Handling Mode Effects in the CLS Cohort Studies

User Guide Executive Summary

November 2024

Contact

Questions and feedback about this user guide: clsdata@ucl.ac.uk.

Authors

Liam Wright, George Ploubidis, Richard Silverwood.

How to cite this guide

Wright, L., Ploubidis, G., Silverwood, R. (2024) *Handling Mode Effects in the CLS Cohort Studies: Executive Summary*. London: UCL Centre for Longitudinal Studies.

This guide was published in November 2024 by the UCL Centre for Longitudinal Studies.

Data citation and CLS acknowledgement

You should cite the data and also acknowledge CLS following the guidance from cls.ucl.ac.uk/data-access-training/citing-our-data/

Centre for Longitudinal Studies

UCL Centre for Longitudinal Studies (CLS)

UCL Social Research Institute

University College London

20 Bedford Way, London WC1H 0AL

www.cls.ucl.ac.uk

The UCL Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre based at the UCL Social Research Institute, University College London. It is home to a unique series of UK national cohort studies. For more information, visit www.cls.ucl.ac.uk.

This document is available in alternative formats. Please contact the Centre for Longitudinal Studies:

Email: clsdata@ucl.ac.uk

Introduction

Surveys are increasingly moving to mixed mode data collection[1] – for instance, carrying out interviews via face-to-face, telephone, video and/or web. The potential advantages of mixed mode data collection are lower costs, increased efficiency, and higher participation rates. These advantages do not come without drawbacks, however. Specifically, responses may differ systematically between survey modes. For instance, the presentation of a survey item either aurally or visually can influence responses, sensitive information may be reported more accurately when given anonymously (e.g., by web survey compared with face-to-face interview) and complex information may be reported more accurately when an interviewer is present. Differences in responses arising from differences in measurement between surveys modes are termed *mode effects*.

Unaccounted for, mode effects can be a problem for obtaining accurate and unbiased estimates, both for descriptive and inferential statistics. Simply adding an indicator variable for survey mode into analyses may not be sufficient to remove bias (and in some cases could *increase* it) as selection into mode is typically non-random. Observed differences between modes are a combination of mode effects and selection effects. Selection effects may *confound* mode effects. Adding an indicator variable for mode does not account for this.

Whether mixing modes generates bias depends on the particular analysis being carried out. Some characteristics, such as age, may be reported similarly regardless of mode used, while subjective and socially sensitive characteristics, like mental health, may not be. Further, the confounding of mode effects by selection effects depends on the variables being analysed, their relationships with each other, and their relationships with the causes of mode selection.

Given this, it is not possible to give concise and definitive recommendations for handling mode effects that apply in all situations. Instead, in the main document we

---

[1] Matt Brown and Lisa Calderwood, 'Mixing Modes in Longitudinal Surveys: An Overview', CLS Working Paper (London: UCL Centre for Longitudinal Studies, 2020), https://cls.ucl.ac.uk/wp-content/uploads/2020/04/CLS-working-paper-2020-3-Mixing-modes-in-longitudinal-surveys-an-overview.pdf.

provide frameworks and relevant empirical evidence to help researchers think about the possible consequences of mode effects in their own analyses of Centre for Longitudinal Studies' (CLS) cohort data. In that document, we also describe methods for handling mode effects, discussing their strengths and limitations, and highlighting *sensitivity analysis* as a particularly promising approach. Unlike other methods, sensitivity analysis does not require appropriately modelling mode selection – to our minds, not a straightforward matter.

In the next section, we enumerate the elements of mixed mode data collection that have appeared in CLS's cohorts. Then, we provide recommendations for accounting for mode effects.

Mixed Mode Elements of the Centre for Longitudinal Studies Cohort Data Collections

Several of the studies run by CLS have included elements of mixed mode data collection, both between individuals within a sweep and within individuals across sweeps. Within sweeps, these are:

1. Millennium Cohort Study (MCS): Sweep 6 Time Use Diaries (Age 14y)

   - Cohort members were offered the choice of web and mobile app data collection modes, with paper version also available if the cohort member was unable to use these modes. The cohort members were informed of this in advance of the Sweep 6 interview.

2. 1970 British Cohort Study (BCS70): Sweep 3 Cognitive Ability Tests (Age 16y)

   - Data collection at the age 16y sweep was conducted during a teacher strike. Consequently, some participants completed cognitive tests at home rather than at school and almost half of participants did not complete the tests at all.

3. Next Steps: Sweeps 5-9 Interviews (Ages 17/18y – 32y)

   - Sequential mixed mode approaches were used in each sweep. In Sweeps 5-8, a web survey was offered initially, followed by telephone and, finally, face-to-face interview. In Sweep 9, a web survey was offered initially, with non-respondents then able to choose from face-to-face, telephone, or video interview or web survey. A shortened web survey was then offered to remaining non-respondents.

4. 1958 National Child Development Study (NCDS): Sweep 9 Interview (Age 55y)

   - A sequential mixed mode approach was used primarily (web initially offered followed by telephone interview), but the survey also embedded a mode experiment, with 1,499 of 10,558 cohort members randomly allocated to telephone-only data collection.[2]

---

[2] Alissa Goodman et al., 'The Impact of Using the Web in a Mixed-Mode Follow-up of a Longitudinal Birth Cohort Study: Evidence from the National Child Development Study', *Journal of the Royal Statistical Society Series A: Statistics in Society* 185, no. 3 (1 July 2022): 822–50, https://doi.org/10.1111/rssa.12786.

5. CLS COVID-19 Surveys: Sweep 3 (February – March 2021)

- During the COVID-19 pandemic, participants of the MCS (parents and cohort members), BCS70, NCDS and Next Steps were invited to complete surveys on their pandemic experiences. The third sweep was conducted using a sequential mixed mode approach with a subset of participants offered web initially followed by telephone interview. Other participants were allocated to web interview only.

The (between person) within sweep elements for mainstage survey data in (3)-(5) above means that there is also mixed mode data collection for these studies between sweeps. Additional between-sweep mixed mode data collection includes:

1. MCS

- In Sweeps 1-6, parents completed interviews via face-to-face interview. In Sweep 7, web survey was used, though interviewers encouraged parents to complete this while the interviewers were in the house interviewing cohort members.

2. Next Steps

- Sweeps 1-4 (13/14y – 16/17y) were carried out by face-to-face interview. Sequential mixed mode approaches were used in Sweep 5-9 (17/18y-32y), with participants able to respond via web survey, telephone or face-to-face interview and, latterly, alternately video interview (age 32y).

3. BCS70

- Sweeps 1-4 (0y-16y), 6-7 (30y-34y) and 9-10 (42y-46y) were carried out by face-to-face interview. Sweep 5 (26y) was carried out by postal paper questionnaire and Sweep 8 (38y) was carried out by telephone.

4. NCDS

- The Sweeps 0-6 (0y-42y), the Biomedical sweep, and Sweep 8 (50y) were carried out by face-to-face interview. Sweep 7 (46y) was carried out by telephone and Sweep 9 (55y) was carried out by telephone or web (see above).

5. CLS COVID-19 Surveys

- The initial two sweeps were carried out by web interview, while Sweep 3 was carried out by telephone or web interview. The use of web or telephone interview differs from pre-COVID-19 sweeps in each of the

surveys, except Sweeps 5-8 (17/18y – 25y) of Next Steps, Sweep 8 (38y) of the BCS70, and Sweeps 7 (46y) and 9 (55y) of the NCDS.

Future and ongoing data collections including mixed mode elements are also planned. Specifically:

- MCS Sweep 8 Interview (Age 23y)
  - Web and face-to-face
- BCS70 Sweep 10 Interview (Age 51y)
  - Face-to-face and video interview
- NCDS Sweep 10 Interview (Age 62y)
  - Face-to-face and video interview

Recommendations for Accounting for Mode Effects

Below we list nine recommendations for handling mode effects. These apply in analyses where data were collected using multiple modes, either within sweep (i.e., where a mixed mode design was used) or between sweeps (e.g., where a telephone survey was preceded by face-to-face sweeps).

First steps

1. Investigate whether a variable is likely to suffer mode effects, a priori.
   - Section 3 of the main document outlines the characteristics of items susceptible to mode effects. d'Ardenne et al. (2017) provide a checklist for scoring items on their likelihood of exhibiting mode effects. If the variable is unlikely to suffer mode effects, disregard Recommendations 2-8, unless it is possible to test this directly (e.g., using data from the NCDS Sweep 9 mode experiment).

2. Determine the likely size of the mode effect based on previous literature.
   - Several experimental studies have been carried out to determine the size and nature of mode effects (see Section 5 of the main document for examples). These span multiple samples and characteristics measured (e.g., sociodemographic, attitudes and physical and mental health). To our knowledge, for continuous variables, these almost always do not exceed 0.3 SD, though a judgment about the transportability of results needs to be made.

3. Draw out your assumptions of the mode effect and mode selection processes relevant to your substantive analysis using directed acyclic graphs (DAGs; see Section 4 of main document).
   - Determine from the DAG whether it is possible to unbiasedly estimate mode effects.

Analysis

4. Report descriptive statistics on survey mode, including the proportions of the sample in each mode and the characteristics of participants in each.

- For the descriptive statistics, use measures that are unlikely to suffer from mode effects, if possible. Otherwise, mode differences may reflect mode effects rather than mode selection. These descriptive statistics should be additional to the standard "Table 1" descriptive statistics produced for the analysis.

5. Run a 'naïve' analysis not accounting for survey mode.

6. (If supported by the DAG), run the substantive analysis accounting for mode effects.

   - Use either the (a) statistical control, (b) estimating the counterfactual value directly, or (c) multiple imputation approaches.

   - The loss of statistical power from using the multiple imputation approach can be gauged by comparing standard errors from the naïve and MI analyses and by examining the number of participants in each survey mode.

7. Run a sensitivity analysis positing values for the mode effects.

   - Assume plausible or implausible values (ideally both) and examine how results change according to mode effects assumed. Perform an analysis even if mode effects can in principle be estimated unbiasedly as unbiased estimates may only be obtainable in practice in a subsample of participants (e.g., those eligible in a mode experiment) and may not geenralise to other groups. Reviewers and other researchers may also disagree with the arguments supporting the belief that mode effects can be estimated unbiasedly given available data.

Reporting

8. Report the results of sensitivity analysis and (if applicable) other analyses performed accounting for mode effects.

   - Describe whether results change quantitatively over the range of mode effect parameters examined. If possible, state what level of mode effect would be required to change results qualitatively.

9. Discuss mode effects in strengths and limitations sections.

   - Describe transparently the likelihood that mode effects are (a) present and (b) could alter substantive conclusions.