Received: 30 September 2020 | Accepted: 3 November 2021

DOI: 10.1111/rssa.12786

ORIGINAL ARTICLE

ROYAL STATISTICAL SOCIETY
Series A Statistics in Society

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

Alissa Goodman[1] | **Matt Brown**[1] | **Richard J. Silverwood**[1] | **Joseph W. Sakshaug**[2,3,4] | **Lisa Calderwood**[1] | **Joel Williams**[5] | **George B. Ploubidis**[1]

[1]Centre for Longitudinal Studies, UCL Social Research Institute, University College London, UK

[2]Institute for Employment Research (IAB), Nuremberg, Germany

[3]LMU Munich, Munich, Germany

[4]University of Mannheim, Mannheim, Germany

[5]Kantar, London, UK

Correspondence
Richard J. Silverwood, Centre for Longitudinal Studies, UCL Social Research Institute, 20 Bedford Way, London, WC1H 0AL, UK.
Email: R.Silverwood@ucl.ac.uk

Funding information
Economic and Social Research Council, Grant/Award Number: ES/M001660/1

Abstract

A sequential mixed-mode data collection, online-to-telephone, was introduced into the National Child Development Study for the first time at the study's age 55 sweep in 2013. The study included a small experiment, whereby a randomised subset of study members was allocated to a single mode, telephone-only interview, in order to test for the presence of mode effects on participation and measurement. Relative to telephone-only, the offer of the Web increased overall participation rates by 5.0 percentage points (82.8% vs. 77.8%; 95% confidence interval for difference: 2.7% to 7.3%). Differences attributable to mode of interview were detected in levels of item non-response and response values for a limited number of questions. Most notably, response by Web (relative to telephone) was found to have increased the likelihood of non-response to questions relating to pay and other financial matters, and increased the likelihood of 'less desirable' responses. For example, response by Web resulted in the reporting of more units of alcohol consumed, and more negative responses to subjective

questions such as self-rated health, self-rated financial status and well-being. As there was evidence of mode effects, there is the potential for biases in some analyses, unless appropriate techniques are utilised to correct for these.

KEYWORDS

longitudinal birth cohort study, mixed-mode data collection, mode effects, non-response, web survey

1 | INTRODUCTION

The availability of the Web for the collection of survey data raises important questions for longitudinal studies, which need to balance potentially conflicting priorities including maximising participation, and the quality and longitudinal integrity of the collected data, while at the same time minimising participant burden and costs. As technologies for facilitating online data collection of complex survey instruments have improved and reduced in cost, a growing number of studies, both large and small, now incorporate online data collection into their designs, a trend that was accelerated during the pandemic where face-to-face data collection was not possible. In this paper, we describe and evaluate the introduction of data collection by Web, via a sequential mixed-mode Web-to-telephone approach, that was adopted in one of Britain's renowned national birth cohort studies, the National Child Development Study (NCDS), at its age 55 sweep in 2013. This was the first birth cohort study in the United Kingdom to use online data collection as a primary tool in one of its data collection sweeps.

NCDS is a national longitudinal study which takes as its subjects ('cohort members') all those living in England, Scotland and Wales who were born in a single week in 1958 (Power & Elliott, 2006). Cohort members have been periodically interviewed as part of the study since 1958, with the ninth follow-up in 2013, when they were aged 55. Historically, the data collection mode for NCDS has been face-to-face, except for the study's age 46 sweep (in 2004) which was conducted by telephone. By contrast, the 2013 survey adopted a sequential mixed-mode design (online, followed by telephone). This was the first time in the history of the cohort that a mixed-mode design had been adopted, and the first time that online data collection had been used.

The primary motivation for the introduction of the mixed-mode was to reduce costs. However, there were other positive reasons to offer the Web to participants, including optimism about response rates, driven by evidence from other studies in the United Kingdom that have shown that those aged between 50 and 65 are the most likely to respond to requests to complete a survey online (Fong & Williams, 2011; Wood & Kunz, 2014). There were also perceived limitations of the alternative interview mode available to the study at this sweep, which was by telephone (telephone was the only alternative option due to budget constraints). The age 55 survey was also to be relatively short: a 30-min survey as opposed to the 60 min or longer typical of face-to-face sweeps, and this was considered likely to encourage greater uptake of the online option. The cohort member was to be the sole respondent, avoiding complications arising from introducing mixed-mode approaches within multiple respondent settings (Jäckle et al., 2015). Finally, the flexibility and convenience offered to study respondents were also seen as positive. Possible drawbacks included the fact that mixed-mode designs may lead to so called 'mode effects', in which differences in

survey responses arise simply from differences in the mode of data collection. Such mode effects can cause biases in analyses if not dealt with adequately by researchers, and consequently can create additional analytical complexity for potential users.

The introduction of the mixed-mode Web-to-telephone approach was an important methodological innovation in the study, and so a key priority was to build in mechanisms that would enable the effectiveness of the sequential mixed-mode approach to be fully and robustly assessed. Of particular interest were evaluations of the effects of the offer of the mixed-mode on overall (wave specific) response rates, on the final composition of the sample, and on the extent of mode effects in item response and item values. To this end, a random subgroup of around 1 in 7 members of the NCDS issued sample were allocated directly to the telephone as a single mode, rather than to the sequential mixed-mode. This embedded experiment enabled an evaluation of how the sequential mixed-mode approach compared to the counterfactual of a telephone-only study design on these dimensions.

In this paper, we provide a first assessment of the success of the sequential mixed-mode approach adopted in NCDS, based on the results from the embedded experiment. Although the experiment is embedded within a sweep of a long-running cohort study, our focus in the present paper is on the impact of using the Web within that sweep only (i.e. cross-sectional rather than explicitly longitudinal implications). The structure of the paper is as follows. First, we provide a literature review and outline our research questions. Next, we describe the NCDS study and provide further details of the sequential mixed-mode design, describing the experiment that was embedded within it and setting out the methodology used in the evaluation. We then provide evidence on the balance of the samples in the mixed-mode and telephone-only arms of the random assignment, consider response rates to the survey in both groups, and set out the characteristics of responders of different types. The main findings from the randomised experiment are then presented, namely the effect of assignment to mixed-mode, and the effects of response by Web on survey response, item response and item values. Finally, we conclude by setting out potential lessons for NCDS and for other studies considering introducing the Web into their design.

2 | LITERATURE REVIEW AND RESEARCH QUESTIONS

In response to decreasing response rates and rising costs associated with implementing large-scale face-to-face surveys, longitudinal and cross-sectional surveys are making increasing use of mixed-mode data collection strategies, especially strategies which involve the Web (De Leeuw, 2018; Jäckle et al., 2017). Long running longitudinal surveys, such as the UK Understanding Society, the UK Next Steps Cohort study, the US Panel Study of Income Dynamics and the US Health and Retirement Study have begun (or are planning) to use Web in a mixed-mode design. This trend accelerated during the pandemic where face-to-face data collection was not possible, with a particularly relevant example being the series of coordinated COVID-19 surveys conducted in NCDS and other UK national longitudinal studies (Brown et al., 2021). The potential for improved response rates, reduced risk of non-response bias and cost savings are key motivations behind the shift towards mixing modes. However, evidence on the actual impact of introducing Web as part of a mixed-mode design within longitudinal surveys is limited. Our main focus lies with sequential mixed-mode designs (as opposed to concurrent designs), which deploy multiple modes of data collection in a specified order. Sequential mixed-mode designs can be cost-effective when they start with a less expensive mode, such as Web, and switch to a more expensive mode, such

as telephone or face-to-face, for non-response follow-up (Hochstim, 1967; McHorney et al., 1994; McMorris et al., 2009; Siemiatycki, 1979; Wagner et al., 2014).

Although Web surveys tend to produce lower response rates than other modes (Daikeler et al., 2019; Manfreda et al., 2008), there is evidence that combining Web with an interviewer-administered mode in a sequential mixed-mode design can produce higher response rates relative to an otherwise equivalent design without Web (Elliott et al., 2009; Greene et al., 2008; Kappelhof, 2015; Sakshaug et al., 2019). However, this result has not been replicated in the few experiments implemented within large-scale longitudinal studies. For instance, an experimental mode design study implemented in the Understanding Society Innovation Panel wave 5 found that sample members assigned to a sequential mixed-mode design with Web followed by face-to-face interviews participated at a lower rate compared to sample members assigned to the unimode face-to-face design (Jäckle et al., 2015). This effect dissipated in subsequent waves as there were no differences in attrition when the same experiment was implemented in waves 6 and 7 of the Innovation Panel (Bianchi et al., 2017). The authors also reported only minimal differences in respondent composition between the two mode designs. The same experiment, implemented in wave 8 of the main Understanding Society survey, also showed no increase in attrition rates between the sequential Web-face-to-face and face-to-face designs (Carpenter & Burton, 2018). However, it should be noted that sample members assigned to the mixed-mode group were offered higher incentives than those in the unimode group. Gaia (2017) reports that this strategy was indeed effective in increasing participation in the mixed-mode group to a level that was comparable to that of the unimode group.

While introducing a sequential mixed-mode design with Web in a longitudinal study may not substantially improve response rates, there is suggestive evidence that it can yield significant cost savings. Cost savings can arise through high Web take-up rates, which preclude interviewer involvement. Bianchi et al. (2017) report an increasing share of respondents who participated via Web in the mixed-mode treatment design of waves 5 (42.7%), 6 (55.6%), and 7 (57.5%) of the Understanding Society Innovation Panel. Given that these households did not require an interviewer in the mixed-mode group, the estimated cost savings were around 10%, 14% and 23% in the respective waves after accounting for incentive costs. In the Next Steps age 25 survey, a sequential mixed-mode design with Web followed by telephone and face-to-face was implemented which resulted in about 61% of respondents participating via Web (Calderwood, 2016). The use of additional incentives for Web completion boosted Web response rates and led to some cost savings due to fewer cases being issued to interviewer-administered modes, though savings were a very small proportion of total fieldwork costs.

Despite their purported cost savings, mixed-mode strategies that involve the Web are susceptible to data quality issues, including item non-response and differential measurement error. Item non-response tends to be higher in Web and other self-administered modes than in interviewer-administered modes (de Leeuw, 2005; Greene et al., 2008; Heerwegh & Loosveldt, 2008; Heerwegh, 2009; Hope et al., 2014; Scott et al., 2011). Consequently, adding Web to an otherwise interviewer-administered design has the potential to increase item non-response. Jäckle et al. (2015) report significantly higher rates of 'don't know' and refusals under the mixed-mode design in wave 5 of the Innovation Panel. Across 1,055 items, the average item non-response rate was about 65% higher in the mixed-mode group than in the face-to-face group.

On the measurement error side, it is well-known that mode can influence the way in which people answer survey questions (De Leeuw, 2005). That is, respondents might give different answers to the same question depending on their mode of interview (Jäckle et al., 2010). For example, it is well known that respondents interviewed in self-administered modes provide fewer

socially desirable responses (Greene et al., 2008; Heerwegh, 2009; Kreuter et al., 2008; Laaksonen & Heiskanen, 2014) and contribute less positivity bias (Hope et al., 2014; Ye et al., 2011) compared to respondents interviewed in interviewer-administered modes. Survey modes are also susceptible to presentation effects. For instance, self-administered modes tend to elicit more primacy effects due to their visual presentation while interviewer-administered modes are more prone to recency effects due to the aural administration of the questionnaire items (Krosnick & Alwin, 1987). Moreover, complex questions involving detailed instructions or definitions may be challenging to administer in self-administered modes due to lack of interviewer support. All of these mode-related measurement effects could lead to potential differences in response distributions between mixed-mode and unimode designs. Only Jäckle (2016) has explored this issue experimentally, finding differences for about 3% of items collected under the Web-face-to-face and face-to-face treatment groups in Understanding Society.

The paucity of experimental evidence on the effects of switching to a mixed-mode design involving Web in a longitudinal study represents a clear research gap in the literature. Several open questions remain regarding the impact of using Web in conjunction with interviewer-administered modes. For example, the reviewed literature suggests that introducing a Web-face-to-face design in a longitudinal study may have a negative effect on participation and item non-response rates, and could come as a shock to panel members who have grown accustomed to being interviewed face-to-face. Whether this finding is consistent across other studies, involving different mode combinations (e.g. Web-telephone), is unclear. Furthermore, it is unclear the extent to which a mixed-mode design with Web yields different response distributions compared to a unimode design without Web. While introducing Web to an interviewer-administered survey is expected to reduce social desirability bias, we do not know whether such reductions apply in a sequential mixed-mode design, or are counter balanced by other factors that influence responses under a given mode design (e.g. selection effects, aural vs. visual presentation).

To shed further light on these issues, we make use of a mixed-mode design experiment implemented in the NCDS. We use these data to address the following research questions:

1. Does introducing a sequential Web–telephone design yield a similar (or higher) response rate, relative to a telephone-only design? Is the likelihood of participation in the Web–telephone design similar across respondent subgroups?
2. Does the Web–telephone design result in higher rates of item non-response relative to the telephone-only design?
3. To what extent are survey responses affected by introducing a Web-telephone design relative to the telephone-only design? Is there evidence that social desirability bias is reduced under the mixed-mode design?

3 | METHODS

3.1 | The NCDS study and the sequential mixed-mode design in its age 55 sweep

The NCDS is an ongoing multidisciplinary cohort study of all babies born in Great Britain in a single week in 1958 (Power & Elliott, 2006). The initial birth survey was conducted by midwives in hospitals across Great Britain, and participants have subsequently been followed up at 7, 11,

16, 23, 33, 42, 44, 46, 50 and 55 years of age, with an age 62-63 sweep ongoing and three COVID-19 surveys conducted in 2020–2021 (Brown et al., 2021). The initial sample of 17,415 individuals was augmented during childhood by immigrants into Great Britain, with a resulting total sample of 18,558. The initial sample at birth contained 98.1% of all babies born in Great Britain in the study week, and even after more than 5 decades, retention remains very high, with 9,137 study members taking part at the age 55 sweep in 2013.

From the original focus on the circumstances and outcomes of birth, the study broadened in scope to map all aspects of health, education and social development as the cohort passed through childhood and adolescence, while in adult life the information collected has covered education and training, labour market activity, housing, family formation, income, health and well-being. At a biomedical sweep at age 44, physical measurements and biological samples were also taken. Most previous sweeps of the study have been conducted face-to-face by interviewers in the cohort members' homes. One exception to this was the age 46 sweep in 2004, which was a telephone interview. Initial plans for the age 55 sweep were to conduct a telephone interview. However, in order to cut costs, and in consideration of some potential benefits to the study of adopting an online approach, the study opted for a sequential Web-to-telephone design.

The sequential mixed-mode design was implemented as follows: initially, all cohort members were asked to complete the questionnaire online. Non-responders (after 6 weeks and three letters/emails) were contacted by telephone (where possible) and asked to do a telephone interview instead. When designing the survey every effort was made to ensure equivalence between the Web and computer-assisted telephone interviewing (CATI) instruments, drawing extensively on the Unimode design principles set out by Dillman et al. (2009). In addition, the great majority of the content of the survey was factual in nature, and such questions are generally acknowledged as being less prone to mode effects (e.g. Lozar & Vehovar, 2002; Schonlau et al., 2003). A full account of the design decisions taken when developing the Web and telephone questionnaires is provided by Brown (2016). The resulting questionnaire was approximately 30 min long and covered household composition, housing, economic activity, qualifications, help and care provided to parents and grandchildren, earnings, income and housing wealth, retirement plans and pensions, self-reported health and health conditions, smoking, drinking, well-being, and the updating of job and partnership event histories.

While the large majority of cohort members were allocated to the mixed-mode protocol, a subset of cohort members was randomly allocated to a single mode, telephone-only protocol, which we describe below.

3.2 | Measuring mode effects: the mixed-mode experiment, and methodology for its evaluation

The use of different data collection modes both over time (in a longitudinal study), and within a study sweep using a sequential mixed-mode approach, as adopted by NCDS at age 55, introduces the possibility of mode effects in the data. Mode effects are present if responses to items differ across individuals, or within individuals over time, solely due to the mode in which the response is given, rather than due to differences in the underlying constructs which the questionnaires or other data collection instruments are designed to capture. Where different mediums are adopted at different data collection sweeps, mode effects may occur longitudinally, i.e. they may affect the measurement of change over time for the same individuals. While these may be important, we do not consider evidence for these in this paper. The sequential mixed-mode design adopted for the

age 55 sweep of NCDS also raises the possibility of mode effects occurring cross-sectionally, within a given sweep. Here mode effects may affect the measurement of differences between individuals at a point in time, since by design individuals within the same survey provide their responses by different modes, with each individual choosing just one of the possible response modes offered. If such mode effects exist and are not dealt with through appropriate statistical methods, they may lead to biases in analyses.

Typically within sequential mixed-mode settings, it is difficult to detect and to correct for such mode effects robustly, simply because it may be impossible to distinguish differences in responses between individuals that are due to *measurement*, and those that are due to *selection*, the latter occurring because individuals choose (or 'select into') their mode of response. For example, it has been demonstrated in this and other sequential mixed-mode data collection contexts, that those choosing to answer by Web are wealthier, more likely to live with a partner and, unsurprisingly, more likely to be regular Web users (Wood & Kunz, 2014) than those opting to respond in subsequent modes. In NCDS, such observable differences are captured in data drawn from the long history of participants' prior participation in the study and thus can potentially be controlled for. However, even with a rich set of prior controls, there may yet be further unobserved differences between individuals choosing between response modes that make attributing differences in responses to either measurement or selection difficult.

In order to investigate fully the extent of mode effects within the age 55 sweep, and furthermore to be able to assess the extent to which any mode effects detected may bias analyses and to enable users to robustly correct for any such biases, a random subset of cohort members was therefore allocated to a telephone-only data collection protocol. We refer to this design as the 'mixed-mode experiment'. In the main stage of data collection at the age 55 sweep of NCDS a total of 11,553 addresses were issued to the fieldwork agency. The experiment included 10,586 cohort members with UK telephone numbers (thus the experiment excluded 967 cohort members, most of whom were emigrants from the UK who were allocated to a Web-only protocol, and others with no phone number). Among those included in the experiment, 1,476 cohort members—or around 1 in 7—were allocated to the telephone-only group, with the remaining 9,110 allocated to the mixed-mode group. The proportion allocated to the telephone-only group was chosen as an adequate sample size to be able to detect any substantial mode effects that might bias inferences. This design is sufficient to detect a difference of 0.08 standard deviations for continuous variables or a maximum 3.9 percentage point difference (e.g. 53.9% vs. 50.0%, 12.5% vs. 10.0%) for binary variables (alpha = 0.05, power = 0.8).

Embedding an experiment of this type allows us to disentangle measurement from selection effects, and specifically to estimate the overall impact on outcomes of interest of employing a mixed-mode data collection approach compared to a telephone interview data collection approach (called the 'intention to treat' effect within the evaluation literature). We can also, under some credible assumptions (spelled out further below), estimate the impact of responding by Web for the subgroup that completed the survey online (here referred to as the 'complier average causal effect'), which is the estimand of most interest for the understanding of mode effects.

3.3 | Methodological approach to the evaluation of mode effects

One simple methodological framework we can use to obtain estimates of these mode effects is to conceptualise the mixed-mode experiment as a *randomised experiment with one-way non-compliance* (Imbens & Rubin, 2015). The treatment of principal interest here, W_i, is defined as a response by Web ($W_i = 1$), as compared to a response by telephone ($W_i = 0$), for cohort

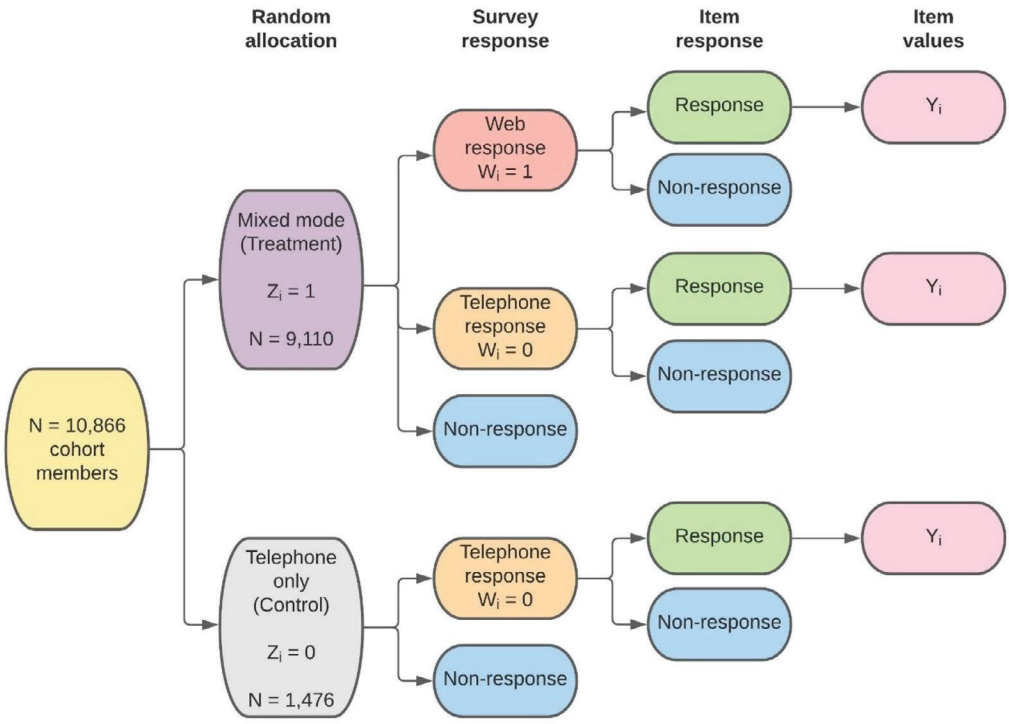

FIGURE 1 Structure of the mixed-mode experiment

member $i = 1, \ldots, N$. We are interested in the causal effect of W_i on outcomes of interest, denoted Y_i. Outcomes of interest for us include overall participation in the study sweep, as well as the individual item responses (as opposed to non-responses) in the sweep, and item values. Additionally, the embedded mode experiment gives us an instrument, Z_i, which is defined as the random assignment to the mixed-mode treatment group ($Z_i = 1$), rather than the telephone-only group ($Z_i = 0$). This instrument is a priori known to have a causal effect on W_i (since only those in the mixed-mode group can respond by Web). The structure of the experiment and the outcomes to be evaluated are illustrated further in Figure 1.

In any experimental setting where individuals are randomised into a treatment group, some may choose to comply with their allocated treatment (compliance), while others may not (non-compliance). In our context, those who are randomly assigned to the mixed-mode group may respond by Web or by telephone to the survey. Compliers can be thought of as those who respond by Web, while those choosing the telephone option within the mixed-mode group are non-compliers. Among those randomly assigned to the telephone-only group, the only response option is by telephone and hence it is not possible to observe non-compliance. Overall, observed non-compliance in this context is therefore one-way only. Since compliance with Web response in the mixed-mode arm is voluntary rather than enforced, the group of compliers is a selected group, and despite the experimental setting, additional assumptions are required to estimate the effect of the treatment on outcomes of interest.

Within this framework, two different treatment effects on outcomes Y_i can be clearly uncovered. The first is a simple intention to treat (ITT) estimand:

$$\mathrm{ITT}_Y = E(Y(Z = 1) - Y(Z = 0)),$$

where $Y(Z = a)$ is the potential outcome (value of Y) when Z is equal to a. This ITT effect is not the effect of the treatment, but only assignment. As such, this captures the effects of the offer of the mixed-mode, relative to a counterfactual of a telephone-only survey design, on outcomes of interest. The main drawback of this ITT analysis is that it does not answer questions about causal effects of Web response itself, only about causal effects of the overall assignment to the mixed-mode group.

One important insight given by Imbens and Ruben is that this overall ITT effect (ITT_Y) can be understood as consisting of two parts, the overall treatment effect on the compliers (those who would have responded by Web if invited to respond by Web and by telephone if not invited to respond by Web) and the non-compliers (all other individuals), weighted by their population proportions.

The second estimand that we consider is the complier average causal effect (CACE), defined as the treatment effect among the compliers:

$$CACE_Y = E(Y(Z = 1) - Y(Z = 0) \mid \text{complier}).$$

This estimand, under a plausible assumption (the 'exclusion restriction'), provides us with an estimate of the average causal effect of responding by Web (compared to by telephone) among the compliers. The exclusion restriction implies that allocation to treatment group (mixed-mode vs. telephone-only) only has an effect on the outcome through its effect on the treatment actually taken (survey participation by Web vs. telephone). This means that among those in the mixed-mode group who chose to answer by telephone the offer of the Web did not affect the answers that they gave in the telephone survey. While we cannot directly test this assumption, this seems at face value to be a reasonable assumption to make.

While the CACE therefore does not apply to the whole population, it does allow us to move beyond examining the effect of treatment assignment to consider the effect of the treatment actually taken. If interest is in the choice of survey design (mixed-mode vs. telephone-only) then the ITT estimand may be the substantively most interesting comparison, but the CACE may be the more relevant if the target of investigation is the causal effect of mode on item non-response and item values. Presenting both ITT and CACE estimands allows a more thorough exploration of the topic.

Given the experimental nature of our setting, the estimator for the ITT_Y effect on overall survey participation is simply the difference in the expected value of Y between the two randomised groups:

$$E(Y \mid Z = 1) - E(Y \mid Z = 0).$$

Non-response to individual items (see Figure 1) is a further option for survey respondents among both the mixed-mode and telephone-only groups, which means that the experimental comparison between the two groups may be violated when examining item responses and individual survey items. We examined the extent of such selection into response by comparing the wave responders and non-responders in each experimental group in terms of key pre-treatment characteristics (listed in Table 1). Subsequently, we controlled for the same pre-treatment characteristics in analyses examining item responses and individual survey items. Due to the necessity to control for pre-treatment characteristics, ITT effects were in practice estimated using regression approaches and CACEs using instrumental variable regression approaches.

TABLE 1 Respondent profiles—all mixed-mode vs all telephone-only (n = 10,586 cohort members randomised into experimental groups)

	Number (%) non-missing in observed data	Number non-missing in MI analysis	Mean		Mean difference		
			Mixed-mode ($Z = 1$) (n = 9110)	Telephone-only ($Z = 0$) (n = 1476)	Estimate	95% CI	*p*-value
Whether participated at 50	10,586 (100.0)	10,586	0.88	0.88	0.00	−0.02, 0.02	0.93
Whether email provided to study	10,586 (100.0)	10,586	0.72	0.72	0.00	−0.02, 0.03	0.94
Computer use—age 50							
Computer at home	10,327 (97.6)	10,586	0.89	0.89	−0.01	−0.03, 0.01	0.31
Use home computer >2 times per week	10,327 (97.6)	10,586	0.62	0.61	0.01	−0.02, 0.04	0.56
Personal access to internet	10,327 (97.6)	10,586	0.76	0.77	−0.01	−0.03, 0.02	0.60
Computer skills excellent or good	10,327 (97.6)	10,586	0.34	0.35	−0.02	−0.04, 0.01	0.24
Demographics—age 50							
Sex	10,586 (100.0)	10,586	0.49	0.49	0.00	−0.03, 0.03	0.93
White British	9,327 (88.1)	10,586	0.96	0.96	0.00	−0.01, 0.02	0.45
Homeowner	9,327 (87.7)	10,586	0.82	0.84	−0.02	−0.04, 0.01	0.17
In work	9,319 (88.0)	10,586	0.84	0.84	0.00	−0.02, 0.02	0.95
Professional/managerial occupation	7,918 (74.8)	–[†]	0.44	0.44	0.00	−0.03, 0.04	0.80
Weekly net pay	8,932 (84.4)	–[‡]	404	404	0	−50, 50	1.00
Living with partner	9,307 (87.9)	10,586	0.79	0.78	0.00	−0.02, 0.03	0.78
Has degree	9,328 (88.1)	10,586	0.33	0.35	−0.02	−0.05, 0.01	0.20

TABLE 1 (Continued)

	Number (%) non-missing in observed data	Number non-missing in MI analysis	Mean		Mean difference		
			Mixed-mode ($Z = 1$) (n = 9110)	Telephone-only ($Z = 0$) (n = 1476)	Estimate	95% CI	*p*-value
Health and health behaviours-age 50							
Fair or poor health	9,294 (87.8)	10,586	0.19	0.18	0.02	−0.01, 0.04	0.17
Regular smoker	9,296 (87.8)	10,586	0.24	0.23	0.01	−0.01, 0.04	0.38
Problematic drinker	8,595 (81.2)	10,586	0.19	0.17	0.02	0.00, 0.04	0.11
CASP−12 Quality of Life Score	8,257 (78.0)	10,586	25.9	26.1	−0.3	−0.6, 0.1	0.14
Cognitive function-age 50							
Animal naming test score	9,221 (87.1)	10,586	22.1	22.2	−0.2	−0.5, 0.2	0.39
Word list recall score	9,221 (87.1)	10,586	6.5	6.5	−0.1	−0.1, 0.0	0.20
Delayed word list recall score	9,167 (86.6)	10,586	5.4	5.4	−0.1	−0.2, 0.0	0.16
Letter cancellation speed score	9,029 (85.3)	10,586	25.9	26.2	−0.3	−0.7, 0.2	0.27
Letter cancellation accuracy score (low score=greater accuracy)	9,029 (85.3)	10,586	4.5	4.6	−0.1	−0.4, 0.2	0.46

Notes: CI, Confidence interval; MI, Multiple imputation.

[†]'Professional/managerial employment' is only defined when the variable 'In work' suggests that the cohort member is in employment. Since 'In work' is itself imputed, the number of cohort members for which 'Professional/managerial occupation' is defined differs by imputed dataset between 8,867 and 8,934.

[‡]'Weekly net pay' is only defined when the variable 'In work' suggests that the cohort member is in employment and the cohort member has non-zero reported weekly pay. Since 'In work' is itself imputed, the number of cohort members for which 'Weekly net pay' is defined differs by imputed dataset between 7,593 and 7,659.

Furthermore, since pre-treatment characteristics were not observed for all cohort members included in the mixed-mode experiment, conducting a complete case analysis would have reduced the analysis sample size and potentially introduced bias. We therefore utilised a multiple imputation (MI) approach in which the imputation model included the experimental group assignment and wave response status (Web vs. telephone vs. non-response) at age 55, and the pre-treatment characteristics to be used as control variables. Further variables known to be predictive of wave non-response (social class at birth, Rutter behavioural score at age 7, social environment at age 7, region of residence at age 11, special educational needs at age 11, English ability at age 11, general ability test score at age 11) were included as auxiliary variables (Mostafa et al., 2021). We used MI by chained equations and generated 20 imputed datasets. For analyses concerning item non-response, a different imputation model was used which additionally included the item non-response indicators for the variables of interest (i.e. the outcome variables in the corresponding analysis models) to inform the imputation of the missing covariate values, though imputed values of the indicators were not used in the analyses. Similarly, for analyses concerning mode effects on measurement, the imputation model also included the outcome variables whose measurement we were subsequently analysing, though again imputed values of these variables were not used in the analyses.

All analyses were conducted using Stata version 15 (StataCorp, College Station, TX).

4 | RESULTS

4.1 | Balance of experimental groups

The balance of the experimental groups was examined by comparing a set of pre-treatment characteristics taken from the age 50 NCDS sweep. Table 1 shows the means of these pre-treatment characteristics by randomly assigned experimental group status. There was no evidence of differences between the mixed-mode and telephone-only groups, as would be expected given randomisation and a relatively large sample.

4.2 | Survey response, compliance status, and complier/non-complier and responder/non-responder characteristics

For the total issued sample for the mainstage fieldwork of 11,553, Table 2 presents the issued sample and response rates, according to whether cohort members were part of the experiment—and hence allocated to either the mixed-mode or telephone-only groups—or were not included in the experiment. Assignment to the mixed-mode generated a higher response rate (82.8%) compared to the telephone-only survey protocol (77.8%). The table also shows response rates for the group that (for various reasons) were not included in the experiment, and who were allocated to a Web-only protocol. This shows a low response rate particularly among cohort members who are not known to have emigrated but for whom no valid UK telephone number was held.

Table 3 provides information about compliance to the treatment by showing the overall proportion of those allocated to the mixed-mode who chose to respond by Web. In total, 5,612 of the 9,110 cohort members allocated to mixed-mode data collection completed the questionnaire online (61.6%). A further 1,935 completed interviews by telephone, representing an additional 21.2% of the mixed-mode group. Web-responders represent 74.4% of

TABLE 2 Survey response

	Experiment (n = 10,586)		Excluded from experiment (n = 967)			
	Mixed-mode ($Z = 1$)	Telephone-only ($Z = 0$)	Emigrants (no UK telephone number)	UK Cases with no telephone number	Other	Total
Issued	9,110	1,476	367	572	28	11,553
Interviewed	7,547	1,149	194	44	25	8,959
Response Rate	82.8%	77.8%	52.9%	7.7%	89.3%	77.5%

TABLE 3 Compliance status—response by Web or telephone (n = 10,586 cohort members randomised into experimental groups)

	Mixed-mode ($Z = 1$) (n = 9,110)			Telephone-only ($Z = 0$) (n = 1,476)	
	Responded			Responded by Telephone	
	Web ($W = 1$)	Telephone ($W = 0$)	Non-response	($W = 0$)	Non-response
N	5612	1935	1563	1149	327
% of all in group	61.6%	21.2%	17.2%	77.8%	22.2%
% of responders	74.4%	25.6%	–	100%	–

all responders in the mixed-mode group—this is the population proportion of compliers, ITT_W.

Among those allocated to the mixed-mode approach, the pre-treatment characteristics of those who chose to complete via the Web (so-called 'compliers' with the treatment) were markedly different than those who participated via telephone—thus confirming that there is strong selection on observable characteristics into response by Web (Table 4). There was no evidence of gender difference in response by Web, but there was evidence of a difference in all other characteristics considered. Those who chose to complete by Web were more likely to have participated at age 50, to have provided the study with an email address, to have had a computer at home, to have been regular home computer users, to have had internet access and to rate their computer skills positively. In terms of socio-economic characteristics, Web completers were more likely to have been in work at age 50, to have been in professional/managerial occupations and to have had higher net earnings. Web completers were also more highly qualified, more likely to have lived with a partner at age 50, reported better health and higher levels of well-being at age 50, were less likely to smoke or have alcohol problems, and achieved higher scores in each of the four cognitive assessments administered at age 50.

A final set of comparisons in this section compares the pre-treatment profiles of responders and non-responders within the mixed-mode and telephone-only groups (Table 5). Despite the fact that response to the survey was higher among the mixed-mode than the telephone-only group, there was little evidence of differences between the profiles of the responders in the two groups. There was some evidence that respondents in the telephone-only group were very slightly more likely to have participated in the age 50 sweep than respondents in the mixed-mode group (95%

TABLE 4 Respondent profiles—mixed-mode Web vs mixed-mode telephone (n = 7,547 mixed-mode respondents)

		Mean		Mean difference		
	Number non-missing	Mixed-mode Web ($Z = 1, W = 1$) (n = 5,612)	Mixed-mode telephone ($Z = 1, W = 0$) (n = 1,935)	Estimate	95% CI	*p*-value
Whether participated at 50	7,547	0.95	0.90	0.04	0.03, 0.06	<0.001
Whether email provided to study	7,547	0.92	0.45	0.47	0.44, 0.49	<0.001
Computer use—age 50						
Computer at home	7,547	0.95	0.79	0.16	0.14, 0.18	<0.001
Use home computer >2 times per week	7,547	0.74	0.43	0.32	0.29, 0.34	<0.001
Personal access to internet	7,547	0.88	0.59	0.29	0.27, 0.31	<0.001
Computer skills excellent or good	7,547	0.43	0.20	0.24	0.21, 0.26	<0.001
Demographics—age 50						
Sex	7,547	0.49	0.49	0.00	−0.03, 0.02	0.82
White British	7,547	0.96	0.95	0.01	0.00, 0.02	0.03
Homeowner	7,547	0.89	0.74	0.15	0.13, 0.17	<0.001
In work	7,547	0.88	0.79	0.09	0.07, 0.11	<0.001
Professional/managerial occupation	–[†]	0.52	0.32	0.21	0.18, 0.23	<0.001
Weekly net pay	–[‡]	441	338	103	48, 158	<0.001
Living with partner	7,547	0.84	0.73	0.11	0.09, 0.14	<0.001
Has degree	7,547	0.41	0.23	0.18	0.15, 0.20	<0.001

TABLE 4 (Continued)

		Mean		Mean difference		
	Number non-missing	Mixed-mode Web ($Z = 1, W = 1$) (n = 5,612)	Mixed-mode telephone ($Z = 1, W = 0$) (n = 1,935)	Estimate	95% CI	*p*-value
Health and health behaviours—age 50						
Fair or poor health	7,547	0.14	0.27	−0.13	−0.15, −0.11	<0.001
Regular smoker	7,547	0.18	0.31	−0.13	−0.16, −0.11	<0.001
Problematic drinker	7,547	0.17	0.21	−0.04	−0.07, −0.02	<0.001
CASP-12 Quality of Life Score	7,547	26.6	25.0	1.6	1.3, 1.9	<0.001
Cognitive function—age 50						
Animal naming test score	7,547	23.0	20.9	2.1	1.7, 2.4	<0.001
Word list recall score	7,547	6.7	6.2	0.5	0.4, 0.6	<0.001
Delayed word list recall score	7,547	5.6	5.0	0.6	0.5, 0.7	<0.001
Letter cancellation speed score	7,547	26.2	25.4	0.8	0.4, 1.2	<0.001
Letter cancellation accuracy score (low score=greater accuracy)	7,547	4.2	4.9	−0.8	−1.0, −0.6	<0.001

Notes: CI, Confidence interval.

[†]'Professional/managerial employment' is only defined when the variable 'In work' suggests that the cohort member is in employment. Since 'In work' is itself imputed, the number of cohort members for which 'Professional/managerial occupation' is defined differs by imputed dataset between 6,475 and 6,511.

[‡]'Weekly net pay' is only defined when the variable 'In work' suggests that the cohort member is in employment and the cohort member has non-zero reported weekly pay. Since 'In work' is itself imputed, the number of cohort members for which 'Weekly net pay' is defined differs by imputed dataset between 5,538 and 5,573.

TABLE 5 Responders versus non-responders in the mixed-mode and telephone-only groups (n = 10,586 cohort members randomised into experimental groups)

		Mean				Mean difference (1)–(3)		
		Mixed-mode ($Z = 1$) (n = 9,110)		Telephone-only ($Z = 0$) (n = 1,476)				
	Number non-missing	(1) Responders ($W = 0,1$) (n = 7547)	(2) Non-responders ($W \neq 0,1$) (n = 1563)	(3) Responders ($W = 0$) (n = 1149)	(4) Non- responders ($W \neq 0$) (n = 327)	Estimate	95% CI	*p*-value
Whether participated at 50	10,586	0.94	0.62	0.95	0.63	−0.02	−0.03, 0.00	0.02
Whether email provided to study	10,586	0.80	0.36	0.79	0.46	0.00	−0.02, 0.03	0.84
Computer use—age 50								
Computer at home	10,586	0.91	0.77	0.92	0.79	−0.01	−0.03, 0.00	0.9
Use home computer >2 times per week	10,586	0.66	0.43	0.66	0.46	0.00	−0.02, 0.03	0.77
Personal access to internet	10,586	0.81	0.54	0.82	0.58	−0.01	−0.04, 0.01	0.22
Computer skills excellent or good	10,586	0.37	0.17	0.40	0.20	−0.03	−0.06, 0.00	0.09
Demographics—age 50								
Sex	10,586	0.49	0.53	0.48	0.54	0.01	−0.02, 0.04	0.60
White British	10,586	0.96	0.95	0.96	0.94	0.00	−0.01, 0.01	0.78
Homeowner	10,586	0.85	0.67	0.87	0.73	−0.02	−0.04, 0.01	0.18
In work	10,586	0.86	0.75	0.87	0.75	−0.01	−0.03, 0.02	0.56
Professional/managerial occupation	–[†]	0.48	0.27	0.48	0.30	0.00	−0.03, 0.03	0.99
Weekly net pay	–[‡]	417	334	415	363	2	−58, 63	0.94
Living with partner	10,586	0.81	0.67	0.81	0.70	0.00	−0.02, 0.03	0.73
Has degree	10,586	0.36	0.17	0.38	0.23	−0.02	−0.05, 0.01	0.21

TABLE 5 (Continued)

		Mean				Mean difference (1)–(3)		
		Mixed-mode ($Z = 1$) (n = 9,110)		Telephone-only ($Z = 0$) (n = 1,476)				
	Number non-missing	(1) Responders ($W = 0,1$) (n = 7547)	(2) Non-responders ($W \neq 0,1$) (n = 1563)	(3) Responders ($W = 0$) (n = 1149)	(4) Non- responders ($W \neq 0$) (n = 327)	Estimate	95% CI	*p*-value
Health and health behaviours–age 50								
Fair or poor health	10,586	0.17	0.29	0.16	0.23	0.01	−0.01, 0.04	0.36
Regular smoker	10,586	0.21	0.38	0.19	0.35	0.02	−0.01, 0.04	0.17
Problematic drinker	10,586	0.18	0.26	0.15	0.25	0.03	0.00, 0.05	0.02
CASP-12 Quality of Life Score	10,586	26.2	24.3	26.4	25.1	−0.2	−0.6, 0.1	0.21
Cognitive function—age 50								
Animal naming test score	10,586	22.4	20.4	22.5	21.3	−0.1	−0.5, 0.3	0.70
Word list recall score	10,586	6.6	6.1	6.6	6.3	0.0	−0.1, 0.1	0.41
Delayed word list recall score	10,586	5.5	4.9	5.5	5.2	0.0	−0.2, 0.1	0.43
Letter cancellation speed score	10,586	26.0	25.6	26.2	26.0	−0.2	−0.7, 0.2	0.30
Letter cancellation accuracy score (low score=greater accuracy)	10,586	4.4	5.1	4.5	5.0	−0.1	−0.4, 0.2	0.43

Notes: CI, Confidence interval.

[†]'Professional/managerial employment' is only defined when the variable 'In work' suggests that the cohort member is in employment. Since 'In work' is itself imputed, the number of cohort members for which 'Professional/managerial occupation' is defined differs by imputed dataset between 8,890 and 8,931.

[‡]'Weekly net pay' is only defined when the variable 'In work' suggests that the cohort member is in employment and the cohort member has non–zero reported weekly pay. Since 'In work' is itself imputed, the number of cohort members for which 'Weekly net pay' is defined differs by imputed dataset between 7,616 and 7,655.

vs. 94%) and that mixed-mode respondents were more likely to have been defined as problematic drinkers at age 50 (18% vs. 15%), though in the context of multiple testing these results should not be over-interpreted. Taken as a whole, these results suggest that the two data collection strategies did not affect the overall balance of observable characteristics of those who chose to participate. This is important since it gives some face validity to the experimental comparisons of item non-response and item values between responders across the two treatment arms.

4.3 | Survey participation – ITT analysis

The previously noted higher response rate among those assigned to the mixed-mode group compared to the telephone-only group equates to a 5.0% point ITT effect (95% confidence interval [CI] 2.7, 7.3) (Table 6). This was little changed upon control for pre-treatment characteristics to 5.3 (95% CI 3.3, 7.3), as would be expected given randomisation and the observed balance of these covariates between the experimental groups.

4.4 | Item non-response – ITT and CACE analyses

Comparisons of some item non-response rates are shown in Table 7. For some items, there was strong evidence of differences between the mixed-mode group and the telephone-only group, but for many items no difference was apparent. Where differences occurred it was typically the case that higher item non-response rates were found among the mixed-mode group (as shown by the ITT estimates), and we thus conclude that Web response (relative to telephone) causes higher non-response in those items (as shown by the CACE estimates). Most of the largest differences related to variables where a numeric value had to be entered—value of home, amount left to pay off on mortgage and gross weekly pay. These questions are all fairly sensitive meaning that one might have expected that the anonymity of the Web would have led to a lower item non-response rate in the mixed-mode group than in the interviewer administered telephone-only group. However, accurately answering these questions would also require a considerable degree of cognitive effort and so it seems that telephone interviewers may have encouraged telephone respondents to provide an answer. Somewhat surprisingly, the item non-response rate on the 'type of employer provided pension' question was higher among the telephone-only group than the mixed-mode group. This is a complex question and one might have assumed that the presence of an interviewer who could potentially have provided clarification to any queries raised by the respondent would have led to a lower item non-response rate among the telephone-only group, but this clearly was not the case.

4.5 | Mode effects in item values – ITT and CACE analyses

Table 8 shows the extent of any mode differences for a selection of socioeconomic characteristics, with similar results presented for health, health behaviour, well-being, leisure and voting variables in Table 9, all of which are likely to be widely used by analysts. In terms of factual socio-economic variables, there was little evidence of mode effects. There was no difference between the mixed-mode group and the telephone-only group in terms of reporting being in work, having a professional/managerial occupation, number of hours worked per week, weekly pay (gross and net) or housing wealth. The small number of socio-economic variables where a mode

TABLE 6 Response to the survey—estimates of randomised group effects (n = 10,586 cohort members randomised into experimental groups)

		Response %		ITT analysis					
				Marginal effect of mixed-mode (no control)			Marginal effect of mixed-mode (controlling for pre-treatment characteristics[†])		
	Number non-missing	Mixed-mode ($Z = 1$)	Telephone-only ($Z = 0$)	Estimate	95% CI	*p*-value	Estimate	95% CI	*p*-value
Probability of response to survey	10,586	82.8	77.8	5.0	2.7, 7.3	<0.001	5.3	3.3, 7.3	<0.001

Notes: CI, Confidence interval; ITT, Intention to treat.

[†]Controlling for pre-treatment variables listed in Table 1.

TABLE 7 Item non-response—estimates of mode effects (n = 8,696 responders)

		Item non-response (%)		ITT analysis						CACE analysis		
				Marginal effect of mixed mode (no control)			Marginal effect of mixed mode (controlling for pre-treatment characteristics[†])					
	Potential respondents	Mixed mode Responders ($Z = 1, W = 0, 1$) (n = 7,547)	Telephone-only Responders ($Z = 0, W = 0$) (n = 1,149)	Estimate	95% CI	*p*-value	Estimate	95% CI	*p*-value	Estimate	95% CI	*p*-value
Expected value of property	7,469	12.9	7.9	5.0	3.1, 6.8	<0.001	4.9	3.1, 6.7	<0.001	6.3	3.6, 9.1	<0.001
Amount to pay off on property	3,391	15.5	11.1	4.5	1.3, 7.7	0.01	4.3	1.1, 7.4	0.01	5.4	1.0, 9.8	0.01
Gross weekly income	5,579	13.9	10.5	3.5	1.0, 5.9	0.01	3.4	1.0, 5.7	0.01	4.3	0.9, 7.7	0.01
Number of cigarettes a day usually smoked	1,206	2.3	0.0	–[‡]	–[‡]	–[‡]	–[‡]	–[‡]	–[‡]	–[‡]	–[‡]	–[‡]
Weight (kg)	8,696	7.2	5.7	1.5	0.0, 2.9	0.05	1.4	−0.1, 2.8	0.07	1.9	−0.2, 4.0	0.08
Whether voted in last general election	8,600	0.8	0.6	0.2	−0.3, 0.7	0.52	0.1	−0.4, 0.6	0.67	0.2	−0.5, 0.9	0.56
Units of alcohol consumed in last 7 days	6,595	1.7	1.6	0.1	−0.7, 1.0	0.74	0.0	−0.9, 0.9	0.94	0.0	−1.2, 1.1	0.97
Likelihood of working at the age of 60	8,649	2.2	2.0	0.1	−0.7, 1.0	0.74	0.0	−0.8, 0.9	0.93	0.1	−1.1, 1.3	0.84
Frequency of alcohol consumption	8,609	0.1	0.1	–[‡]	–[‡]	–[‡]	–[‡]	–[‡]	–[‡]	–[‡]	–[‡]	–[‡]
Likelihood of working at the age of 66	8,645	2.2	2.4	−0.2	−1.1, 0.8	0.72	−0.2	−1.1, 0.7	0.65	−0.3	−1.6, 0.9	0.58
Party voted for in 2010 general election	6,305	7.7	8.0	−0.4	−2.3, 1.6	0.72	−0.4	−2.4, 1.6	0.69	−0.5	−3.0, 2.0	0.69
Employer provided pension type	3,753	5.5	13.1	−7.6	−10.7, −4.6	<0.001	−8.0	−11.0, −5.0	<0.001	−9.7	−12.5, 7.0	<0.001

Notes: CACE, Complier average causal effect; CI, Confidence interval; ITT, Intention to treat.

[†]Controlling for pre-treatment variables listed in Table 1.

[‡]Inestimable due to very low item non-response among one or both groups.

TABLE 8 Mean estimates—mixed-mode vs telephone-only (n = 8,696 responders)

		Mean		ITT analysis						CACE analysis		
				Marginal effect of mixed-mode (no control)			Marginal effect of mixed-mode (controlling for pre-treatment characteristics[†])					
	Number non-missing	Mixed-mode responders ($Z = 1, W = 0,1$) (n = 7547)	Telephone-only responders ($Z = 0, W = 0$) (n = 1149)	Estimate	95% CI	*p*-value	Estimate	95% CI	*p*-value	CACE	95% CI	p-value
Socio-economic characteristics												
Whether working at 55	8,574	0.81	0.82	−0.01	−0.04, 0.01	0.30	−0.01	−0.03, 0.01	0.41	−0.01	−0.04, 0.01	0.38
Whether professional/managerial occupation at 55	8,696	0.35	0.35	0.00	−0.03, 0.03	0.91	0.01	−0.01, 0.04	0.30	0.02	−0.02, 0.05	0.34
Hours worked per week in main job	5,503	37.1	37.7	−0.6	−1.5, 0.4	0.27	−0.6	−1.5, 0.2	0.15	−0.8	−2.0, 0.3	0.15
Gross weekly pay (if has a job)—capped £4000	4,828	623	593	30	−18, 77	0.23	54	−11, 119	0.11	67	−11, 146	0.09
Net weekly pay (if has a job)—capped £2500	4,775	425	411	14	−12, 40	0.29	23	−16, 61	0.24	25	−19, 69	0.26
Age 55—self-rated financial situation—reverse coded	8,586	3.83	4.00	−0.17	−0.23, −0.10	<0.001	−0.15	−0.21, −0.09	<0.001	−0.20	−0.28, −0.12	<0.001
Any qualifications reported	8,604	0.15	0.14	0.01	−0.01, 0.03	0.29	0.01	−0.01, 0.04	0.21	0.02	−0.01, 0.05	0.23

(Continues)

TABLE 8 (Continued)

		Mean		ITT analysis						CACE analysis		
				Marginal effect of mixed-mode (no control)			Marginal effect of mixed-mode (controlling for pre-treatment characteristics†)					
	Number non-missing	Mixed-mode responders ($Z = 1, W = 0,1$) (n = 7547)	Telephone-only responders ($Z = 0, W = 0$) (n = 1149)	Estimate	95% CI	*p*-value	Estimate	95% CI	*p*-value	CACE	95% CI	p-value
Number of qualifications reported	8,696	0.17	0.15	0.02	0.00, 0.05	0.08	0.03	0.00, 0.05	0.06	0.04	0.00, 0.07	0.06
Likelihood of working at the age of 60	8,464	68.7	66.7	2.0	−0.3, 4.4	0.09	2.1	0.0, 4.3	0.05	2.8	0.0, 5.6	0.05
Likelihood of working at the age of 66	8,454	39.0	35.4	3.6	1.2, 6.0	0.004	3.5	1.2, 5.9	0.003	4.7	1.6, 7.8	0.003
Expected value of property	6,558	373,000	404,000	−31,000	−72,000, 10,000	0.14	−33,000	−73,000, 7000	0.10	−43,000	−94,000, 8000	0.10
Amount yet to pay off on property	2,884	121,000	106,000	15,000	−24,000, 53,000	0.46	14,000	−24,000, 53,000	0.47	18,000	−30,000, 66,000	0.47
Number of relationships reported	8,692	1.07	1.07	0.00	−0.01, 0.02	0.68	0.00	−0.02, 0.02	0.99	0.00	−0.02, 0.02	0.99
Number of addresses reported	1,566	1.39	1.35	0.04	−0.07, 0.14	0.50	0.03	−0.07, 0.14	0.53	0.05	−0.10, 0.19	0.53
Number of economic activities	2,728	1.52	1.51	0.01	−0.10, 0.12	0.90	0.00	−0.11, 0.11	0.95	0.00	−0.14, 0.15	0.95

Notes: CACE, Complier average causal effect; CI, Confidence interval; ITT, Intention to treat.

†Controlling for pre-treatment variables listed in Table 1.

TABLE 9 Mean estimates–mixed-mode vs telephone-only (n = 8,696 responders)

		Mean		ITT analysis						CACE analysis		
				Marginal effect of mixed-mode (no control)			Marginal effect of mixed-mode (controlling for pre-treatment characteristics[†])					
	Number non-missing	Mixed-mode Responders ($Z = 1, W = 0, 1$) (n = 7547)	Telephone-only responders ($Z = 0, W = 0$) (n = 1149)	Estimate	95% CI	*p*-value	Estimate	95% CI	*p*-value	CACE	95% CI	*p*-value
Health												
Self-rated general health (1 = worst health to 5 = best health)	8,612	3.33	3.43	−0.10	−0.17, −0.04	0.003	−0.07	−0.13, −0.02	0.01	−0.09	−0.17, −0.02	0.01
Whether classified as disabled	8,579	0.20	0.17	0.03	0.01, 0.06	0.004	0.03	0.01, 0.05	0.01	0.04	0.01, 0.07	0.01
Whether has a long–standing illness	8,590	0.33	0.30	0.03	0.00, 0.06	0.03	0.03	0.00, 0.05	0.04	0.04	0.00, 0.07	0.05
Number of health problems reported	8,696	1.24	1.11	0.13	0.05, 0.21	0.001	0.11	0.04, 0.18	0.002	0.15	0.06, 0.24	0.002
Weight (kg)	8,085	79.4	79.8	−0.4	−1.5, 0.8	0.51	−0.6	−1.6, 0.4	0.23	−0.8	−2.2, 0.5	0.23
Specific health problems												
Asthma or wheezy bronchitis	8,593	0.12	0.11	0.01	−0.01, 0.03	0.23	0.01	−0.01, 0.03	0.40	0.01	−0.01, 0.04	0.35
Diabetes	8,598	0.07	0.06	0.00	−0.01, 0.02	0.61	0.00	−0.01, 0.02	0.80	0.00	−0.02, 0.02	0.69
Backache, sciatica, disc prolapse	8,596	0.26	0.23	0.03	0.00, 0.05	0.04	0.03	0.00, 0.05	0.05	0.03	0.00, 0.07	0.06
Cancer or leukaemia	8,599	0.04	0.04	0.00	−0.01, 0.01	0.90	0.00	−0.01, 0.01	0.79	0.00	−0.02, 0.01	0.84
Problems with hearing	8,598	0.11	0.09	0.02	0.00, 0.04	0.03	0.02	0.00, 0.04	0.04	0.02	0.00, 0.05	0.06

(Continues)

TABLE 9 (Continued)

		Mean		ITT analysis						CACE analysis		
				Marginal effect of mixed-mode (no control)			Marginal effect of mixed-mode (controlling for pre-treatment characteristics[†])					
	Number non-missing	Mixed-mode Responders ($Z = 1, W = 0,1$) (n = 7547)	Telephone-only responders ($Z = 0, W = 0$) (n = 1149)	Estimate	95% CI	*p*-value	Estimate	95% CI	*p*-value	CACE	95% CI	*p*-value
High blood pressure	8,595	0.22	0.22	0.00	−0.02, 0.03	0.78	0.00	−0.03, 0.02	0.94	0.00	−0.04, 0.03	0.93
Heart problems	8,586	0.06	0.04	0.01	0.00, 0.03	0.04	0.01	0.00, 0.03	0.05	0.01	0.00, 0.03	0.12
Depression, emotional and psychiatric	8,590	0.16	0.13	0.02	0.00, 0.05	0.03	0.02	0.00, 0.04	0.03	0.03	0.00, 0.05	0.07
Health behaviours												
Number of units of alcohol consumed in last 7 days	6,484	10.6	8.1	2.5	1.7, 3.4	<0.001	1.9	1.2, 2.7	<0.001	2.5	1.5, 3.5	<0.001
Whether drinks most days	8,597	0.18	0.17	0.01	−0.01, 0.03	0.37	0.00	−0.02, 0.03	0.78	0.01	−0.02, 0.04	0.73
Whether regular smoker	8,597	0.14	0.13	0.01	−0.01, 0.03	0.46	−0.01	−0.02, 0.01	0.46	0.00	−0.03, 0.02	0.64
Well-being[‡]												
My age prevents me from doing the things I would like to do	8,580	3.13	3.22	−0.09	−0.15, −0.04	0.001	−0.08	−0.13, −0.02	0.004	−0.10	−0.17, −0.03	0.004
I feel what happens to me is out of my control	8,566	3.07	3.13	−0.06	−0.12, −0.01	0.03	−0.05	−0.11, 0.00	0.05	−0.07	−0.14, 0.00	0.05
I feel left out of things	8,572	3.28	3.42	−0.15	−0.20, −0.09	<0.001	−0.14	−0.19, −0.09	<0.001	−0.19	−0.25, −0.12	<0.001
I feel full of energy these days	8,584	2.91	3.00	−0.10	−0.15, −0.04	0.001	−0.08	−0.13, −0.03	0.001	−0.11	−0.17, −0.04	0.001

TABLE 9 (Continued)

		Mean		ITT analysis						CACE analysis		
				Marginal effect of mixed-mode (no control)			Marginal effect of mixed-mode (controlling for pre-treatment characteristics[†])					
	Number non-missing	Mixed-mode Responders ($Z = 1, W = 0,1$) (n = 7547)	Telephone-only responders ($Z = 0, W = 0$) (n = 1149)	Estimate	95% CI	*p*-value	Estimate	95% CI	*p*-value	CACE	95% CI	*p*-value
I feel that life is full of opportunities	8,573	3.03	3.15	−0.12	−0.17, −0.06	<0.001	−0.10	−0.15, −0.05	<0.001	−0.14	−0.20, −0.07	<0.001
I feel that the future looks good for me	8,538	3.20	3.36	−0.16	−0.21, −0.11	<0.001	−0.14	−0.18, −0.10	<0.001	−0.19	−0.24, −0.13	<0.001
Leisure												
Play sport or go walking or swimming at least once a week	8,585	0.63	0.69	−0.07	−0.09, −0.04	<0.001	−0.06	−0.09, −0.03	<0.001	−0.08	−0.12, −0.04	<0.001
Have a meal in a pub or restaurant at least once a week	8,589	0.19	0.27	−0.07	−0.10, −0.05	<0.001	−0.07	−0.10, −0.05	<0.001	−0.10	−0.13, −0.07	<0.001
Voting												
Whether voted Conservative 2010	5,818	0.37	0.40	−0.02	−0.06, 0.01	0.20	−0.02	−0.06, 0.01	0.25	−0.03	−0.07, 0.02	0.26
Whether voted Labour 2010	5,818	0.32	0.35	−0.03	−0.06, 0.01	0.12	−0.03	−0.07, 0.00	0.09	−0.04	−0.08, 0.01	0.09
Whether voted Liberal Democrat 2010	5,818	0.19	0.15	0.04	0.01, 0.07	0.01	0.04	0.01, 0.06	0.01	0.05	0.01, 0.08	0.01

Notes: CACE, Complier average causal effect; CI, Confidence interval; ITT, Intention to treat.

[†]Controlling for pre-treatment variables listed in Table 1.

[‡]Higher scores = better well-being.

difference was found tended to be based on questions which employed a rating scale and it was typically the case that telephone-only respondents responded more positively than those in the mixed-mode group. For example, telephone-only respondents gave a more positive rating of their current financial situation and reported a lower likelihood of working at the age of 66.

These relatively more positive responses among the telephone-only group were also apparent with relation to health variables. Telephone-only respondents reported better self-rated general health on average, were less likely to give responses which led them to be classified as disabled or having a long-standing illness, and tended to report a lower number of specific health problems. Mixed-mode respondents were more likely to report suffering from a subset of specific health problems, namely back problems, hearing problems and depression. There was no evidence of a difference between the mixed-mode and telephone-only group in terms of the prevalence of reporting being a regular smoker, nor in the reported frequency of consuming alcohol, but the average reported number of units of alcohol consumed in the last 7 days was higher among the mixed-mode group. The telephone-only group also reported better well-being across all six well-being variables and higher levels of different leisure activities.

In terms of voting, the sole difference was that the mixed-mode group were more likely to have reported voting Liberal Democrat in the 2010 election.

The above-described differences all persisted after adjustment for pre-treatment characteristics in the ITT analysis, albeit with occasional limited attenuation. The CACE analysis results were in line with the ITT findings.

5 | CONCLUSIONS

The higher response rate obtained for the sample allocated to the sequential mixed-mode data collection protocol (82.8% vs. 77.8% in the telephone-only group) is a clear benefit to the adoption of the mixed-mode approach. The higher than expected participation by Web, at over 60%, showed that the majority of the cohort was willing (and indeed preferred) to complete the questionnaire online rather than respond by telephone, a finding subsequently confirmed by the participation of the cohort in the Centre for Longitudinal Studies COVID-19 Web Surveys during the pandemic (Brown et al., 2021).

Online and telephone data quality appears to be comparable in the majority of cases—as evidenced by the lack of detected mode effects in most variables collected in the study sweep. However, there were some clear exceptions to this—most notably in self-assessed ratings of financial status and health, self-reports of a number of health conditions and indices of well-being, and in item non-response to financial variables such as income/earnings and wealth, all of which are heavily used survey items of central interest to many users of the study. Where mode effects have been detected, the inclusion of an experimental element to the survey has enabled us identify exactly which variables present mode effects, as well as allowing the potential for robust methodologies to correct for these.

One motivation for introducing the mixed-mode design with the age 55 sweep of NCDS was due to the potential for cost savings. While we did not collect any experimental data that would robustly enable us to assess whether adopting the sequential mixed-mode Web-to-telephone approach relative to a short telephone-only survey saved costs, indicative evidence from tenders received to a competitive tendering exercise for fieldwork suggests that there was little difference in cost between a Web-to-telephone approach and telephone-only. In part, this was due to high initial development costs for the Web, meaning that Web costs would most likely be relatively lower in the future.

There were many strengths to this study. The randomised design allowed us to robustly examine several outcomes of interest, namely wave response, item non-response and mode effects in item values. The use of two different estimators, the ITT and CACE, allowed us to estimate both the effects of the offer of the mixed-mode protocol relative to a counterfactual of a telephone-only survey design and the causal effect of answering by Web relative to answering by telephone among the compliers. By conducting the analysis in a long-running cohort study we were able to incorporate a wide range of potentially important pre-treatment characteristics in the study. Missing data in pre-treatment characteristics was robustly handled using a series of analysis-specific imputation models within a MI framework (Mostafa et al., 2021).

There were also a number of limitations to the study. Although we did consider item non-response and measurement differences across a large number of variables that were considered of a priori importance, there may have been variables which were subject to such differences which we did not consider. While it is not possible to assert the extent to which the findings of the study are generalisable beyond the population considered, it would seem reasonable to assume that the broad findings would relate to samples with similar age and geographical profiles surveyed at a similar point in time. However, as this was a study of members of a long-standing cohort who will be familiar with the survey and its previous interview mode, it is unclear to what extent generalisability will extend to genuinely cross-sectional settings. Ongoing studies of mode effects in different populations are therefore of great importance. Our focus in the present paper was on the cross-sectional impact of the mixed-mode approach. Such cross-sectional considerations naturally affect longitudinal analyses, but future work could explicitly explore longitudinal implications of the mixed-mode approach.

What implications do these findings have for future data collections in NCDS or other similar studies? In order to employ the mixed-mode approach judiciously, we mainly restricted data collection to more factual topics. Telephone- and Web-collected data are not usually compatible if the question asks for a judgment on a topic (especially non-salient ones), asks about non-norm behaviours or attitudes, or uses a multi-point response scale. Our findings of mode effects in the key subjective questions that were included, and in other questions where social desirability bias was likely to be an issue, confirm this judgement as correct. Perhaps less expected were the differences in item non-response in financial variables, which suggests the Web, and mixed-mode approaches which include the Web, may also not be ideal for these items.

Going forward in the design of future sweeps, these limitations suggest that sequential mixed-mode Web-to-telephone should not become the primary mode of approach, especially for the study's major sweeps where collection of financial variables, for example, is a central part of the study. However, the successful take-up by Web, and its apparent superior data quality to phone for some question types—for example, where social desirability appears to be an issue—suggests potential continued use of the Web as part of a multi-mode approach to future sweeps, perhaps as a supplement to face-to-face interviewing.

ORCID

Richard J. Silverwood https://orcid.org/0000-0002-2744-1194

REFERENCES

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(2), 385–408.

Brown, M. (2016). Going online with the National Child Development Study: Design decisions during the development of the Age 55 web survey. Centre for Longitudinal Studies Working Paper 2016/2. Available from: https://cls.ucl.ac.uk/wp-content/uploads/2017/04/CLS-WP-20162.pdf

Brown, M., Goodman, A., Peters, A., Ploubidis, G.B., Sanchez, A., Silverwood, R. et al. (2021) COVID-19 Survey in Five National Longitudinal Studies: Waves 1, 2 and 3 User Guide (Version 3). London: UCL Centre for Longitudinal Studies and MRC Unit for Lifelong Health and Ageing. Available from: https://cls.ucl.ac.uk/wp-content/uploads/2017/02/UCL-Cohorts-COVID-19-Survey-user-guide.pdf

Calderwood, L. (2016) Effects of Mode on Response Rates. Paper presented at the CLOSER Workshop: Mixing Modes and Measurement Methods in Longitudinal Studies, London.

Carpenter, H. & Burton, J. (2018) Adaptive Push-to-Web: Experiments in a Household Panel Study. Understanding Society Working Paper Series, No. 2018-05. Available from: https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2018-05.pdf

Daikeler, J., Bošnjak, M. & Lozar Manfreda, K. (2019) Web versus other survey modes: an updated and extended meta-analysis comparing response rates. *Journal of Survey Statistics and Methodology*, 8(3), 513–539.

De Leeuw, E.D. (2018) Mixed-mode: past, present, and future. *Survey Research Methods*, 12(2), 75–89.

De Leeuw, E.D. (2005) To mix or not to mix data collection modes in Surveys. *Journal of Official Statistics*, 21(5), 233–255.

Dillman, D., Smyth, J. & Christian, L.M. (2009) *Internet, mail and mixed-mode surveys: the tailored design method*, 3rd edn. Hoboken, New Jersey: Wiley.

Elliott, M.N., Zaslavsky, A.M., Goldstein, E., Lehrman, W., Hambarsoomians, K., Beckett, M.K. et al. (2009) Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. *Health Services Research*, 44(2), 501–518.

Fong, B. & Williams, J. (2011) *British crime survey: feasibility of boosting police force area (PFA) sample sizes using supplementary recontact surveys* Report for the Home Office. London: TNS-BMRB.

Gaia, A. (2017). The Effect of Respondent Incentives on Panel Attrition in a Sequential Mixed-Mode Design. No. 2017-03, UK: Institute of Social and Economic Research.

Greene, J., Speizer, H. & Wiitala, W. (2008) Telephone and web: mixed-mode challenge. *Health Services Research*, 43(1), 230–248.

Heerwegh, D. & Loosveldt, G. (2008) Face-to-face versus web surveying in a high-internet-coverage population: differences in response Quality. *Public Opinion Quarterly*, 72(5), 836–846.

Heerwegh, D. (2009) Mode differences between face-to-face and web surveys: an experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research*, 21(1), 111–121.

Hochstim, J.R. (1967) A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association*, 62(319), 976–989.

Hope, S., Campanelli, P., Nicolaas, G., Lynn, P. & Jäckle, A. (2014) The Role of the Interviewer in Producing Mode Effects: Results from a Mixed-modes Experiment Comparing Face-to-Face, Telephone and Web Administration. No. 2014-20, ISER Working Paper Series. Available from: https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2014-20.pdf

Imbens, G.W. & Rubin, D.B. (2015). Instrumental variables analysis of randomized experiments with one-sided noncompliance. In: Imbens, G. W., & Rubin, D. B. (Eds.) *Causal inference for statistics, social, and biomedical sciences: an introduction*. ch. 24. New York: Cambridge University Press, pp. 542–559.

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. *Survey Research Methods*, 9(1), 57–70.

Jäckle, A., Roberts, C. & Lynn, P. (2010) Assessing the effect of data collection mode on measurement. *International Statistical Review*, 78(1), 3–20.

Jäckle, A. (2016) Identifying and Predicting the Effects of Data Collection Mode on Measurement. Paper presented at the CLOSER Mixing Modes and Measurement Methods in Longitudinal Studies Workshop, London.

Jäckle, A., Gaia, A. & Benzeval, M. (2017) Mixing Modes and Measurement Methods in Longitudinal Studies. Report Submitted to the Cohort & Longitudinal Studies Enhancement Resources (CLOSER). UCL Institute of Education, London. Available from: https://www.closer.ac.uk/wp-content/uploads/Mixing-modes-and-measurement-methods-in-longitudinal-studies.pdf

Kappelhof, J. (2015) The impact of face-to-face vs sequential mixed-mode designs on the possibility of nonresponse bias in surveys among non-western minorities in the Netherlands. *Journal of Official Statistics*, 31(1), 1–31.

Kreuter, F., Presser, S. & Tourangeau, R. (2008) Social desirability bias in CATI, IVR, and web surveys: the effects of mode and question sensitivity. *Public Opinion Quarterly*, 72(5), 847–865.

Krosnick, J.A. & Alwin, D.F. (1987) An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, 51(2), 201–219.

Laaksonen, S. & Heiskanen, M. (2014) Comparison of three modes for a crime victimization survey. *Journal of Survey Statistics and Methodology*, 2(4), 459–483.

Lozar, M.K. & Vehovar, V. (2002). Mode Effect in Web Surveys. In the proceedings from The American Association for Public Opinion Research (AAPOR) 57th Annual Conference, 2002.

Manfreda, K.L., Bosnjak, M., Berzelak, J., Haas, I. & Vehovar, V. (2008) Web surveys versus other survey modes: a meta-analysis comparing response rates. *International Journal of Market Research*, 50(1), 79–104.

McHorney, C.A., Kosinski, M. & Ware, J.E. Jr (1994) Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey. *Medical Care*, 32(6), 551–567.

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.

Mostafa, T., Narayanan, M., Pongiglione, B., Dodgeon, B., Goodman, A., Silverwood, R.J. et al. (2021) Missing at random assumption made more plausible: evidence from the 1958 British birth cohort. *Journal of Clinical Epidemiology*, 136, 44–54.

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

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.

Schonlau, M., Zapert, K., Simon, L., Sanstad, K., Marcus, S., Adams, J. et al. (2003) A Comparison between responses from a propensity-weighted web survey and an identical RDD survey. *Social Science Computer Review*, 22(1), 128–138.

Scott, A., Jeon, S.H., Joyce, C.M., Humphreys, J.S., Kalb, G., Witt, J. & Leahy, A. (2011) A Randomised trial and economic evaluation of the effect of response mode on response rate, response bias, and item non-response in a survey of doctors. *BMC Medical Research Methodology*, 11(1), 126.

Siemiatycki, J. (1979) A comparison of mail, telephone, and home interview strategies for household health surveys. *American Journal of Public Health*, 69(3), 238–245.

Wagner, J., Arrieta, J., Guyer, H. & Ofstedal, M.B. (2014) Does sequence matter in multimode surveys: results from an experiment. *Field Methods*, 26(2), 141–155.

Wood, M. & Kunz, S. (2014). CAWI in a Mixed-mode Longitudinal Design. *Understanding Society Working Paper Series 2014-07*. Colchester: University of Essex.

Ye, C., Fulton, J. & Tourangeau, R. (2011) More positive or more extreme? A meta-analysis of mode differences in response choice. *Public Opinion Quarterly*, 75(2), 349–365.

How to cite this article: Goodman, A., Brown, M., Silverwood, R.J., Sakshaug, J.W., Calderwood, L., Williams, J. et al. (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)*, 1–29. Available from: https://doi.org/10.1111/rssa.12786