Understanding Society
Working Paper Series
2025 – 04
May 2025

Quantifying the impacts of web-first sequential mixed mode survey design on UKHLS COVID-19 Study dataset quality

Jamie C. Moore[1] and Gabriele Durrant[2,3]

[1]Institute for Social and Economic Research, University of Essex, UK
[2]Department of Social Statistics and Demography, University of Southampton, UK [3]National Centre for Research Methods, University of Southampton, UK

Understanding Society

Economic and Social Research Council

Non-technical summary

Many social surveys have adopted web first sequential mixed mode designs in which first a web questionnaire is offered, then non-respondents followed up face-to-face (Computer Assisted Personal Interviewing, CAPI) or by telephone (Computer Assisted Telephone Interviewing, CATI). Such designs may be less costly than CAPI or CATI only designs and may produce datasets of higher quality than web only designs. However, there has been little research evaluating dataset quality in longitudinal surveys at (later) waves after the design is introduced, which are equally important to survey quality but suffer from greater levels of attrition and, in the case of longitudinal datasets including only respondents to all waves, non-response. In addition, work on design refinements to reduce costs and / or improve dataset quality is limited. We address these questions using data from the UK Household Longitudinal Study (UKHLS) COVID-19 Study, which utilised a web with CATI follow-up design at several waves.

Key findings are that: 1) such follow-ups improve response rates and dataset sizes at both waves and for both cross-sectional (all respondents to the wave) and longitudinal datasets; 2) such follow-ups improve dataset representativeness compared to eligible samples for all considered datasets; 3) such follow-ups improve non-response weighted dataset quality in terms of remaining non-response biases and precision loss for all datasets; and 4) design refinements, namely not web sampling non-regular internet users and instead using the savings to expand CATI of such individuals, reduce the quality of some datasets (in the first case) and only improve datasets when more extra respondents can be added than obtainable for the cost of reducing web sampling (in the second). We then discuss the implications of our findings for survey practice.

Quantifying the impacts of web-first sequential mixed mode survey design on UKHLS COVID-19 Study dataset quality

Jamie C. Moore[1] and Gabriele Durrant[2,3]

[1] Institute for Social and Economic Research, University of Essex, UK.

[2] Department of Social Statistics and Demography, University of Southampton, UK.

[3]National Centre for Research Methods, University of Southampton, UK

Corresponding author: Jamie Moore, Institute for Social and Economic Research, University of Essex, Colchester, Wivenhoe Park, Essex CO4 3SQ, United Kingdom, moorej@essex.ac.uk.

Abstract:

This paper investigates the impacts of follow-ups of web non-respondents with CATI (Computer Assisted Telephone Interviewing) on cross-sectional (all respondents to the survey wave) and longitudinal (only respondents to the wave and all waves previous) dataset quality, and of refinements to the survey design on datasets. The analysis uses data from the UK Household Longitudinal Study (UKHLS) COVID-19 Study. The impacts of follow-ups of non-respondents on response rates, representativeness and weighted dataset quality are quantified, along with those of not web sampling non-regular internet users and instead expanding use of CATI. Implications of findings for survey practice are then discussed.

Keywords: follow-up of web non-respondents, web surveys, survey non-response, representativeness, non-response weighting, data quality.

JEL classification: C81, C83

Acknowledgements:

We would like to thank Paul Clarke (ISER) for comments on an earlier version of this manuscript.

This research was supported by 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, benefitting 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.

Research Strand 8 addresses nonresponse and the project led by Prof Gabriele Durrant focusses on non-response follow-ups to web surveys of the UK general population. It is a collaboration between the University of Southampton and the University of Essex, jointly with IPSOS, Verian, NatCen and ONS. The aim of the workstream is to review existing evidence, analyse recent data and trends and to provide guidance to survey agencies and researchers with regards to follow-up of non-respondents and representativeness in web surveys. The work was also supported by the National Centre for Research Methods (NCRM, 2020-2025) (ESRC grant number: ES/T000066/1), and Understanding Society (ESRC grant number: ES/Y010469/1).

The Understanding Society survey is funded by ESRC grants awarded to Nick Buck (RES-586-47-0002, ES/K005146/1) and Michaela Benzeval (ES/N00812X/1, ES/S007253/1, ES/T002611/1, ES/Y003071/1, ES/Y010469/1), with co-funding from various Government Departments. The COVID-19 Study was funded by the Economic and Social Research Council (ES/K005146/1) and the Health Foundation (2076161). The data are collected by the National Centre for Social Research (NatCen) and Verian (formerly Kantar Public). The COVID-19 Study was funded by the Economic and Social Research Council (ES/K005146/1) and the Health Foundation (2076161). Fieldwork for the survey was carried out by Ipsos MORI and Kantar Public (now Verian). The data are distributed by the UK Data Service (main survey: 10.5255/UKDA-SN-6614-20; COVID-19 Study: 10.5255/UKDA-SN-8644-8).

1. Introduction

In recent years, many social surveys have adopted web first sequential mixed mode designs in which first web mode is offered, then non-respondents followed up in interviewer administered modes (face-to-face (Computer Assisted Personal Interviewing, CAPI) or telephone (Computer Assisted Telephone Interviewing, CATI) (Brown & Calderwood 2020). Compared to CAPI or CATI only designs, these designs can reduce survey costs (Dillman 2014, p.401). Compared to web only designs, they can increase dataset quality in terms of dataset size, resemblance to study populations (representativeness) and non-response bias, although they can also cause measurement differences, where responses depend on mode (De Leeuw 2018; Burton & Jackle 2020). However, there is little work on how follow-ups impact on dataset quality in longitudinal surveys at waves after the one at which they are introduced (later waves), which are equally valuable but subject to greater sample attrition and non-response. In addition, research on design refinements to reduce costs and / or increase dataset quality is limited. Answering these important questions for survey designers is the aim of this paper.

To address our questions, we use datasets from the UK Household Longitudinal Study (UKHLS) COVID-19 Study, a high frequency longitudinal web survey of participants in the UKHLS main survey undertaken during the COVID-19 pandemic in which at several waves some non-respondents were followed up by CATI. We define *dataset quality* in terms of response rates and therefore dataset sizes, representativeness, and remaining non-response biases and precision loss after non-response weighting. Whilst we acknowledge the importance of measurement differences, they are not studied in this paper and are left to future work. We consider the following four research questions

RQ1: How do follow-ups affect response rates and dataset sizes at later waves?

RQ2: How do follow-ups affect dataset representativeness at later waves?

RQ3: How do follow-ups affect remaining non-response biases and precision loss after non-response weighting at later waves?

RQ4: Are refinements to the survey design to reduce costs and / or increase dataset quality, namely not web sampling non-regular internet users and using the savings to expand CATI of them, possible?

1.1. Motivation: declining response rates – a challenge for modern survey design

Declining response rates are a challenge for surveys (de Heer & de Leeuw 2002; Luiten et al. 2020). They reduce dataset size, inflating survey estimate variances (precision loss). If non-respondents and respondents differ, they can also cause estimates to deviate from population values (non-response biases), leading to invalid inference. Consequently, significant effort is expended on maximising dataset quality. Measures are undertaken before or during data collection to increase response by under-represented subgroups, for instance by following-up non-respondents (bias prevention measures: Groves et al. 2001; Groves & Heeringa 2006). They are also undertaken post collection to reduce remaining biases, such as producing non-response weights (bias adjustment measures: e.g. Valliant & Dever 2013). Note as well that bias prevention measure effectiveness can increase adjustment ability to reduce biases (Lundquist & Sarndal 2013; Sarndal & Lundquist 2014a, b; Schouten et al. 2016), and, as adjustments are inefficient (Little & Vartivarian 2005), reduce precision loss due to their use.

These efforts to maximise dataset quality increase survey costs. One solution to this issue is to replace CAPI or CATI with less costly modes such as web (Couper et al. 2007; Schonlau et al. 2009; Baker et al. 2010; Olson et al. 2021). Beyond cost, web mode may also

increase response in some subgroups (McGonagle & Sastry 2023). Its disadvantages are that overall response rates are often lower (Fricker et al. 2005; Jackle et al. 2015; Kirchner & Felderer 2016; Daikeler et al. 2020; Wu et al. 2022), and that dataset quality (with the proviso regarding measurement differences noted earlier) tends to be maximised by use of both web and other modes (mixed mode designs: e.g. Cornese & Bosnjak 2018; Burton & Jackle 2020; Peytchev et al. 2022). Hence, many surveys have begun to adopt web first sequential mixed mode designs in which first web mode is offered, then non-respondents followed up by CAPI or CATI (Brown & Calderwood 2020; van Berkel et al. 2020, 2024; Institute for Social and Economic Research 2021, 2024a, 2024b; Lipps & Pekari 2021; Voorpostel et al. 2021; McGonagle & Sastry 2023). These can reduce costs compared to CAPI or CATI only designs (Lipps & Pekari 2021; McGonagle & Sastry 2023) and improve dataset quality compared to web only designs (Dillman et al. 2009; Klausch et el. 2015; Lipps & Pekari 2021; Mackeben & Sakshaug 2023; McGonagle & Sastry 2023; Moore et al. 2024).

So far though, most research on the impacts on dataset quality of such designs compares datasets to those from the single non-web mode designs used previously in surveys (e.g. Bianchi et al. 2017; Voorpostel et al. 2021), or quantifies the impact of follow-ups on cross-sectional survey datasets or longitudinal survey datasets at the wave at which the design is introduced (Klausch et al. 2015; Lipps & Pekari 2021; Mackeben & Sakshaug 2023; McGonagle & Sastry 2023; Moore et al. 2024). Little work exists on how in longitudinal surveys follow ups impact on datasets at waves after the introductory wave (later waves), which are equally important to survey quality. At these waves, in addition to non-response datasets will be affected by greater levels of attrition, where sample members drop out due to death, moving out of scope or refusal to participate (Lynn 2006). Both these factors will impact on the longitudinal datasets that include only respondents to all waves and are the focus of these

surveys, and also (to a lesser extent with non-response) on cross-sectional datasets that include all respondents to the wave irrespective of previous responses, which are often produced from the data as well. Moreover, work on refinements to these web-first designs is in its early stages. Limited information exists on whether it is possible to reduce survey costs and / or improve dataset quality. Longitudinal surveys are a major social science investment, so research to reduce these knowledge gaps is of considerable value.

1.2. Previous research relating to research questions

Concerning RQs 1 to 3, the most relevant work investigates changes over time in the impacts of CAPI follow ups on dataset quality in the UKHLS Innovation Panel (IP: Moore et al. submitted). Since 2012, a third of the annually interviewed sample has been allocated to a web-first with CAPI follow-up design. Moore et al. studied cross-sectional datasets from this sub sample covering the period 2012-2023. In the current context, the period 2012-2017 is of primary interest: after, refreshment samples with no previous experience of web mode entering the analysis samples complicate results. Over this period, follow-ups increased dataset sizes, though to a decreasing extent, with web response rates increasing. Impacts on representativeness compared to issued samples and on non-response biases remaining after non-response weighting declined, with web dataset representativeness and weighted dataset biases improving (note that UK population internet access levels increased considerably over the study period). Precision loss due to weight use was not quantified, and longitudinal datasets were too small to be studied. In addition, after 2017, follow-ups still increased the size of datasets including refreshment samples (at a declining rate), but they did not improve representativeness or weighted dataset biases.

Concerning RQ4, one way to reduce web first design costs is to first offer individuals the mode they are most likely to respond by. In the UKHLS main survey, web-first with CAPI follow-up is used for those predicted to be most likely to respond by web (Lynn 2017). For those less likely to respond by web, CAPI-first with web follow-up is used. Regarding follow-ups, Jackson et al. (2024) predict addresses unlikely to respond to CATI follow-up in the California Health Interview Survey, and report that calling them less often than those more likely to respond reduces costs without affecting dataset quality. In the UKHLS COVID-19 Study, Moore et al. (2024) report that non-regular internet user respondents have no impact on wave 1 web dataset representativeness and weighted dataset quality. This implies that not offering web mode to such individuals could reduce costs without affecting datasets. It also raises the possibility that datasets could be (further) improved by using the savings to instead obtain responses from more of them by CATI. However, so far neither the impact of reduced web sampling on web plus CATI datasets nor that of expanding CATI has been quantified.

2. Data

The UKHLS COVID-19 Study eligible sample is drawn from UKHLS main survey participants, and we use main survey information in our investigations. Hence, we briefly describe both surveys.

2.1. The UKHLS main *Understanding Society* survey

The UKHLS main survey began in 2009 and surveys people in the UK (Institute for Social and Economic Research 2024). Its sample consists of probability samples, including several refreshment samples. Annual interviews are sought from all adults in sample HHs. It has a

sequential mixed-mode design: some sample members are allocated to web and others to CAPI, with follow-up in other modes. Research shows that the survey continues to support valid population inference (Benzeval et al. 2020).

2.2. UKHLS COVID-19 Study

When the pandemic began, individuals in UKHLS main survey wave 10 sample HHs were asked to complete extra web surveys on how it was affecting them. The COVID-19 Study eligible sample consisted of those aged 16+ who had not attrited, died or emigrated prior to its first wave in April 2020. Eight further waves were fielded, the last in September 2021. Ca. 1/3 of web non-respondents identified from main survey information as non-regular internet users (those who use it less than once a week, with those lacking information similarly designated) were followed up by CATI at wave 1, with respondents and 12 others who requested it again followed up in the mode at wave 6 (Institute of Social and Economic Research 2021).

3. Methods

In addition to the wave 6 cross-sectional and longitudinal datasets, we consider the COVID-19 Study wave 1 dataset despite it having been studied by Moore et al. (2024). This is because our analyses differ slightly, and because we use a more recent main survey dataset release that includes non-response weights (used as 'selection' weights in COVID-19 Study non-response weight construction and also to weight eligible sample estimates of respondent characteristics used to evaluate such weights: see section 3.2) for more respondents: see the online Appendix for more details of the datasets used in this paper.

3.1. Methods to evaluate dataset sizes, response rates and representativeness

We quantify COVID-19 Study eligible sample sizes at waves 1 and 6, along with (both cross-sectional and longitudinal dataset at wave 6) respondent numbers and response rates, both for each mode and overall. To assess dataset representativeness, for those with main survey wave 9 weights, for each dataset we also quantify the main survey wave 9 measured sociodemographic characteristics of respondents by each mode and overall and compare them to the characteristics of their eligible samples. We consider the following characteristics: Sex (male, female); Age (16-29, 30-39, 40-49, 50-59, 60-69, 70+); Qualifications (Degree, A level, GCSE or lower); HH structure (Single, no kids, Single, kids, Couple, no kids, Couple, kids, Other); Ethnic minority (Yes, No); Country (England, Wales, Scotland, Northern Ireland); HH Tenure (Owned, Mortgage, Rented, Social Housing); HH income (continuous); and Longstanding illness (Yes, no). We compute their prevalences and 95% confidence intervals (CIs).

3.2. Non-response weight construction and evaluation

3.2.1. Non-response weight estimation

The aim of the COVID-19 Study non-response weights is to map respondents to the UK population at main survey wave 9 (2017-18), updated for mortality and emigration but not immigration. We compute them for each considered dataset as the product of the main survey wave 9 cross-sectional non-response weights and a regression-based adjustment for Study non-response, so estimation depends on having main survey weights. We use a range of sociodemographic, economic, health and survey design variables as predictors in regression models. We describe the methods used to identify final models and estimate non-response adjustments and weights in the online Appendix.

3.2.2. Methods to evaluate non-response weighted dataset quality

A difficulty when evaluating non-response biases is finding benchmark population values (e.g. Hand 2018). Hence, we instead evaluate biases by comparing COVID-19 Study non-response weighted mean estimates of main survey measured respondent characteristics to equivalent eligible sample main survey wave 9 weighted benchmarks. Note that this approach relies on the main survey weights effectively mapping respondents to the population: see Benzeval et al. (2020) for evidence of this. To statistically compare estimates, we use the test of Moore et al. (2024), which accounts for dataset partial dependencies (respondents are subsets of eligible samples): see the online Appendix for details. 15 characteristics are studied: 10 that are included in weighting models (Subjective financial situation (SFS): comfortable or OK; SFS: just about getting by; SFS: finding it quite / very difficult; Tenure: owner occupied; Tenure: mortgage; Tenure: rented, Tenure: social housing; Low skill occupation: yes; Any savings income: yes; Behind with some or all bills: yes), and 5 that are not (Income poverty: yes; Receives core benefit: yes; Visited GP in last year: yes; Smoker: yes; Hospital outpatient in last year: yes). All are binary variables. We also compute mean absolute standardized biases (MASBs), the means of absolute biases between estimates and benchmarks divided by benchmark estimate standard deviations, and their 95% CIs. These means are our main focus in the paper.

To evaluate precision loss due to weight use, we utilise the DEFF (Kish 1965). This metric provides a conservative estimate (weighting variables and outcomes of interest are assumed to be uncorrelated) of the extent to which survey sampling error is expected to depart from that under simple random sampling with a 100% response rate:

$$DEFF = 1 + (SD(weights)/mean(weights))^2, \quad (1)$$

where $SD(weights)$ is the weight standard deviation. A larger value implies greater precision loss. We also estimate effective (weighted) dataset sizes (N_{DEFF} = *N* / DEFF). N_{DEFF} is affected by unweighted dataset size, but quantifies impacts on datasets used by substantive researchers.

3.3. The impacts of not web sampling non-regular internet users on dataset quality

To evaluate the impacts of not web sampling non-regular internet users on dataset quality, we remove them from web only and web plus CATI respondent datasets, use the methods described in sections 3.2 and 3.3 to quantify dataset sizes, representativeness and non-response weighted dataset quality, and compare findings to those for the datasets including such respondents.

3.4. The impacts of expanding CATI on dataset quality

Empirical work investigating the impacts of expanding CATI on dataset quality is not possible due to the COVID-19 Study having ended. Hence, we utilise a simulation approach.

3.4.1. Simulated dataset generation

Simulated dataset generation is fully described in the online Appendix. We provide a brief overview here. We simulate extra CATI respondents from main survey wave 9 weighted non-regular internet users not issued to CATI, including those that responded by web. We predict synthetic CATI response probabilities for these individuals given modelling of those of non-regular internet users that were issued to CATI (note that this assumes that such probabilities are same for both groups of individuals, a point we return to in section 7). Then, we use these

probabilities to simulate extra CATI respondents, utilising a multinomial sampling without replacement procedure. At wave 1, we simulate datasets with *n* = 104, 250, 500, 750 and 1000 extra respondents. 104 is the rounded number interviewable for the cost of web sampling non-regular internet users at wave 1 given that 21.7 successful web responses were obtained for the cost of one successful CATI response in the COVID-19 Study (Moore et al. 2024) i.e. the number responding, 2247 (see Table 2), divided by 21.7. Note that a further breakdown of these costs is not available, and also that we could have added to the 104 figure the 33 CATI respondents that would have been obtained by not web sampling the real CATI respondents: we did not do so to provide the lowest possible figure given the experienced cost ratio. The other *n*'s reflect scenarios in which this cost ratio is reduced. For each *n*, we generate 1000 datasets. In the COVID-19 Study, only wave 1 CATI respondents and a few others are issued to CATI at wave 6 (see section 2.1), and not all respond (see Table 1). Hence, utilising a procedure analogous to that used for wave 1, we select extra wave 6 respondents from the extra wave 1 respondents only, and add the same individuals to both the cross-sectional and longitudinal datasets (see online Appendix, Table 3 for simulated dataset sizes).

3.4.2. Methods to evaluate simulated dataset quality

We evaluate simulated datasets in terms of non-response weighted dataset quality (we do not consider dataset representativeness due to space constraints). We estimate weights for each dataset as in section 3.2.1. Concerning biases, we compare weighted estimates of respondent characteristics to eligible sample benchmarks and compute MASBs as in section 3.2.2. We then compare means across simulated datasets for each dataset / number of extra wave 1 respondents and 95% CIs (prediction intervals (PIs) in these scenarios) to relevant benchmark empirical web plus CATI dataset MASBs. Concerning precision loss, we compute

DEFFs and N_{DEFF}s and for each dataset / extra respondent number and compare means and 95% PIs to relevant empirical dataset benchmarks.

4. Results

4.1. RQ1: How do follow-ups affect response rates and dataset sizes at later waves?

We report response rates and dataset sizes in Table 1. Eligible sample size at wave 1 is 43,981, 33,951 with main survey wave 9 weights. 18,479 respond, 16,680 with a main survey wave 9 weight, a response rate of 42.01%. Of these, 17,761 respond by web, 16,009 with a main survey wave 9 weight, a response rate of 40.40%. 3,398 non-regular internet users that do not respond by web are issued to CATI. 718 respond, 671 with a main survey wave 9 weight. Hence, follow-ups increase dataset size by 4.04%.

Eligible sample size at wave 6 is 43,862, 33,847 with main survey wave 9 weights. Regarding the wave 6 cross-sectional dataset, 12,424 respond, 11,620 with a main survey wave 9 weight, a response rate of 28.33%. This is less than at wave 1, an expected result due to increased attrition (see Introduction). Of these, 12,035 respond by web, 11,248 with a main survey wave 9 weight. a response rate of 27.44%. 730 individuals are issued to CATI, the wave 1 CATI respondents plus 12 others who requested it. 391 respond, 374 with a main survey wave 9 weight. Hence, follow-ups increase dataset size by 3.25%.

Regarding the longitudinal dataset, of the wave 6 eligible sample 11,784 respond having also responded at wave 1, 11,200 with a wave 9 weight, a response rate of 26.87%. This is less than for the wave 6 cross-sectional dataset, an expected result due to the greater impact of non-response when sample members must respond to all (both) waves (see Introduction). Of these, 11,392 respond by web, 10,683 with a main survey wave 9 weight, a

response rate of 25.97%. 383 individuals respond by CATI, 366 with a main survey wave 9 weight. Hence, follow-ups increase dataset size by 3.44%.

4.2. RQ2: How do follow-ups affect dataset representativeness at later waves?

We report the representativeness of respondents compared to the eligible sample in terms of main survey measured characteristics at wave 1 in Table 2. Regarding web respondents only, the following characteristics are significantly (i.e. estimate 95% CIs do not overlap) under-represented: Sex: male; Age: 20-39; Age: 80-89; Age: 90+; Qualifications: GCSE or lower; HH structure: single, no kids; Ethnic minority: yes; Country: Northern Ireland; Tenure: rented; Tenure: social housing; and Long standing illness: yes. The following characteristics are significantly over-represented: Age: 40-49; Age: 50-59; Age: 60-69; Qualifications: degree; HH structure: couple, no kids; HH structure: couple, kids; Country: England; Tenure: owned; and Tenure: mortgage. In addition, Household income is significantly higher. Patterns are similar for web plus CATI respondents, except estimates for Couple: kids and Long standing illness: yes no longer differ from the eligible sample, and Age: 70-79 becomes significantly over-represented. In addition, though differences remain significant those for some other characteristics are closer to eligible sample estimates, for example, Age: 80-89; Age: 90+; HH structure: single, kids; Qualifications: GCSE or lower; Tenure: social housing; HH income. Hence, follow-ups slightly improve the representativeness of respondents compared to the eligible sample.

We report the representativeness of respondents in the wave 6 cross-sectional dataset in Table 3. Very much similar differences compared to the eligible sample to those observed at wave 1 exist for both web respondents only and web plus CATI respondents, although they are often slightly larger in magnitude. This pattern is repeated, with differences

between the eligible sample and respondents even larger, in the wave 6 longitudinal dataset (Table 4). Hence, follow-ups slightly improve the representativeness of respondents compared to the eligible sample in the wave 6 datasets.

4.3. RQ3: How do follow-ups affect remaining non-response biases and precision loss after non-response weighting at later waves?

We report non-response weighted estimate mean absolute standardized non-response biases (MASBs) for datasets in Table 4, and biases for each considered characteristic in the online Appendix, Table 2. The wave 1 web respondent MASB is slightly larger than that for web plus CATI respondents, but differences are not statistically significant (i.e. estimate 95% CIs overlap). Regarding individual biases (variables are binary, so estimates are prevalences), four are significant for web respondents, and none for the web plus CATI respondents. The wave 6 cross-sectional web respondent MASB is slightly larger than that for web plus CATI respondents, but differences are not significant. Regarding individual biases, four are significant for web respondents, and none for web plus CATI respondents. The wave 6 longitudinal web respondent MASB is the same as that for web plus CATI respondents. Regarding individual biases, two are significant for web respondents, and none for web plus CATI respondents.

We report precision loss in Table 5. For all three datasets, web plus CATI respondent DEFFs are smaller than web respondent DEFFs, implying increased precision. Effective dataset sizes (N_{DEFF}) are also larger for web plus CATI respondents. When comparing datasets, a strategy of 'dataset *x* is of higher quality than dataset *y* only if its MASB is smaller or similar and its N_{DEFF} larger' is used, because the non-response weights are estimated to facilitate

analyses without (further) correction. Hence, given also the bias results, follow-ups improve non-response weighted datasets.

4.4. RQ4: Are refinements to the survey design to reduce costs and / or increase dataset quality possible?

4.4.1. The impacts of not web sampling non-regular internet users on dataset quality

4.4.1.1. Dataset sizes

We report web and web plus CATI dataset sizes with and without non-regular internet user web respondents in Table 1 (see also section 4.1 for datasets with non-regular internet user web respondents). At wave 1, 15,514 regular internet users respond by web (15,332 with a main survey wave 9 weight). Hence, given 17,761 respondents when they are included, not including non-regular internet user web respondents reduces web dataset size by 12.60%. 16,232 respondents are regular internet users who respond by web or CATI respondents (16,003 with a main survey wave 9 weight). Hence, given 18,479 respondents when they are included, not including non-regular internet user web respondents reduces web plus CATI dataset size by 12.16%.

In the wave 6 cross-sectional dataset, 10,875 regular internet users respond by web (10,754 with a main survey wave 9 weight). Hence, given 12,035 respondents when they are included, not including non-regular internet user web respondents reduces web dataset size by 9.64%. 11,266 respondents are regular internet users who respond by web or CATI respondents (11,128 with a main survey wave 9 weight). Hence, given 12,426 respondents when they are included, not including non-regular internet user web respondents reduces web plus CATI dataset size by 9.34%.

In the wave 6 longitudinal dataset, 10,350 regular internet users respond by web having also responded at wave 1 (10,233 with a main survey wave 9 weight). Hence, given 11,392 respondents when they are included, not including non-regular internet user web respondents reduces web dataset size by 9.15%. 10,733 respondents to waves 1 and 6 are regular internet users who respond by web or CATI respondents (10,599 with a main survey wave 9 weight). Hence, given 11,784 respondents when they are included, not including non-regular internet user web respondents reduces web plus CATI dataset size by 8.92%.

4.4.1.2. Dataset representativeness

We report the representativeness of respondents with or without non-regular internet user web respondents compared to the eligible sample at wave 1 in Table 2 (see also section 4.2 for datasets with non-regular internet user web respondents). At wave 1, regular internet user web respondent characteristics significantly differing from those for the eligible sample are the same as for all web respondents. Some estimates are slightly more similar to eligible sample estimates than those for all web respondents (for example, Age: 20-29, HH structure: couple, no kids, Tenure: owned and Tenure: rented), but others are slightly less so (for example, Qualifications: degree, Qualifications: GCSE or lower and HH income). Three more regular internet user plus CATI respondent characteristics significantly differ from those for the eligible sample than for web plus CATI respondents (Age: 40-49, HH structure: couple, kids and Longstanding illness: yes), though differences for age: 70-79 become non-significant. Hence, excluding non-regular internet users very slightly improves web respondent representativeness, but reduces that of web plus CATI respondents.

Differences compared to the eligible sample are also similar for regular internet user web respondents and all web respondents in the wave 6 cross-sectional dataset, though those

for Tenure: mortgage are significant for the former only (Table 3). Differences for regular internet user web plus CATI respondents and web plus CATI respondents are similar as well, though those for Age: 40-49 are significant for the latter only. Hence, excluding non-regular internet users slightly reduces the representativeness of web respondents, but improves that of web plus CATI respondents.

For the wave 6 longitudinal dataset, similar significant differences compared to eligible sample exist for regular interest user web and the all web respondents, although Tenure: mortgage is significant for the former only (Table 4). Note however, that estimates for a number of characteristics are slightly closer to the eligible sample for regular internet user web respondents. Differences for web plus CATI respondents are slightly greater than for regular internet user web plus CATI respondents, with those for Age: 40-49 significant for the former only. Hence, excluding non-regular internet users has little impact on the representativeness of web respondents, but reduces that of web plus CATI respondents.

4.4.1.3. Non-response weighted dataset quality

We report non-response weighted estimate mean absolute standardized non-response biases (MASBs) for datasets with and without non-regular internet user web respondents in Table 4, and biases for each characteristic in the online Appendix, Table 2 (see also section 4.3 for datasets with non-regular internet user web respondents). The wave 1 regular internet user web respondent MASB is slightly larger than that for all web respondents, but differences are not statistically significant (estimate 95% CIs overlap), with three, rather than four, individual biases significant. Similarly, the regular internet user web plus CATI respondent MASB is slightly larger than for web plus CATI respondents, but differences are not significant, with two, rather than no, individual biases significant. The same pattern exists with the wave

six cross-sectional dataset MASBs, with five regular internet user web respondent individual biases significant (four are for all web respondents), and no regular internet user web plus CATI respondent individual biases significant (similar to with web plus CATI respondents). With the wave six longitudinal dataset, the regular internet user web respondent MASB is slightly larger than that for all web respondents, but differences are not significant, and two, rather than three, individual biases are significant. The regular internet user web plus CATI respondent MASB is slightly smaller than that for web plus CATI respondents, but differences are not significant, with one, rather than no, individual biases significant.

We report precision loss in Table 5 (see also section 4.3 for datasets with non-regular internet user web respondents). At wave 1, the regular internet user web respondent DEFF is smaller than that for all web respondents, implying greater precision, and effective dataset size (N_{DEFF}) is very slightly (0.15%) larger. The regular internet user web plus CATI respondent DEFF is larger than that for all web plus CATI respondents, implying reduced precision, and effective dataset size (N_{DEFF}) is 28.73% smaller. Similar patterns exist for the wave 6 cross-sectional and longitudinal datasets, with effective dataset sizes for regular internet user web respondents respectively 0.96% and 5.80% larger than for all web respondents, and those for regular internet user plus CATI respondents 14.60% and 20.95% smaller than for all web plus CATI respondents. Hence, given our criteria for determining comparative dataset quality (see section 4.3), these results and the bias results imply that non-regular internet user web respondents improve web plus CATI, but not web, datasets.

4.4.2. The impacts of expanding CATI on dataset quality

We report the quality of regular internet user plus CATI datasets with extra CATI respondents in terms of non-response biases in Table 7. For all three considered datasets, mean MASBs

are small and decrease slightly as extra respondents increase (note that 104, 250, 500, 750 and 1000 extra wave 1 respondents translate into ~60, ~130, ~265, ~395 and ~530 extra wave 6 respondents respectively: see online Appendix, Table 1), with sometimes 95% PIs for neighbouring scenarios that do not overlap, implying increasing dataset quality. Mean MASBs and their 95% PIs are also compared to benchmark empirical web plus CATI dataset MASBs (see Table 4 for benchmarks). The wave 1 dataset benchmark MASB is 0.009. At 104 (the number obtained by not issuing non-regular internet users to web: see section 3.4), 250 and 500 extra respondents, mean MASBs and their 95% PIs are slightly larger than this benchmark, but at 750 extra they are similar and at 1000 extra they are smaller. With the wave 6 cross-sectional dataset, mean MASBs and their 95% PIs are all larger than the benchmark (= 0.010). With the wave 6 longitudinal dataset, at 104, 250 and 500 extra respondents they are similar to the benchmark (= 0.014), but at higher numbers they are smaller.

We report precision loss for these datasets in Table 8. For all three datasets, mean DEFFs decrease as extra respondents increase, with non-overlapping 95% PIs, implying reduced precision loss. Mean effective dataset sizes (N_{DEFFs}) increase, again with non-overlapping 95% PIs, implying increased dataset quality. Mean DEFFs and N_{DEFFs} are also compared to empirical web plus CATI dataset benchmarks (see Table 6 for benchmarks). For the wave 1 dataset, at 104, 250 and 500 extra respondents DEFFs and their 95% PIs are larger and N_{DEFFs} and their 95% PIs smaller than the benchmarks (= 2.299 and 7252.608 respectively: see Table 6), but at higher numbers they are respectively smaller and larger (mean effective dataset size increases = 2.09% to 7.12%). Hence, given also the bias results and our criteria for determining comparative dataset quality (see section 4.3), expanding CATI only improves the dataset when 750 or more extra respondents are added.

For the wave 6 cross-sectional dataset, at 104, 250, 500 and 750 extra wave 1 respondents DEFFs and their 95% PIs are slightly smaller than the benchmark (= 2.701), implying greater precision loss, but at 1000 extra they are slightly larger, implying reduced precision loss. N_{DEFFs} and their 95% PIs are always smaller than the benchmark (= 4303.478). Hence, given also the bias results, expanding CATI does not improve the dataset at any of the studied extra respondent numbers. For the wave 6 longitudinal dataset, at 104 and 250 extra wave 1 respondents DEFFs and their 95% PIs are larger than the benchmark (= 3.089), implying greater precision loss, but at higher numbers they are smaller, implying reduced precision loss. The N_{DEFF} and its 95% PI are smaller than the benchmark (= 3579.862) at 104 and 250 extra respondents, but at higher numbers they are larger (mean effective dataset size increases = 2.63% to 14.64%). Hence, given also the bias results, expanding CATI improves the dataset only when 500 or more extra wave 1 respondents are added.

7. Summary and conclusions

Summary: We examined the impacts of web first sequential mixed mode designs with CATI follow-ups on longitudinal survey dataset quality. We quantified response rates and dataset sizes, dataset representativeness and non-response weighted dataset quality in terms of remaining biases and precision loss, both at the wave the design was introduced and at the wave following, which was likely to be subject to greater levels of attrition. We considered cross-sectional (all respondents to the wave) and longitudinal (only respondents to all waves) datasets. the latter of which was likely to be more affected by non-response. In addition, we examined refinements to the design, namely not web sampling non-regular internet users to reduce costs, and, utilising simulation methods, using the savings to instead expand CATI of

them to improve dataset quality. We used data from the UKHLS COVID-19 Study, which at waves 1 and 6 followed-up a subset of web non-respondents by CATI.

Key findings: CATI follow-ups of web non-respondents improved dataset quality. Dataset sizes increased by 3-4%, and representativeness compared to eligible samples slightly increased. Non-response biases remaining after weighting remained similar, but precision loss was reduced and effective dataset sizes were larger. Not web sampling non-regular internet users sometimes slightly reduced the representativeness of web and web plus CATI datasets, but also sometimes improved it. It had little impact on biases in weighted web datasets, and reduced precision loss and increased effective dataset sizes even though it decreased numbers of individuals included. However, though it had little impact on weighted dataset biases, it increased precision loss and reduced effective dataset size in web plus CATI datasets. Using the savings from not web sampling non-regular internet users to expand CATI did not improve datasets at the number of extra respondents that would have been obtained given the web: CATI interview cost ratio in the survey, but did improve the wave 1 and wave 6 longitudinal datasets in terms of precision loss and effective dataset size at higher extra respondent numbers.

These findings are the first concerning the impact of interviewer administered (CAPI or CATI) follow-ups of web non-respondents at waves after the one at which web-first sequential mixed mode designs are introduced in longitudinal datasets from longitudinal surveys. Most previous work has considered impacts at introductory waves or in cross-sectional surveys, with similar findings to our own on wave 1 of the COVID-19 Study (Klausch et al. 2015; Lipps & Pekari 2021; Mackeben & Sakshaug 2023; McGonagle & Sastry 2023). The exception is that of Moore et al. (submitted), who quantified the impact of CAPI follow-ups on cross-sectional datasets from a longitudinal survey over an 11 year period, and found that

dataset improvements had become negligible by the time of the COVID-19 Study (2020-21). Possibly, differences between these findings and ours on the wave 6 cross-sectional COVID-19 Study dataset are due to CATI and CAPI obtaining responses from individuals with different characteristics, or most individuals in Moore et al's. datasets having had prior experience of web mode (only ~60% had in our datasets). In addition, our findings concerning design refinements are the first to quantify the impacts on dataset quality of selecting individuals whom to offer web-first (see Lynn 2017 for an example of such a refinement without quantifying impacts on dataset quality, and Jackson et al. 2024 for an example of reducing follow-ups of some individuals without affecting dataset quality).

Implications of findings for survey practice: Our findings concerning how CATI follow-ups of web respondents increased the quality of the wave 6 COVID-19 Study datasets as well as the wave 1 dataset imply that these designs can improve longitudinal survey dataset quality. This is especially true given that follow-ups were only undertaken for a subset of web non-respondents: greater improvements might be expected if all non-respondents are included. Therefore, while extensive testing should be undertaken before final implementation (such surveys are often a major social science investment), we recommend their use in other longitudinal surveys.

Our findings concerning refinements to the COVID-19 Study design imply that not web sampling non-regular internet users to reduce costs could not have been undertaken without reducing web plus CATI respondent dataset quality in terms of precision loss and effective dataset size. They also imply that such reductions could not have been made up for by using the savings to instead expand CATI. Hence, the implemented survey design was a good choice, which is a testament to the UKHLS team, who developed and fielded the survey in a short time span during a difficult period for humanity.

Our findings concerning design refinements are though, of wider relevance. Datasets sometimes improved even with reduced web sampling if more CATI respondents were added i.e. when the web: CATI interview cost ratio was smaller than in the COVID-19 Study. Regarding this ratio, the fixed cost element will likely be comparable in other surveys with similar designs. However, potentially indicating a scenario where the costs of CATI will be reduced sufficiently that enough extra respondents to improve datasets can be obtained by not web sampling non-regular internet users, costs per interview will be lower in larger surveys. In addition, the cost ratio itself may smaller when other follow-up modes such as mail (a common design: see Olson et al. 2021) are used. Hence, reducing web sampling and instead expanding use of the follow-up mode may be a useful design refinement in some circumstances.

Limitations: One limitation of our research is that the non-response weighted estimates of respondent characteristics were not compared to actual population values. This is because, as in many studies (see Hand 2018), such population values were not available. Instead, we compared estimates to benchmarks computed using the main survey weighted eligible samples: see Benzeval et al. (2020) for evidence that main survey weighted estimates approximate population values. Two other limitations concern the response probabilities assigned to potential extra CATI respondents in the simulation study. The first of these is that such probabilities for non-regular internet users in general are the same as those for individuals that responded to CATI follow-ups. This assumption though, may be valid given that individuals in the UK faced movement restrictions, including being furloughed from their jobs, for periods of the COVID-19 pandemic. The second is that probabilities were predicted using models with only sex, age and education and their interactions as predictors. In reality, there are more response predictors. However, a more complex model including all

interactions between them (which are needed so that extra respondents can be identified: see section 3.4.1) could not be fitted because there were too few individuals in some cells. Should such methods be used in other surveys, including more predictors may be possible if the survey is larger than the COVID-19 Study.

Future research: Our findings indicate two questions that should be pursued in future research on this topic. The first is to investigate whether findings are comparable in other longitudinal surveys. This research should consider both surveys in which follow-ups are interviewer administered (i.e. CATI or CAPI) and surveys in which other modes are used for follow-ups. The latter should especially consider in which mail is used: given that it is less costly, the web: follow-up mode cost ratio will be lower than with interviewer administered follow-ups, increasing the likelihood that the refinements to the web first design evaluated in this paper will be beneficial (see earlier: see Biemer et al. 2022; Peytchev et al. 2022 for work on such designs). It would also gain value if it could include more than the two waves studied in this paper, as most longitudinal surveys are fielded for longer periods. The second question also concerns the refinements to the web first design. Research is needed on how to quantitatively identify when the follow-up mode should be used instead of web in empirical situations, to replace the informal approach based on degree of internet use utilised here. It must take into account: a) that returns from the follow-up mode may exceed those from web, rather than web failing to improve datasets (non-regular internet user web respondents often improved the COVID-19 Study datasets), and b) that in longitudinal surveys multiple datasets include the same respondents, so that decisions made regarding one may impact on others. a) has parallels with designs in which individuals are prioritized for interview based on their impact on datasets (see Tourangeau et al. 2017 for discussion). Regarding both aspects of the problem, a likely start point is work on when to switch between data collection methods in

simpler scenarios (e.g. Groves & Heeringa 2006, Wagner & Raghunathan 2010; Lewis 2017, 2019).

References

Baker, R., Blumberg, S., Brick, M.J., Couper, M.P., Courtright, M., Dennis, M., Dillman, M., Frankel, M.R., Garland, P., Groves, R.M., Kennedy, C., Krosnick, J., Lee, S., Lavrakas, P.J., Link, M., Piekarski, L., Rao, K., Rivers, D., Thomas, R.K. & Zahs, D. (2010) *AAPOR report on online panels.* American Association for Public Opinion Research.

Benzeval, M., Bollinger, C. R., Burton, J., Crossley, T.F. & Lynn, P. (2020) *The representativeness of Understanding Society.* Understanding Society Working Paper Series 2020–08, Institute for Social and Economic Research.

Bianchi, A., Biffignandi, S. & Lynn, P. (2017) Web-face-to-face mixed-mode design in a longitudinal survey: effects on participation rates, sample composition, and costs. *Journal of Official Statistics*, 33: 385-408.

Biemer, P.P., Mullan Harris, K., Burke, B.J., Liao, D. & Tucker Halpern, C. (2022) Transitioning a Panel Survey from in-person to Predominantly Web Data Collection: Results and Lessons Learned, *JRSSSA*, 185: 798–821. https://doi.org/10.1111/rssa.12750

Brown, M., Goodman, A., Peters, A., Ploubidis, G.B., Sanchez, A., Silverwood, R. & Smith, K. (2020) *COVID-19 Survey in Five National Longitudinal Studies: Waves 1 and 2 User Guide (Version 2).* London: UCL Centre for Longitudinal Studies and MRC Unit for Lifelong Health and Ageing.

Burton, J. & Jäckle, A. (2020) *Mode Effects*, Understanding Society Working Paper 2020- 05. Colchester: University of Essex.

Cornesse, C. & Bosnjak, M. (2018) Is there an association between survey characteristics and representativeness? A meta-analysis. *Surv. Res. Meth.*, 12: 1-13. https://doi.org/10.18148/srm/2018.v12i1.7205

Couper, M. P., Kapteyn, A., Schonlau, M. & Winter, J. (2007) Noncoverage and nonresponse in an Internet survey. *Soc. Sci. Res.* 36: 131–148. https://doi.org/10.1016/j.ssresearch.2005.10.002

De Heer, W., & De Leeuw, E. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. *Survey nonresponse*, *41*, 41-54.

de Leeuw, E. (2018), Mixed-Mode: Past, Present, and Future. *Survey Research Methods*, 12: 75 – 89.

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J. & Messer, B. L. (2009) Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. Soc. Sci. Res 38: 1-18.

Daikeler, J., Bosnjak, M., & Lozar Manfreda, K. (2020). Web versus other survey modes: An updated and extended meta-analysis comparing response rates. *J. Surv. Stat. Meth.,* **8**, 513–539. https://doi.org/10.1093/jssam/smz008

Fricker, S., Galesic, M., Tourangeau, R. & Yan, T. (2005) An Experimental Comparison of Web and Telephone Surveys. *POQ*, **69**, 370–392. https://doi.org/10.1093/poq/nfi027

Groves, R. M. & Heeringa, S. (2006) Responsive design for household surveys: tools for actively controlling survey errors and costs*. J. Roy. Stat. Soc. A.*, **169**, 439-457. https://doi.org/10.1111/j.1467-985X.2006.00423.x

Groves, R. M., Dillman, D. A., Eltinge, J. L., & Little, R. J. (eds.) (2001) *Survey Nonresponse*. Wiley Series in Survey Methodology.

Institute for Social and Economic Research (2021) *Understanding Society COVID-19 User Guide. Version 10.0.* Colchester: University of Essex.

Institute for Social and Economic Research (2024) *Understanding Society – The UK Household Longitudinal Study, Innovation Panel, Waves 1-16, User Manual.* Colchester: University of Essex.

Institute for Social and Economic Research (2024b) *Understanding Society: Waves 1-14, 2009-2023 and Harmonised BHPS: Waves 1-18, 1991-2009, User Guide*, Colchester: University of Essex.

Jäckle, A., Lynn, P., Burton, J. (2015) Going Online with a Face-to-Face Household Panel: Effects of a Mixed Mode Design on Item and Unit Non-Response, *Surv. Res. Meth.* 9: 57-70. https://doi.org/10.18148/srm/2015.v9i1.5475

Jackson, M.T., Hughes, T. & Fu, J (2024) Improving the Efficiency of Outbound CATI As a Nonresponse Follow-Up Mode in Address-Based Samples: A Quasi-Experimental Evaluation of a Dynamic Adaptive Design, *J. Surv. Stat. Meth.* 12: 712–740, https://doi.org/10.1093/jssam/smae005

Kirchner, A., & Felderer, B. (2016) The Effect of Nonresponse and Measurement Error on Wage Regression across Survey Modes: A Validation Study. In: *Total Survey Error in Practice* (eds. Biemer, P. P., De Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L. E., Tucker, N. C. & West, B. T.), pp. 531–556, Hoboken: John Wiley & Sons. https://doi.org/10.1002/9781119041702

Kish, L. (1965) *Survey Sampling.* Wiley: New York.

Klausch, T., Hox, J. & Schouten, B. (2015) *Assessing the mode dependency of sample selectivity across the survey response process.* Statistics Netherlands, The Hague.

Lewis, T. (2017) Univariate tests for phase capacity: Tools for identifying when to modify a survey's data collection protocol. *J. Off. Stat.* **33**, 601–624. https://doi.org/10.1515/jos-2017-0029

Lewis, T. H. (2019) Multivariate tests for phase capacity. *Surv. Res. Meth.* **13**, 153–165. https://doi.org/10.18148/srm/2019.v13i2.7370

Little, R. J. A. & Vartivarian, S. (2005) Does weighting for nonresponse increase the variance of survey means? *Surv. Meth.* **31**, 161-168.

Lipps, O. & Pekari, N. (2021) Sequentially mixing modes in an election survey. *Survey Methods: Insights from the Field.* Retrieved from https://surveyinsights.org/?p=15281

Luiten, A., Hox, J., & de Leeuw, E. (2020) Survey Nonresponse Trends and Fieldwork Effort in the 21st Century: Results of an International Study across Countries and Surveys. *Journal of Official Statistics*, *36*, 469-487. https://doi.org/10.2478/jos-2020-0025

Lundquist, P. and Sarndal, C.-E. (2013) Aspects of responsive design with applications to the Swedish Living Conditions Survey. *J. Off. Stat.* **29**, 557–582. https://doi.org/10.2478/jos-2013-0040

Lynn, P. (2006) Attrition and non-response. *JSSSA*, 169: 393-394.

Lynn, P. (2017) *Pushing Household Panel Survey Participants from CAPI to Web.* Paper presented at the 28th International Workshop on Household Survey Nonresponse, Utrecht, August-September 2017.

Mackeben J. & Sakshaug J.W. (2023) Transitioning an employee panel survey from telephone to online and mixed-mode data collection. *Stat. J. IAOS*. 39: 213-232. https://doi.org/10.3233/SJI-220088

McGonagle, K.A, & Sastry, N. (2023) Transitioning to a Mixed-Mode Study Design in a National Household Panel Study: Effects on Fieldwork Outcomes, Sample Composition and Costs. *Surv. Res, Meth.*, **17**, 411-427. https://doi.org/10.18148/srm/2023.v17i4.8172

Moore, J.C., Burton, J., Crossley, T. F., Fisher, P., Gardiner, C., Jäckle, A., & Benzeval, M. (2024) *Assessing Bias Prevention and Bias Adjustment in a Sub-Annual Online Panel Survey*. Understanding Society Working Papers Series 2024-04.

Moore, J.C., Alvarez, P.C., Durrant, G., Jackle, A., Smith, P.W.F. & Burton, J. (submitted) Are interviewer administered follow ups of web non-respondents still needed to maximise respondent dataset quality? Evidence from Understanding Society: the UK Household Longitudinal Study *Surv. Res. Meth.*

Olson, K., Smyth, J.D., Horwitz, R., Keeter, S., Lesser, V., Marken, S., Mathiowetz, N.A., McCarthy, J.S., O'Brien, E., Opsomer, J.D. & Steiger, D. (2021) Transitions from telephone surveys to self-administered and mixed-mode surveys: AAPOR task force report. *J. Surv. Stat. Meth.* 9: 381-411. https://doi.org/10.1093/jssam/smz062

Peytchev, A., Pratt, D. & Duprey, M. (2022) Responsive and adaptive survey design: Use of bias propensity during data collection to reduce nonresponse bias. *J. Surv. Stat. Meth.* **10**, 131-148. https://doi.org/10.1093/jssam/smaa013

Sarndal, C.-E. and Lundquist, P. (2014a) Balancing the response and adjusting estimates for nonresponse bias: complementary activities. *J. Soc. Statist.*, **155**, 28–50. http://www.numdam.org/item/JSFS_2014__155_4_28_0/

Sarndal, C.-E. and Lundquist, P. (2014b) Accuracy in estimation with nonresponse: a function of the degree of imbalance and degree of explanation. *J. Surv. Stat. Meth.* **2**, 361–387. https://doi.org/10.1093/jssam/smu014

Schonlau, M., Soest, A. van, Kapteyn, A. & Couper, M. (2009) Selection Bias in Web Surveys and the Use of Propensity Scores: *Soc. Meth. Res.* 37: 291-318. https://doi.org/10.1177/0049124108327128

Schouten, B., Cobben, F., Lundquist, P. & Wagner, J. (2016) Does more balanced survey response imply less non-response bias? *J. Roy. Stat. Soc. A,* **179**, 727-748. https://doi.org/10.1111/rssa.12152

Tourangeau, R., Brick, M.J., Lohr, S. & Li, J. (2017) Adaptive and Responsive Survey Designs: A Review and Assessment, *J. Roy. Stat. Soc. A*, **180**, 203–223. https://doi.org/10.1111/rssa.12186

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples (Vol. 1)*. New York: Springer. https://doi.org/10.1007/978-1-4614-6449-5

van Berkel, K., van der Doef, S., & Schouten, B. (2020) Implementing adaptive survey design with an application to the Dutch health survey. *J. Off. Stat.*, **36**, 609-629. https://doi.org/10.2478/jos-2020-0031

van Berkel, K., van Den Brakel, J., Groffen, D., & Burger, J. (2024) Experiences with mixed-mode surveys in times of COVID-19 at Statistics Netherlands. *Stat. J. IAOS*, **40**, 361-373. https://doi.org/10.3233/SJI-230092

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples (Vol. 1)*. New York: Springer. https://doi.org/10.1007/978-1-4614-6449-5

Vassallo, S., & Sanson, A. (Eds.). (2013). The australian temperament project: The first 30 years. Melbourne: The Australian Institute of Family Studies. https://doi.org/10.1037/e567282013-002

Voorpostel, M., Lipps, O., & Roberts, C. (2021). Mixing modes in household panel surveys: Recent developments and new findings. In: *Advances in longitudinal survey methodology* (ed. P. Lynn), pp. 204-226. Wiley.

Wagner, J. R. (2008) *Adaptive Survey Design to Reduce Nonresponse Bias.* PhD diss., University of Michigan, Michigan.

Wagner, J., & Raghunathan, T. E. (2010). A new stopping rule for surveys. *Stat. Med.* **29**, 1014–1024. https://doi.org/10.1002/sim.3834

Watson, N. and Lynn, P. (2021) *Refreshment sampling for longitudinal surveys*, in P. Lynn (ed.), Advances in Longitudinal Survey Methodology, Wiley, pp. 1-25.

Wu M-J., Zhao, K. & Fils-Aime, F. (2022) Response rates of online surveys in published research: A meta-analysis. *Comp. Human Behav. Rep.*, **7**, 100206. https://doi.org/10.1016/j.chbr.2022.100206.

Table 1. Eligible samples, UKHLS main survey wave 9 weight availability, response rates and respondent dataset sizes for (the components of) the UKHLS COVID-19 Study wave 1 and wave 6 cross-sectional and longitudinal (also responding to wave 1) datasets. 'Eligible' is eligible sample members. 'Reg. int. user' are regular internet users. 'N-reg. int. user' are non-regular internet users. 'CATI' are non-regular internet user web non-respondents who were issued to CATI (hence, they are also counted in (iii)). 'Web' is regular and non-regular internet user web respondents combined. 'All' is web and CATI respondents combined, and 'Reg. int. user + CATI' is regular internet user and CATI respondents combined.

	Eligible				Respondents		
	Eligible	Reg. int. user	N-reg. int. user	CATI	Web	All	Reg. int. user + CATI
Wave 1							
N eligible	43981	29726	14255	3408			
N eligible with w9 weight	33951	29405	4546	2916			
N respondents		15514	2247	718	17761	18479	16232
Response rate (%)		52.19	15.76	21.07	40.38	42.02	36.91
N respondents with w9 weight		15332	677	671	16009	16680	16003
Response rate with w9 weight		52.14	14.89	23.01	47.15	49.13	47.14
Wave 6 cross-sectional							
N eligible	43862	29676	14186	730			
N eligible with w9 weight	33847	29355	4492	678			
N respondents		10875	1160	391	12035	12426	11266
Response rate		36.65	8.18	53.56	27.44	28.33	25.69
N respondents with w9 weight		10754	494	374	11248	11622	11128
Response rate with w9 weight		36.63	11.00	55.16	33.23	34.34	32.88
Wave 6 longitudinal							
N eligible	43862	29676	14186	730			
N eligible with w9 weight	33847	29355	4492	678			
N respondents		10350	1042	383	11392	11784	10733
Response rate		34.88	7.35	52.47	25.97	26.87	24.48
N respondents with w9 weight		10233	450	366	10683	11200	10599
Response rate with w9 weight		34.86	10.02	53.98	31.56	33.09	31.31

Table 2. COVID-19 Study wave 1 dataset member characteristics. 'Eligible' (columns (i) & (vii)) is all eligible sample members. 'Regular internet users' are web respondents who reported using the internet 1-2 times a week or more at main survey wave 9. 'Web' are all web respondents. 'Web plus CATI' are web plus CATI respondents. 'Regular internet user plus CATI' are regular internet users plus CATI respondents. Characteristics are quantified using UKHLS main survey wave 9 information, and reported as means of binary variables indicating individuals have the characteristic or not and their 95% CIs.

	Eligible			Regular internet user			Web			Web plus CATI			Regular internet user plus CATI		
	Estimate	95% CIs		Estimate	95% CIs		Estimate	95% CIs		Estimate	95% CIs		Estimate	95% Cis	
		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper
Sex: Male	0.442	0.436	0.447	0.418	0.411	0.426	0.418	0.411	0.426	0.416	0.409	0.424	0.417	0.409	0.424
Age: 20-29	0.147	0.143	0.151	0.111	0.106	0.116	0.107	0.103	0.112	0.104	0.100	0.109	0.107	0.103	0.112
Age: 30-39	0.132	0.128	0.135	0.136	0.131	0.141	0.131	0.125	0.136	0.127	0.122	0.132	0.132	0.127	0.137
Age: 40-49	0.166	0.162	0.170	0.183	0.177	0.189	0.177	0.171	0.183	0.172	0.166	0.178	0.177	0.171	0.183
Age: 50-59	0.187	0.183	0.191	0.216	0.210	0.223	0.214	0.207	0.220	0.209	0.203	0.215	0.212	0.206	0.218
Age: 60-69	0.163	0.159	0.167	0.195	0.189	0.201	0.199	0.193	0.205	0.198	0.192	0.204	0.195	0.189	0.201
Age: 70-79	0.135	0.132	0.139	0.132	0.126	0.137	0.141	0.135	0.146	0.147	0.141	0.152	0.138	0.133	0.144
Age: 80-89	0.059	0.056	0.061	0.025	0.023	0.028	0.029	0.027	0.032	0.038	0.035	0.041	0.034	0.031	0.037
Age: 90+	0.011	0.009	0.012	0.001	0.001	0.002	0.002	0.002	0.003	0.005	0.004	0.006	0.004	0.003	0.005
Qualifications: Degree	0.397	0.392	0.403	0.507	0.499	0.515	0.496	0.488	0.504	0.484	0.476	0.491	0.495	0.487	0.503
Qualifications: A-level	0.216	0.211	0.220	0.213	0.206	0.219	0.211	0.205	0.218	0.209	0.202	0.215	0.210	0.204	0.216
Qualifications: GCSE or lower	0.387	0.381	0.392	0.280	0.273	0.287	0.293	0.286	0.300	0.308	0.301	0.315	0.295	0.288	0.302
HH structure: Couple, kid(s)	0.245	0.241	0.250	0.267	0.260	0.274	0.260	0.253	0.266	0.251	0.245	0.258	0.259	0.252	0.265
HH structure: Couple, no kid(s)	0.378	0.373	0.383	0.429	0.421	0.437	0.440	0.432	0.448	0.432	0.424	0.439	0.423	0.415	0.431
HH structure: Single, kid(s)	0.037	0.035	0.039	0.033	0.030	0.036	0.032	0.029	0.035	0.031	0.029	0.034	0.032	0.029	0.035
HH structure: Single, no kid(s)	0.339	0.334	0.344	0.271	0.264	0.278	0.268	0.262	0.275	0.285	0.279	0.292	0.286	0.279	0.293
Ethnic minority: Yes	0.184	0.180	0.189	0.126	0.120	0.131	0.123	0.118	0.128	0.124	0.119	0.129	0.125	0.120	0.130
Country: England	0.786	0.782	0.791	0.814	0.808	0.820	0.812	0.806	0.818	0.809	0.803	0.815	0.811	0.805	0.817
Country: Wales	0.065	0.062	0.067	0.058	0.054	0.062	0.058	0.055	0.062	0.059	0.056	0.063	0.059	0.055	0.062
Country: Scotland	0.084	0.081	0.087	0.086	0.082	0.091	0.087	0.083	0.092	0.087	0.083	0.092	0.087	0.082	0.091
Country Northern Ireland	0.065	0.063	0.068	0.042	0.039	0.045	0.043	0.039	0.046	0.044	0.041	0.047	0.043	0.040	0.046
Tenure: Owned	0.357	0.351	0.362	0.370	0.363	0.378	0.384	0.376	0.391	0.388	0.381	0.396	0.377	0.369	0.384
Tenure: Mortgage	0.376	0.371	0.381	0.434	0.426	0.441	0.423	0.415	0.431	0.410	0.403	0.418	0.421	0.413	0.429
Tenure: Rented	0.111	0.107	0.114	0.101	0.096	0.106	0.098	0.093	0.102	0.098	0.093	0.102	0.100	0.095	0.105
Tenure: Social Housing	0.154	0.150	0.158	0.093	0.088	0.097	0.093	0.089	0.098	0.101	0.097	0.106	0.100	0.095	0.105
Household income (£/month)	3505.915	3475.001	3536.829	3783.460	3735.632	3831.289	3753.237	3706.548	3799.926	3680.057	3634.633	3725.481	3713.668	3666.90	3760.43
Long-standing illness: Yes	0.356	0.350	0.361	0.327	0.320	0.335	0.335	0.328	0.343	0.344	0.337	0.351	0.337	0.329	0.344

Table 3. COVID-19 Study wave 6 cross-sectional dataset member characteristics. 'Eligible' (columns (i) & (vii)) is all eligible sample members. 'Regular internet users' are web respondents who reported using the internet 1-2 times a week or more at main survey wave 9. 'Web' are all web respondents. 'Web plus CATI' are web plus CATI respondents. 'Regular internet user plus CATI' are regular internet users plus CATI respondents. Characteristics are quantified using UKHLS main survey wave 9 information, and reported as means of binary variables indicating individuals have the characteristic or not and their 95% CIs.

	Eligible			Regular internet user			Web			Web plus CATI			Regular internet user plus CATI		
	Estimate	95% CIs		Estimate	95% CIs		Estimate	95% CIs		Estimate	95% CIs		Estimate	95% CIs	
		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper
Sex: Male	0.441	0.436	0.447	0.417	0.408	0.427	0.416	0.407	0.425	0.415	0.406	0.424	0.416	0.407	0.425
Age: 20-29	0.147	0.144	0.151	0.073	0.068	0.078	0.070	0.066	0.075	0.069	0.064	0.073	0.071	0.066	0.076
Age: 30-39	0.132	0.129	0.136	0.109	0.103	0.115	0.105	0.100	0.111	0.102	0.097	0.108	0.106	0.101	0.112
Age: 40-49	0.167	0.163	0.171	0.166	0.159	0.173	0.160	0.153	0.167	0.156	0.149	0.162	0.161	0.154	0.168
Age: 50-59	0.188	0.183	0.192	0.225	0.217	0.233	0.221	0.214	0.229	0.217	0.209	0.224	0.221	0.213	0.228
Age: 60-69	0.163	0.159	0.167	0.233	0.225	0.241	0.237	0.229	0.245	0.235	0.227	0.243	0.232	0.224	0.240
Age: 70-79	0.135	0.131	0.139	0.163	0.156	0.170	0.171	0.164	0.178	0.177	0.170	0.184	0.168	0.161	0.175
Age: 80-89	0.058	0.055	0.060	0.029	0.026	0.032	0.033	0.030	0.036	0.041	0.037	0.044	0.037	0.033	0.040
Age: 90+	0.010	0.009	0.011	0.002	0.001	0.002	0.002	0.001	0.003	0.004	0.003	0.005	0.003	0.002	0.004
Qualifications: Degree	0.398	0.393	0.403	0.514	0.505	0.524	0.503	0.494	0.512	0.492	0.483	0.501	0.504	0.494	0.513
Qualifications: A-level	0.216	0.212	0.221	0.206	0.198	0.214	0.205	0.197	0.212	0.202	0.195	0.210	0.204	0.196	0.211
Qualifications: GCSE or lower	0.386	0.381	0.391	0.280	0.271	0.288	0.292	0.284	0.301	0.306	0.297	0.314	0.293	0.284	0.301
HH structure: Couple, kid(s)	0.246	0.242	0.251	0.231	0.224	0.239	0.225	0.217	0.232	0.218	0.210	0.225	0.225	0.217	0.233
HH structure: Couple, no kid(s)	0.378	0.373	0.383	0.485	0.476	0.495	0.495	0.485	0.504	0.487	0.478	0.496	0.479	0.470	0.488
HH structure: Single, kid(s)	0.038	0.036	0.040	0.029	0.026	0.032	0.029	0.025	0.032	0.028	0.025	0.031	0.028	0.025	0.032
HH structure: Single, no kid(s)	0.339	0.334	0.344	0.254	0.246	0.262	0.252	0.244	0.260	0.267	0.259	0.275	0.268	0.260	0.276
Ethnic minority: Yes	0.185	0.181	0.189	0.100	0.095	0.106	0.099	0.094	0.105	0.100	0.095	0.105	0.100	0.095	0.106
Country: England	0.787	0.782	0.791	0.817	0.810	0.824	0.815	0.808	0.823	0.813	0.806	0.820	0.815	0.808	0.822
Country: Wales	0.065	0.062	0.067	0.057	0.052	0.061	0.057	0.053	0.061	0.058	0.054	0.062	0.057	0.053	0.062
Country: Scotland	0.084	0.081	0.087	0.086	0.081	0.091	0.087	0.082	0.092	0.087	0.082	0.092	0.086	0.081	0.092
Country Northern Ireland	0.065	0.063	0.068	0.040	0.036	0.044	0.041	0.037	0.044	0.042	0.038	0.046	0.041	0.038	0.045
Tenure: Owned	0.356	0.351	0.361	0.430	0.420	0.439	0.441	0.431	0.450	0.444	0.435	0.454	0.433	0.424	0.443
Tenure: Mortgage	0.377	0.371	0.382	0.400	0.390	0.409	0.390	0.381	0.399	0.379	0.370	0.388	0.389	0.380	0.398
Tenure: Rented	0.111	0.108	0.114	0.088	0.082	0.093	0.085	0.080	0.090	0.085	0.080	0.090	0.087	0.082	0.093
Tenure: Social Housing	0.154	0.150	0.158	0.080	0.075	0.085	0.081	0.076	0.086	0.089	0.083	0.094	0.087	0.082	0.093
Household income (£/month)	3509.185	3478.193	3540.176	3738.002	3680.592	3795.413	3701.488	3645.681	3757.296	3638.595	3584.069	3693.121	3677.344	3621.09	3733.59
Long-standing illness: Yes	0.354	0.349	0.360	0.348	0.339	0.357	0.355	0.346	0.364	0.362	0.353	0.371	0.355	0.346	0.364

Table 4. COVID-19 Study wave 6 longitudinal dataset member characteristics. 'Eligible' (columns (i) & (vii)) is all eligible sample members. 'Regular internet users' are web respondents who reported using the internet 1-2 times a week or more at main survey wave 9. 'Web' are all web respondents. 'Web plus CATI' are web plus CATI respondents. 'Regular internet user plus CATI' are regular internet users plus CATI respondents. Characteristics are quantified using UKHLS main survey wave 9 information, and reported as means of binary variables indicating individuals have the characteristic or not and their 95% CIs.

	Eligible			Regular internet user			Web			Web plus CATI			Regular internet user plus CATI		
	Estimate	95% CIs		Estimate	95% CIs		Estimate	95% CIs		Estimate	95% CIs		Estimate	95% CIs	
		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper
Sex: Male	0.441	0.436	0.447	0.416	0.406	0.425	0.414	0.405	0.424	0.413	0.404	0.422	0.415	0.405	0.424
Age: 20-29	0.147	0.144	0.151	0.072	0.067	0.077	0.069	0.064	0.074	0.067	0.063	0.072	0.070	0.065	0.075
Age: 30-39	0.132	0.129	0.136	0.110	0.104	0.117	0.107	0.101	0.112	0.103	0.098	0.109	0.107	0.101	0.113
Age: 40-49	0.167	0.163	0.171	0.164	0.157	0.171	0.159	0.152	0.166	0.154	0.147	0.161	0.160	0.153	0.167
Age: 50-59	0.188	0.183	0.192	0.224	0.216	0.232	0.221	0.213	0.229	0.216	0.209	0.224	0.220	0.212	0.228
Age: 60-69	0.163	0.159	0.167	0.236	0.227	0.244	0.238	0.230	0.246	0.237	0.229	0.245	0.234	0.226	0.242
Age: 70-79	0.135	0.131	0.139	0.164	0.157	0.171	0.173	0.165	0.180	0.178	0.171	0.185	0.170	0.162	0.177
Age: 80-89	0.058	0.055	0.060	0.028	0.025	0.031	0.032	0.028	0.035	0.040	0.036	0.044	0.036	0.033	0.040
Age: 90+	0.010	0.009	0.011	0.001	0.001	0.002	0.002	0.001	0.003	0.004	0.003	0.005	0.003	0.002	0.004
Qualifications: Degree	0.398	0.393	0.403	0.517	0.507	0.526	0.506	0.496	0.515	0.494	0.485	0.503	0.505	0.496	0.515
Qualifications: A-level	0.216	0.212	0.221	0.206	0.198	0.214	0.205	0.197	0.213	0.202	0.195	0.210	0.203	0.196	0.211
Qualifications: GCSE or lower	0.386	0.381	0.391	0.278	0.269	0.286	0.289	0.281	0.298	0.304	0.295	0.312	0.291	0.283	0.300
HH structure: Couple, kid(s)	0.246	0.242	0.251	0.231	0.223	0.239	0.224	0.216	0.232	0.217	0.210	0.225	0.224	0.216	0.232
HH structure: Couple, no kid(s)	0.378	0.373	0.383	0.490	0.480	0.500	0.499	0.490	0.509	0.491	0.482	0.500	0.483	0.473	0.492
HH structure: Single, kid(s)	0.038	0.036	0.040	0.028	0.025	0.031	0.027	0.024	0.031	0.027	0.024	0.030	0.027	0.024	0.030
HH structure: Single, no kid(s)	0.339	0.334	0.344	0.251	0.243	0.260	0.249	0.241	0.257	0.265	0.257	0.273	0.266	0.257	0.274
Ethnic minority: Yes	0.185	0.181	0.189	0.095	0.089	0.101	0.094	0.088	0.099	0.095	0.089	0.100	0.096	0.090	0.101
Country: England	0.787	0.782	0.791	0.819	0.811	0.826	0.817	0.810	0.824	0.814	0.807	0.822	0.817	0.809	0.824
Country: Wales	0.065	0.062	0.067	0.056	0.051	0.060	0.056	0.052	0.061	0.057	0.053	0.062	0.057	0.052	0.061
Country: Scotland	0.084	0.081	0.087	0.087	0.082	0.093	0.088	0.083	0.093	0.088	0.083	0.093	0.087	0.082	0.093
Country Northern Ireland	0.065	0.063	0.068	0.038	0.035	0.042	0.039	0.035	0.042	0.040	0.037	0.044	0.039	0.036	0.043
Tenure: Owned	0.356	0.351	0.361	0.432	0.422	0.441	0.443	0.433	0.452	0.446	0.437	0.456	0.435	0.426	0.445
Tenure: Mortgage	0.377	0.371	0.382	0.400	0.391	0.410	0.391	0.382	0.400	0.380	0.371	0.389	0.389	0.380	0.398
Tenure: Rented	0.111	0.108	0.114	0.087	0.081	0.092	0.084	0.079	0.090	0.084	0.079	0.090	0.087	0.081	0.092
Tenure: Social Housing	0.154	0.150	0.158	0.079	0.074	0.084	0.079	0.074	0.084	0.087	0.082	0.092	0.086	0.081	0.092
Household income (£/month)	3509.185	3478.193	3540.176	3737.561	3679.207	3795.915	3704.237	3647.390	3761.083	3638.667	3583.178	3694.157	3674.930	3617.78	3732.07
Long-standing illness: Yes	0.354	0.349	0.360	0.346	0.337	0.355	0.352	0.343	0.361	0.359	0.351	0.368	0.353	0.344	0.363

Table 5. UKHLS COVID-19 Study non-response weighted estimate mean absolute standardised biases in respondent characteristics compared to equivalent eligible sample, UKHLS main survey, weighted estimate benchmarks (MASBs) for wave 1 and wave 6 cross-sectional and longitudinal regular internet user, all web, web plus CATI and regular internet user plus CATI respondent datasets. See text for full explanation.

	Regular internet user			Web			Web plus CATI			Regular internet user plus CATI		
	MASB	95% CIs		MASB	95% CIs		MASB	95% CIs		MASB	95% CIs	
		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper
Wave 1	0.014	0.009	0.019	0.012	0.007	0.017	0.009	0.006	0.011	0.012	0.007	0.017
Wave 6 cross-sectional	0.023	0.012	0.034	0.016	0.009	0.023	0.010	0.005	0.014	0.013	0.009	0.017
Wave 6 longitudinal	0.022	0.012	0.032	0.014	0.007	0.020	0.014	0.009	0.019	0.015	0.010	0.021

Table 6. COVID-19 Study non-response weighted dataset precision loss as quantified by Kish's DEFF, and effective dataset sizes (N_{DEFF}) for the wave 1 and wave 6 cross-sectional and longitudinal datasets.

		Regular internet user	Web	Web plus CATI	Regular internet user plus CATI
Wave 1	DEFF	2.634	2.754	2.299	3.095
	N_{DEFF}	5821.017	5812.028	7256.608	5171.301
Wave 6 cross-sectional	DEFF	2.627	2.774	2.701	3.028
	N_{DEFF}	4094.261	4054.938	4303.478	3675.43
Wave 6 longitudinal	DEFF	3.172	3.503	3.089	3.747
	N_{DEFF}	3226.497	3049.477	3579.862	2829.826

Table 7: Non-response weighted estimate means of mean absolute biases standardized by benchmark estimate standardised deviations and their 95% prediction intervals (PIs) across simulated COVID-19 Study datasets with different numbers of extra wave 1 CATI respondents for the wave 1, wave 6 cross-sectional and wave 6 longitudinal datasets.

	Number of extra wave 1 respondents														
	104			250			500			750			1000		
	95% PI			95% PI			95% PI			95% PI			95% PI		
	Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper
Wave 1	0.012	0.012	0.012	0.012	0.012	0.012	0.011	0.011	0.011	0.009	0.009	0.009	0.008	0.008	0.008
Wave 6 cross-sectional	0.014	0.014	0.014	0.014	0.014	0.014	0.013	0.013	0.013	0.012	0.012	0.012	0.011	0.011	0.012
Wave 6 longitudinal	0.014	0.013	0.014	0.014	0.014	0.014	0.014	0.013	0.014	0.013	0.013	0.013	0.012	0.012	0.012

Table 8. Non-response weighted dataset precision loss as quantified by mean DEFFs and mean effective dataset sizes (N_{DEFF}) and their 95% prediction intervals (PIs) across simulated COVID-19 Study datasets with different numbers of extra wave 1 CATI respondents for the wave 1, wave 6 cross-sectional and wave 6 longitudinal datasets.

		Number of extra wave 1 respondents														
		104			250			500			750			1000		
		95% PI			95% PI			95% PI			95% PI			95% PI		
		Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper
Wave 1	DEFF	3.019	3.017	3.021	2.760	2.758	2.763	2.385	2.383	2.387	2.174	2.173	2.176	2.072	2.071	2.073
	N_{DEFF}	5335.598	5332.216	5338.979	5836.634	5831.582	5841.686	6755.227	6748.888	6761.566	7409.822	7404.830	7414.814	7773.729	7770.010	7777.447
Wave 6 cross-sectional	DEFF	3.338	3.336	3.340	3.183	3.181	3.185	2.931	2.929	2.934	2.719	2.716	2.721	2.620	2.617	2.622
	N_{DEFF}	3353.157	3351.278	3355.036	3516.650	3514.087	3519.213	3818.820	3815.185	3822.455	4117.210	4113.409	4121.011	4273.566	4269.235	4277.897
Wave 6 longitudinal	DEFF	3.297	3.295	3.300	3.160	3.156	3.163	2.905	2.901	2.910	2.701	2.698	2.705	2.604	2.601	2.607
	N_{DEFF}	3235.593	3233.411	3237.775	3377.343	3373.914	3380.772	3673.674	3668.302	3679.045	3950.870	3945.901	3955.839	4098.813	4094.040	4103.585

Supplementary information

1. Datasets used

We use the 18th edition of the UKHLS main survey datasets (DOI: http://doi.org/10.5255/UKDA-SN-6614-19), and the 8th edition of the UKHLS COVID-19 Study datasets (DOI: http://doi.org/10.5255/UKDA-SN-8644-8). Both these datasets are available from the UK Data Service (UKDS: https://ukdataservice.ac.uk). Note however, that they are not the currently released versions of the datasets. They are though, available upon application to the UKDS.

2. COVID-19 Study non-response weight estimation

As noted in the main text, Study IP-NR weights were the product of the wave 10 cross-sectional non-response weight and a regression-based adjustment for Study non-response. For cross-sectional weights, this adjustment was the inverse of the estimated response probability to the Study wave. For longitudinal weights, the main survey weight was adjusted by the product of a chain of weights estimated given the probability of wave 1 response conditional on having a main survey weight, of wave 2 response conditional on wave 1 response, and so on. Main survey wave 10 predictors were used to estimate probabilities.

To identify the final regression models used to response probabilities, Lasso procedures are used. Lasso procedures (Tibshirani 1996; Steyerberg et al. 2001) are regularised regression methods. As with other regularised regression methods for binary data (i.e. 0 = non-response, 1 = response), they maximise the joint probability of the model parameters given the observed data similar to maximum likelihood methods, but in addition

impose a regularisation penalty on model complexity (Ahrens et al. 2020). Due to the imposition of this penalty, such methods tend to outperform maximum likelihood methods in terms of out of sample prediction, as reducing model complexity and inducing shrinkage bias decreases prediction error. In doing so, they also address the problem of model overfitting: high in-sample fit, but poor prediction performance on unseen data.

Regularised regression methods incorporate tuning parameters that determine the amount and form of regularisation penalty. Several techniques exist to choose the value of these parameters. The first is cross-validation, which explicitly evaluates out of sample prediction performance. The data are split into training and validation datasets. The models for different values of the tuning parameters are then estimated and variables selected using the training dataset. Next, they are applied to the validation dataset, and performance quantified (Ahrens et al. 2020). The second technique is the use of information criteria. These are interpretable as likelihood methods that penalise the number of parameters in models. Again, models for different tuning parameters are estimated and variables selected, then the best performing is chosen based on information criteria value. When producing the sample inclusion weights, we use information criteria techniques to choose tuning parameter values and identify models for estimating inclusion probabilities. Specifically, we utilise the Extended Bayesian Information Criterion (EBIC: Chen & Chen 2008), because simulations show that in the majority of scenarios they perform better than other similar options in terms of model identification (see Ahrens et al. 2020). We do not use cross validation methods because the size of analysis datasets prevents their division into training and validation datasets (see Moore et al. 2024 for further justification of these methods in the current context). We use the Stata 18 package 'lassologit' (Ahrens et al. 2020) to perform analyses.

The above techniques require that predictors are standardised so that they have unit variance. Hence, when modelling response probabilities for weight estimation we first convert all multi-category predictors into dummy variables. Once the selected model is identified, we then extend it to all selected covariate categories whether they were selected or not: in previous work, we have found that this approach reduces biases (relative to benchmarks) in weighted estimates (unpublished results). After final model identification, we use post-Lasso estimation to estimate response probabilities for weight estimation, because Lasso estimated coefficients are subject to attenuation bias (Ahrens et al. 2020). Specifically, we use probit models, with response probabilities predicted using model coefficients and sample member characteristics.

Once response probabilities are predicted, non-response weights for individuals are estimated as the product of their (main survey or previous COVID-19 Study wave) 'input' weight and the inverse of their predicted response probability. Then, in a final step, weights are trimmed to reduce precision loss: those more than 25 times the normalized median are replaced with the threshold value, which limits precision loss to acceptable levels while still reducing biases (Moore et al. 2024).

3. Moore et al.'s (2024) test of the equality of weighted estimate means

To evaluate biases, Moore et al. (2024) proposed a test that compares UKHLS main survey wave 9 measured, COVID-19 Study IP-NR weighted mean estimates of respondent characteristics to main survey wave 9 weighted benchmarks. To formalize this test, consider a "quasi-randomization" setup (Valliant and Dever, 2018). Let $I_i = 1$ indicate that individual i is selected into the eligible set for UKHLS, and $I_i = 0$ if not. Let $R_i^t = 1$ indicate that individual

i responds to wave t (here, wave 9 of the main survey), and $R_i^t = 0$ if not, conditional on being in the eligible set. Denote the probability that individual i is in the eligible set by $\Pr(I_i = 1) = \pi^i$ and probability that individual i responds at wave t, given they are in the eligible set, by $\Pr(R_i^t = 1|I_i = 1) = \phi_i^t$. Let U be the set of individuals in the population and r^t be the set of respondents at wave t (that is, the set of individuals for whom $R_i^t I_i = 1$).

A1. Assume that $\pi_i > 0\ \forall i,\ \ \phi_i^t > 0\ \forall i$, and weights for wave t, w_i^t, are available such that $w_i^t = (\pi_i \phi_i^t)^{-1}$.

For a quantity, y_i^t, observed in wave t, an estimator of the population total is:

$$\hat{T}(y^t) = \sum_{i \in r^t} w_i^t y_i^t = \sum_{i \in U} R_i^t I_i w_i^t y_i^t \tag{4}$$

Again, in the application in this paper wave *t* is wave 9 of the main study, so this is just the weighted total using wave 9 respondents and the associated wave 9 weights. It is a standard result that $\hat{T}(y^t)$ is unbiased under **A1** (see, e.g., Valliant and Dever, 2018, Chapter 3). To see this take expectations over both the sampling and response processes:

$$E_I E_{R^t}[\sum_{i \in U} R_i^t I_i w_i^t y_i^t] = \sum_{i \in U} w_i^t y_i^t E_I E_{R^t}[R_i^t I_i] = \sum_{i \in U} y_i^t \tag{5}$$

The last equality uses the fact that $E_I E_{R^t}[R_i^t I_i] = E_I\left[I_i\left[E_{R^t}[R_i^t|I_i]\right]\right] = \pi_i \phi_i^t$, and **A1**.

Now consider wave 1 of the COVID-19 Study, which is treated simply as a subsequent wave, $t + k$ of the panel. Let $R_i^{t+k} = 1$ indicate that individual i responds to panel wave $t + k$, and $R_i^{t+k} = 0$ if not, conditional on being in the eligible set *and* responding to wave t. This is termed *retention.* Let r^{t+k} be the set of respondents retained at wave $t + k$ (in the current application, from wave 9 of the main survey, retained into Wave 1 of the COVID-19 Study).

The probability that individual i responds at wave $t+k$, given they are in the eligible set and responded at time t (that is they are retained), is $\Pr\left(R_i^{t+k}=1\middle|I_i=1,R_i^t=1\right)=\theta_i^{t+k}$. Thus, the probability that they respond to wave $t+k$ is $\pi_i\phi_i^t\theta_i^{t+k}$.

A2. Assume that $\pi_i>0\ \forall i,\ \phi_i^t>0\ \forall i, \theta_i^{t+k}>0\ \forall i$, and weights for wave $t+k$, w_i^{t+k}, are available, such that $w_i^{t+k}=\left(\pi_i\phi_i^t\theta_i^{t+k}\right)^{-1}$

Consider an alternative estimator of population total of y^t, the quantity of interest at wave t:

$$\tilde{T}(y^t)=\sum_{i\in r^{t+k}} w_i^{t+k}y_i^t=\sum_{i\in U} R_i^{t+k}R_i^tI_iw_i^{t+k}y_i^t \quad (5)$$

By similar arguments to those above, $\tilde{T}(y^t)$ is unbiased under **A2**. To see this take expectations of the sampling, response *and* retention processes.

$$E_IE_{R^t}E_{R^{T+k}}\left[\sum_{i\in U} R_i^{t+k}R_i^tI_iw_i^ty_i^t\right]=\sum_{i\in U} w_i^ty_i^tE_IE_{R^t}E_{R^{T+k}}\left[R_i^{t+k}R_i^tI_i\right]=\sum_{i\in U} y_i^t \quad (6)$$

The last equality uses the fact that $E_IE_{R^t}E_{R^{T+k}}\left[R_i^{t+k}R_i^tI_i\right]=E_I\left[I_i\left[E_{R^t}\left[R_i^tE[R_i^{t+k}|R_i^t,I_i]|I_i\right]\right]\right]=\pi_i\phi_i^t\theta_i^{t+k}$, and **A2**. This result simply says that under **A2** the population total of y^t can alternatively be estimated using the subset of wave t respondents who are retained at wave $t+k$, and the appropriate wave $t+k$ weights.

Note that $\hat{T}(y^t)$ is unaffected by the retention process, so that $E_IE_{R^t}E_{R^{T+k}}\left[\hat{T}(y^t)\right]=E_IE_{R^t}\left[\hat{T}(y^t)\right]=\sum_{i\in U} y_i^t$, and together these results imply:

$$E_IE_{R^t}E_{R^{T+k}}\left[\hat{T}(y^t)-\tilde{T}(y^t)\right]=0 \quad (7)$$

This is the joint implication of **A1** and **A2** that the test evaluates.

Note that:

$$\tilde{T}(y^t) = \sum_{i \in r^{t+k}} w_i^{t+k} y_i^t = \sum_{i \in s^t} R_i^{t+k} w_i^{t+k} y_i^t. \quad (8)$$

This allows one to proceed as follows:

$$\hat{T}(y^t) - \tilde{T}(y^t) = \left(\sum_{i \in s^t} w_i^t y_i^t - \sum_{i \in s^t} R_i^{t+k} w_i^{t+k} y_i^t \right) = \sum_{i \in s^t} y_i^t \left(w_i^t - R_i^{t+k} w_i^{t+k} \right)$$

$$= \sum_{i \in s^t} y_i^t \omega_i \quad (9)$$

Where the composite weight ω_i is observed for all $i \in s^t$ because $R_i^{t+k} w_i^{t+k} = 0$ when $R_i^{t+k} = 0$. This means that there is no need to observe w_i^{t+k} for attritors (those not retained from wave t to $t+k$), although in practice it often is.

This formulation of $\hat{T}(y^t) - \tilde{T}(y^t)$ takes advantage of the fact that each retained individual (wave *t+k* respondent) is also a wave t respondent and so their weights can be "paired." Working with $\hat{T}(y^t) - \tilde{T}(y^t) = \sum_{i \in s^t} y_i^t \omega_i$ means that inferences only need to be made about a weighted total, which is done using standard methods for inference with complex survey samples. The null that $\hat{T}(y^t) - \tilde{T}(y^t) = 0$ is tested. A rejection of the null would suggest either A1 or A2 (or both) do not hold. As the main survey weights have been extensively evaluated in previous work, a rejection of this null would lead one to doubt A2, that is, the adequacy of the COVID-19 Study weights.

In practice, a version of this test based on weighted means is implemented. This test compares

$$\hat{M}(y^t) = \sum_{i \in s^t} w_i^t y_i^t \Big/ \sum_{i \in s^t} w_i^t \quad , \quad (10)$$

And:

$$\tilde{M}(y^t) = \sum_{i \in r^t} w_i^{t+k} y_i^t \Big/ \sum_{i \in r^t} w_i^{t+k} \quad . \quad (11)$$

These ratio estimators are not generally unbiased, but the bias is typically small (Cochran 1977) and given **A1** and **A2,** they are consistent estimators of the population mean. Thus, $\widehat{M}(y^t) - \widetilde{M}(y^t)$ will converge to zero with probability one as the sample size goes to infinity (formally this requires either that the population size goes to infinity, as in, for example, DuMouchel and Duncan (1983); or that the sample size goes to infinity via sampling with replacement, as in, for example, Deaton (1997). The advantage of working in means is that the magnitude of departures from the null are often easier to interpret. For example, it may be easier to assess the importance of differences in weighted estimates of mean age or mean income across the two alternative estimates, than it is to assess differences in totals. A version of this test can also be derived in a model-based framework, see Crossley et al. (2021), who draw on results for model-based inverse probability-weighted estimators in Wooldridge (2002, 2007).

4. Simulated dataset generation

We simulate extra CATI respondents from non-regular internet users not issued to CATI. At wave 1, firstly we label all such individuals with main survey wave 9 information and weights, the prospective simulated respondents (PSRs), as non-respondents. Second, we predict synthetic CATI response propensities for PSRs given modelling of those of non-regular internet users that were issued to CATI. We use a probit model (as with the Study non-response weights: see online Appendix), with sex (2 categories), age (7 categories) and education (3 categories) and their interactions as predictors. We could not include more predictors because interactions between all of them, needed for the following procedure step, could not be fitted due to few or no (responding) individuals in some cells.

Third, we use the PSR synthetic response propensities to simulate extra respondents. In a multinomial sampling without replacement procedure, we sum propensities for PSRs in each sex * age * education category and use these sums to calculate category cumulative cut off points between 0 and 1. To understand this element, consider an example with a single variable with two categories. Category 1 includes 50 individuals, with a response propensity of 0.5, and category 2 100 individuals with the same response propensity. Category 1 cut off points are hence 0 and (50 * 0.5) / ((50 * 0.5) + (100 * 0.5)) = 0.33. Category 2 cut off points are 0.33 and 1 (= 0.33 + ((100 * 0.5) / ((50 * 0.5) + (100 * 0.5))). Next, we identify the category the extra respondent belongs to by generating a random number between 0 and 1, identify the cut off points between which it falls, and select one member at random as the respondent. We undertake this *n* times. We simulate datasets with 104, 250, 500, 750 and 1000 extra respondents. 104 is the rounded number interviewable for the cost of web sampling non-regular internet users at wave 1 given that 21.7 successful web responses were obtained for the cost of one successful CATI interview in the COVID-19 Study (Moore et al. 2024) i.e. the

number responding, 2247 (see main paper, Table 1), divided by 21.7. The others reflect scenarios in which this cost ratio is reduced. For each *n*, we generate 1000 datasets: this limited number is due to the use of a time consuming Lasso procedure to identify response propensity models when estimating weights (see online Appendix).

In the COVID-19 Study, only wave 1 CATI respondents plus 12 others who requested it were similarly issued at wave 6 (see main paper, section 2.1). Not all responded (see main paper, Table 1). Hence, we select extra wave 6 respondents from extra wave 1 respondents only (the PSRs in this step). First, we use a probit model of wave 6 response propensities for real wave 1 CATI respondents with sex, age and education and all interactions as predictors to predict wave 6 PSR synthetic propensities. Second, we generate a random number between 0 and 1 for each PSR, and if it falls below or equals its synthetic propensity, designate them a wave 6 respondent. We undertake this for each dataset at each value of *n*, and add the same respondents to both the cross-sectional and longitudinal datasets (see this document, Table 3 for dataset sizes).

The algorithm makes several assumptions. The first is that the CATI response probabilities of non-regular internet users in general are the same as for those that did so having not responded by web. Second is that the survey estimate of interest *y* is independent of response outcome *R* given mode *M* and the auxiliary variables *Z*. This is a weaker variant of the ignorable non-response assumption: the stronger version, without conditioning, is assumed when estimating non-response weights. The third is that among non-regular internet users *y* is independent of *M* given *Z* i.e. that there are no mode effects.

5. References

Ahrens, A., Hansen, C. B., & Schaffer, M. E. (2020) lassopack: Model selection and prediction with regularized regression in Stata. *The Stata Journal*, **20**, 176-235. https://doi.org/10.1177/1536867X20909697

Chen, J. & Chen, Z. (2008) Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, **95**, 759–771. https://doi.org/10.1093/biomet/asn034

Cochran, W. G. (1977). Sampling techniques. *John Wiley & Sons Inc*.

Crossley, T. F., Fisher, P. & Low, H. (2021) The Heterogeneous and Regressive Consequences of COVID-19: Evidence from High Quality Panel Data. *J. Pub. Econ.* **193**, 104344. https://doi.org/10.1016/j.jpubeco.2020.104334

Deaton, A. (1997). *The analysis of household surveys: a microeconometric approach to development policy*. World Bank Publications.

DuMouchel, W. H., & Duncan, G. J. (1983). Using sample survey weights in multiple regression analyses of stratified samples. *J. Amer. Statist. Assoc.*, **78**, 535-543. https://doi.org/10.1080/01621459.1983.10478006

Moore, J.C., Burton, J., Crossley, T. F., Fisher, P., Gardiner, C., Jäckle, A., & Benzeval, M. (2024) *Assessing Bias Prevention and Bias Adjustment in a Sub-Annual Online Panel Survey*. Understanding Society Working Papers Series 2024-04.

Steyerberg, E.W., Eijkemans, M.J.C. & Habbema, J.D.F. (2001) Application of shrinkage techniques in Logistic regression analysis: a case study. *Stat. Neerl.,* **55**, 76–88. https://doi.org/10.1111/1467-9574.00157

Tibshirani, R. (1996) Regression and shrinkage via the Lasso, *J. Roy. Stat. Soc. Ser. B,* **58**, 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Valliant, R., Dever, J. A., & Kreuter, F. (2018) *Practical tools for designing and weighting survey samples*. New York: Springer.

Wooldridge, J. M. (2002). Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese economic journal*, **1**, 117-139. https://doi.org/10.1007/s10258-002-0008-x

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *J. Econometrics*, **141**, 1281-1301. https://doi.org/10.1016/j.jeconom.2007.02.002

6. Tables

Appendix: Table 1. Simulated dataset sizes for each dataset type at each number of extra CATI respondents. Given simulation design (see main text), wave 1 dataset sizes at each number of extra CATI respondents are the same, so only these values are presented. Wave 6 dataset sizes are smaller than wave 1 values due to attrition and non-response, and vary, so their means and minimum and maximum sizes are presented.

	Number of extra wave 1 respondents														
	104			250			500			750			1000		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Wave 1	16108.00			16254.00			16504.00			16754.00			17004.00		
Wave 6 cross-sectional	11192.01	11174	11210	11277.88	11254	11302	11423.73	11390	11466	11571.01	11531	11619	11696.88	11621	11764
Wave 6 longitudinal	10668.01	10650	10686	10753.88	10730	10778	10899.73	10866	10942	11047.01	11007	11095	11172.88	11097	11240

Appendix Table 2: UKHLS COVID-19 Study wave 1 dataset non-response weight biases. Whether differences between Study eligible sample main survey wave 9 weighted survey variable means ('wt. est.') and similar for Study non-response weighted respondents ('wt. diff.') equal zero is tested. `Core benefits' include Income Support, Job Seekers Allowance and Universal Credit. * equals *P*<0.05.

	Eligible	Reg. int. user	All Web	Web + CATI	Reg. int. user + CATI
Variable	wt. est.	wt. diff	wt. diff	wt. diff	wt. diff
In weighting model:					
Subjective financial situation (SFS):	0.71	0.00	0.01	0.00	0.00
Comfortable or OK	(0.00)				
SFS: Just about getting by	0.21	0.00	0.00	-0.00	-0.00
	(0.00)				
SFS: Finding it quite/very difficult	0.08	-0.01	-0.01*	-0.00	0.00
	(0.00)				
Tenure: Owner occupied	0.35	0.02*	0.01*	0.00	0.00
	(0.00)				
Tenure: Mortgage	0.35	-0.02*	-0.02*	0.01	0.01*
	(0.00)				
Tenure: Rented	0.12	-0.01	-0.00	0.00	-0.00
	(0.00)				
Tenure: Social Housing	0.18	0.01	0.00	-0.01	-0.01
	(0.00)				
Low skill Occupation: Yes	0.35	0.00	0.00	-0.00	-0.01
	(0.00)				
Any savings income: Yes	0.36	-0.00	-0.00	0.01	0.01*
	(0.00)				
Behind with some or all bills: Yes	0.06	-0.00	-0.00	0.00	0.00
	(0.00)				
Not in weighting model:					
Income poverty: Yes	0.14	0.01	0.01	-0.00	-0.01
	(0.00)				
Receives core benefit: Yes	0.05	-0.00	-0.00	-0.00	-0.01
	(0.00)				
Visited GP in last year: Yes	0.78	-0.00	-0.00	-0.01	-0.01
	(0.00)				
Smoker: Yes	0.15	0.01*	0.01*	0.01	0.00
	(0.00)				
Hospital outpatient in last year: Yes	0.46	-0.00	-0.01	-0.01	-0.01

Appendix Table 3: UKHLS COVID-19 Study wave 6 cross-sectional dataset non-response weight biases. Whether differences between Study eligible sample main survey wave 9 weighted survey variable means ('wt. est.') and similar for Study non-response weighted respondents ('wt. diff.') equal zero is tested. `Core benefits' include Income Support, Job Seekers Allowance and Universal Credit. * equals *P*<0.05.

	Eligible	Reg. int. user	All Web	Web + CATI	Reg. int. user + CATI
Variable	wt. est.	wt. diff	wt. diff	wt. diff	wt. diff
In weighting model:					
Subjective financial situation (SFS):	0.71	-0.00	0.00	-0.00	-0.01
Comfortable or OK	(0.00)				
SFS: Just about getting by	0.21	0.00	0.00	0.00	0.00
	(0.00)				
SFS: Finding it quite/very difficult	0.08	-0.00	-0.00	0.00	0.00
	(0.00)				
Tenure: Owner occupied	0.35	0.03*	0.01*	0.01	0.01
	(0.00)				
Tenure: Mortgage	0.35	-0.04*	-0.02*	0.01	0.01
	(0.00)				
Tenure: Rented	0.12	-0.01	-0.00	-0.00	-0.00
	(0.00)				
Tenure: Social Housing	0.18	0.02*	0.01	-0.01	-0.01
	(0.00)				
Low skill Occupation: Yes	0.35	0.02	0.01	0.00	0.00
	(0.00)				
Any savings income: Yes	0.36	-0.01	-0.00	0.00	0.01
	(0.00)				
Behind with some or all bills: Yes	0.06	-0.00	0.00	0.00	0.00
	(0.00)				
Not in weighting model:					
Income poverty: Yes	0.14	0.02*	0.02*	0.00	-0.00
	(0.00)				
Receives core benefit: Yes	0.05	-0.00	-0.00	-0.01	-0.01
	(0.00)				
Visited GP in last year: Yes	0.78	0.00	0.00	-0.00	-0.00
	(0.00)				
Smoker: Yes	0.15	0.02*	0.02*	0.01	0.01
	(0.00)				
Hospital outpatient in last year: Yes	0.46	-0.00	-0.01	-0.01	-0.02
	(0.00)				

Appendix Table 4: UKHLS COVID-19 Study wave 6 longitudinal dataset non-response weight biases. Whether differences between Study eligible sample main survey wave 9 weighted survey variable means ('wt. est.') and similar for Study non-response weighted respondents ('wt. diff.') equal zero is tested. `Core benefits' include Income Support, Job Seekers Allowance and Universal Credit. * equals *P*<0.05.

	Eligible	Reg. int. user	All Web	Web + CATI	Reg. int. user + CATI
Variable	wt. est.	wt. diff	wt. diff	wt. diff	wt. diff
In weighting model:					
Subjective financial situation (SFS):	0.71	0.01	0.01	0.01	-0.00
Comfortable or OK	(0.00)				
SFS: Just about getting by	0.21	0.00	0.00	-0.00	0.00
	(0.00)				
SFS: Finding it quite/very difficult	0.08	-0.01	-0.01	-0.00	0.00
	(0.00)				
Tenure: Owner occupied	0.35	0.03*	0.02	0.01	0.01
	(0.00)				
Tenure: Mortgage	0.35	-0.04*	-0.02*	0.01	0.02*
	(0.00)				
Tenure: Rented	0.12	-0.01	-0.00	-0.00	-0.01
	(0.00)				
Tenure: Social Housing	0.18	0.01	0.00	-0.02	-0.02
	(0.00)				
Low skill Occupation: Yes	0.35	0.01	0.00	-0.00	-0.00
	(0.00)				
Any savings income: Yes	0.36	0.00	-0.00	0.01	0.01
	(0.00)				
Behind with some or all bills: Yes	0.06	-0.00	-0.00	0.00	0.01
	(0.00)				
Not in weighting model:					
Income poverty: Yes	0.14	0.01	0.01	-0.00	-0.01
	(0.00)				
Receives core benefit: Yes	0.05	-0.00	-0.01	-0.01	-0.01
	(0.00)				
Visited GP in last year: Yes	0.78	0.00	0.00	-0.00	0.00
	(0.00)				
Smoker: Yes	0.15	0.02	0.02*	0.01	0.00
	(0.00)				
Hospital outpatient in last year: Yes	0.46	0.00	-0.01	-0.01	-0.02
	(0.00)				