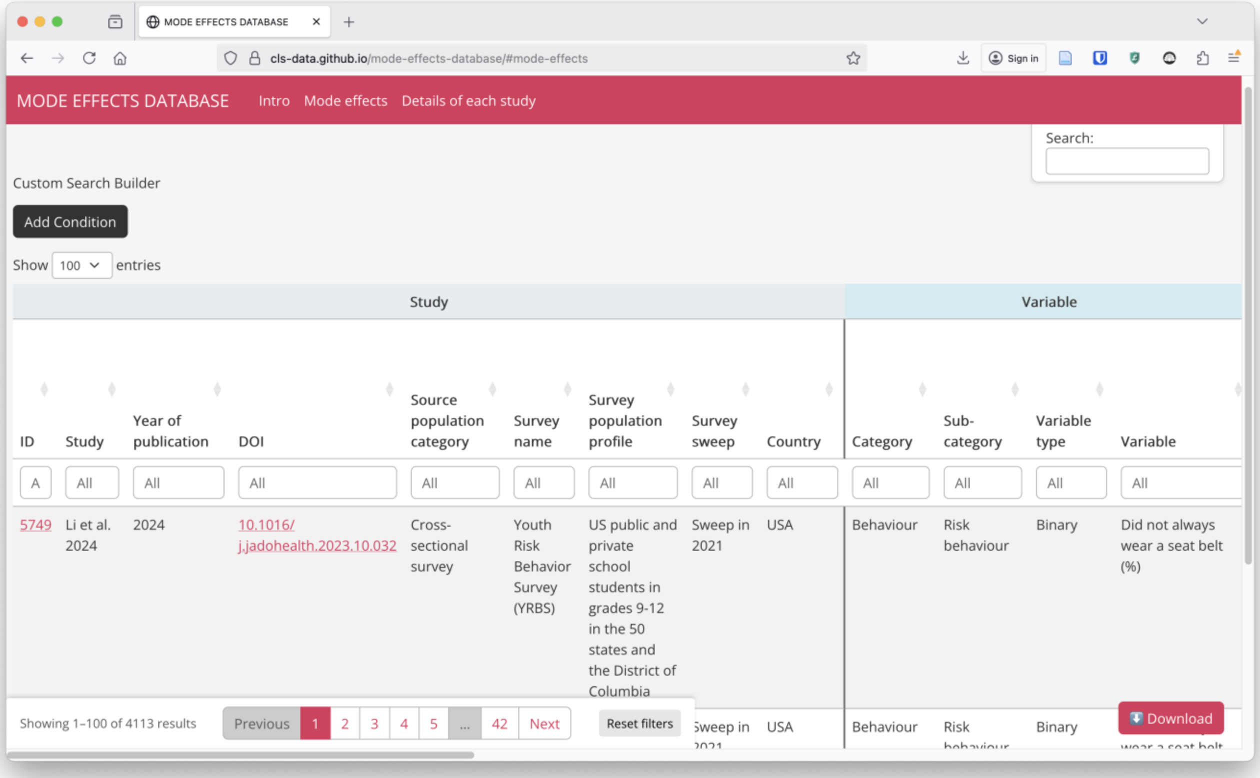
- ✓ Consider using existing reporting guidelines
  e.g. CONsolidated Standards Of Reporting Trials (CONSORT)
  e.g. Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA)

Limitations

- Despite focus on experimental studies, mode selection is still possible
  - extent difficult to assess due to reporting
- Only able to synthesise standardised mode effects
  - although 84.2% were available, the rest were excluded
- Classification of modes into categories was not author-reported
- Due to the number of data extraction items, errors are possible

Database

https://cls-data.github.io/mode-effects-database/



Contact

Email: g.tomova@ucl.ac.uk

BlueSky: @georgiatomova.bsky.social

CENTRE FOR LONGITUDINAL STUDIES

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

Adaptive mixed-mode survey design

Barry Schouten (Statistics Netherlands)

RSS session on Handling Survey Mode Effects

November 12, 2025

Outline

1. Historical background
2. Adaptive mixed-mode survey design
   - Optimizing representation
   - Minimizing the total mode effect
3. Adaptive mixed-mode survey design including measurement
   - Response style propensities
   - Adding constraints on subgroup comparability
4. Discussion

Background

Some history of mixed-mode survey design at SN

- 2005 – 2009: Preparations/pilots for adding web to telephone and F2F
- 2007 – 2010: Development of R/CV indicators and first ASD pilots by methodology department
- 2009 – 2013: Migration to sequential MM designs, incl parallel runs

- 2011 – 2014: Attempts to bring ME error within ASD scope by methodology department plus two PhD projects
- 2017 – 2021: SN goes AMMSD. New designs/redesigns are ASD by default

- 2019 – 2023: The Smokers phenomenon and Health survey re-interview
- 2023 – now: Growing concern about ME. Investigation potential strategies

Some history of mixed-mode survey design at SN

- Mixed-mode survey design is mostly driven by budget reduction
- Adaptive survey design is about doing different/tailored
- Together they may lead to a focus on ‘damage control’
- Given that web/digital communication is now the standard, there may be a renewed look that includes measurement equivalence and comparability

Adaptive mixed-mode survey design

Adaptive mixed-mode survey design

Migration to ASD strategies at Stat Netherlands

Survey	Mode strategy	Period
Health	cawi → capi	2018 – present
Public Opinion	cawi → cati / capi	2018 – present
Lifestyle	cawi → cati / capi	2019 – present
Leisure	cawi → capi	2019 – present
Labour Force	cawi → cati / capi	2021 – present
Social Coherence	cawi → cati / capi	2022 – present

Adaptive mixed-mode survey design

Context: Relatively rich administrative data, strict rules on refusal conversion, strict rules on incentive strategies, repeated general population surveys.

All social surveys at StatNL are adaptive by default and the design feature to tailor is the survey mode.

The objective in ASD optimization is the coefficient of variation relative to relevant background characteristics linked from admin data. Constraints come from precision and budget.

Most common choices: age categories, HH income quintiles, migration background, household composition

ASD optimization

$$\min_{\{\pi, p_g : g \in G\}} CV\left(\rho\left(p_g\right)\right)\left(= \frac{\sqrt{\sum_{g=1}^{G} w_g \left(R_g^W + p_g\left(1 - R_g^W\right) R_g^F - RR\left(p_g\right)\right)^2}}{RR\left(p_g\right)}\right),$$ coefficient of variation

subject to

$$r\left(\pi, p_g\right) = N\pi \sum_{g=1}^{G} w_g \left(R_g^W + p_g\left(1 - R_g^W\right) R_g^F\right) \geq A,$$ size of the response

$$F2F\left(\pi, p_g\right) = N\pi \sum_{g=1}^{G} w_g p_g \left(1 - R_g^W\right) \leq B,$$ face-to-face workload

$$n(\pi) = N\pi \leq C,$$ sample size

$$0 \leq p_g \leq 1, \forall g, 0 \leq \pi \leq 1$$

with strata g specified prior to fieldwork based on relevance for survey variables and nonresponse prediction.

ASD optimization for the Health Survey 2024

Nine strata based on age, household income and migration background

group	p cawi	f capi	p capi	p total
1	20	100	17	33
2	20	100	24	38
3	26	100	32	48
4	33	100	24	48
5	37	67	30	49
6	40	49	33	49
7	44	22	43	49
8	44	31	29	49
9	62	0	38	62
total	38	63	26	48

$$CV(\rho) = 0.146$$

$$\bar{\rho} = 0.478$$

ASD aiming at minimizing the total mode effect

Study based on parallel runs between 2010 and 2012

Idea: Minimize the method effect relative to F2F

Example: LFS

1. *Registered unemployed*: 7.5%
2. *65+ households without employment*: 19.8%
3. *Young household members and no employment*: 2.4%
4. *Non-western without employment*: 1.5%
5. *Western without employment*: 11.0%
6. *Young household member and employment*: 15.6%
7. *Non-western and employment*: 3.9%
8. *Western and employment*: 33.5%
9. *Large households*: 4.9%

- *Single web*
- *Telephone short*
- *Telephone extended*
- *F2F short*
- *F2F extended*
- *Web followed by telephone short*
- *Web followed by telephone extended*
- *Web followed by F2F short*
- *Web followed by F2F extended*

ASD aiming at minimizing the total mode effect

Costs per strategy per stratum relative to F2F extended

Strategy	Stratum								
	1	2	3	4	5	6	7	8	9
Web	0.03	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03
Phone2	0.11	0.15	0.10	0.09	0.13	0.11	0.09	0.12	0.14
Phone2+	0.13	0.17	0.11	0.10	0.15	0.14	0.11	0.16	0.19
F2F3	0.84	0.89	0.83	0.83	0.86	0.84	0.81	0.84	0.89
F2F3+	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Web→Phone2	0.08	0.11	0.09	0.09	0.09	0.08	0.08	0.07	0.07
Web→Phone2+	0.09	0.12	0.10	0.10	0.10	0.09	0.09	0.08	0.07
Web→F2F3	0.60	0.66	0.64	0.70	0.59	0.56	0.65	0.51	0.61
Web→F2F3+	0.72	0.71	0.80	0.84	0.73	0.68	0.81	0.62	0.71

ASD aiming at minimizing the total mode effect

Stratum response propensity per strategy

	Stratum								
	1	2	3	4	5	6	7	8	9
Web	23.2%	23.6%	15.5%	10.8%	27.9%	27.7%	17.5%	36.7%	22.4%
Phone2	12.2%	31.4%	8.5%	4.7%	19.7%	13.3%	7.2%	18.1%	21.2%
Phone2+	20.8%	41.3%	15.2%	8.6%	31.1%	23.8%	14.3%	33.3%	37.5%
F2F3	43.5%	53.5%	42.2%	34.1%	45.1%	45.3%	35.9%	46.7%	54.6%
F2F3+	52.4%	58.3%	51.0%	41.2%	51.2%	54.9%	46.0%	56.8%	61.4%
Web→Phone2	28.3%	41.0%	20.2%	13.9%	36.3%	34.0%	20.8%	44.5%	23.1%
Web→Phone2+	32.8%	48.4%	23.8%	17.5%	42.1%	41.1%	25.8%	52.1%	24.4%
Web→F2F3	46.3%	57.7%	38.6%	32.7%	50.0%	51.0%	39.3%	58.9%	50.0%
Web→F2F3+	49.8%	58.3%	43.4%	36.6%	52.6%	54.7%	44.3%	62.0%	54.2%

ASD aiming at minimizing the total mode effect

Optimization problem
(with N population size, N_g stratum sizes)

$\min_{\{p_g(s)\}} \left\| \sum_{g,s} \frac{N_g}{N} \frac{p_g(s)\rho_X(g,s)\Delta y_x(g,s)}{\sum_{\tilde{s}} p_g(\tilde{s})\rho_x(g,\tilde{s})} \right\|$	Adjusted method effect
subject to	
$\sum_{g,s} N_g p_g(s) c_X(g,s) \leq B;$	Total variable costs
$\sum_s N_g p_g(s)\rho_X(g,s) \geq R_g;$	Stratum quota for precision
$\frac{1}{N}\sum_g N_g \left(\sum_s p_g(s)\rho_X(g,s) - \rho(s)\right)^2 \geq \alpha;$	R-indicator
$\sum_g N_g p_g(\emptyset) \leq n;$	Total sample size
$0 \leq p_g(s) \leq 1;\ \sum_s p_g(s) = 1.$	

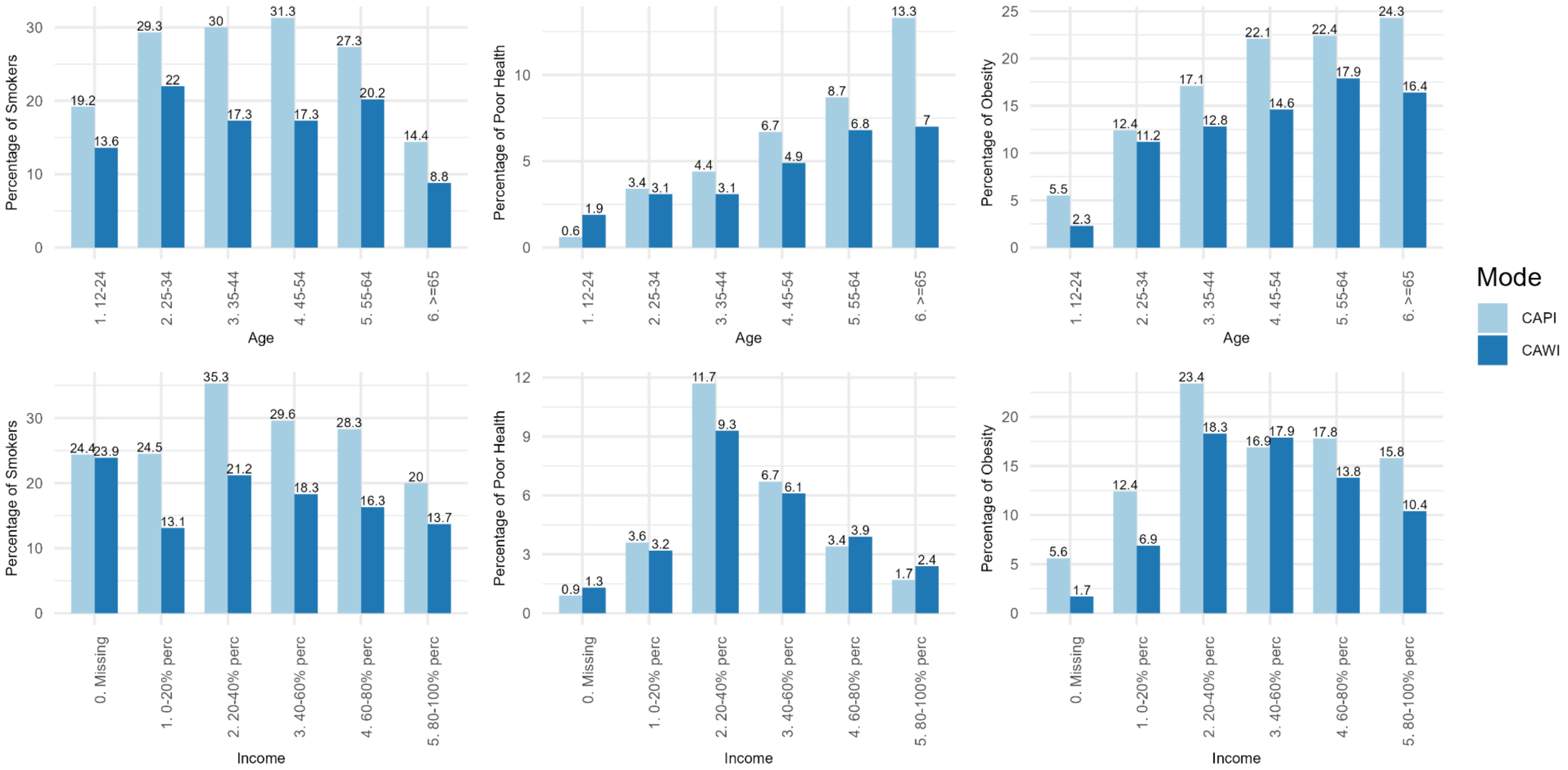
ASD aiming at minimizing the total mode effect

Stratum allocation probabilities

Strategy	Strata based on age, employment, ethnicity and hh size								
	1	2	3	4	5	6	7	8	9
Web	0	0	0	0	0	0	0	0	0
Tel	0	0	0.30	0	0	0	0	0	0
TelE	0.48	1	0.13	0.08	0.63	0	0.86	0.48	0.22
F2F	0.45	0	0.43	0	0	0	0.14	0	0.15
F2FE	0	0	0.07	0.92	0	0	0	0	0
Web-tel	0	0	0	0	0	0	0	0	0.63
Web-telE	0	0	0	0	0.37	0.83	0	0.52	0
Web-F2F	0.07	0	0.07	0	0	0.17	0	0	0
Web-F2FE	0	0	0	0	0	0	0	0	0

Adaptive mixed-mode survey design including measurement error

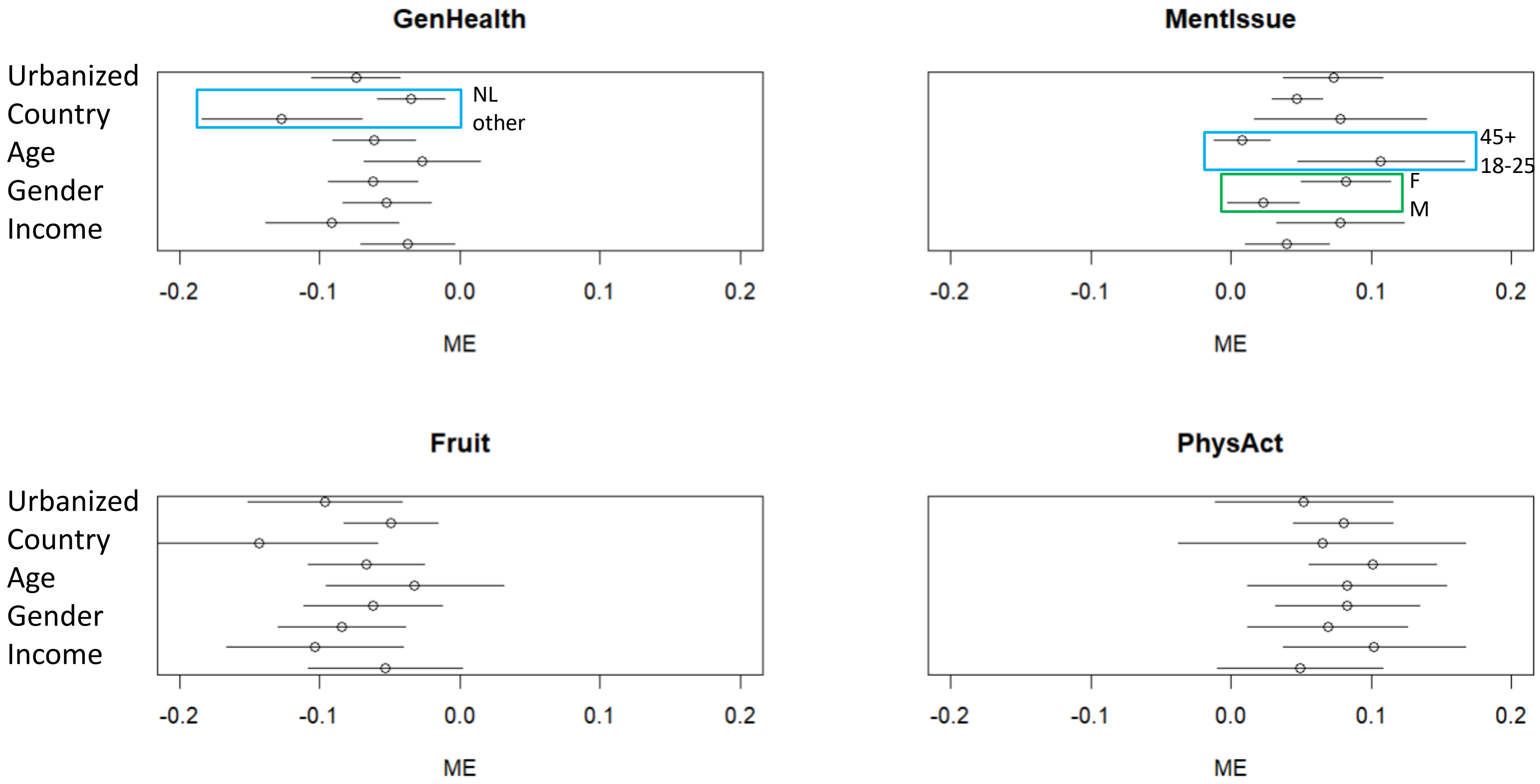
The Smokers phenomenon in the Health Survey



Re-interview – Measurement effects for the Health Survey

Type	Statistic/question	ME effect		ME
		Estimate	SE	adjustment
Robust	Dentist visit (last 6 months)	0.6%	1.1%	+0.4%
	Doctor/GP visit (last 6 months)	2.0%	2.0%	+1.6%
	Physiotherapy (last 6 months)	1.4%	1.5%	+1.1%
	Type 2 diabetes	0.0%	0.3%	+0.0%
	Active in sports club	0.7%	1.7%	+0.6%
Complex	Satisfies norm eating fruits	7.2%	1.9%	+5.7%
	Satisfies norm eating vegetables	2.2%	2.0%	+1.8%
	Satisfies general fit norm	-7.7%	2.2%	-6.1%
	Satisfies fit norm muscle-bone	0.5%	1.8%	+0.4%
	Use of non-prescribed medicine(s)	15.2%	2.2%	+12.1%
Sensitive	Smoking	-0.6%	1.0%	-0.5%
	Obese	-1.8%	0.9%	-1.5%
	Heavy drinker (CASI)	-0.5%	1.2%	-0.4%
	Ever used drugs (CASI)	-0.5%	0.6%	-0.4%
	Use of prescribed medicine	3.1%	1.7%	+2.5%
Subjective	Self-perceived health as (very) good	5.8%	1.4%	+4.6%
	Self-perceived psychological issues	-5.4%	1.3%	-4.3%
	Self-perceived physical barriers	-1.2%	1.0%	-1.0%

Re-interview – Measurement effects for the Health Survey



Objectives

- Strategies to account for ME effects in MM designs:
  - Prevent through design of data collection instruments/questionnaires
  - Avoid through adaptive survey design
  - Adjust through estimation of mode-specific ME biases
  - Stabilize through calibration to fixed mode distributions
- Questions:
  - How to include constraints on ME effects in ASD optimization?
  - What is the impact on ASD optimization?
  - How to handle multi-purpose surveys?
  - How does a modified ASD optimization compare to other strategies?

Adaptive survey design and measurement error

- ASD focus primarily on nonresponse:
  - Impact ME error on costs is indirect at best
  - ME error requires advanced data collection designs
  - No tradition in social surveys to intervene/change strategy for ME error
- Some work has been done:
  - Single-purpose surveys: Limit mode-effect in migrating to MM designs
  - Multi-purpose surveys: Limit the impact of ‘response style’ propensities on representation
- Why include also ME error?
  - ME and NR have common causes that strengthen design choices
  - ME and NR also have specific causes that may cancel each other out

Minimize or constrain 'response style' propensities

- Define response styles based on:
  - Paradata (time stamps, audit trails)
  - Gold standard data
  - Underlying constructs (MTMM, MCFA)
- Example study (Calinescu & Schouten, SRM 2013):
  - Motivated underreporting given mode and yes/no proxy reporting
- Did not work
  - Too early
  - Too abstract (for management and survey PI's)
  - Response styles are also multi-dimensional

Options for an additional constraint on ME effects

Mode-specific measurement biases affect:

- Accuracy
- Comparability between subpopulations
- Comparability in time

Given estimates for ME effects are available for subpopulations and main survey variables, what to do?

- Natural approach: Constraint on ME effect w.r.t. an ME benchmark mode
- Complication: ME benchmark mode in multi-purpose surveys may vary
- Solution: Assume the ‘true’ value lies in between the modes.

Options for an additional constraint on ME effects

Midway ME effect

$$\mathrm{ME}_{MID,g}(y, p_g) = \tfrac{1}{2}\mathrm{ME}_g(y) q_g(p_g) - \tfrac{1}{2}\mathrm{ME}_g(y)\left(1 - q_g(p_g)\right) = \mathrm{ME}_g(y)\left(q_g(p_g) - \tfrac{1}{2}\right),$$

with $\mathrm{ME}_g(y)$ the estimated ME effect between the modes for stratum g and variable y and $q_g(p_g)$ the share of the mode in stratum g.

Additional constraint in the ASD optimization, comparability subpopulations,

$$\sum_{m=1}^{M} v_m \sqrt{\sum_{g=1}^{G} w_g \left(\mathrm{ME}_{MID,g}(y_m, p_g) - \overline{ME}_{MID}(y_m, p_g)\right)^2} \leq D,$$

with v_m a weight assigned to variable m.

Optimization

Scenario’s

1. Web only
2. Full mixed-mode (not feasible within budget)
3. ASD without ME constraint (the current default strategy)
4. ASD with ME constraint
    a. 2%
    b. 1%
5. Changing other constraints:
    a. Relaxing precision
    b. Relaxing sample size
    c. F2F budget: -5%, +5%, +10%

Optimization results - allocation

Input parameters	Stratum					
	1	2	3	4	5	6
Share in population	22%	16%	14%	9%	33%	6%
Response web	20%	14%	25%	34%	43%	44%
Response F2F	32%	34%	39%	36%	44%	48%
ME effect gen health	-13%	-13%	-11%	-5%	-4%	-3%

Scenario	Stratum					
	1	2	3	4	5	6
ASD without ME	0.81	0.99	0.82	0.44	0.36	0.42
ASD with ME at 2%	0,80	0,84	0,77	0,74	0,41	0,30
ASD with ME at 1%	0,68	0,71	0,69	0,75	0,57	0,51
precision↓	0,63	0,65	0,61	0,61	0,39	0,32
sample size ↑	0,59	0,61	0,56	0,53	0,27	0,20
-5% F2F budget	0,49	0,65	0,75	0,37	0,72	0,91
+5% F2F budget	0,69	0,70	0,69	0,74	0,58	0,46
+10% F2F budget	0,68	0,71	0,68	0,74	0,58	0,49

Strata definition:
- Income HH
- Age
- Migration background

Optimization results - indicators

Scenario	Indicator		
	CV	RR	ME
Web only	0.386	30.0%	2.0%
Full mixed-mode	0.194	54.9%	2.0%
ASD without ME	0.116	46.1%	2.6%
ASD with ME at 2%	0.140	46.1%	2.0%
ASD with ME at 1%	0.207	46.1%	1.0%
precision↓	0.191	43.4%	1.0%
sample size ↑	0.179	41.5%	1.0%
-5% F2F budget	0.280	46.1%	1.0%
+5% F2F budget	0.207	46.1%	1.0%
+10% F2F budget	0.207	46.1%	1.0%

Optimization results – survey estimates (unweighted)

Estimator		Self-reported health	Smoking
Web only	No adjustment	78.5%	20.2%
	ME adjustment (BM = F2F)	85.3%	19.2%
	ME adjustment (BM = midway)	81.9%	19.7%
Full MM	No adjustment	77.7%	25.3%
	ME adjustment (BM = F2F)	81.4%	24.8%
	ME adjustment (BM = midway)	77.6%	23.1%
ASD without ME		77.4%	24.6%
ASD with ME at 2%		77.9%	24.6%
ASD with ME at 1%		77.9%	24.3%
	precision↓	78.0%	24.0%
	sample size ↑	78.0%	23.7%
	-5% F2F budget	77.5%	23.8%
	+5% F2F budget	78.0%	24.3%
	+10% F2F budget	77.9%	24.3%

Discussion

Discussion

Conclusions:

- Mode-specific ME can be sizeable and cannot be ignored
- ME stratum comparability can be added as constraint to ASD, but
  - Strong impact on representation and allocation
  - Mild impact on estimates in example
  - ASD optimization sensitive to budget, sample size and precision thresholds
  - Can be ‘manipulated’ by the choice of survey variables
- Preliminary conclusion: Combine ME adjustment with ASD