SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

Working Paper 27:

Experimental Evidence on the Effect of Mixed-Mode Survey Designs on Item Measurement and Variable Correlations

Liam Wright, Georgia D. Tomova, Matt Brown, Konstantinos Tsigaridis, Morag Henderson, Richard J. Silverwood

Centre for Longitudinal Studies, University College London

August 2026

www.surveyfutures.net

Acknowledgements

Survey Futures is an Economic and Social Research Council (ESRC)-funded initiative (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.

The research reported here forms part of Survey Futures Research Strand 6: Reducing and evaluating mode effects.

Citation

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., Brown M., Tsigaridis, K., Henderson M. & Silverwood R.J. (2026) 'Experimental Evidence on the Effect of Mixed-Mode Survey Designs on Item Measurement and Variable Correlations', *Survey Futures Working Paper* no. 27. Colchester, UK: University of Essex. Available at https://surveyfutures.net/working-papers/.

Contents

Experimental Evidence on the Effect of Mixed-Mode Survey Designs on Item Measurement and Variable Correlations

Liam Wright *, Georgia D. Tomova, Matt Brown, Konstantinos Tsigaridis, Morag Henderson, Richard J. Silverwood

Centre for Longitudinal Studies, Institute of Education, University College London

Abstract

Background: Mixed-mode designs can generate bias in all types of analysis of survey data due to systematic differences in how survey participants may respond according to mode – a phenomenon referred to as a mode effect. There is widespread evidence of mode effects upon item measurement (e.g., item means and proportions), but less on associations between survey items, despite estimating associations being a primary use of survey data.

Methods: We estimated average differences (means and proportions) in item responses according to mode using data from an experiment in which participants were randomized to respond via web questionnaire, video interview, or in-person interview. We then used these estimates to simulate (counterfactual) data from two web-then-in-person sequential mixed-mode surveys of young adults in the United Kingdom (Next Steps and the Millennium Cohort Study, MCS) as if they were collected using a single mode, web. We estimated associations between variables using this simulated data and compared these against associations obtained using the observed (i.e., factual) data; the difference between the two provides a measure of variation in variable correlations reflecting the mixed-mode survey design.

Results: Mode effects were observed for an array of items comparing web with in-person and video modes; differences between video and in-person were more muted. Participants in web were more likely to report negatively valent values for sensitive items, in particular (e.g., poorer physical and mental health, lower life satisfaction, and greater financial difficulty), though effect sizes were generally small with few exceeding 0.2 SD and all below 0.3 SD. The impact of mode effects on variable associations in the mixed-mode surveys was also generally small: a 0.004 SD change in absolute effect sizes on average in both Next Steps and the MCS.

Conclusion: Items related to sensitive characteristics showed small mode effects when comparing web and in-person or video interview modes. Mode effects did not appear to translate into substantial changes to associations between variables as assessed in a sequential web-then-in-person mixed-mode survey compared with a counterfactual web-only survey.

Keywords: survey design; mixed-mode surveys; mode effects; mode selection; survey experiment

1 Introduction

In recent years, many surveys have prioritised digital modes of data collection over traditional in-person interviews – notably, by offering participants the chance to respond via web questionnaire or, more recently, video interview. This includes the latest waves of data collection in several of the UK's major longitudinal data infrastructures, such as the Millennium Cohort Study, Next Steps, and the United Kingdom Household Longitudinal Study. Each of these has adopted sequential mixed-mode designs, with participants initially invited to respond by web and offered alternate modes (e.g., video, telephone, or in-person interview), if not responding.

Digital modes can reduce costs and increase coverage and response rates, but idiosyncratic features of survey modes can influence the data participants provide. For instance, in non-anonymous modes, such as in-person interview, participants may feel less comfortable accurately revealing sensitive information, such as on physical or mental ill health and financial insecurity, relative to a web questionnaire (d'Ardenne et al., 2025). Differences in responses between modes due to *how* items are presented are termed *mode effects*.

Mode effects can introduce bias into all types of analysis of mixed-mode survey data, including univariate statistics, such as item means and proportions, and multivariate statistics, such as measures of association between variables. Most obviously, mixed-mode designs induce correlations between items subject to mode effects as these share a common source of variation, mode. Mixed-mode designs also induce correlations between items subject to mode effects and variables related to *mode selection* – differences between modes due to *who* is being measured in each mode. For instance, in a concurrent web or in-person mixed-mode survey – modes that vary on interviewer presence – factors related to choosing one mode over the other (e.g., age or computer skills) would become correlated with responses given to sensitive items.

Though it may at first glance seem appealing, removing the bias from mixing survey modes is generally not as simple as conditioning upon mode (e.g., by including a mode indicator variable in regression models or stratifying by mode). This is because of a statistical phenomenon known as *collider bias* (Tomova, Silverwood, Tennant, et al., 2026). To give intuition, first note that mode can be conceptualized as a consequence of variables that cause mode selection (e.g., age and computer skills in a concurrent web or in-person mixed-mode survey). If we know a participant's response mode (e.g., web) and that the participant lacks a trait that normally increases the likelihood of response via that mode (e.g., they in fact have poor computer skills), then they are more likely to possess some other trait that made them choose that mode (e.g., they are old). In other words, conditional on mode, independent causes of mode selection (as well as consequences of these causes; Tomova, Silverwood, Tennant, et al., 2026) will appear inversely correlated. Controlling for mode when using mixed-mode data can therefore introduce bias, rather than reduce it.

A large literature has used experimental data to estimate mode effects (Tomova, Silverwood, & Wright, 2026a); since randomization of mode means that the mode selection process is random by design, differences between modes should reflect mode effects and not mode selection. These studies have broadly focused on obtaining estimates of mode effects on univariate statistics – for instance, differences in likelihood of reporting poor mental health. Less examined is the effect of mixing modes upon associations between variables, a notable limitation given estimating different types of association is a key purpose to which (mixed-mode) survey data are employed (Vannieuwenhuyze, 2015). Recent simulation work shows that in the presence of strong mode effects and/or mode selection, substantial bias may be introduced in the observed associations, both as a result of mode effects themselves and as a result of inappropriate controlling for mode (Tomova, Silverwood, & Wright, 2026b). But, existing experimental studies show generally small differences in association due to mode effects, in practice (Clarke & Bao, 2022; Dolan & Kavetsos, 2016; Jäckle et al., 2006; Martin & Lynn, 2011; Piccitto et al., 2022; Vannieuwenhuyze, 2015; Villar & Fitzgerald, 2017).

However, these empirical studies have several limitations. First, they predominantly assess differences resulting from joint mode effects (e.g., by comparing correlations in different single mode surveys), rather than arising from the combination of mode effects and mode selection that can occur in mixed-mode surveys (Vannieuwenhuyze, 2015, an exception). Second, they do not assess the frequency with which conditioning upon mode increases differences in association (relative to a single-mode survey), rather than reduces them. Third, several operationalize differences as statistically significant changes in the size of a coefficient, but this is a function of both effect size and statistical power, rather than effect size alone (Greenland et al., 2016; Sullivan & Feinn, 2012), the quantity which is of substantive interest. Finally, they broadly focus only on survey items related to political attitudes and subjective wellbeing.

In this study, we begin to close these gaps by using data from a recent mixed-mode survey experiment in which individuals were randomly allocated to complete a survey by (1) in-person interview, (2) video interview, or (3) web questionnaire. The experiment was carried out to examine the potential for mode effects in recent data collections of two major UK cohort studies, each of which collected data on a wide range of topics and also used sequential mixed-mode designs: the Millennium Cohort Study (MCS) and Next Steps. We use the experiment to estimate mode effects for many of the survey items that appeared in these cohorts' data collections and then also use these estimates to simulate (counterfactual) data for the MCS and Next Steps as if they had been collected in a single mode. With this, we are able to – assumptions being met – estimate the degree to which associations between observed variables reflect the mixed-mode survey design. As the analysis proceeds in two steps using different datasets, each step is presented separately in the following report.

2 The Mode Effects Experiment

2.1 Methods

2.1.1 Participants

The experiment data used in the present study were part of the Centre for Longitudinal Studies Mode Effects Study (CLSMES) 2024, a project aiming to evaluate mode effects in key measures included in the latest data collections of Next Steps (age 32y) and the MCS (23y). CLSMES targeted people aged 20-40 and resident in England to roughly match the characteristics of the Next Steps and MCS cohorts.

Participants were recruited at home and in-street from clustered sample areas on a 'free-find' basis – i.e., they were approached without pre-screening. Soft quotas were used to ensure participants were broadly representative of age, sex, and region. In total, 1,800 participants were recruited to the study from 200 sample areas.

At recruitment, participants were asked to complete two surveys to take place approximately two weeks apart and were given a £20 voucher for each survey taken. Surveys were completed between 15th January 2024 and 21st February 2024 and were conducted using web questionnaire, video interview (Microsoft Teams), or in-person interview. According to the protocol, participants were randomized, using a rotation plan, to one of the nine possible combinations of these modes (web-web, web-video, web-in-person, …), but participants were not aware of mode assignment at the time of recruitment. Participants answered the complete questionnaire at Wave 1, but the Wave 2 questionnaire was much shorter. Consequently, in this study, we analysed data from Wave 1 only.

2.1.2 Measures

The survey questionnaire contained items on household and social relationships, housing, employment and finances, physical and mental health and wellbeing, health behaviours, identity, and social capital and civic behaviour (e.g., voting); these items also appeared in the Next Steps and MCS surveys, though the Next Steps and MCS questionnaires were longer. Some items (especially those relating to sensitive information) were included in a self-complete module: in the in-person interview, participants were handed a computer to complete the module with; in the video interview, participants were provided a website link to click; and in the web mode, items were integrated within the wider questionnaire. Variables collected as part of the self-complete module are indicated with asterisks ('*') in the results.

Descriptions of the variables used in the present study are shown in Table 1. The survey also contained items on occupation and a cognitive test, but given the complexity of these data, we do not analyse these here (see Koerber et al., Forthcoming and; Tsigaridis et al., 2026 for results from these). Note, the data we do analyse have previously been used to estimate mode effects (Asensio et al., 2025). However, Asensio et al. focus on comparisons with video, specifically (i.e., they do not compare web with in-person, a common mixed-mode design), and do not assess mode effects on variable associations, the primary aim here, warranting re-estimation of mode effects, *de novo*.

2.1.3 Statistical Analysis

We performed several analyses. First, we examined covariate balance across Wave 1 modes for to a set of background demographic and socioeconomic characteristics: gender, ethnicity, region, education and social grade. As we found some differences across mode (gender, education and social grade), we adjusted for these in subsequent analyses. Note, gender and education were captured as part of the survey rather than at recruitment. However, each is easy-to-recall and objective and is arguably unlikely to exhibit mode effects (d'Ardenne et al., 2025).

Second, as survey mode can lead to differences in item response rates, for each variable, we calculated differences in item missingness according to mode using linear probability models, adjusting for background characteristics. Third, we calculated differences in responses for each variable by mode; for continuous variables, this was differences in means, and for categorical variables, this was differences in proportions. Again, we used linear regression models, adjusting for background characteristics. For (unordered) categorical variables, we repeated linear probability models for each possible category (i.e., calculating the difference in proportion reporting a particular category vs. any other). Variables on Likert or similar ordered scales (e.g., self-rated health) were instead treated as continuous variables in these analyses (such variables are indicated in Table 1). Given the large number of variables, in summarizing results, we use 0.2 SD as a cut-off for a small effect size, corresponding to the heuristic proposed by Cohen (1988). We used bootstrapping as this gave us a set of 'random draws' of these estimates that were potentially correlated in size between different variables and could then be used to produce counterfactual values in the MCS and Next Steps analyses (Section 3).

We finally performed a series of robustness checks and sensitivity analyses. First, we estimated mode effect estimates for categorical items using alternate statistical estimators – logistic regression for binary variables, multinomial regression for multinomial categorical variables, and ordered logistic regression for ordered categorical variables. Second, we repeated analyses using mode allocation for the Wave 2 interview as a 'negative control exposure' (Lipsitch et al., 2012); future mode is unlikely to influence current responses, so any non-zero associations should indicate violations of (conditional) exchangeability (we add the same covariates given these are used in the main analysis).

There was only a small amount of missing data overall, so we used (variable specific) complete cases for our analysis. We carried out all analyses in R version 4.5.2 (R Core Team, 2025).

2.2 Results

2.2.1 Descriptive Statistics

1,692 individuals took part in the (Wave 1) survey; 582 (34.4%) by web, 565 (33.4%) by in-person interview, and 545 (32.2%) by video interview. Table 2 shows the distribution of background characteristics by survey mode. There was some imbalance in these characteristics – males and participants whose highest qualification was GCSEs or lower were more likely to complete the in-person survey. There were also differences between modes in the dates, days of the week and time of day individuals responded (Supplementary Figure S1)[1]: individuals responding by web survey were more likely to complete the survey later at night or early in the morning, while video interviews were more likely to take place at lunchtime or in the evening.

2.2.2 Mode Effects on Item Response Rates

Mode effects upon items responses rates at Wave 1 are shown in Figure 1 (full results, including from unadjusted models, are tabulated in Supplementary Table S1). Item missingness rates were generally small and zero in several cases (gender, age, region, social grade, receives Universal Credit, life satisfaction, smoking status). Participants left 0.80 items (2.42%) items incomplete on average and 45.2% of individuals had at least one item missing. Item missingness rates (across all modes) were 20.4% for childhood financial difficulties (this item had a specific category for 'Don't know/can't remember'), 11.6% and 9.4% for GAD-2 and PHQ-2 Likert scores (measures of anxiety and depression), and 5.1% for income, respectively.

1.7 pp (95% CI = -4.0, 7.5) and 6.3 pp (0.5, 12.1) more individuals had at least one missing item in in-person than web and video modes, respectively. However, for in-person vs. web this difference was mainly driven by differences in reporting of income: 13.6 pp (11.1, 16.1) higher missingness in the in-person mode. Those in web were ~ 5% more likely to report parental education than in in-person and video modes. Corresponding differences for completing an item on childhood financial difficulties were 7.0 pp (2.4, 11.7) and 4.2 pp (-0.5, 9.0). Participants in video mode completed 0.3 (0.1, 0.4) fewer items than those in the in-person mode, on average. This was driven by differences in reporting of fruit consumption (-3.9 pp; 95% CI = -5.7 pp, -2.1 pp), voting (-3.7 pp; -6.2 pp, -1.1 pp), and exercise (-3.2 pp; -5.2 pp, -1.6 pp).

2.2.3 Mode Effects on Means and Proportions

Mode effects upon item proportions and means are shown, split by domain, in Figures 2-6 (adjusted models). Results for continuous variables are standardized into units of SD for comparability (using SDs from web participants). Full – including unstandardized – results are tabulated in Supplementary Table S2. In general, participants in the web mode were more likely to report difficulties for sensitive

[1] Supplementary files are available at the Open Science Framework website: https://osf.io/756vd.

items (e.g., mental health and financial situation), with differences between video and in-person modes more muted. Only eight (of 111; 7.2%) mode comparisons exceeded the 0.2 SD threshold of a small effect, seven of which were for the web vs. in-person comparison, though confidence intervals were often wide.

There were no sizeable differences in data quality measures or demographic variables (Figure 2; after adjustment for the potential confounders). For measures of SEP (Figure 3), individuals in the web mode reported 0.17 SD (0.06, 0.28) greater current financial difficulties than those in the in-person mode. The equivalent figure for web vs. video comparison was 0.16 SD (0.05, 0.27). Individuals in the web mode also reported lower income (0.24 SD; 0.13, 0.35) compared with those in in-person mode, though differences between web and video were small (-0.09 SD; -0.20, 0.01), which could be explained by higher item missingness in the in-person mode. For childhood circumstances, web participants were more likely to report financial difficulties during childhood, as well as lower trust in others and having lower social capital (e.g., having someone they feel close to, Figure 4), with mode effects similar comparing web with in-person or with video. Effect sizes exceeded 0.2 SD only for having someone close (web vs. video and in-person) and trust in others (web vs. in-person only).

Compared with in-person and video modes, participants in the web mode reported lower life satisfaction and self-rated health, greater depression (PHQ-2 Likert) and anxiety (GAD-2 Likert), and were more likely to report suffering from any chronic condition, a limiting chronic condition, and a chronic psychological condition, specifically (Figure 5). Again, differences between video and in-person modes were muted with the exception that video participants were more likely to report having a chronic condition. The largest effect size was for life satisfaction in the web vs in-person comparison: -0.27 SD (-0.38, -0.16). Differences in PHQ-2 Likert were 0.25 SD (0.14, 0.37). There were few differences between modes in items on health behaviours (Figure 5). Exceptions included reporting of greater number of cigarettes smoked per day in the video (vs. in-person) mode.

2.2.4 Negative Control Analysis

Results for mode effects on item missingness and items means and proportions using Wave 2 allocated mode as a negative control are displayed in Tables S2-S6. There were few differences between modes in this analysis (exceptions were father's education, having children, and being an ex-smoker), and the differences were not systematic within a domain (i.e., no systematic underreporting of sensitive information in one mode).

2.3 Summary

Mode effects were observed for several items. In general, participants were more likely to report sensitive information (e.g., lower life satisfaction) in the web mode. Differences between the video and in-person modes were more muted. Variables for which we observed mode effects > 0.2 SD between the web and in-person modes were life satisfaction, PHQ-2 Likert, GAD-2 Likert, income, trust in others, having a close person, and having a limiting illness. We take mode effect estimates for the five continuous variables (life satisfaction, PHQ-2 Likert, GAD-2 Likert, income, and trust in others) forwards to assess the potential effect of mode upon variable correlations in the MCS and Next Steps.

3 Analysis of Cohort Data

3.1 Methods

3.1.1 Participants

3.1.1.1 The Millennium Cohort Study

The MCS follows a sample of individuals from across the UK born between 2000/02 (Connelly & Platt, 2014). Participants were recruited using a two-stage stratified sampling design and sampled from selected postcode areas, with oversampling of households from Northern Ireland, Scotland and Wales, ethnic minority backgrounds, and disadvantaged areas. Almost all cohort members were recruited at 9 months, with a small number (~ 4%) entering the study at 3y. Cohort members have been surveyed on eight occasions (9 months, 3y, 5y, 7y, 11y, 14y, 17y, and – most recently – 23y); some survey waves have involved data collection from parents and other family members, too.

Between 9 month and 17y, data collection largely comprised in-person interviews, but at 23y, a sequential mixed-mode design was used, with fieldwork split into two stages. In the first, ‘soft launch’ stage, a random subsample of 3,000 cohort members were randomized to in-person interview only or sequential web-then-in-person interview mixed-mode designs; the purpose of the soft launch was to assess the effect of mixed-mode on response rates. Based on the results of the soft launch, the sequential mixed-mode design (which increased response rates; Pople et al., Forthcoming) was rolled out to remaining ‘main stage’ cohort members. Non-respondents in both soft launch and main stages were finally offered a short mop-up web questionnaire. 78.8% of total respondents participated by web (4.7% mop-up questionnaire).

3.1.1.2 Next Steps

Next Steps (formerly known as the Longitudinal Study of Young People in England, LSYPE) follows a cohort of individuals who were in Year 9 (age 13/14) of secondary school in England in 2003/04 (Calderwood & Sanchez, 2016; Wu et al., 2024). Cohort members were recruited using a two-stage stratified sampling design with oversampling of ethnic minority and economically disadvantaged children. 15,770 individuals were originally recruited to the study, with a sample boost of 352 ethnic minority participants added at Wave 4 (16/17y). Cohort members were surveyed annually for 7 years to age 19/20y, then again at ages 25y and (most recently) 32y. Earlier data collections also included interviews of cohort members’ parents or carers.

Next Steps employed a single-mode (in-person interview) design for Waves 1-4, with sequential mixed-mode designs employed thereafter. At 32y, cohort members were first invited to respond via web questionnaire. After three weeks, non-respondents were then offered a choice of responding by in-person, video, or (in exceptional circumstances) telephone interview or by web questionnaire via a device provided by the fieldwork agency, with the original web questionnaire invitation remaining open during this time. After this, remaining non-respondents were invited to a shorter 'mop-up' web questionnaire. 88.1% of participants responded via web (5.7% short 'mop-up' questionnaire) and 9.9% by in-person interview – the remaining 2% were via telephone (1.9%) or video (0.1%). Given so few observations were from telephone or video (and the former is not covered in the CLSMES experiment), we restrict to web and in-person observations.

3.1.2 Measures

As noted in the *Introduction*, mixed-mode survey designs can induce correlations between variables subject to mode effects and between variables subject to mode effects and variables related to mode selection, and this correlation is larger the stronger mode effects and mode selection are (Tomova, Silverwood, & Wright, 2026b). Variables subject to mode effects in the MCS 23y and Next Steps 32y surveys were identified by selecting continuous variables where the estimated mode effect obtained in the prior section exceeded 0.2 SD. We restricted to continuous measures subject to mode effects because the method we used to simulate counterfactual data does not work for categorical variables. Note, the variable *trust in others* was not available in the MCS, and the bands used to define income differed in this study, too.

Variables potentially related to mode selection (but not subject to mode effects) were identified from single-mode waves of Next Steps (Waves 1-4) and the MCS (Waves 1-7). Given these waves contain data from thousands of survey items, we focused on variables we thought would (a) be of wide interest in the health and social sciences (e.g., family socioeconomic background, smoking) and (b) highly related to mode selection, basing this on work on the predictors of mode selection in another cohort study (Goodman et al., 2022; Wright et al., 2025). The variables we selected spanned: socio-demographics (sex, highest qualification, economic activity, maternal and paternal age at birth, family type [two or single parent household], ever been in care), family background (social class, parental highest education, low income, eligibility for free school meals, eligibility for educational maintenance allowance), and health and behaviour (self-rated health, chronic conditions, depressive symptoms, alcohol use, cannabis use, smoking habits, special educational needs). Note, for highest qualification, we used data from the latest survey (i.e., 23y or 32y) given that we required measurement from an age when education was likely to be completed and that we did not expect education to be subject to mode effects. More detail on the variables is displayed in Table 3.

3.1.3 Statistical Analysis

The aim of the analysis was to assess the extent to which bivariate associations between variables reflected the mixed-mode designs in MCS and Next Steps, compared those that would be obtained from the sample participants in a single-mode web survey. We estimated linear regressions where the dependent variable was subject to mode effects and the independent variable was subject to mode effects or related to mode selection, repeating for every combination of such variables.

For each pair of these variables, we obtained estimates from three regressions. The first was a simple regression between data as observed. The second included a mode indicator as a control variable: as discussed in the *Introduction*, this may introduce bias but is sometimes naïvely done anyway. The third used simulated counterfactual values for a single mode using the mode effect estimates from CLSMES in Section 2 and used these in regressions instead, in order to account for mode effects; counterfactuals were simulated for any variable that may have been subject to a mode effect – simulation involved subtracting the mode effect from the observed value for those in the in-person interview to obtain a prediction of how the same individuals would have responded if surveyed via web, given an assumption about the size of the mode effect (Wright et al., Forthcoming). (We used web as the reference [i.e., counterfactual] mode, given the preponderance of MCS and Next Steps data was from web.) This assumes that CLSMES mode effect estimate is unbiased and generalizes to in-person respondents in MCS and Next Steps samples.

We used the regression estimates to answer several questions. First, what are the average and maximum difference in effect sizes when accounting for mode effects via counterfactual simulation? Small differences would suggest mode effects do not typically large impact on associations. Second, for what proportion of variable-pairs does counterfactual simulation move associations towards the null? This provides evidence on whether associations not accounting for mode effects are anti-conservative on average. Third, for what proportion of variable-pairs does adding a mode indicator make associations more dissimilar to those that would be obtained in a survey using a single (web) mode? This provides evidence on whether adding a mode indicator may serve to increase bias rather than reduce it.

To obtain estimates of uncertainty, we used bootstrapping with replacement (1,000 repetitions) – note the bootstraps from the CLSMES experiment were used for the mode effect estimates to retain correlations between estimates for separate variables. Again, as there were generally low amounts of missing data, we used complete cases in the analysis of cohort data and did not weight for non-response. For comparability, in some presentations, we standardized effect sizes using SDs from the web mode in a given cohort.

3.2 Results

3.2.1 Descriptive Statistics

Descriptive statistics for each cohort are displayed in Supplementary Table S4. 78.8% (N = 7,626) of included MCS cohort members participated via web, with a corresponding figure of 89.9% (N = 6,410) for Next Steps. There was mode selection according to several variables, including education level, socioeconomic background, sex and cognitive ability: generally, those with higher education, skills, and socioeconomic advantage were more likely to appear in the web mode.

3.2.2 Associations Between Variables

(Standardized) associations are displayed in Figure 7 (full results tabulated in Supplementary Table S5). In general, estimating associations with counterfactual simulated data did not change estimates much (left panel). This was also true when adding a mode indicator to the regression model (right panel).

In the MCS, on average, adding an indicator for mode changed effects sizes by 0.002 SDs, while the counterfactual simulation approach changed effect sizes by 0.004 SDs. The maximum change in effect sizes were 0.007 SDs for the mode indicator and 0.024 SDs for the counterfactual simulation. The mode indicator moved estimates towards the null 32.8% of the time, while counterfactual simulation moved estimates towards the null 30.4% of the time. Adding a mode indicator increased the difference from the (hypothetical) single mode survey 4.5% of the time.

In Next Steps, adding an indicator for mode changed effects sizes by 0.001 SDs, on average, while the counterfactual simulation approach changed effect sizes by 0.004 SDs, on average. The maximum change in effect sizes were 0.005 SDs for the mode indicator and 0.017 SDs for the counterfactual simulation. The mode indicator moved estimates towards the null 49.4% of the time, while the counterfactual simulation moved estimates towards the null 18.4% of the time. Adding a mode indicator increased the difference from the (hypothetical) single mode survey 51.6% of the time, though in practice the extent of change was very small.

4 Discussion

Using data from a web, in-person, or video mode experiment, we found evidence that mode effects were only present in a small minority (~ 10%) of survey items. Few mode effects exceeded 0.2 SD – a common heuristic for a small effect size (Cohen, 1988) – and the largest mode effect observed was 0.27 (life satisfaction, comparing web and in-person modes). Differences were most common when comparing web to in-person or video modes (in-person and video modes were more similar), where items reflected sensitive information (e.g., on health, finances and wellbeing), and were also observed for items collected in self-complete modules (e.g., anxiety and depression). Mode effects, however, appeared to generate only negligible differences in association between study variables, relative to the association that would have been obtain using a web-only survey of the same participants (and assuming the CLSMES mode effect estimates to the MCS and Next Steps in-person responders). Adding a mode indicator to regressions (which can induce collider bias; Tomova, Silverwood, Tennant, et al., 2026) increased difference in a minority or slight majority of cases (4.5% in MCS, 51.6% in Next Steps), but again the size of this change was negligible.

Results for mode effects on item measurement (means and proportions) were broadly consistent with previous experimental studies which broadly show larger mode effects for items on sensitive information (e.g., mental health) and when comparing anonymous (e.g., web) and non-anonymous (in-person, video) modes (Tomova, Silverwood, & Wright, 2026a). Effect sizes were also similar to previous studies – even for items subject to mode effects, mode effects were between 0.2-0.3 SD (see, e.g., Goodman et al., 2022). Interestingly, mode effects were also observed for sensitive information collected during a self-complete module (e.g., on life satisfaction), despite responses not being reported to interviewers. This is unlikely to reflect device effects as broadly similar numbers of video and web respondents responded via smartphone. Income was also subject to a mode effect, though item missingness rates were substantially larger in the in-person mode, which may have biased estimates. Note, previous frameworks for predicting mode effects (e.g., d'Ardenne et al., 2025) also suggest responses to cognitively complex questions or effortful tasks can vary between (online and offline) modes, but no such items (except perhaps income) were examined here. See Tsigaridis (2026) for results from the CLSMES on a cognitive measure, the backwards digit span.

Results for mode effects on associations between variables were also consistent with previous studies, which broadly show limited effects, at least when assessed as magnitude of effect, rather than statistically significant differences (Clarke & Bao, 2022; Dolan & Kavetsos, 2016; Jäckle et al., 2006; Martin & Lynn, 2011; Piccitto et al., 2022; Vannieuwenhuyze, 2015; Villar & Fitzgerald, 2017). One explanation is likely the relatively low proportions of individuals using in-person interview (~ 20% in MCS and ~ 10% in Next Steps) in the data we used, which limits any resulting effect of the mixed-mode design, though we should note this level of mode split is common among many sequential mixed-mode surveys. Another reason is that variables subject to mode effects appeared as outcome variables in regression models, whereas treated these as predictors would have induced regression dilution bias (Tomova, Silverwood, Tennant, et al., 2026), though in practice this would very likely be small, too.

Another explanation is the generally small mode effects observed in CLSMES. Though this may suggest the effect of mixed-mode on variable associations is small in practice and can be ignored, there are circumstances (e.g., strong mode selection, more even mode split, and true associations close to the null) where it could have a material impact on results (Tomova, Silverwood, & Wright, 2026b) – consideration of these possibilities is necessary when using mixed-mode data. Note, the finding that adding a mode indicator *increased* difference from a single mode survey in a non-negligible number of cases (though by a small amount, on average) suggests that relying on this method to reduce bias from mixing modes could be misguided, in line with previous advice (Understanding Sociey, 2026; Wright et al., 2024). However, in practical terms, estimates were little changed.

4.1 Strengths and Limitations

This study had several strengths. We assessed mode effects across a wide range of items appearing in many major cohort studies. We also assessed the consequence of mixing-modes on variable correlations – a primary use of survey data – in two real-world data sources. There were also limitations. Though an experimental design was used, there were some baseline differences between participants in each mode, and differences in item response could have further generated bias in the mode effect estimates. However, estimates were similar when adding covariates to model and results overall were in line with previous experiments. We were also able to use the two-wave design as a potential negative control, which provided reassuring results – unlike the results using Wave 1 mode used, differences were not systematic within a domain (i.e., no systematic underreporting of sensitive information in one mode). However, randomization could have been an issue with Wave 1 allocation, specifically. Though the CLSMES sample was designed to be (roughly) comparable to the MCS and Next Steps samples, there are differences and counterfactual simulation relied on generalising the mode effect estimates to these. Finally, we compared association in a web-then-telephone mixed-mode to a web-only survey *of the same participants* but, in practice, survey modes can have effects on the response and 'coverage' of a survey (Daikeler et al., 2020; Edwards & Perkins, 2024; Schouten et al., 2013), which could lead to differences in associations due to sample differences (i.e., due to Type 1 and 2 selection bias; Lu et al., 2024).

5 Conclusions

Mode effects impact average responses to sensitive items in web vs in-person or video interviews, including in self-completion modules. Differences between in-person and video interview are more muted. Compared with a single web mode survey, mixing web and in-person modes does not appear to substantially change associations between survey items. Adding a mode indicator variable to (attempt to) account for mode effect can increase bias in a non-negligible number of cases, though, practically, by a negligible amount.

6 Figures

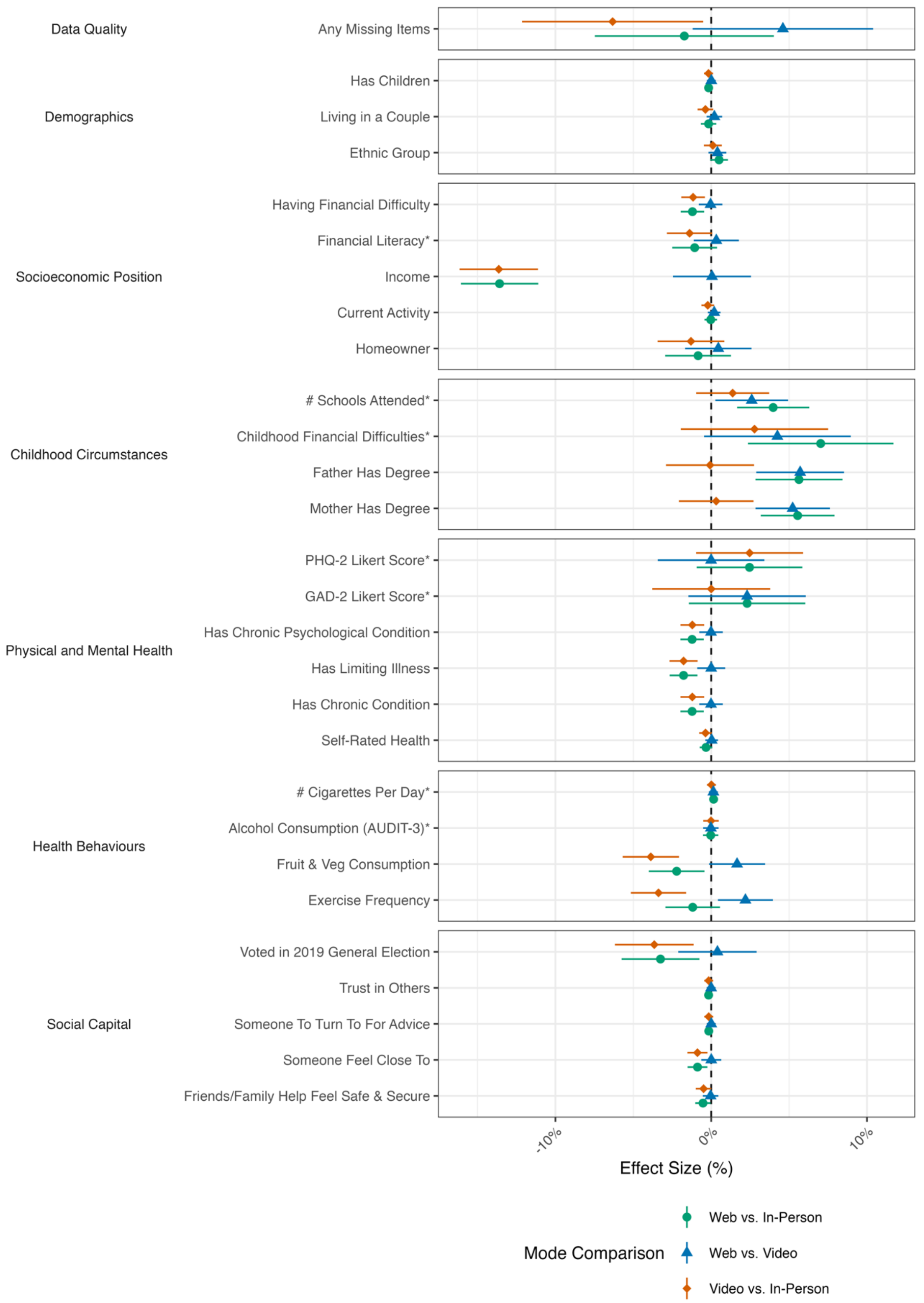

***Figure 1: Mode effects on item missingness (%) by variable and modes compared**. Derived from OLS regressions with adjustment for age, sex, education, and social grade. Variables with asterisks ('*') were collected as part of a self-complete module in each mode.*

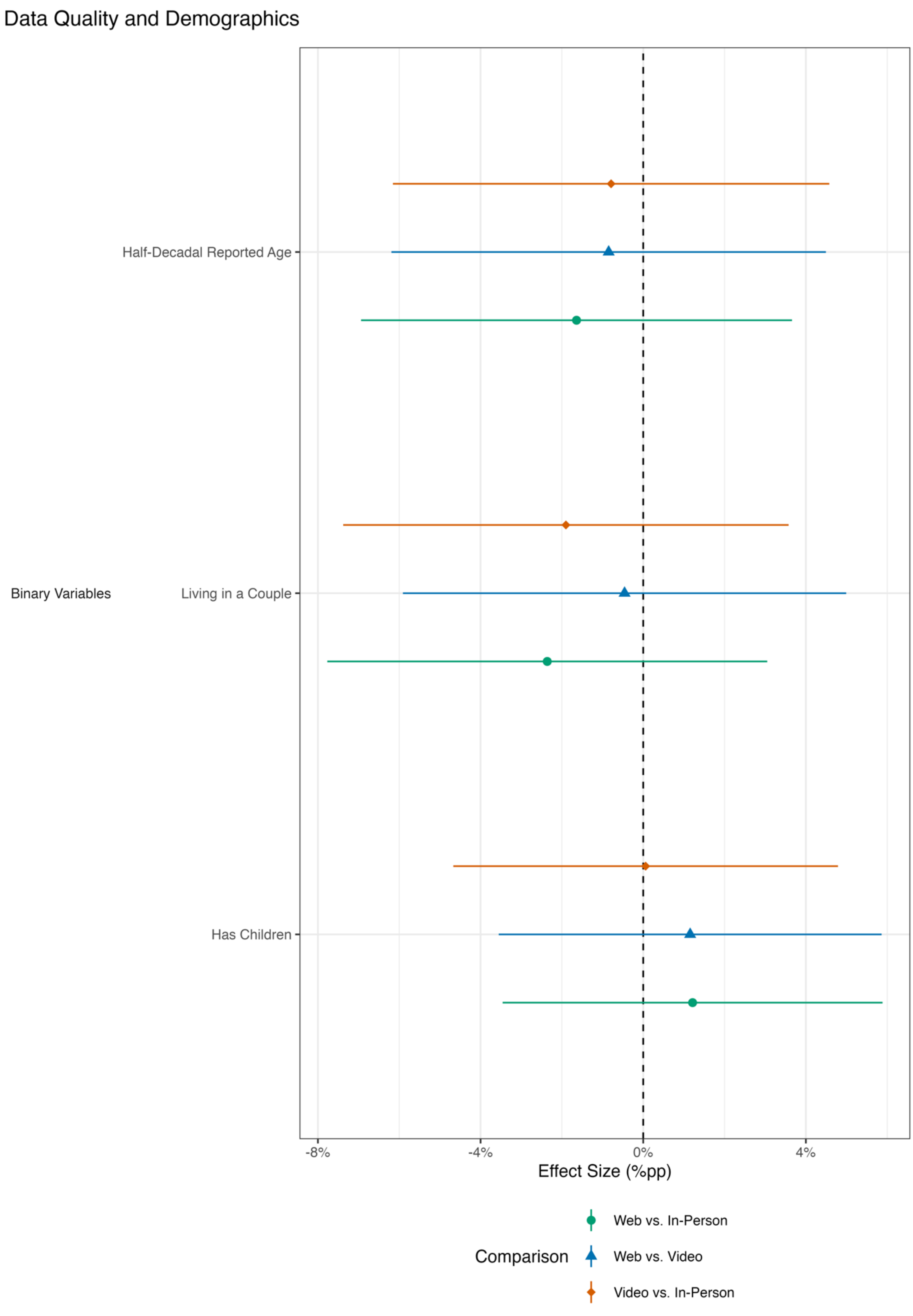

***Figure 2: Mode effects on item means (SD) and proportions (%) by variable and modes compared**. Variables from data quality and demographics domain. Derived from OLS regressions with adjustment for age, sex, education, and social grade. Variables with asterisks ('*') were collected as part of a self-complete module in each mode.*

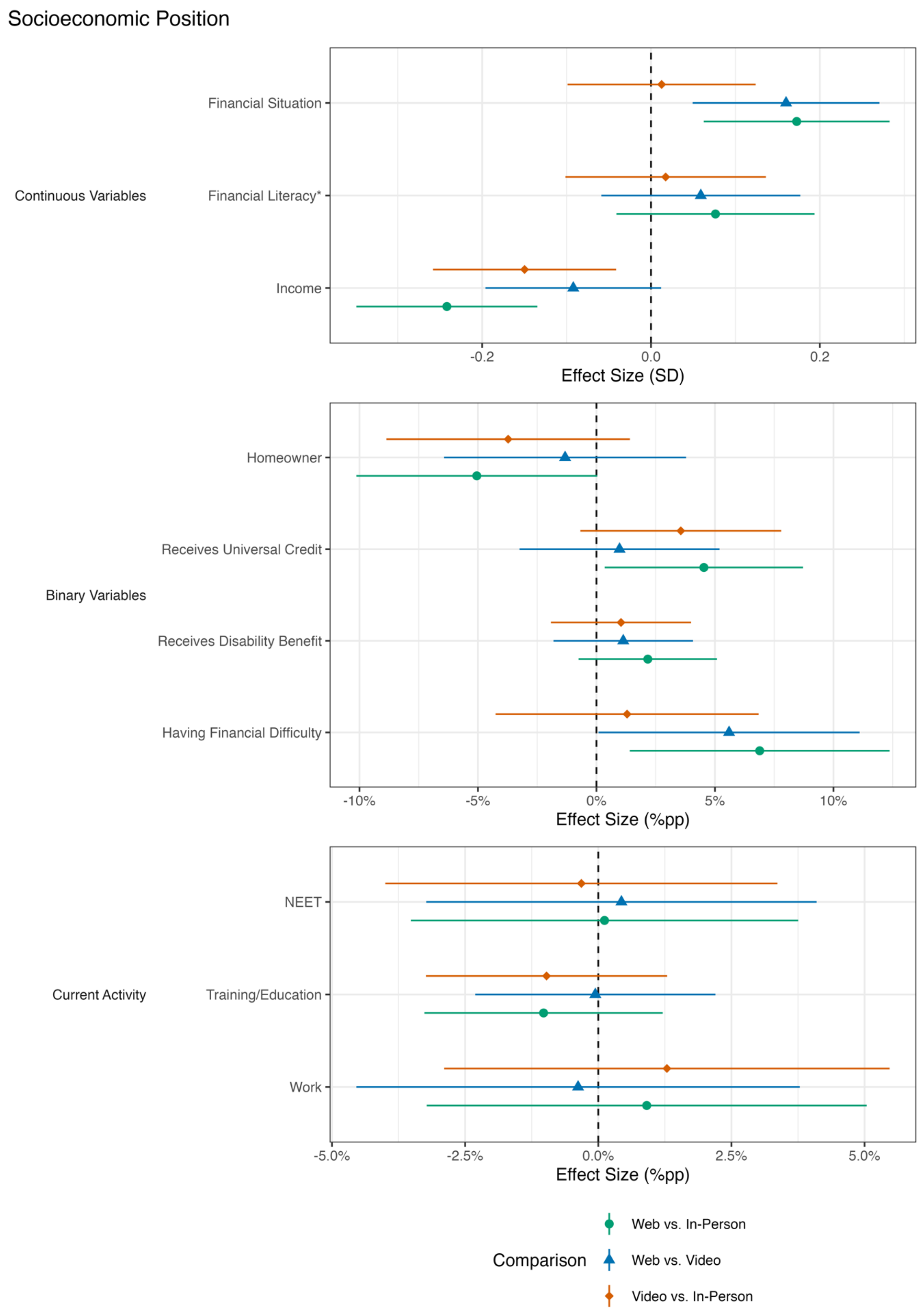

***Figure 3: Mode effects on item means (SD) and proportions (%) by variable and modes compared**. Variables from (adult) socioeconomic position domain. Derived from OLS regressions with adjustment for age, sex, education, and social grade. Variables with asterisks ('*') were collected as part of a self-complete module in each mode.*

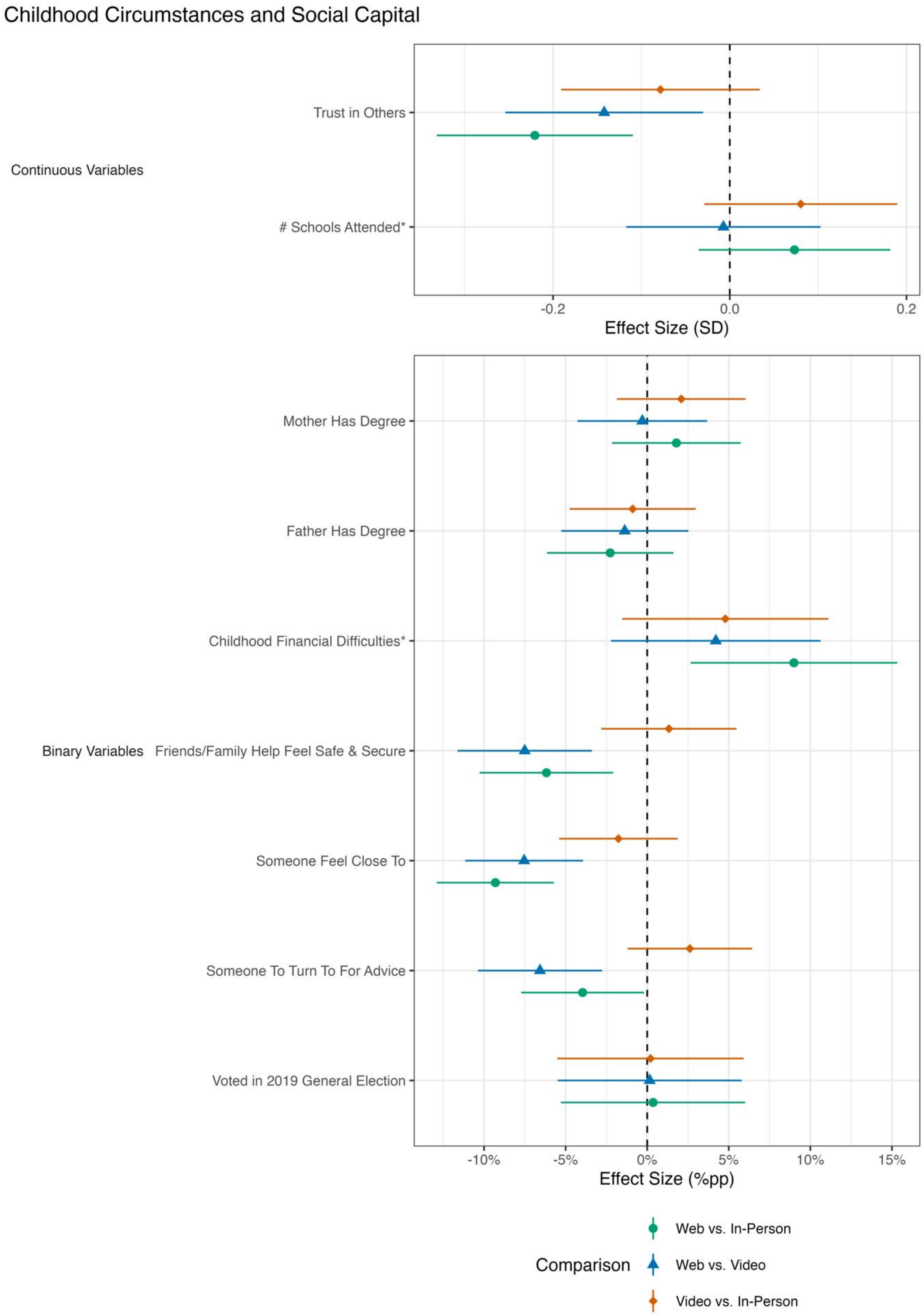

Figure 4: Mode effects on item means (SD) and proportions (%) by variable and modes compared. *Variables from childhood circumstances and adult social capital domain. Derived from OLS regressions with adjustment for age, sex, education, and social grade. Variables with asterisks ('*') were collected as part of a self-complete module in each mode.*

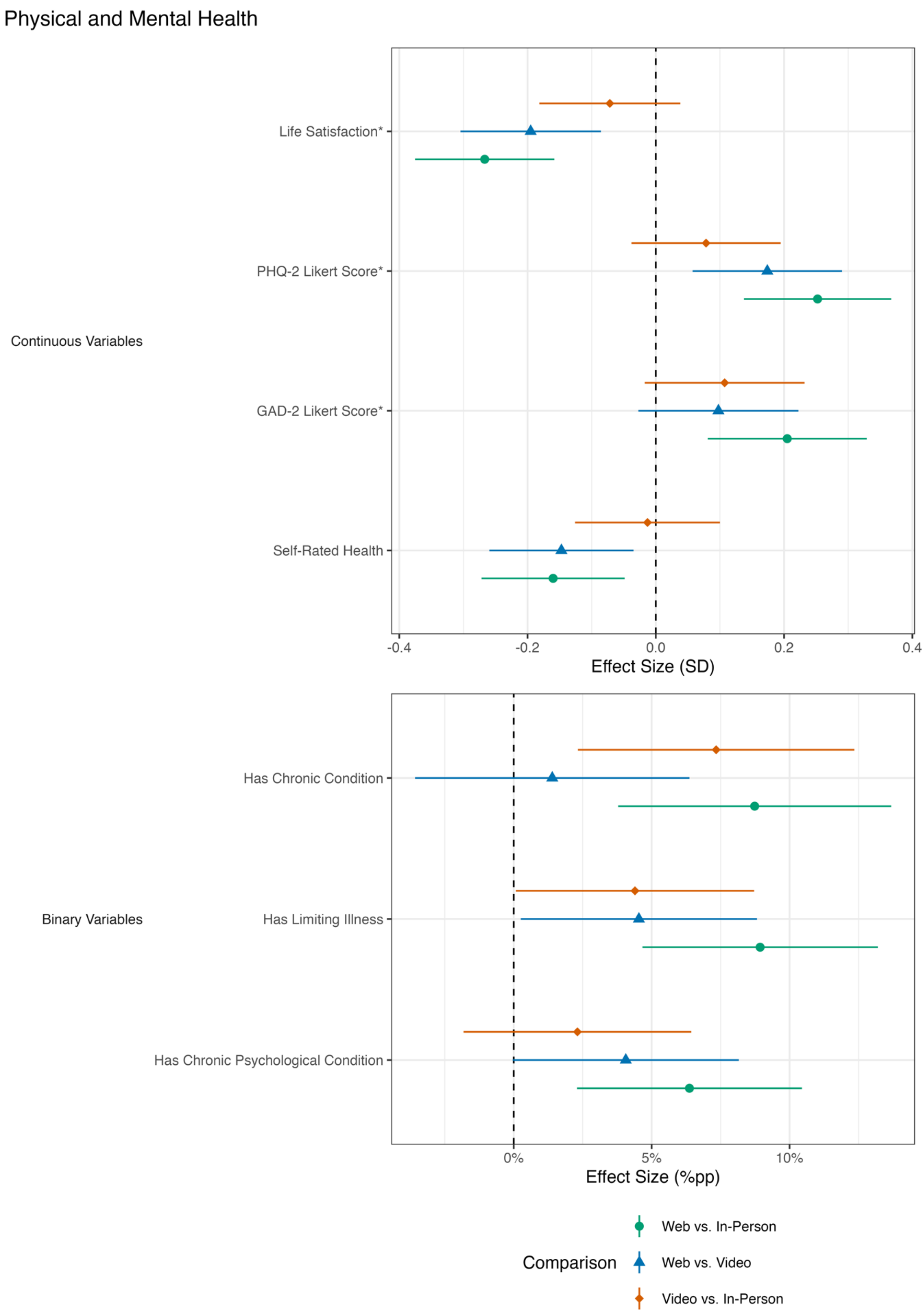

Figure 5: Mode effects on item means (SD) and proportions (%) by variable and modes compared. *Variables from physical and mental health domain. Derived from OLS regressions with adjustment for age, sex, education, and social grade. Variables with asterisks ('*') were collected as part of a self-complete module in each mode.*

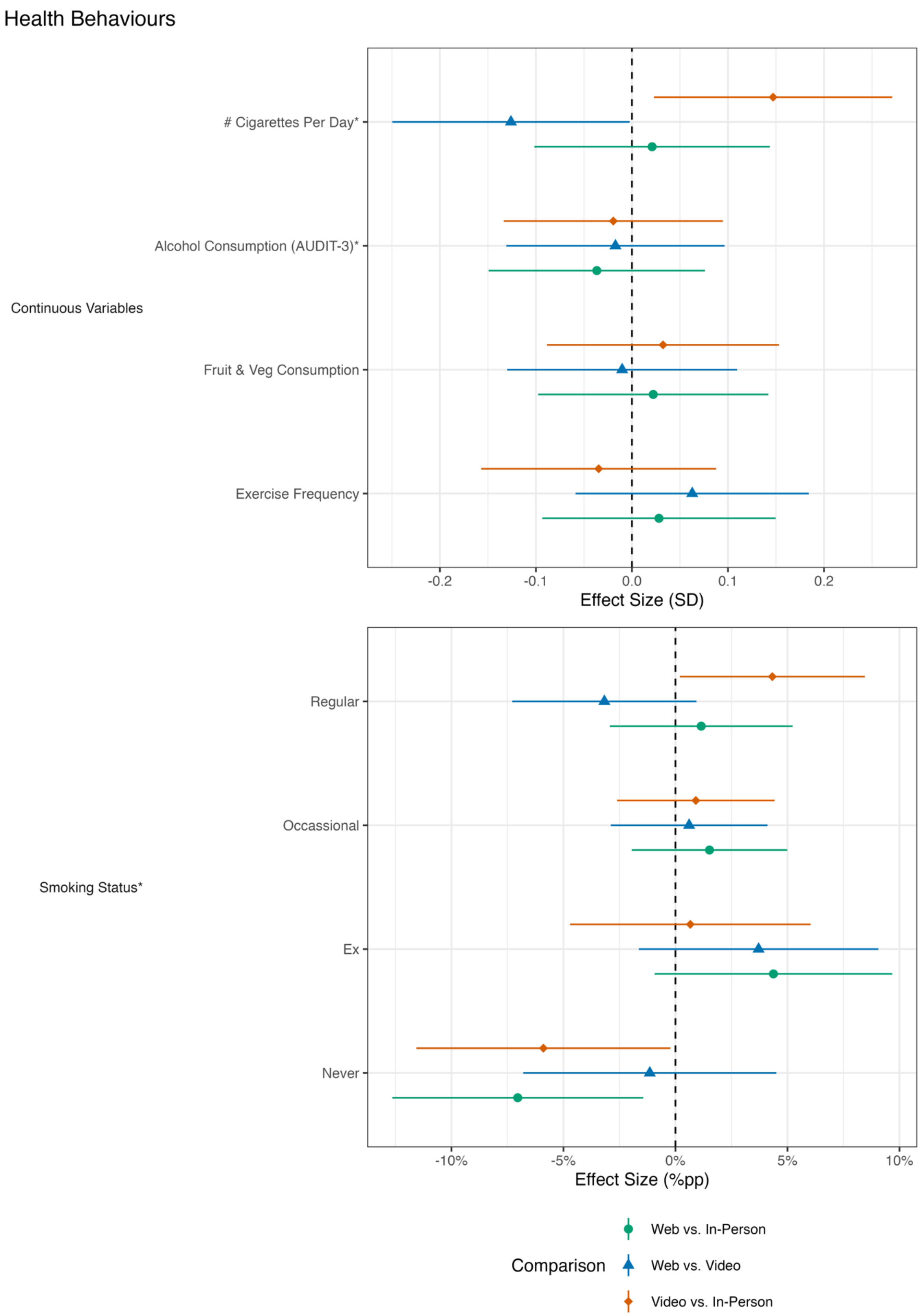

Figure 6: Mode effects on item means (SD) and proportions (%) by variable and modes compared. *Variables from behaviours domain. Derived from OLS regressions with adjustment for age, sex, education, and social grade. Variables with asterisks ('*') were collected as part of a self-complete module in each mode.*

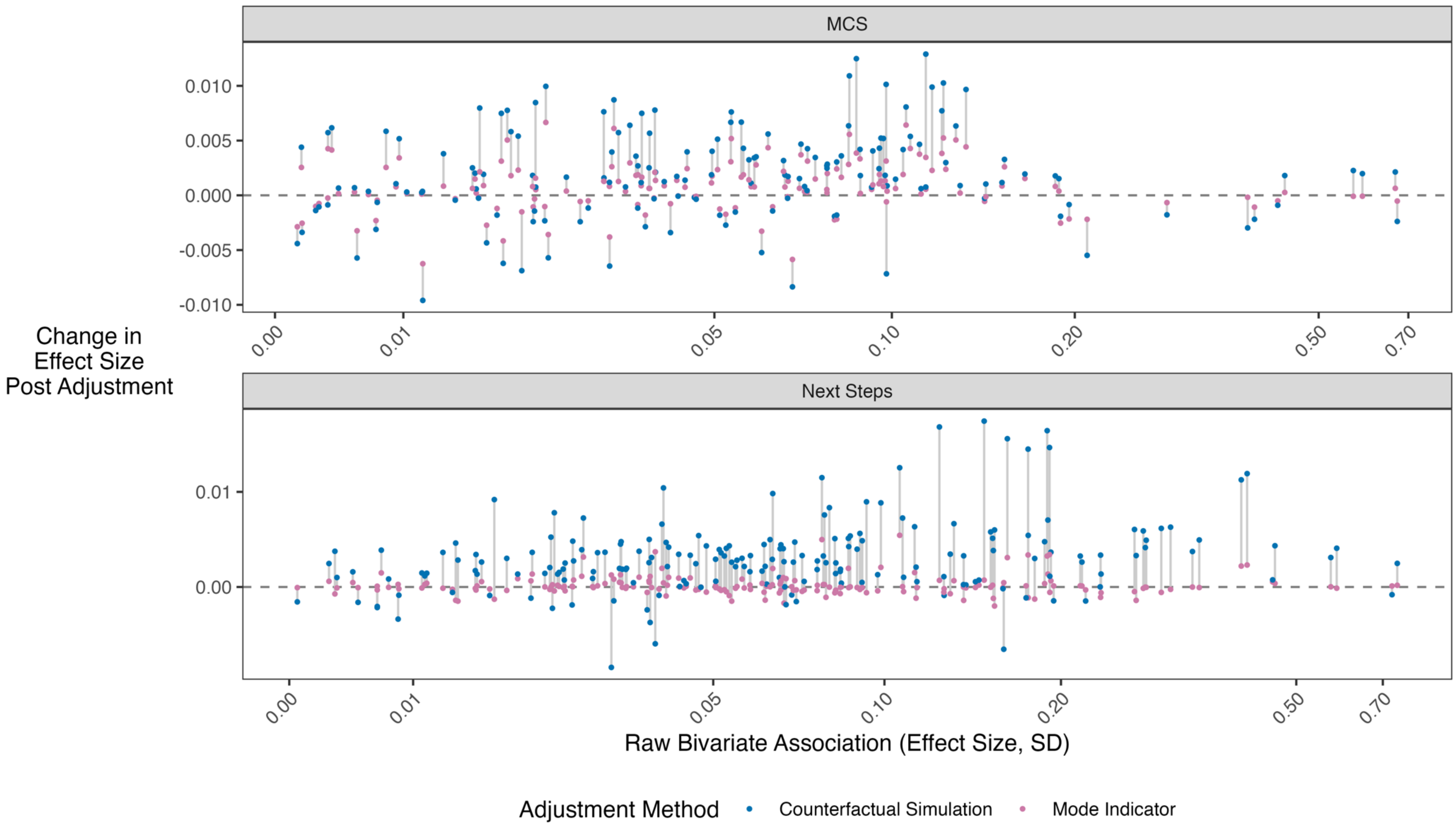

Figure 7: Change in size of association between pairs of variables in MCS and Next Steps when attempting for account for mixed-mode survey design. *X-axis shows the bivariate association on effect size (SD) scale calculated with raw (observed) data (and rescaled to ensure positive signed association). Y-axis shows the change in effect size when conditioning on mode with mode indicator variable (pink dots) or using counterfactual simulation using mode effect estimates from CLSMES Experiment (Section 3). Grey vertical bars added to clarify estimates using same pairs of variables. Note, logarithmic scale on x-axis is used for presentational purposes. Note the magnitudal difference in the scale of the x-axis and y-axis.*

7 Tables

*Table 1: **Description of variable assessed in the mode effects experiment.** Variables indicated with an asterisk ('*') were captured in a self-completion module.*

Domain	Variable	Description	Alternate Estimator
Data Quality	Any Missing Items	Range: 0 - 1.	-
	Number of Missing Items	Range: 0 - 17.	-
	Half-Decadal Reported Age	Categories: No; Yes.	Logit
Demographics	Gender	Categories: Male; Female; Other.	Multinomial Logit
	Age	Range: 20 - 41.	-
	Ethnic Group	Categories: White; Non-White.	Logit
	Region	Categories: North West; North East; Yorkshire & Humberside; West Midlands; East Midlands; South East; East of England; South West; Greater London.	Multinomial Logit
	Living in a Couple	Categories: No; Yes.	Logit
	Has Children	Categories: No; Yes.	Logit
Socioeconomic Position	Highest Education	Categories: GCSE or Lower; Further Education; Higher Education.	Multinomial Logit
	Social Grade	Categories: AB; C1; C2; DE.	Multinomial Logit
	Homeowner	Categories: No; Yes.	Logit
	Current Activity	Categories: Work; Training/Education; NEET.	Multinomial Logit
	Income	Ordered. Categories: < £2k; £2k-£4k; £4k-£11k; £11k-£15k; £15k-£20k; £20k-£23k; £23k-£27k; £27k-£31k; £31k-£35k; £35k-£40k; £40k-£46k; £46k-£50k; £50k-£57k; £57k-£68k; £68k-£83k; £83k+.	Ordered Logit
	Receives Universal Credit	Categories: No; Yes.	Logit
	Receives Disability Benefit	Categories: No; Yes.	Logit
	Financial Literacy*	Ordered. Categories: 0; 1; 2.	Ordered Logit
	Financial Situation	Ordered. Categories: Living comfortably; Doing alright; Just about getting by; Finding it quite difficult; Finding it very difficult.	Ordered Logit
	Having Financial Difficulty	Categories: No; Yes.	Logit
Childhood Circumstances	Mother Has Degree	Categories: No; Yes.	Logit
	Father Has Degree	Categories: No; Yes.	Logit
	Childhood Financial Difficulties*	Categories: No; Yes.	Logit
	# Schools Attended*	Range: 1 - 10.	-

Domain	Variable	Description	Alternate Estimator
Physical and Mental Health	Self-Rated Health	Ordered. Categories: Poor; Fair; Good; Very good; Excellent.	Ordered Logit
	Has Chronic Condition	Categories: No; Yes.	Logit
	Has Limiting Illness	Categories: No; Yes.	Logit
	Has Chronic Psychological Condition	Categories: No; Yes.	Logit
	GAD-2 Likert Score*	Range: 2 - 8.	-
	PHQ-2 Likert Score*	Range: 2 - 8.	-
	Life Satisfaction*	Ordered. Categories: 0 - Not at all; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10 - Completely.	Ordered Logit
Health Behaviours	Exercise Frequency	Range: 0 - 7.	-
	Fruit & Veg Consumption	Range: 0 - 10.	-
	Alcohol Consumption (AUDIT-3)*	Range: 0 - 15.	-
	Smoking Status*	Categories: Never; Ex; Occassional; Regular.	Multinomial Logit
	# Cigarettes Per Day*	Range: 0 - 60.	-
Social Capital	Friends/Family Help Feel Safe & Secure	Categories: No; Yes.	Logit
	Someone Feel Close To	Categories: No; Yes.	Logit
	Someone To Turn To For Advice	Categories: No; Yes.	Logit
	Trust in Others	Ordered. Categories: 0 - Not at all trusting; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10 - Extremely trusting.	Ordered Logit
	Voted in 2019 General Election	Categories: No; Yes.	Logit

Table 2: Descriptive statistics by survey mode in the mode effects experiment

	Variable	Total N (%) / Mean (SD)	Total % Missing	Web N (%) / Mean (SD)	Web % Missing	F2F N (%) / Mean (SD)	F2F % Missing	Video N (%) / Mean (SD)	Video % Missing	Mode Difference p-value
	Sample Size	1,692	-	582 (34.4%)	-	565 (33.4%)	-	545 (32.2%)	-	-
Gender	Male	753 (44.5%)	0.0%	294 (50.5%)	0.0%	226 (40.0%)	0.0%	233 (42.8%)	0.0%	<0.01
	Female	919 (54.3%)		278 (47.8%)		337 (59.6%)		304 (55.8%)		
	Other	20 (1.2%)		10 (1.7%)		2 (0.4%)		8 (1.5%)		
	Age	31.2 (6.1)	0.0%	30.8 (6.0)	0.0%	31.2 (6.2)	0.0%	31.6 (6.0)	0.0%	0.08
Ethnic Group	White	1,549 (91.8%)	0.2%	525 (90.7%)	0.5%	523 (92.6%)	0.0%	501 (92.1%)	0.2%	0.48
	Non-White	139 (8.2%)		54 (9.3%)		42 (7.4%)		43 (7.9%)		
Region	North West	246 (14.5%)	0.0%	82 (14.1%)	0.0%	84 (14.9%)	0.0%	80 (14.7%)	0.0%	1.00
	North East	82 (4.8%)		30 (5.2%)		28 (5.0%)		24 (4.4%)		
	Yorkshire & Humberside	153 (9.0%)		52 (8.9%)		52 (9.2%)		49 (9.0%)		
	West Midlands	185 (10.9%)		61 (10.5%)		62 (11.0%)		62 (11.4%)		
	East Midlands	141 (8.3%)		49 (8.4%)		48 (8.5%)		44 (8.1%)		
	South East	255 (15.1%)		87 (14.9%)		88 (15.6%)		80 (14.7%)		
	East of England	157 (9.3%)		56 (9.6%)		50 (8.8%)		51 (9.4%)		
	South West	150 (8.9%)		53 (9.1%)		49 (8.7%)		48 (8.8%)		
	Greater London	323 (19.1%)		112 (19.2%)		104 (18.4%)		107 (19.6%)		
Highest Education	GCSE or Lower	468 (27.7%)	0.2%	152 (26.2%)	0.3%	177 (31.4%)	0.2%	139 (25.5%)	0.0%	<0.01
	Further Education	827 (49.0%)		291 (50.2%)		280 (49.6%)		256 (47.0%)		
	Higher Education	394 (23.3%)		137 (23.6%)		107 (19.0%)		150 (27.5%)		
Social Grade	AB	298 (17.6%)	0.0%	99 (17.0%)	0.0%	88 (15.6%)	0.0%	111 (20.4%)	0.0%	<0.01
	C1	631 (37.3%)		210 (36.1%)		205 (36.3%)		216 (39.6%)		
	C2	389 (23.0%)		159 (27.3%)		131 (23.2%)		99 (18.2%)		
	DE	374 (22.1%)		114 (19.6%)		141 (25.0%)		119 (21.8%)		

Table 3: Description of (mode selection) variables used in the analysis of cohort data, by cohort

Domain	Variable	MCS	Next Steps
Demographics	Sex	Categories: Male, Female. Sweep of collection: 9mon-3y	Categories: Male, Female. Sweep of collection: 32y
	Ethnic Group	Categories: White, Non-White. Sweep of collection: 9mon-3y	Categories: White, Non-White. Sweep of collection: 13y-16y
	Mother's Age at Birth	Range: 14-48, Mean: 28.9, SD: 5.8. Sweep of collection: 9mon-3y	Range: 14-49, Mean: 28.3, SD: 5.3. Sweep of collection: 16y
	Father's Age at Birth	Range: 15-50, Mean: 32.0, SD: 5.9. Sweep of collection: 9mon-3y	Range: 13-50, Mean: 31.1, SD: 6.1. Sweep of collection: 16y
Socioeconomic Position	Highest Education	Categories: None/Other/NVQ Levels 1-2, NVQ Level 3, NVQ Level 4, NVQ Level 5. Sweep of collection: 23y	Categories: None/Other/NVQ Level 1, NVQ Level 2, NVQ Level 3, NVQ Level 4, NVQ Level 5. Sweep of collection: 32y
	Economic Activity	-	Categories: Education, Training, Employment, NEET. Sweep of collection: 16y
	Verbal Ability	Range: -6.5-3.4, Mean: 0.1, SD: 1.0. Sweep of collection: 11y	-
	Special Educational Needs	Categories: No, Yes. Sweep of collection: 14y	Categories: No, Yes. Sweep of collection: 13y-16y
Childhood Circumstances	Family Structure	Categories: Two Parent, Single Parent. Sweep of collection: 14y	Categories: Two Parent, Single Parent, No Parent. Sweep of collection: 16y
	Family Social Class (NS-SEC)	Categories: Managerial & Professional, Intermediate, Routine & Manual, Not Working. Sweep of collection: 9mon-3y	Categories: Managerial & Professional, Intermediate, Routine & Manual, Not Working. Sweep of collection: 16y
	Parental Highest Qualification	Categories: None/Other, NVQ Level 1, NVQ Level 2, NVQ Level 3, NVQ Level 4/5. Sweep of collection: 9mon-3y	Categories: Degree, HE Below Degree, A Level, GCSE A-C, Level 1 or Below/Other, None. Sweep of collection: 16y
	Family Income	OECD Equivalised Income Above/Below 60% Median. Categories: Not in Poverty, In Poverty. Sweep of collection: 14y	-
	Educational Financial Support	Eligibility for Free School Meals. Categories: No, Yes. Sweep of collection: 14y	Receives Educational Maintenance Allowance. Categories: No, Yes. Sweep of collection: 16y
	Ever Been in Care	-	Categories: Not Been in Care, Been in Care. Sweep of collection: 13y-16y
Health	Self-Rated Health	Categories: Excellent, Very Good, Good/Fair/Poor. Sweep of collection: 17y	Categories: Excellent, Fairly Good, Not very good/not good at all. Sweep of collection: 16y
	Long-Standing Health Condition	Categories: No, Yes. Sweep of collection: 17y	Categories: No, Yes. Sweep of collection: 16y
	Mental (Ill) Health Score	Kessler Depression Score. Range: 0-24, Mean: 7.4, SD:	GHQ-12 Caseness score. Range: 0-12, Mean: 2.2, SD: 2.8. Sweep of collection: 16y

Domain	Variable	MCS	Next Steps
		4.9. Sweep of collection: 17y	
	Body Mass Index	Range: 11.0-49.4, Mean: 23.4, SD: 4.8. Sweep of collection: 17y	-
Health Behaviours	Smoking Status	Categories: Never Tried, Tried Once, Former, Occasional, Daily. Sweep of collection: 17y	-
	Alcohol Use	Occassions drunk alcohol in last 12 months. Categories: Never, 1-2 times, 3-5 times, 6-9 times, 10-19 times, 20+ times. Sweep of collection: 17y	Frequency alcohol consumption in last 12 months. Categories: Never, Occasionally, Every few months, Monthly, 2-3 times per month, 1+ times per week. Sweep of collection: 16y
	Cannabis Use	Occassions smoked cannabis in last 12 months. Categories: Never, 1-2 times, 3-10 times, 10+ times. Sweep of collection: 17y	Ever tried cannabis. Categories: No, Yes. Sweep of collection: 16y
	School Truancy	Truant in last 12 months. Categories: No, Yes. Sweep of collection: 14y	Truant in last 12 months. Categories: No, Yes. Sweep of collection: 15y

8 Statements

8.1 Declaration of Interest

All authors declare no conflicts of interest.

8.2 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 final 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.

8.3 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.

8.4 Data Availability

Data from the Millennium Cohort Study (http://doi.org/10.5255/UKDA-Series-2000031) and Next Steps (http://doi.org/10.5255/UKDA-Series-2000030) are available from the UK Data Service (UKDS). Data from the CLS Mode Effects Experiment is being prepared and will be added to the UKDS in due course. The supplementary files and code used to run the analysis are available at https://osf.io/756vd/.

9 References

Asensio, M., Brown, M., Silverwood, R., Hanson, T., & Durrant, D. (2025). *Assessing the impact of video interviewing on measurement quality: Evidence from an experimental study on mode effects* (Working Paper No. 8). Survey Futures. https://surveyfutures.net/wp-content/uploads/2025/11/working-paper-08-video-interviews.pdf

Calderwood, L., & Sanchez, C. (2016). Next Steps (formerly known as the Longitudinal Study of Young People in England). *Open Health Data*, *4*, e2. https://doi.org/10.5334/ohd.16

Clarke, P. S., & Bao, Y. (2022). Estimating mode effects from a sequential mixed-mode experiment using structural moment models. *The Annals of Applied Statistics*, *16*(3), 1563–1585. https://doi.org/10.1214/21-AOAS1557

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. ed., reprint). Psychology Press.

Connelly, R., & Platt, L. (2014). Cohort Profile: UK Millennium Cohort Study (MCS). *International Journal of Epidemiology*, *43*(6), 1719–1725.

d'Ardenne, J., Bull, R., Das, A., Perera, Z., & Sexton, O. (2025). *How to mitigate against measurement effects when surveys move online: A Measurement Effect Risk Framework (MERF) and web questionnaire guidance* (Survey Practice Guide No. 2; p. 34). Survey Futures. https://surveyfutures.net/wp-content/uploads/2025/06/survey-practice-guide-2-mitigate-measurement-effects-surveys-move-online.pdf

Daikeler, J., Bošnjak, M., & Lozar Manfreda, K. (2020). Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates. *Journal of Survey Statistics and Methodology*, *8*(3), 513–539. https://doi.org/10.1093/jssam/smz008

Dolan, P., & Kavetsos, G. (2016). Happy Talk: Mode of Administration Effects on Subjective Well-Being. *Journal of Happiness Studies*, *17*(3), 1273–1291. https://doi.org/10.1007/s10902-015-9642-8

Edwards, P., & Perkins, C. (2024). Response is increased using postal rather than electronic questionnaires—New results from an updated Cochrane Systematic Review. *BMC Medical Research Methodology*, *24*(1), 209. https://doi.org/10.1186/s12874-024-02332-0

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., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Jäckle, A., Roberts, C., & Lynn, P. (2006). *Telephone versus face-to-face interviewing: Mode effects on data quality and likely causes: Report on phase II of the ESS-Gallup mixed mode methodology project*.

Koerber, H., Brown, M., & Calderwood, L. (Forthcoming). *Evaluating a Look-Up Tool for Occupation Coding: Evidence on Feasibility, Validity, and Reliability*.

Lipsitch, M., Tchetgen Tchetgen, E., & Cohen, T. (2012). Negative control exposures in epidemiologic studies. *Epidemiology (Cambridge, Mass.)*, *23*(2), 351–352. https://doi.org/10.1097/ede.0b013e3182460c23

Lu, H., Gonsalves, G. S., & Westreich, D. (2024). Selection Bias Requires Selection: The Case of Collider Stratification Bias. *American Journal of Epidemiology*, *193*(3), 407–409. https://doi.org/10.1093/aje/kwad213

Martin, P., & Lynn, P. (2011). The effects of mixed mode survey designs on simple and complex analyses. *ISER Working Paper Series, ISER Working Paper Series*, Article 2011–28. https://europeansocialsurvey.org/sites/default/files/2023-06/The%20effect%20of%20mixed%20mode%20survey%20designs.pdf

Piccitto, G., Liefbroer, A. C., & Emery, T. (2022). Does the Survey Mode Affect the Association Between Subjective Well-being and its Determinants? An Experimental Comparison Between Face-to-Face and Web Mode. *Journal of Happiness Studies*, *23*(7), 3441–3461. https://doi.org/10.1007/s10902-022-00553-y

Pople, L., Brown, M., Haselden, L., Gilby, N., Burston, K., & Fitzsimons, E. (Forthcoming). *Web-First or Face-to-Face? Experimental Evidence from the Millennium Cohort Study at the Transition to Adulthood.*

R Core Team. (2025). *R: A Language and Environment for Statistical Computing* (Version 4.5.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Schouten, B., van den Brakel, J., Buelens, B., van der Laan, J., & Klausch, T. (2013). Disentangling mode-specific selection and measurement bias in social surveys. *Social Science Research*, *42*(6), 1555–1570. https://doi.org/10.1016/j.ssresearch.2013.07.005

Sullivan, G. M., & Feinn, R. (2012). Using Effect Size—Or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*, *4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Tomova, G. D., Silverwood, R. J., Tennant, P. W. G., & Wright, L. (2026). How can the use of different modes of survey data collection introduce bias? An introduction to mode effects using directed acyclic graphs (DAGs). *American Journal of Epidemiology*, kwag017. https://doi.org/10.1093/aje/kwag017

Tomova, G. D., Silverwood, R. J., & Wright, L. (2026a). *Mode effects on survey item measurement: A systematic review of the experimental evidence* (Bc7qn_v1). SocArXiv. https://osf.io/preprints/socarxiv/bc7qn_v1/

Tomova, G. D., Silverwood, R. J., & Wright, L. (2026b). *When, and how much, might effect estimates be biased when analysing mixed-mode survey data? The roles of mode effects, mode selection, and mode split.* SocArXiv. https://doi.org/10.31235/osf.io/6sxbq_v1

Tsigaridis, K. G., Gaia, A., Moulton, V., Wright, L., Brown, M., & Silverwood, R. J. (2026). *Mode effects in cognitive assessment: Analysis of an experiment using in-person, video and web data collection*. SocArXiv. https://doi.org/10.31235/osf.io/ue4ta_v1

Understanding Sociey. (2026). *Analysis advice for mixed mode data*. Understanding Society. https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/analysis-advice-for-mixed-mode-data/

Vannieuwenhuyze, J. T. A. (2015). Mode Effects on Variances, Covariances, Standard Deviations, and Correlations. *Journal of Survey Statistics and Methodology*, *3*(3), 296–316. https://doi.org/10.1093/jssam/smv009

Villar, A., & Fitzgerald, R. (2017). Using Mixed Modes in Survey Research: Evidence from Six Experiments in the ESS. In *Values and Identities in Europe* (p. 38). Routledge.

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

Wright, L., Tomova, G. D., Pina-Sánchez, J., & Silverwood, R. J. (Forthcoming). *Quantitative Bias Analysis for Assessing Bias from Survey Mode Effects*.

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?* SocArXiv. https://doi.org/10.31235/osf.io/df4c5_v1

Wu, A. F.-W., Henderson, M., Brown, M., Adali, T., Silverwood, R. J., Peycheva, D., & Calderwood, L. (2024). Cohort Profile: Next Steps—the longitudinal study of people in England born in 1989–90. *International Journal of Epidemiology*, *53*(6), dyae152. https://doi.org/10.1093/ije/dyae152