SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

Working Paper 9:
Can The Continuum of Resistance Model Be Applied to Understand Mode Selection in Sequential Mixed-Mode Surveys?

Liam Wright[1], Georgia Tomova[1], Richard J. Silverwood[1]

[1] Centre for Longitudinal Studies, University College London

November 2025

www.surveyfutures.net

Survey Futures is an Economic and Social Research Council (ESRC)-funded initiative (grant grant ES/X014150/1) aimed at bringing about a step change in survey research to ensure that high quality social survey research can continue in the UK. The initiative brings together social survey researchers, methodologists, commissioners and other stakeholders from across academia, government, private and not-for-profit sectors. Activities include an extensive programme of research, a training and capacity-building (TCB) stream, and dissemination and promotion of good practice. The research programme aims to assess the quality implications of the most important design choices relevant to future UK surveys, with a focus on inclusivity and representativeness, while the TCB stream aims to provide understanding of capacity and skills needs in the survey sector (both interviewers and research professionals), to identify promising ways to improve both, and to take steps towards making those improvements. 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.

This paper is a product of the project "Assessing and disseminating methods for handling mode effects", led by Liam Wright (UCL), which forms part of *Survey Futures* Research Strand 6, "Reducing and evaluating mode effects."

Prior to citing this paper, please check whether a final version has been published in a journal. If so, please cite that version. In the meanwhile, the suggested form of citation for this working paper is:

Wright L, Tomova G D & Silverwood R J (2025) 'Can the Continuum of Resistance Model be Applied to Understand Mode Selection in Sequential Mixed-Mode Surveys?', *Survey Futures Working Paper* no. 9. Colchester, UK: University of Essex. Available at https://surveyfutures.net/working-papers/.

Contents

Table of Contents

Abstract

Background: Analyses of sequential mixed-mode survey data can be biased due to non-random selection of participants into mode. Understanding this bias requires knowledge of predictors of mode selection, but research on this is sparse compared with survey non-response. The 'continuum of resistance' model of survey response predicts that delayed responders – who, in sequential mixed-mode surveys, use later modes – and non-responders share similar characteristics. If correct, research on non-response could generalize to mode selection.

Methods: We used data from a major UK birth cohort study (the 1958 National Child Development Study) which embedded a sequential web-then-telephone survey at the age 55y sweep. We assessed whether (a) late (telephone) responders and non-responders share characteristics and (b) whether non-response models also predict telephone response. For (a), we calculated univariate descriptive statistics and performed cluster analysis to compare participant characteristics across web, telephone, and non-response group. For (b), we estimated random forest models for non-response and telephone response (conditional on response) then compared predictive accuracy (Area Under the Receiver Operating Characteristic curve [AUC ROC] and Brier scores) when using the non-response model to instead predict telephone response.

Results: Telephone and non-respondents were similar on almost all (measured) characteristics, and dissimilar in most regards to web respondents. Predictions from non-response models had similar predictive accuracy to predictions from models trained on telephone response, specifically – AUC ROC values in hold-out samples were 0.72 (95% CI = 0.70, 0.74) and 0.74 (95% CI = 0.72, 0.75), respectively.

Conclusions: The characteristics of late- and non-responders in a sequential mixed-mode survey were very similar, consistent with the 'continuum of resistance' model of survey response. This suggests that research on non-response could transport to understanding mode selection in sequential mixed-mode surveys, though replications in other surveys with different mixed-mode designs is required.

Keywords: mixed-mode surveys; mode effects; mode selection; non-response; continuum of resistance

Introduction

Surveys continue to adopt mixed-mode designs. The potential benefits of mixed-mode surveys are lower costs and higher response rates as participants can be initially offered cheaper modes or be given the option to respond via the mode that is most convenient for them. A drawback, however, is that the measurement of survey items can differ between modes (d'Ardenne et al., 2017). This may affect how participants respond to individual items; difference in response between modes due to how items are measured are termed *mode effects* (Jäckle et al., 2010). Mode effects can introduce bias into analyses of mixed-mode survey data. This includes inducing correlations between items subject to mode effects and between items subject to mode effects and factors that predict mode used (Wright et al., 2024); difference between modes due to *who* is responding in each mode is termed *mode selection*.

Take a hypothetical web-then-telephone sequential mixed-mode survey where males are more likely to respond by telephone. Web questionnaires are anonymous, but telephone surveys usually require participants give their answers directly to an interviewer (albeit remotely), even for sensitive questions. Respondents in the telephone survey may be unwilling to give accurate responses, which would lead to underreporting of, for example, mental and physical health conditions (Goodman et al., 2022). A naïve analysis of data from this survey would identify sex differences in these conditions, but this could be entirely spurious or an otherwise biased reflection of underlying reality. A Causal Directed Acyclic Graph (DAG; Greenland et al., 1999) reflecting this scenario is shown in Figure 1.

Where there is mode selection, simply controlling for mode is unlikely to fully ameliorate bias from mode effects (see Figure 1 and also Wright et al., 2024). This is because, in the presence of mode selection, mode may be a "collider" variable (Munafò et al., 2018): conditioning on mode induces a correlation between the causes of mode selection – if you know one cause of mode selection isn't present, one or more of the other causes must be. Introducing this bias can exceed the reduction in bias from controlling for mode effects. Predicting whether – and how – a specific analysis of mixed-mode data could be biased therefore requires an understanding of factors related to mode selection. However, this likely sits outside a researcher's usual area of expertise. Further, this information may not be readily obtainable – to our knowledge, research on the factors influencing mode selection is relatively sparse.

Selection into mode can be intentional or result from other concerns. In some surveys, participants may actively choose – or be actively offered – a specific mode, but in the sequential mixed-mode design, alternate modes are offered only after non-response to previously offered modes. Factors that predict delayed response to survey invitations in general should therefore be more prevalent in latter-offered modes. Late response has been used to understand the characteristics of survey non-responders, with researchers adopting a 'continuum of resistance' model (Lin & Schaeffer, 1995) whereby the factors determining late response are assumed to be similar to those determining non-response (see, for example, Boniface et al., 2017). Importantly, this reasoning can be applied in reverse: if the continuum of resistance model is correct, the characteristics of non-respondents could be used to understand characteristics of people in later offered modes. This is fortunate given burgeoning research on the predictors of survey non-response (e.g., Katsoulis et al., 2024; Mostafa et al., 2021; Silverwood et al., 2024; Watson & Wooden, 2009).

However, the ‘continuum of resistance’ model may not be valid as delayed response and non-response may reflect distinct decisions and decision-making processes, each with their own unique determinants. This may be especially the case in sequential mixed-mode surveys as the characteristics of a mode itself may induce a person to participate in the survey, where they otherwise would not; for instance, a person may not have the requisite skills to complete a web questionnaire but still complete a telephone interview when called.

Longitudinal studies offer an opportunity to test the continuum of resistance model as data from (sweep) non-responders may be available from prior sweeps which they did respond to. In this paper, we use data from a major birth cohort study, the 1958 National Child Development Study, which used a sequential mixed mode design in a recent sweep, to examine whether the factors predicting mode selection are similar to those predicting non-response. Our aim is to determine whether the literature on non-response may plausibly be applied to understand mode selection in sequential mixed-mode surveys.

Methods

Participants

Participants were from the 1958 National Child Development Study (NCDS; UCL Social Research Institute, 2024), a British birth cohort tracking all individuals born in mainland Britain in a single week of March 1958 (Power & Elliott, 2006). Immigrants to the UK born in the same week were later added to the cohort using school enrolments. Cohort members have been followed intermittently from birth to the present day, with surveys capturing a wide variety of demographic, socioeconomic, and health data. The ninth major sweep (age 55y, 2013) embedded a survey mode experiment with cohort members (for whom a telephone number was available) randomly allocated to either telephone only or a web-then-telephone sequential mixed-mode survey design. In the web-then-telephone arm, individuals who did not respond to an invitation to participate by web after six weeks and three written contact attempts (letters or emails) were then contacted by telephone for an interview (see Goodman et al., 2022, for more information).

To understand whether the factors predicting new (rather than ongoing) sweep non-response and telephone response are similar, we restricted our analysis to cohort members who (a) participated at the prior sweep (age 50y, 2008) and (b) were eligible for the experiment and allocated to the web-then-telephone sequential mixed-mode arm (n = 8,031, 43.3% of the total NCDS cohort). Of these, 5,345 (66.6% of eligible) cohort members participated by web, 1,826 (22.7%) participated by telephone, and 860 (10.7%) did not respond.

Measures

Factors predicting mode selection at the age 55y sweep have been examined in a prior study (Goodman et al., 2022). We examined the same set of factors here. These were a mix of categorical and continuous variables, each of which was collected at the age 50y sweep (a single mode survey). Variables were: sex, ethnic group, resident partner, cognitive ability (PCA of measures of word recognition and memory), highest qualification (NVQ), housing tenure, occupational social class, weekly net income, self-rated health, smoking status, alcohol use (AUDIT group), health related quality of life (CASP-12 score), computer use at home, internet use outside work, and computer skills. For comparability, continuous variables were standardized (mean = 0; SD = 1) using means and standard deviations observed in the web sample. More detail on the variables is provided in Table 1.

Statistical Analysis

Our aim was to understand whether (a) the average characteristics of telephone respondents were similar to non-respondents and, relatedly, whether (b) the predictors of telephone response were similar to the predictors of non-response. To this end, we performed a series of descriptive and predictive analyses. First, we produced univariate descriptive statistics (means and proportions) for each variable for web, telephone, and non-respondents, separately, and examined visually whether telephone and non-respondents were more similar to each other than they were with web respondents, a prediction of the continuum of resistance model. Second, as these univariate descriptives did not account for the interrelation between variables, we supplemented this with a cluster analysis, examining whether the proportion of cohort members in each cluster was similar for telephone and non-respondents and dissimilar to the proportions of web respondents. Given the measures were a mix of categorical and continuous variables, we performed the cluster analysis using the Partitioning Around Medoids algorithm with Gower's Distance measure. We performed cluster analysis for 2 to 8 clusters, with the final number of clusters chosen by examining an elbow plot of average silhouette width – silhouette width compares distances between observations within a cluster (which should be small) to distances from nearest neighbouring clusters (which should be large).

Third, we assessed whether a multivariate machine learning model trained to predict non-response (vs response [telephone or web]) also gave accurate predictions for whether a cohort member responded by telephone (vs. web), including in comparison to a model trained to predict telephone response (vs. web; i.e., among respondents), specifically. Models were estimated using the random forest algorithm (Breiman, 2001; James et al., 2021), which we expected to perform better than generalized linear (logit) regression models given the large number of (collinear) predictors, which potentially have interacting and non-linear relationships with telephone and non-response; random forests use decision trees that make successive splits of the data to generate predictions which allows for non-linear relationships without explicit parametrization. We evaluated the accuracy and transportability of model predictions in four ways. First, among web and telephone respondents, we examined the (Spearman's rank) correlation between model predictions for probabilities of telephone response and non-response – high correlations would suggest that those who were likely to answer by telephone were also more likely to not respond at all. Second, we compared (permutation-based) variable importance across models to assess whether most predictive variables were shared. Third, we calculated the Brier score in the web and telephone respondents applying the non-response model as if it had predicted telephone response; the Brier score, which ranges 0-1, is conceptually similar to mean square error for prediction models. Lower values indicate better prediction. Fourth, we calculated the Area Under Receiving Operating Characteristic Curve (AUC ROC), which can be conceptually understood as calculating, for random pairs of web and telephone respondents, the proportion of pairs where the telephone respondent had the higher predicted probability of non-response. Given the proportions of telephone respondents and non-respondents differed in our data (22.7% vs 10.7%), we used Bayes' Theorem to rescale non-response probabilities when applying these to predict telephone response. We compared Brier and AUC ROC scores against those estimated from telephone response models specifically, as well as against 'baseline' models where (for Brier scores) a single probability equal to the proportion of the sample responding by telephone was used to predict telephone response (22.7%) and (for AUC ROC) a random cohort member in the pair was assumed to respond by telephone (i.e., a coin toss – AUC ROC = 0.5). To avoid overfitting, we calculated Brier and AUC ROC scores in holdout samples not used to estimate the random forest models (further described below). As a preliminary step, we also used ten-fold cross-validation to choose values for the following model 'hyperparameters': intermediate and terminal node size (the minimum number of observations at a split or used to make a prediction, respectively) and maximum tree depth (the number of layers of nodes used in a given model).

There was a small amount of item missingness in the measures that we used. Accordingly, we imputed missing data under the Missing at Random assumption using multiple imputation by chained equations (20 imputations, burn-in 10) (van Buuren, 2018). We used classification and regression trees for each imputation model to allow for flexibility in the relationships between variables. No auxiliary variables were added. For the univariate descriptive statistics, we used the imputed data and combined estimates using Rubin's Rules (Rubin, 1987). For analyses using random forest models, we used the impute-then-bootstrap method (100 bootstraps per imputed dataset; Bartlett & Hughes, 2020), calculating confidence intervals with the centile method (smallest 2.5th and largest 97.5th values). Observations not included in a given bootstrap (36.8% of observations, in expectation) were used as a holdout (assessment) sample to evaluate that model.

As the cluster analysis solutions can vary across imputations, we carried out the cluster analysis for each imputed dataset separately, and then (a) calculated χ^2 statistics for cluster membership by response type (separately web vs. telephone, web vs. non-response, and telephone vs. non-response) for each cluster analysis model separately, and (b) visually examined whether elbow plots and average silhouette width were similar in each imputed dataset. Thankfully, cluster analysis results were very similar across imputed datasets.

Results

Descriptive Statistics

5,345 (66.6%) eligible cohort members participated by web, 1,826 (22.7%) by telephone, and 860 (10.7%) did not respond. Univariate descriptive statistics by response type are displayed in Figure 2. Telephone participants and non-respondents differed on several characteristics from web participants and were very similar (on average) to each other along various dimensions, including cognitive ability, use of computers and the internet, physical and mental health, and measures of socio-economic position (income, education, housing tenure and occupational class). However, some small differences between telephone respondents and non-respondents were observed, including smoking status. Supplementary Figure S1 displays the results as between group standardized mean differences or odds ratios. Differences between telephone vs. web respondents were typically only slightly smaller in size than differences between non-respondents vs. web respondents.

Cluster Analysis

We chose a three-cluster solution (mean average silhouette width across imputations = 0.100; elbow plot shown in Supplementary Figure S2). The proportions of telephone respondents and non-respondents in each cluster were very similar (p-values > 0.05 in each imputed dataset) and differed from the proportions from the web response groups (p-values < 0.01 in each imputed dataset; Figure 3). Relative to telephone and non-respondents, web respondents were much more likely to appear in Cluster 1 and much less likely to appear in Cluster 3. Cluster 1 was characterised by having a degree (NVQ4), higher than average cognitive ability, and employment in a professional occupation. Cluster 3 was characteristed by lower-than-average cognitive ability, worse quality of life and self-rated health, and limited computer use and skills. Average characteristics among cohort members in each cluster are shown in Supplementary Figure S3.

Predicting Telephone Response Using Predictions of Non-Response

A scatter plot of predicted telephone response (x-axis) against predicted non-response (y-axis) among web and telephone respondents is shown in Figure 4. The (Spearman's) correlation between the predicted probabilities was $\rho = 0.82$ (95% CI = 0.77, 0.86). Predictors ordered by variable importance are shown in Figure 5. There was concordance between the two – six of the eight most important variables in the non-response model were in the top eight for the telephone response mode, and (Spearman's rank) correlation in the order of variable importance was $\rho = 0.72$ (95% CI = 0.52, 0.89).

Some discrepancies were notable, however. The three most important variables in the telephone response model were computer use, internet use and computer skills, while, in the non-response models, the three most important variables were education, computer use and cognitive ability. The correlation between computer skills and internet use and education/cognitive ability may explain the strong correlation between predicted probabilities in telephone- and non-response models.

The Brier score when applying the model predictions from non-response to telephone response in the holdout sample was 0.17 (95% CI = 0.16, 0.17) (Table 2). This improved upon a 'baseline' prediction using the raw percentage of telephone respondents in the sample (22.7%; Brier score 0.19, 95% CI = 0.19, 0.19) and was only slightly worse than the Brier score when using predictions from the telephone response model specifically (0.16, 95% CI = 0.15, 0.16). The AUC ROC when using non-response predictions in the holdout sample was 0.72 (95% CI = 0.70, 0.74; Table 2). This was similar to the AUC ROC for telephone specific prediction of 0.74 (95% CI = 0.72, 0.75) and considerably better than using at random guess (i.e., AUC ROC = 0.50).

Discussion

Using data from a web-then-telephone sequential mixed-mode survey embedded within a longitudinal cohort study, we found that (a) the characteristics of cohort members responding in the latter mode (telephone) were broadly similar to non-respondents and (b) the characteristics of both were dissimilar to cohort members responding in the earlier mode (web). We further found that models produced to predict non-response (vs. response) were also able to predict telephone (vs. web) response to a similar degree of accuracy as models produced to predict telephone response specifically.

The results are broadly consistent with the 'continuum of resistance' model – late (telephone) respondents and non-respondents were more similar to each other than to early (web) respondents. There were, however, some differences between the telephone-respondent and non-respondent groups. Specifically, computer skills and internet use were among the three most important predictors of telephone response, whereas for non-response, education and cognitive ability were more predictive. This could reflect specific decisions not to participate by web, rather than against responding altogether, though computer use was the second most important variable in the non-response model. It is possible that, for some individuals, the first offered mode determines non-response altogether, regardless of which modes are offered later. Similarities between telephone respondents and non-respondents may therefore reflect the initial offer of web rather than a 'continuum of resistance'. This would contradict the results of two previous studies which randomised the sequence of offered modes and did not find an effect on overall responses rates, though the modes offered were not identical to those used here (McMorris et al., 2009; Sakshaug et al., 2019; Wells et al., 2024). Further, in the NCDS 55y sweep, the sequential mixed-mode design only had a relatively small effect on response rates, increasing response by ~ 5% points, relative to a single telephone-only design (Goodman et al., 2022). The analysis here requires replication in surveys using different mixed-mode designs, including different modes and sequences. Ultimately, those who respond by a later mode may do so due to responding late, due to having a preference against the first mode offered, or a combination of the two, and the exact reasons may differ across surveys and populations.

If the 'continuum of resistance' model is correct, studies assessing the predictors of non-response may generalize to late response in sequential mixed-mode surveys, too. The converse would also be true, which could have utility to research attempting to understanding potential bias from non-response. Specifically, observations (for variables not subject to mode effects) in individuals appearing in latter modes could be artificially set to missing and then imputed to assess the plausibility of the Missing at Random Assumption (van Buuren, 2018) used in multiple imputation.

Using the results of studies on the predictors of non-response to understand mode selection is beneficial given the growing literature on non-response (Katsoulis et al., 2024; Mostafa et al., 2021; Silverwood et al., 2024). However, a few notes of caution are required, in addition to the possibility – discussed above – that mode itself influences response. First, evidence from the literature testing the continuum of resistance model has not been uniformly consistent (Clarsen et al., 2021). Second, in our data, web respondents made up the majority (74.5%) of respondents. Prediction models for non-response were therefore weighted towards web respondents. A more equal split of modes may have generated different results, though this would not explain the descriptive similarity between telephone- and non-respondents (i.e., as observed in Figure 2).

Third, we limited our analysis to NCDS cohort members who were eligible for the telephone survey and had responded at the prior sweep. Sequential mixed-mode surveys often include a subsample of participants only eligible for a particular mode (typically due to ability or contact information available) – these criteria again reflect a distinct decision-making process, which may have unique determinants and predictors. However, in practice, these subsamples may be small. Continued non-response likely also reflects idiosyncratic processes – relevant factors at *first* non-response may be important (e.g., contemporary mental health) and less so later, given non-response is the status quo action. Fourth, (non-random) attrition can lead to sample biases, which mean that predictors of late- or non-response at a later wave do not translate to late- or non-response at an earlier wave (and vice versa). For instance, if males are more likely to drop-out early in a study, remaining males will likely be unrepresentative with regards to other characteristics that influence non-response (i.e., there will be selection [collider] bias). Again, replication in other surveys with different designs and other idiosyncrasies is required.

Strengths and Limitations

Strengths of this study included the longitudinal design which enabled the use of a wide array of potential predictors of late- and non-response. Other strengths were the use of multiple imputation to address missing data concerns and the assessment of predictive accuracy out-of-sample to avoid overfitting. Limitations included the use of a cohort which had already undergone attrition and the possibility discussed above that results do not generalize to other groups that could appear in a (longitudinal) sequential mixed-mode surveys – those allocated to a single mode and those who drop out in earlier survey sweeps. We only compared web and telephone response; other modes, or the modes offered in reverse sequence, may have larger effects on response, which could influence the transportability of findings. Finally, results for variable importance metrics are ‘local’ and reflect the set of predictors that are (and are not) included in a model. Importance does not imply a variable has a causal effect on response. The inclusion of other variables, including those that have direct causal effects, could yield different results.

Conclusions

In line with the continuum of resistance model, the characteristics of late (i.e., telephone) respondents in a web-then-telephone sequential mixed-mode survey were broadly similar to non-respondents, though some differences plausibly determining mode preference (i.e., internet use and computer skills) were also observed. The results suggest that findings from the burgeoning literature on the determinants of survey non-response could be applied to understand mode selection in sequential mixed-mode surveys, though replication in other surveys, using alternate sequences is required.

Statements

Declaration of interest

All authors declare no conflicts of interest.

Funding

This work was funded as part of *Survey Futures*, a project funded by the Economic and Social Research Council (ES/X014150/1). The UCL Centre for Longitudinal Studies is supported by the Economic and Social Research Council (ES/W013142/1). The funders had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. All researchers listed as authors are independent from the funders and all final decisions about the research were taken by the investigators and were unrestricted.

Acknowledgements

Author contributions

All authors contributed to and agreed upon the analysis plan. LW carried out the analysis and wrote the first draft of the manuscript. All author provided critical revisions and read and approved the final manuscript.

Data availability

Data from the National Child Development Study are available from the UK Data Service: https://doi.org/10.5255/UKDA-Series-2000032. Variables for the 55y sweep mixed-mode experiment treatment assignment, which were used to determine eligible non-respondents, were acquired directly from the Centre for Longitudinal Studies, which manages the NCDS: https://cls.ucl.ac.uk/data-access-training/data-access/. The code used to run the analysis is available at https://osf.io/pz645/.

Figures

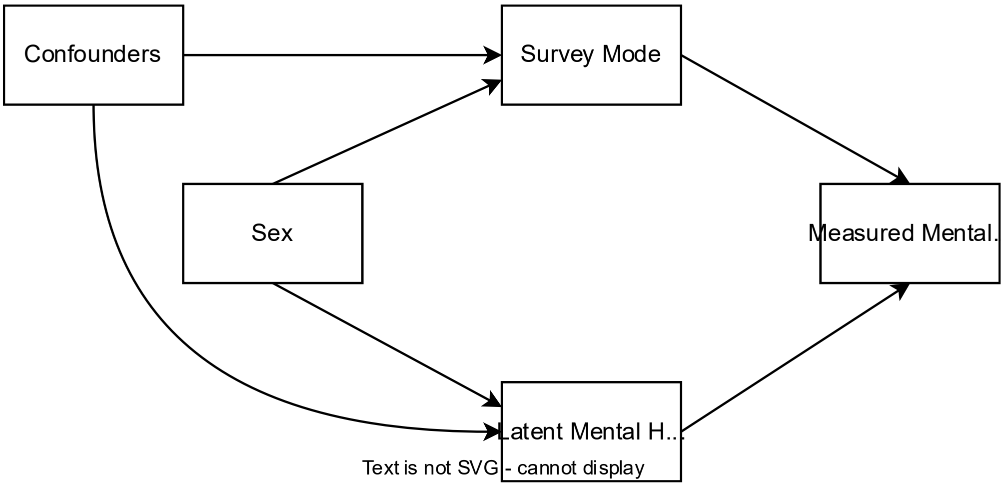

Figure 1: DAG representing an analysis of data from a mixed mode survey where (a) a mental health measure is subject to a* mode effect*, (b) sex is a cause of* mode selection*, and (c) there are other causes of mode selection that are also causes of (latent) mental health. *In this data, sex will be associated with measured mental health due to its causal effect upon latent mental health ('Sex' → 'Latent Mental Health' → 'Measured Mental Health') but also due to its association with survey mode (i.e., the path 'Sex' → 'Survey Mode' → 'Measured Mental Health'). Adjusting for survey mode (e.g., by stratification or statistical control) closes this latter path but opens another biasing path from sex to measured mental health as survey mode is a collider between sex and other causes of mode selection and latent mental health ('Sex' → 'Survey Mode'← 'Confounders' → 'Latent Mental Health' → 'Measured Mental Health'). See Wright et al. (2024) for further discussion of representing bias in mixed mode surveys using DAGs.*

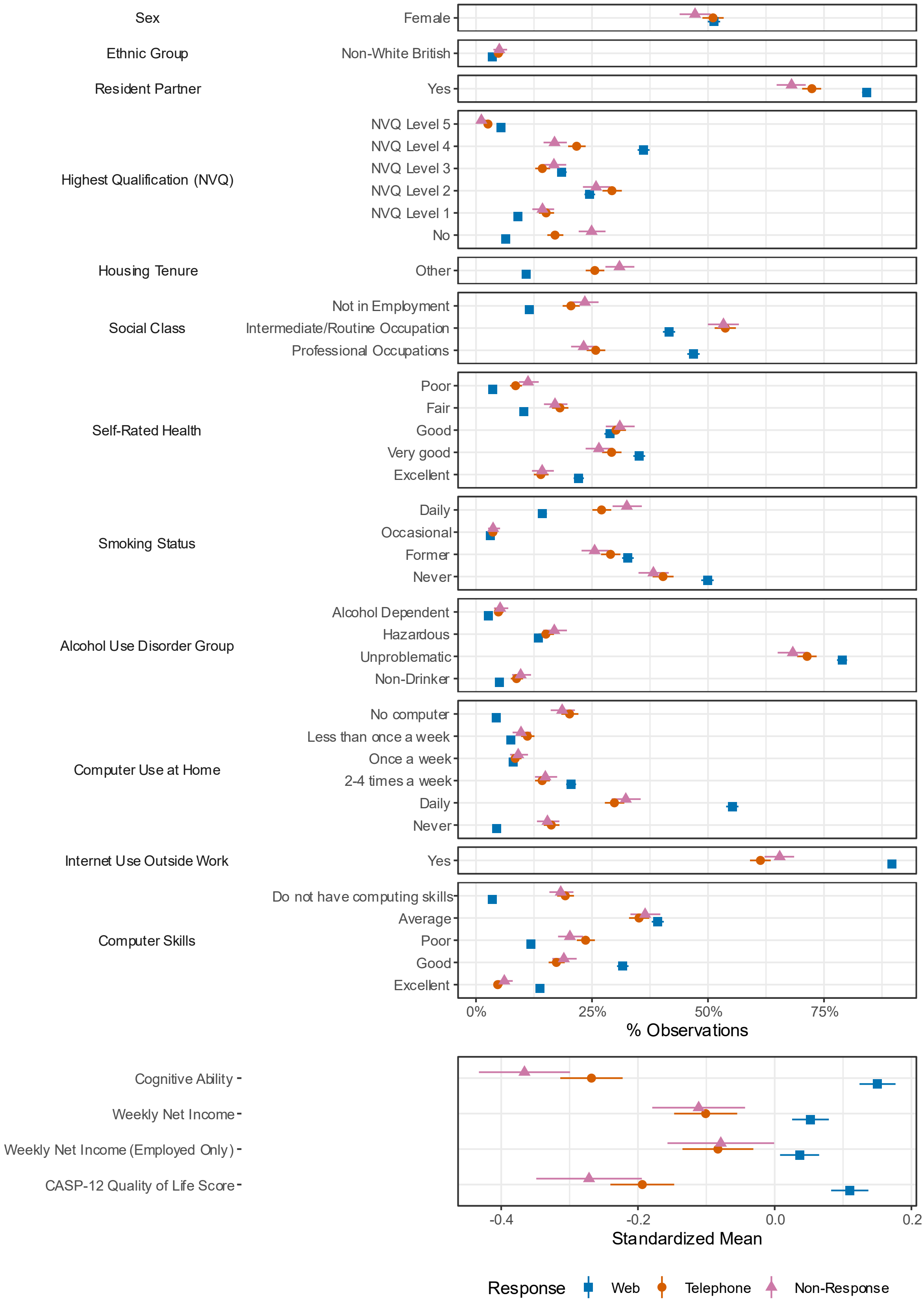

Figure 2: Average (mean or %) characteristics (+ 95% CIs) of cohort members according to response type at Age 55y (web, telephone, or non-response). *Figures calculated using multiply imputed data (m = 20) combined using Rubin's Rules (Rubin, 1987).*

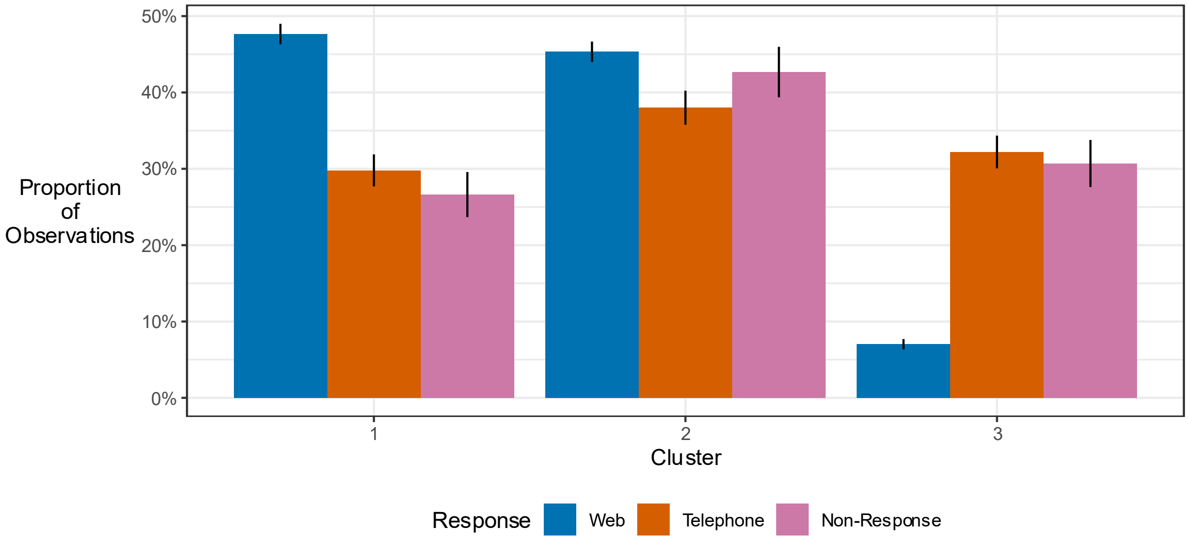

Figure 3: Proportion of cohort members in each cluster (+ 95% CIs) according to response type at Age 55y (web, telephone, or non-response). *Cluster analysis performed using Gower's distance and the partitioning around medoids algorithm with observed data. Final number of clusters selected using average silhouette width metric (elbow plot displayed in Supplementary Figure S2). Average characteristics of cohort members in each cluster are shown in Supplementary Figure S3. For clarity, proportions are shown for the cluster analysis performed using the first imputed dataset. Results were very similar across imputations*

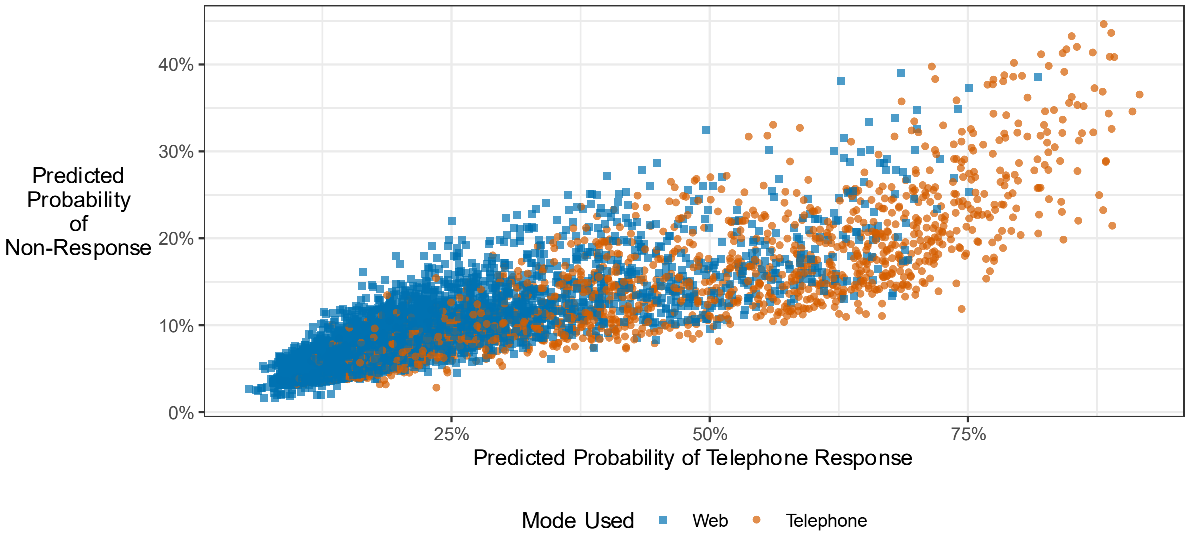

Figure 4: Predicted probabilities of telephone response (x-axis) and non-response (y-axis) among cohort members who, in fact, participated by web or telephone. *Derived from random forest models using multiply imputed data (imputed-then-bootstrap method; m = 20, bootstraps = 500). Probabilities averaged (mean) across bootstrap models.*

Figure 5: Rank of variable importance for predictors including in the telephone response and non-response random forest models. *Derived from random forest models using multiply imputed data (imputed-then-bootstrap method; m = 20, bootstraps = 500). Variable importance calculated using permutation-based approach and averaged (mean) across bootstrap models prior to ranking. Nine of the top ten most important variables in the non-response model appear in the ten most important variables in the telephone response model.*

Tables

Table 1: Variables used in the present study

Variable	Description	Source Variables
Sex	Categorical: Male (49.2%); Female (50.8%)	N622
Ethnic Group	Categorical: White British (96%); Non-White British (4%)	ND8ETHNC
Resident Partner	Categorical: No (20.2%); Yes (79.8%)	ND8SPPHH
Cognitive Ability	Mean (SD): 0 (1); Range: -4.27 - 3.4	PCA of: N8CFANI, N8CFLISN, N8CFLISD, N8CFMIS
Highest Qualification (NVQ)	Categorical: No (10.8%); NVQ Level 1 (11%); NVQ Level 2 (25.7%); NVQ Level 3 (17.3%); NVQ Level 4 (30.8%); NVQ Level 5 (4.3%)	ND8HNVQ
Housing Tenure	Categorical: Home Owner (83.8%); Other (16.2%)	N8TEN
Social Class	Categorical: Professional Occupations (39.6%); Intermediate/Routine Occupation (45.6%); Not in Employment (14.8%)	N8ECON02, ND8NS3P
Weekly Net Income	Mean (SD): 337.87 (810.38); Range: 0 - 57696.75	N8ECON02, N8CNETWK
Weekly Net Income (Employed Only)	Mean (SD): 412.15 (877.77); Range: 0.02 - 57696.75	N8ECON02, N8CNETWK
Self-Rated Health	Categorical: Excellent (19.4%); Very good (32.9%); Good (29.4%); Fair (12.7%); Poor (5.5%)	N8HLTHGN
Smoking Status	Categorical: Never (46.5%); Former (31.1%); Occasional (3.2%); Daily (19.1%)	N8SMOKIG
Alcohol Use Disorder Group	Categorical: Non-Drinker (6.3%); Unproblematic (76.3%); Hazardous (14.1%); Alcohol Dependent (3.3%)	ND8AUDG
CASP-12 Quality of Life Score	Mean (SD): 26.13 (5.76); Range: 2 - 36	ND8CSP12
Computer Use at Home	Categorical: Daily (47%); 2-4 times a week (18.5%); Once a week (8.2%); Less than once a week (8.5%); Never (8.3%); No computer (9.4%)	N8HPCUSE, N8PCHOME
Internet Use Outside Work	Categorical: No (19.4%); Yes (80.6%)	N8INTACC
Computer Skills	Categorical: Excellent (10.9%); Good (27%); Average (38%); Poor (15.4%); Do not have computing skills (8.7%)	N8CMPSK1

Table 2: Accuracy of predictions for response models

Sample	Prediction Model	AUC ROC	Brier Score
Assessment	Telephone Response Model	0.74 (0.72, 0.75)	0.16 (0.15, 0.16)
	Non-Response Model	0.72 (0.70, 0.74)	0.17 (0.16, 0.17)
	Base Prediction	0.5	0.19 (0.19, 0.19)
Analysis	Telephone Response Model	0.85 (0.84, 0.86)	0.13 (0.12, 0.13)
	Non-Response Model	0.71 (0.70, 0.73)	0.17 (0.17, 0.17)
	Base Prediction	0.5	0.19 (0.19, 0.19)
Original	Telephone Response Model	0.80 (0.79, 0.80)	0.14 (0.14, 0.14)
	Non-Response Model	0.72 (0.71, 0.72)	0.17 (0.17, 0.17)
	Base Prediction	0.5	0.19 (0.19, 0.19)

Accuracy (and 95% confidence intervals) of prediction for response to the NCDS Age 55y survey via telephone (compared with web response). 'Telephone response model' refers to predictions derived from random forest model predicting telephone response using variables described in the 'Methods: Measures' section. 'Non-response model' refers to predictions derived from random forest models predicting non-response using the same predictor variables but then used to predict telephone response instead – to account for differences in the base probability of telephone response (22.7%) and non-response (10.7%), predicted probabilities for the non-response model were rescaled using Bayes' Theorem. 'Base prediction' for AUC ROC is 0.5 – the probability that a random cohort member in a telephone/web response pair in fact responded by telephone. 'Base prediction' for Brier score is the Brier score where the overall probability of telephone response was used to as a prediction for all individuals. Analyses were performed using multiply imputed data using impute-then-bootstrap method (m = 20, bootstraps = 100). 95% confidence intervals were calculated using the centile method. Predictive accuracy is shown in three separate samples. 'Original' refers to the n = 7,171 cohort members who responded via web or telephone. 'Analysis' refers to the cohort members (including repetitions) used in a specific bootstrap sample. 'Assessment' refers to the holdout sample of cohort members not included in a specific bootstrap (approximately 36.8% of observations) – the assessment sample are therefore not used to estimate a given model and so do not suffer from potential overfitting as the 'Analysis' and 'Original' samples do.

References

Bartlett, J. W., & Hughes, R. A. (2020). Bootstrap inference for multiple imputation under uncongeniality and misspecification. *Statistical Methods in Medical Research*, 096228022093218. https://doi.org/10.1177/0962280220932189

Boniface, S., Scholes, S., Shelton, N., & Connor, J. (2017). Assessment of Non-Response Bias in Estimates of Alcohol Consumption: Applying the Continuum of Resistance Model in a General Population Survey in England. *PLOS ONE*, *12*(1), e0170892. https://doi.org/10.1371/journal.pone.0170892

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Clarsen, B., Skogen, J. C., Nilsen, T. S., & Aarø, L. E. (2021). Revisiting the continuum of resistance model in the digital age: A comparison of early and delayed respondents to the Norwegian counties public health survey. *BMC Public Health*, *21*(1), 730. https://doi.org/10.1186/s12889-021-10764-2

d'Ardenne, J., Collins, D., Gray, M., Jessop, C., & Pilley, S. (2017). *Assessing the risk of mode effects: Review of proposed survey questions for waves 7-10 of Understanding Society* (Working Paper Nos 2017–04; Understanding Society Working Paper Series, p. 18). University of Essex. https://www.understandingsociety.ac.uk/research/publications/524254

Goodman, A., Brown, M., Silverwood, R. J., Sakshaug, J. W., Calderwood, L., Williams, J., & Ploubidis, G. B. (2022). The Impact of Using the Web in a Mixed-Mode Follow-up of a Longitudinal Birth Cohort Study: Evidence from the National Child Development Study. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(3), 822–850. https://doi.org/10.1111/rssa.12786

Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. *Epidemiology*, *10*(1), 37–48.

Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the Effect of Data Collection Mode on Measurement. *International Statistical Review*, *78*(1), 3–20. https://doi.org/10.1111/j.1751-5823.2010.00102.x

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (Second edition). Springer. https://doi.org/10.1007/978-1-0716-1418-1

Katsoulis, M., Narayanan, M., Dodgeon, B., Ploubidis, G., & Silverwood, R. (2024). *A data driven approach to address missing data in the 1970 British birth cohort* (p. 2024.02.01.24302101). medRxiv. https://doi.org/10.1101/2024.02.01.24302101

Lin, I.-F., & Schaeffer, N. C. (1995). Using Survey Participants to Estimate The Impact of Nonparticipation. *Public Opinion Quarterly*, *59*(2), 236–258. https://doi.org/10.1086/269471

McMorris, B. J., Petrie, R. S., Catalano, R. F., Fleming, C. B., Haggerty, K. P., & Abbott, R. D. (2009). Use of Web and In-Person Survey Modes to Gather Data From Young Adults on Sex and Drug Use: An Evaluation of Cost, Time, and Survey Error Based on a Randomized Mixed-Mode Design. *Evaluation Review*, *33*(2), 138–158. https://doi.org/10.1177/0193841X08326463

Mostafa, T., Narayanan, M., Pongiglione, B., Dodgeon, B., Goodman, A., Silverwood, R. J., & Ploubidis, G. B. (2021). Missing at random assumption made more plausible: Evidence from the 1958 British birth cohort. *Journal of Clinical Epidemiology*, *136*, 44–54. https://doi.org/10.1016/j.jclinepi.2021.02.019

Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M., & Davey Smith, G. (2018). Collider scope: When selection bias can substantially influence observed associations. *International Journal of Epidemiology*, *47*(1), 226–235. https://doi.org/10.1093/ije/dyx206

Power, C., & Elliott, J. (2006). Cohort profile: 1958 British birth cohort (National Child Development Study). *International Journal of Epidemiology*, *35*(1), 34–41. https://doi.org/10.1093/ije/dyi183

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470316696

Sakshaug, J. W., Cernat, A., & Raghunathan, T. E. (2019). Do Sequential Mixed-Mode Surveys Decrease Nonresponse Bias, Measurement Error Bias, and Total Bias? An Experimental Study. *Journal of Survey Statistics and Methodology*, *7*(4), 545–571. https://doi.org/10.1093/jssam/smy024

Silverwood, R. J., Calderwood, L., Henderson, M., Sakshaug, J. W., & Ploubidis, G. B. (2024). A data-driven approach to understanding non-response and restoring sample representativeness in the UK Next Steps cohort. *Longitudinal and Life Course Studies*, *15*(2), 227–250. https://doi.org/10.1332/17579597Y2024D000000010

UCL Social Research Institute. (2024). *National Child Development Study* (Version 14th Release) [Data set]. UK Data Service. https://doi.org/10.5255/UKDA-SERIES-2000032

van Buuren, S. (2018). *Flexible imputation of missing data* (Second edition). CRC Press, Taylor & Francis Group.

Watson, N., & Wooden, M. (2009). Identifying Factors Affecting Longitudinal Survey Response. In *Methodology of Longitudinal Surveys* (pp. 157–181). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470743874.ch10

Wells, B. M., Christian, L., Bautista, R., Lafia, S., & Davern, M. (2024). *Mode-Sequencing Insights from the General Social Survey*. NORC. https://www.norc.org/research/library/mode-sequencing-insights-from-the-general-social-survey.html

Wright, L., Ploubidis, G. B., & Silverwood, R. J. (2024). *Handling Mode Effects in CLS's cohort Studies*. UCL Centre for Longitudinal Studies.

Supplementary Figures

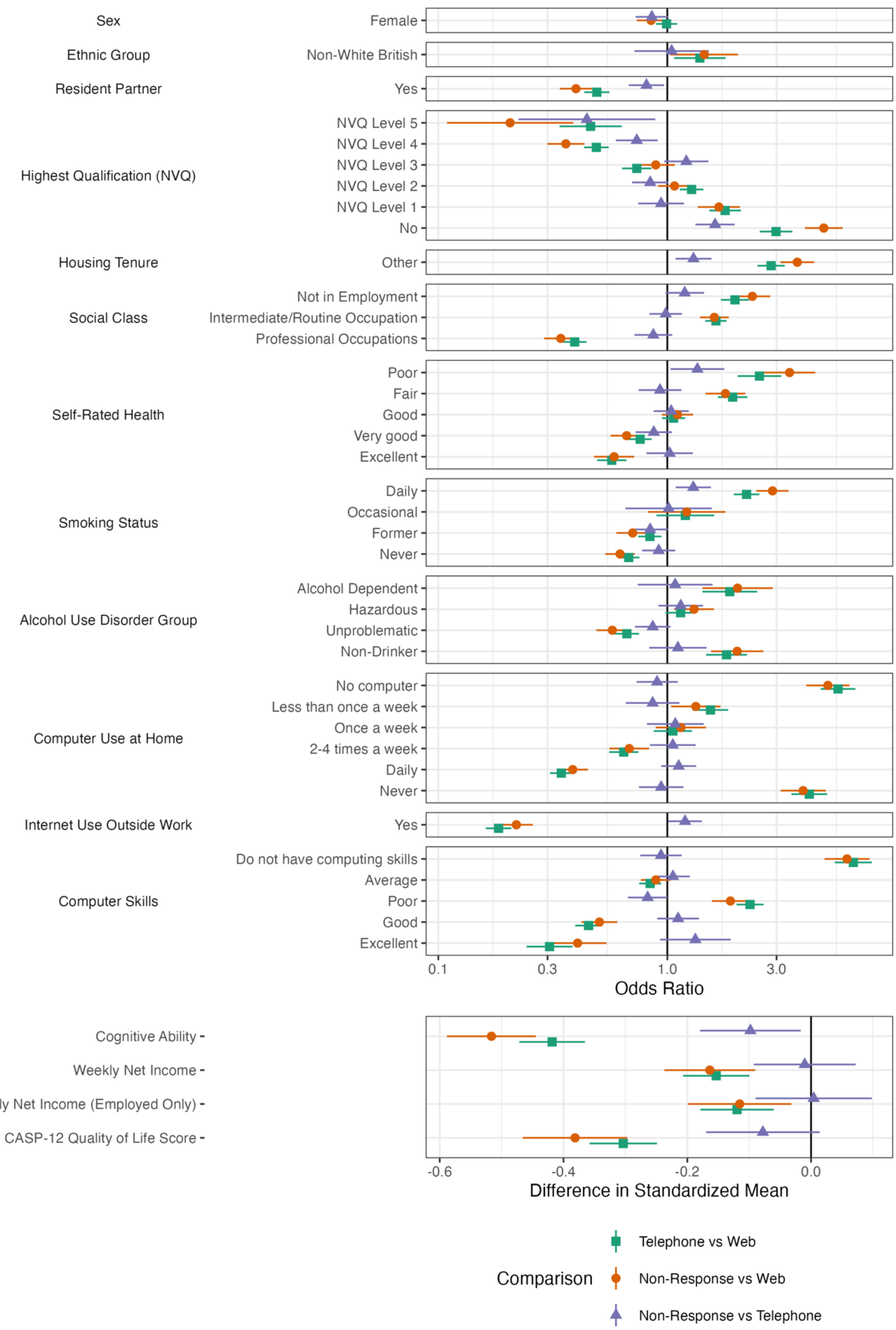

Figure S6: Differences in participant characteristics (+ 95% CIs) between response types/modes (web, telephone, non-response). Differences in binary variables shown as odds ratios and difference in continuous variables as difference in standardized means (standardization performed using means and standard deviation in web response groups). Figures calculated using generalised linear models (logistic or Gaussian) with multiply imputed data (m = 20) combined using Rubin's Rules (Rubin, 1987).

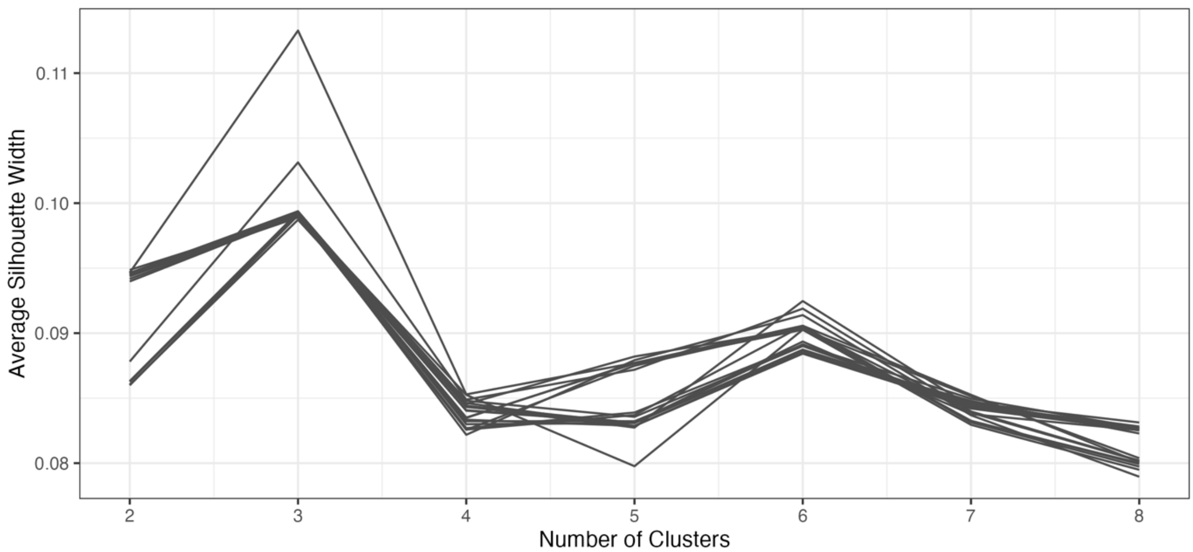

Figure S7: Elbow plot of average silhouette width for cluster analysis solutions generation using partition around medoid algorithm with Gower's distance metric by imputed dataset.

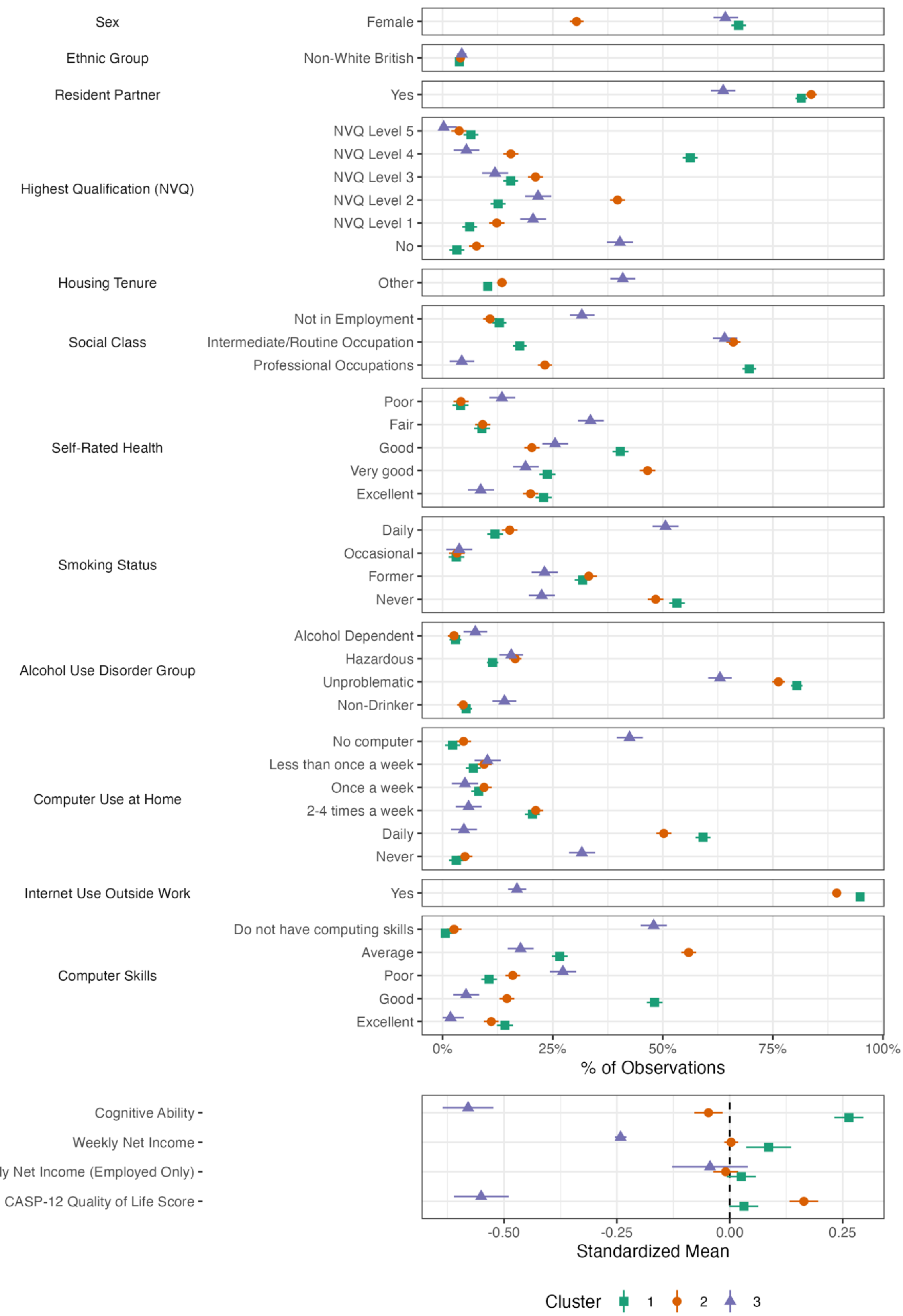

Figure S8: Average cohort member characteristics by cluster. Averages for categorical variables shown as probabilities, averages for continuous variables shown as standardized means (standardization performed using means and standard deviation in web response groups). Variables ordered (left to right) by difference between clusters. For clarity, proportions are shown for the cluster analysis performed using the first imputed dataset. Results were similar across imputations.

www.surveyfutures.net