SURVEY FUTURES

SURVEY DATA COLLECTION METHODS COLLABORATION

Considering the impacts of mixed mode survey design on UKHLS COVID-19 Study dataset performance

Jamie C. Moore

ISER, University of Essex

University of Essex | University of Southampton | UKRI Economic and Social Research Council

NCRM National Centre for Research Methods | Office for National Statistics | UCL | Warwick The University of Warwick | Manchester 1824 The University of Manchester | City University of London Est 1894 | National Centre for Social Research | LSE The London School of Economics and Political Science | Ulster University | Unil UNIL | Université de Lausanne | Ipsos | Kantar Public

Web first sequential mixed mode survey designs

- Offering web mode then following up non-respondents face to face (CAPI) or by telephone (CATI) aims to balance survey costs against non-response biases and precision loss.

- Should reduce the amount of costly in person interviewing while still retaining its benefits in terms of minimising biases and precision loss.

- Such designs are becoming increasingly popular.

- However, are refinements possible?

- Can costs / biases / precision loss be further reduced?

UKHLS COVID-19 Study

- Sample people in all UKHLS main survey HHs eligible for next annual interview, 9 waves Mar 2020 – Oct 2021.
- Web based: all issued to web, whether a regular internet user (RNU) or not.
- At waves 1 and 6 some non-regular net user (NRNU) web non-respondents followed up by CATI i.e. a web first sequential mixed mode design.
- CATI follow ups improved the wave 1 web dataset in terms of non-response weight performance (biases and precision loss).
- NRNU respondents did not improve the web dataset.
- Hence, could costs have been reduced by not allocating NRNUs to web, and datasets improved by instead expanding CATI?

Research questions

1. Did the web-first design perform better than web only at wave 6 as well?

- Study longitudinal, so datasets including respondents to all waves only must be evaluated as well.
- Quantify weighted dataset biases and precision loss.

2. Did allocating NRNUs to web improve datasets?

- If no, not allocating them to web to reduce costs would not have affected performance.
- Evaluate similarly to RQ 1.

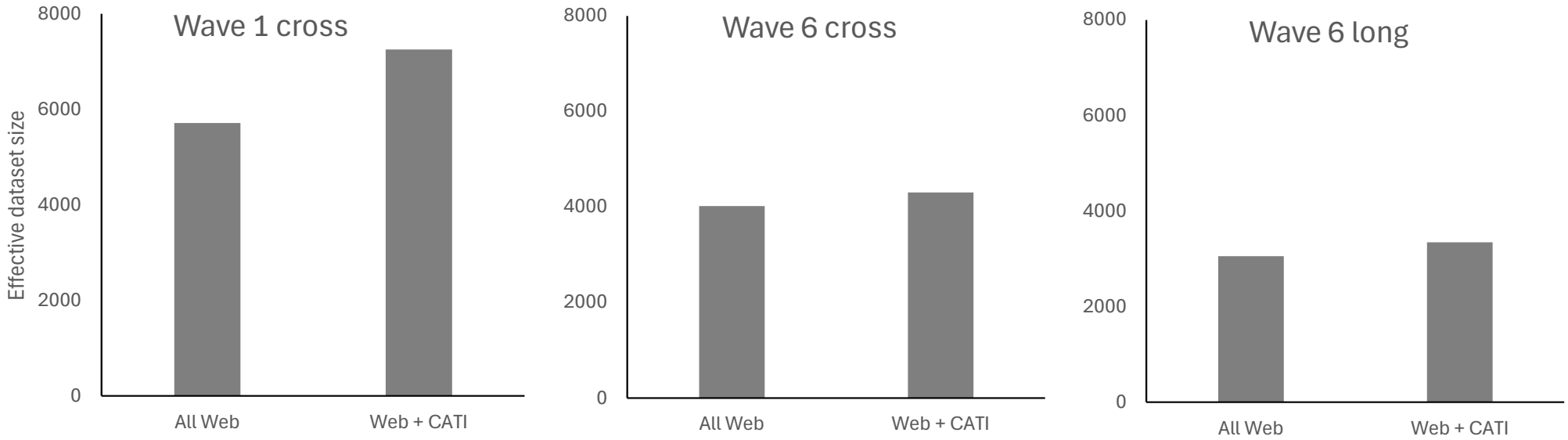
3. Could expanding CATI have improved datasets?

- Approaches limited by Study having ended.

RQ 1 & 2: Dataset sizes and response rates (RRs)

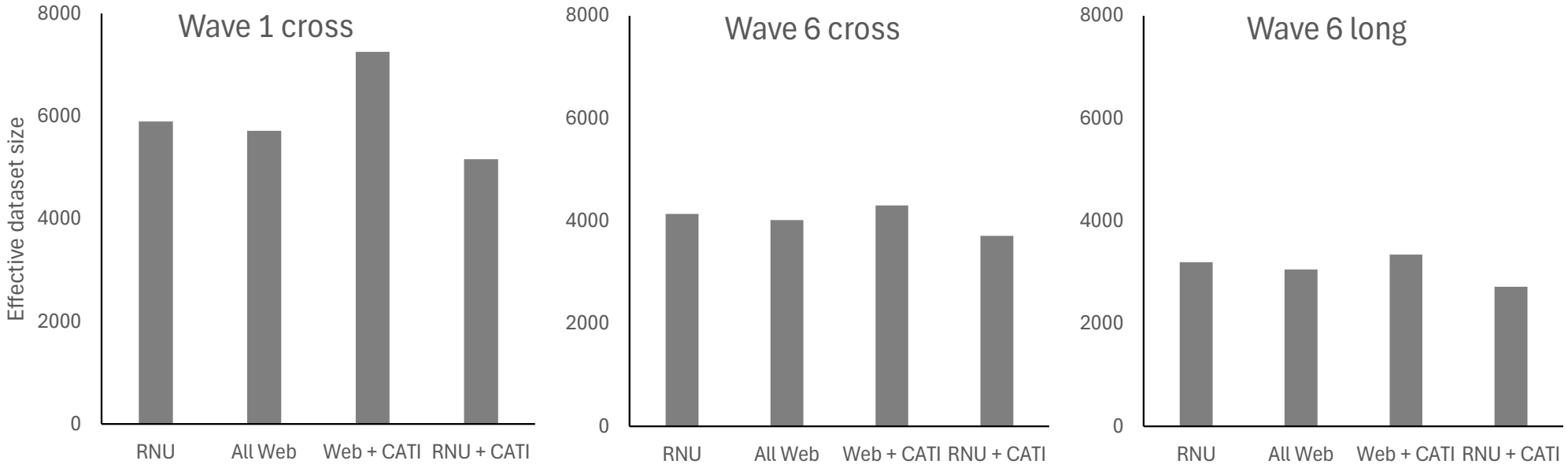
- Dataset sizes: web plus CATI > all web (just) > RNU plus CATI > RNU.
- RRs: RNU > CATI > NRNU web.
- Most respondents RNUs; NRNU web & CATI ~ 6% each.
- RRs lower at wave 6, NRNU and CATI contributions decrease slightly.

RQ 1: Weight performance



- Baises studied for 15 main survey characteristics, differences between Study non-response weighted estimates and similar main survey weighted sample benchmarks quantified.
- Means of absolute biases standardised by benchmark standard deviations (MASBs) reported.
- **Not shown, but web plus CATI MASBs similar to web (largest = 0.023).**
- Precision loss quantified with effective dataset sizes (N_{DEFF} = N / DEFF (= variance inflation compared to SRS with 100% RR).
- **Precision loss less (DEFFs lower, N_{DEFF} larger) for web plus CATI.**

RQ 2: Weight performance



- **Not shown, but RNU and RNU plus CATI MASBs similar to all web and web plus CATI MASBs respectively.**
- **Precision loss less (DEFFs larger, N_{DEFF} smaller) when NRNU web respondents not included in web datasets, but greater when they are not included in web plus CATI datasets.**

RQ 3: Would expanding CATI have improved datasets?

- A simulation study undertaken.

- Extra CATI respondents simulated by randomly selecting NRNUs not issued to CATI given sample sex * age * region grouping frequency and estimated response propensity using same predictor set.
- These added to empirical RNU plus CATI datasets, then performance in terms of biases and precision loss compared to that of benchmark empirical web plus CATI datasets.

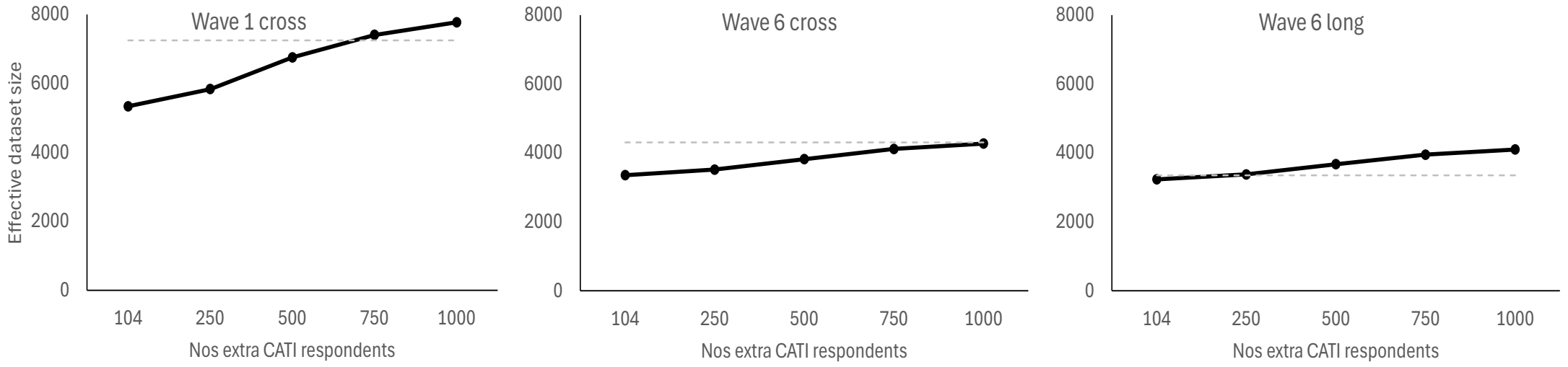
- 104*, 250, 500, 750 or 1000 extra respondents added to wave 1 dataset.

- In Study, only wave 1 respondents sampled again at wave 6.
- Incorporated into algorithm, so with non-response fewer added to wave 6 datasets.

- 1000 simulations for each of the 15 scenarios.

* The number obtained for the cost of allocating NRNUs to web given the Study 21.7: 1 web: CATI mode cost ratio.

RQ3: Simulated dataset weight performance



- Biases evaluated using simulated dataset MASBs similar to RQ 1 and RQ2.
- **Not shown, but simulated dataset MASBs similar to for benchmark datasets.**
- Precision loss evaluated using mean DEFFS and N_{DEFF}s.
- **Precision loss less (DEFFs lower, N_{DEFF} larger) than for the benchmark dataset with => 750 extra respondents at wave 1, and with => 250 extra with the wave 6 longitudinal dataset.**

Research implications

- That the utilised COVID-19 Study design performed better than web only.

- Web datasets had limited biases and acceptable precision loss, but CATI follow up improved them.
- Suggests that follow up of web non-respondents should continue.

- That the design also performed better than several refinements.

- Not allocating NRNUs to web often reduced performance, so could not have been used to reduce costs.
- Simulations showed that not allocating NRNUs to web and instead expanding CATI would not have improved datasets at the Study web: CATI cost ratio.
- Suggests the design was a good choice given resources available and Study web: CATI mode cost ratio.

- That expanding CATI may be optimal when the web: CATI cost ratio is smaller.

- Sometimes it improved dataset performance when more respondents added.
- Larger surveys (per unit fixed costs lower)? Surveys where (CATI) questionnaire has a (less costly) self-completion component?

Thanks for listening!

- This research was supported by the UKRI-ESRC grant **ES/T002611/1** for “**Understanding Society: The UK Household Longitudinal Survey Waves 13-15”** and the UKRI-ESRC strategic research **grant ES/X014150/1** for “**Survey data collection methods collaboration: securing the future of social surveys**”, known as **Survey Futures**. Survey Futures is directed by Professor Peter Lynn, University of Essex, and is a collaboration of twelve organisations, benefiting from additional support from the Office for National Statistics and the ESRC National Centre for Research Methods. Further information can be found at www.surveyfutures.net.