SURVEY FUTURES

SURVEY DATA COLLECTION METHODS COLLABORATION

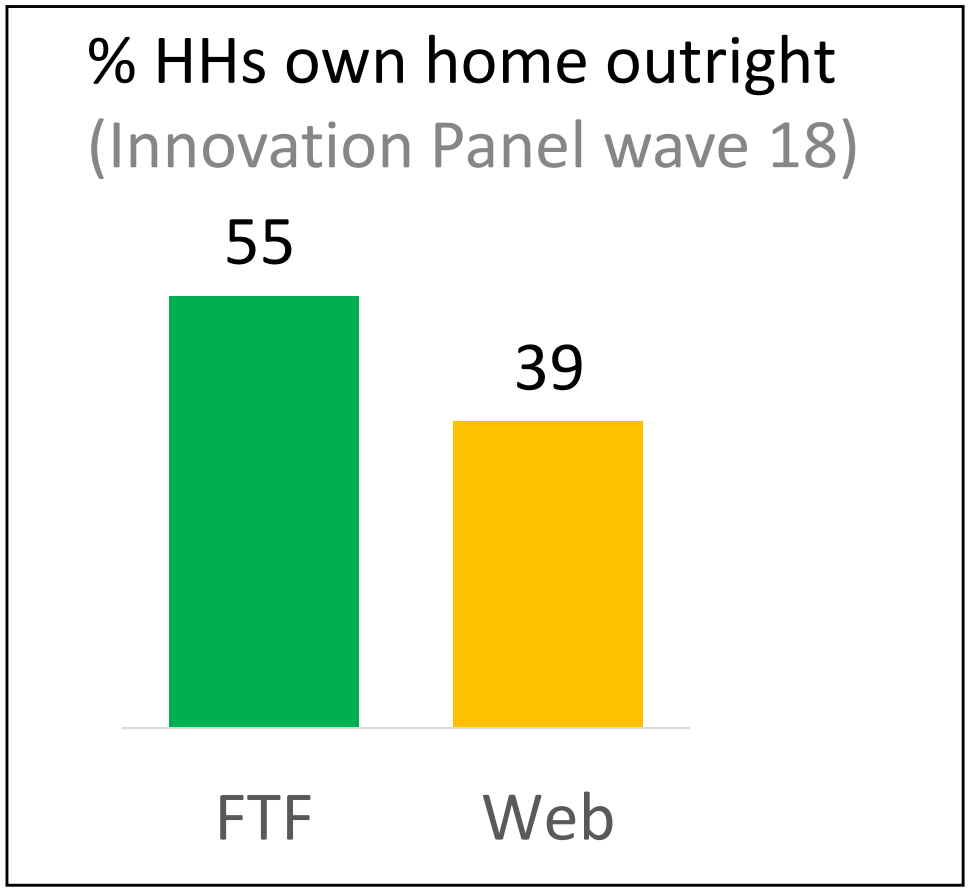
A review of research designs to disentangle measurement and selection effects in mixed mode surveys

Annette Jäckle[1], Jonathan Burton[1], Mick P. Couper[2]
([1]University of Essex, [2]University of Michigan)

Survey Futures conference, 17-18 June 2026 (London)

The Problem

- Survey using multiple modes
- How do we know whether the data are comparable?

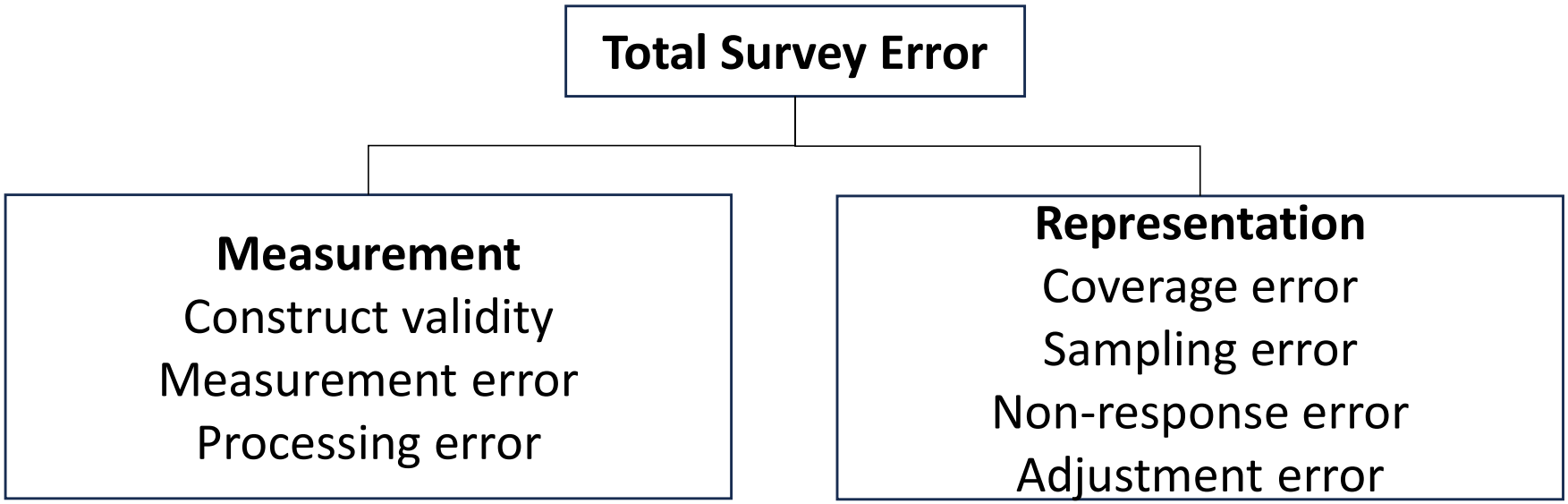


- Real difference?
- Artifact of the mode?

How can we know?

The Problem: Confounds

- If response distributions differ between modes, e.g. $\overline{Y_{\text{FTF}}} \neq \overline{Y_{\text{WEB}}}$
- Multiple possible reasons
- Total Survey Error in FTF ≠ web

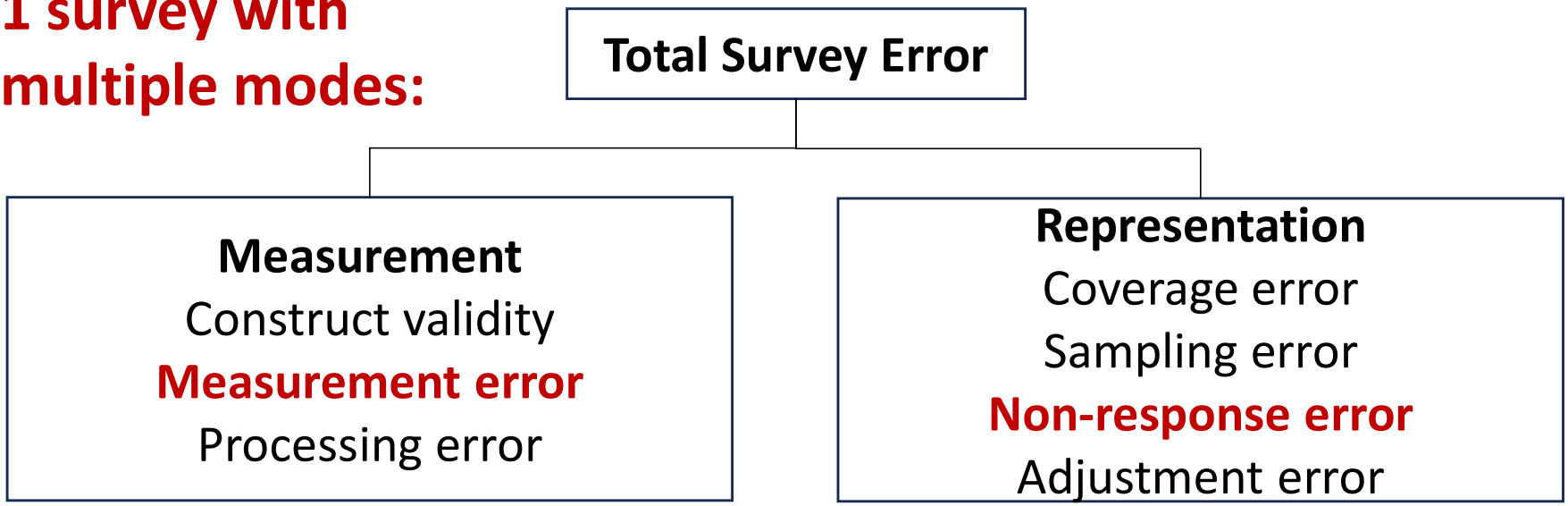


The Problem: Confounds

- If response distributions differ between modes, e.g. $\overline{Y}_{\text{FTF}} \neq \overline{Y}_{\text{WEB}}$
- Multiple possible reasons
- Total Survey Error in FTF ≠ web

How can we identify mode effects on measurement?

1 survey with multiple modes:



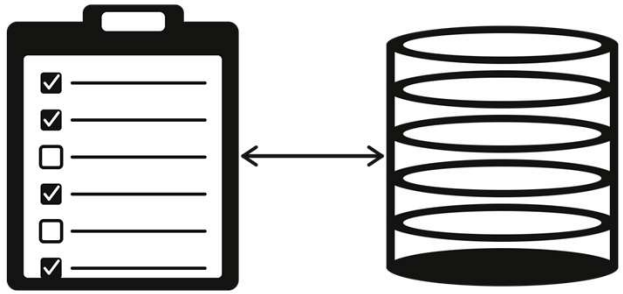
Project Output: Practical Guide

- Review of research designs
  - To disentangle selection and measurement effects of modes
- Focus on observed variables
  - Methods for latent variables: see Hox et al (2017)
- Record linkage
- Lab experiment
- Field experiment
- Repeated measures

Record Linkage: "Gold Standard"

Requires:
- External data about Rs / NRs
- Same concepts & time

Assumes:
- External data are 'truth'
- Linked sample ~ full sample



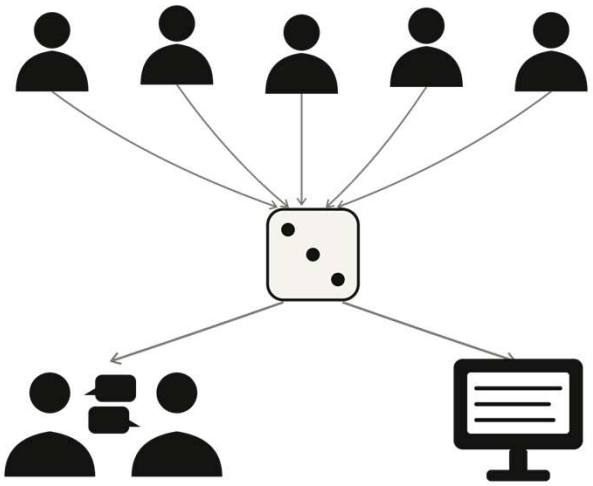
⊕ Identify measurement error: mode differences?
Identify non-response error

⊖ Limited number of variables can be tested

Examples: Kreuter et al (2008), Felderer et al (2019)

Lab Experiment

- Recruit participants, then
- Randomize allocation to modes



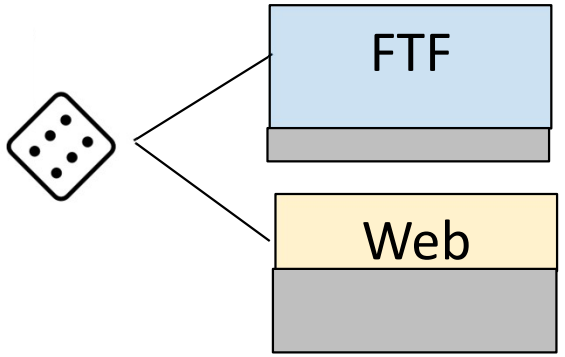
⊕ Mode is only difference between samples
All other sources of TSE held constant

⊖ Do results generalise to surveys in the field?

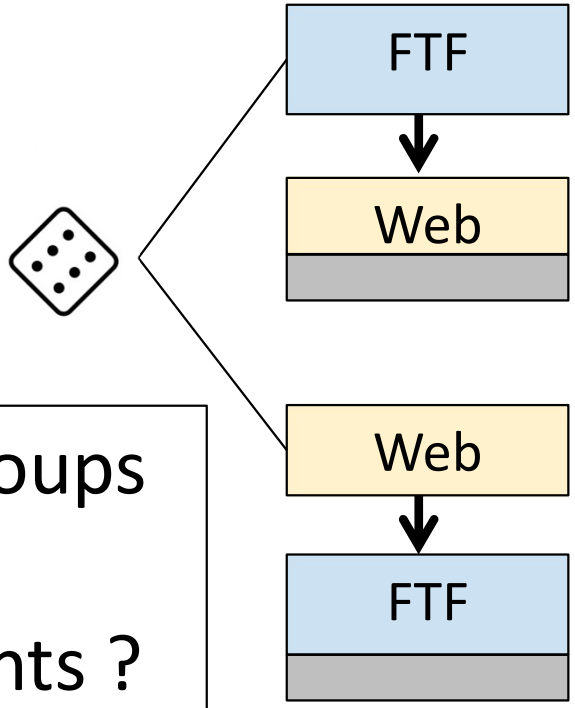
Examples: Chang & Krosnick (2010), Lind et al (2013)

Field Experiment

Allocation: 1 mode



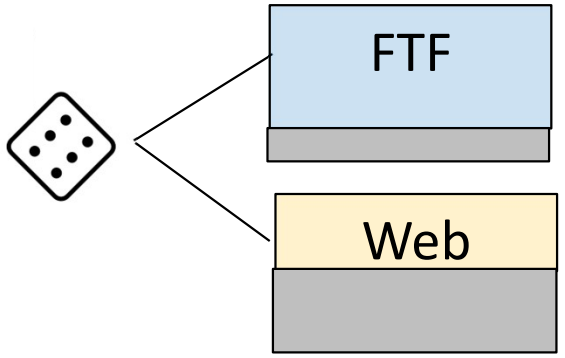
Allocation: mode sequence



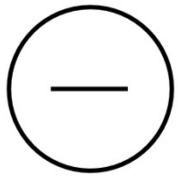
- Randomisation ➔ comparable groups
- But: non-response
- FTF respondents ≠ web respondents ?
- Differences in answer distributions:
  - Different types of respondents?
  - Different measurement?

Field Experiment

Allocation: 1 mode



- Observed characteristics of Rs
- Adjust for differences in sample composition
  - Covariates in models
  - Weights
  - Propensity score matching
- Assume - covariates
  - Fully explain selection into modes
  - Measurement not affected by mode

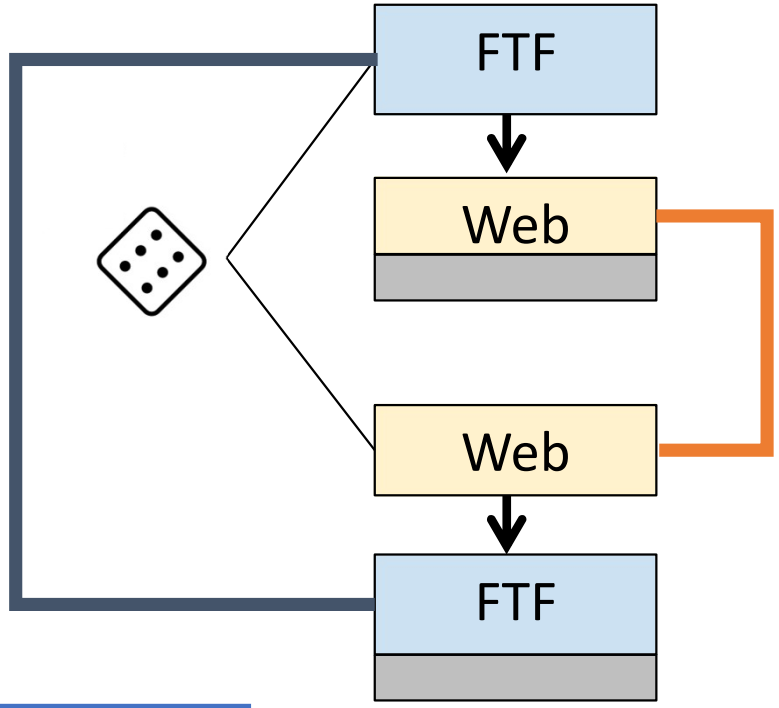


Cannot test assumptions
Unobserved differences between groups likely

Field Experiment

- As treated analysis
  - FTF vs
  - Web

Allocation: mode sequence

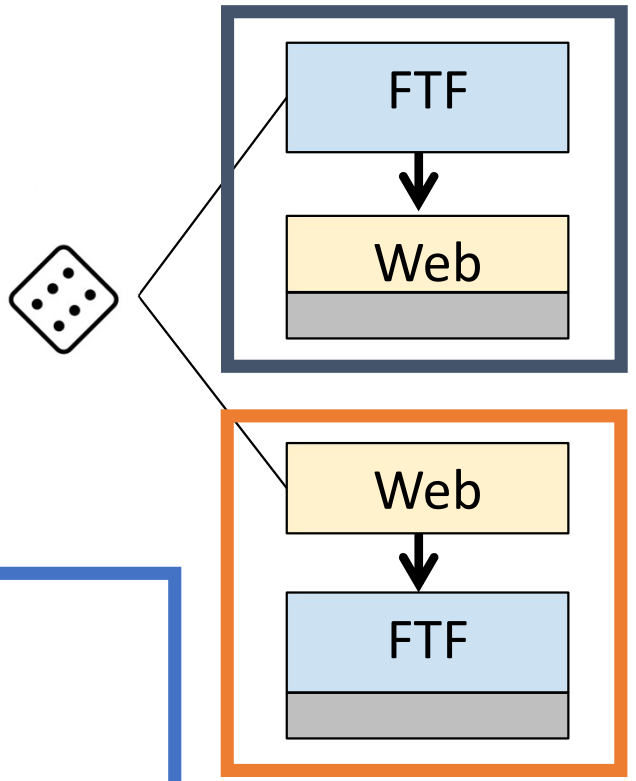


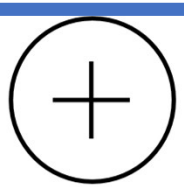
Self-selection into modes

Field Experiment

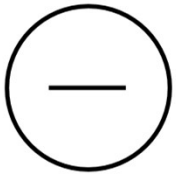
- Intention to treat analysis (ITT)
  - FTF-first vs
  - Web-first

Allocation: mode sequence





Groups more comparable

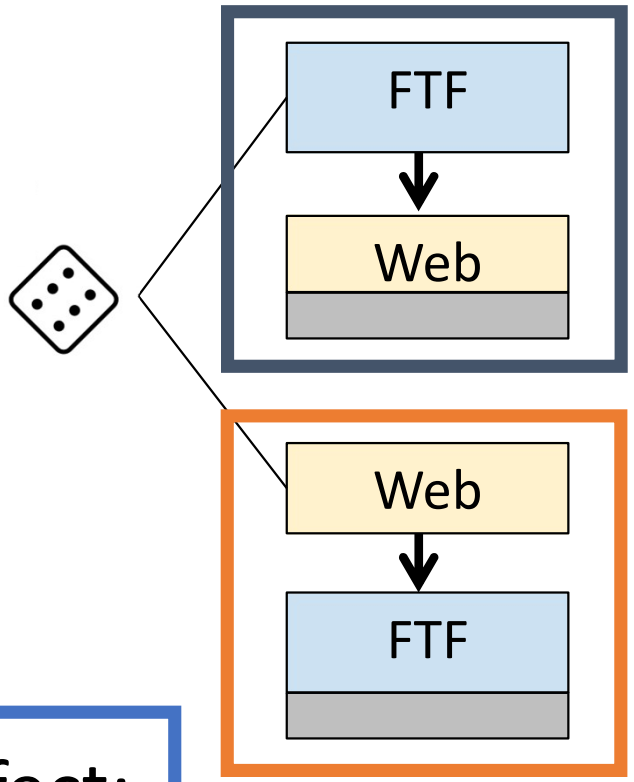


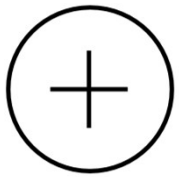
Not effect of individual modes

Field Experiment

- Instrumental variable
  - Use random allocation
  - As instrument for mode
- Rescales the ITT estimate
  - By prop of web “converts”
- Assume – instrument
  - Correlated with mode
  - Not corr with Y / R chars

Allocation: mode sequence



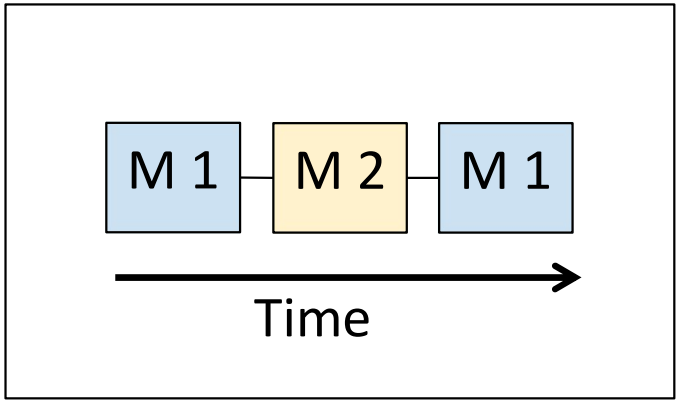


Local Average Treatment Effect:
Effect of mode for web converts

See: Greenland (2000), Vannieuwenhuyze (2015), Clarke & Bao (2022)

Repeated Measures

- Respondents complete survey in different modes over time
- Within respondent estimator
  - Mode effect estimated from "mode switchers"
  - Nets out unobserved differences between Rs



- Assume – mode switch
  - Not related to changes in Rs life that also affect their answers

⊖ If most Rs switch modes in one direction
Time trends counted as mode effect

Does the Method Matter?

- Understanding Society Innovation Panel
- FTF-first vs web-first
- Waves 5-8 pooled

Design	Analysis	N Questions	% Padj<0.05
Experiment	By mode of interview	503	17
	Covariate adjustment	503	8
	Instrumental variable	502	3
Repeated	Within R estimator	405	10

Padj: P-value adjusted for clustering in respondents and multiple testing

Does the Method Matter?

Most differences between modes:
Due to (unobserved) selection into modes
Difficult to identify

Design	Analysis	N Questions	% Padj<0.05
Experiment	By mode of interview	503	17
	Covariate adjustment	503	8
	Instrumental variable	502	3
Repeated	Within R estimator	405	10

Padj: P-value adjusted for clustering in respondents and multiple testing

Project Output: Practical Guide

- Research designs to disentangle selection and measurement effects of modes
  - Assumptions
  - Analysis methods
  - Limitations
  - Worked examples
  - Annotated bibliography
- Focus on observed variables
  - Methods for latent variables: see Hox et al (2017)

Thank you.

Annette Jäckle

aejack@essex.ac.uk

References

- Chang & Krosnick (2010) Comparing oral interviewing with self-administered computerized questionnaires: An experiment, *Public Opinion Quarterly*, 74(1):154–167.
- Clarke & Bao (2022) Estimating mode effects from a sequential mixed-mode experiment using structural moment models, *The Annals of Applied Statistics*, *16*(3):1563–1585.
- Felderer, Kirchner, Kreuter (2019) The effect of survey mode on data quality: Disentangling nonresponse and measurement error bias, *Journal of Official Statistics*, 35(1):93–115.
- Greenland (2000) An Introduction to instrumental variables for epidemiologists, *International Journal of Epidemiology*, 29 (4) 722-729.
- Hox, de Leeuw & Klausch (2017) Mixed mode research: Issues in design and analysis, ch. 23 in *Total Survey Error in Practice*, Biemer et al (eds.), New York: Wiley.
- Kreuter, Presser & Tourangeau (2008) Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity, *Public Opinion Quarterly*, 72(5): 847–865.
- Lind, Schober, Conrad & Reichert (2013) Why do survey respondents disclose more when computers ask the questions? *Public Opinion Quarterly*, 77(4):888–935.
- Vannieuwenhuyze (2015) Mode effects on variances, covariances, standard deviations, and correlations. *Journal of Survey Statistics and Methodology* 3(3): 296-316.