Survey Research Methods (2021)
Vol. 15, No. 3, pp. 281-298
doi:10.18148/srm/2021.v15i3.7669
European Survey Research Association

Understanding the patterns of mode switching in longitudinal studies

Alexandru Cernat
School of Social Sciences
University of Manchester
United Kingdom

Joseph W. Sakshaug
Institute for Employment Research (IAB),
Ludwig Maximilian University of Munich, and
University of Mannheim

Using mixed-mode data collection is becoming a mainstream way of conducting longitudinal surveys. However, interviewing the same units in a mixed-mode longitudinal design can lead to respondents switching between modes over time. As a result, mode switching behaviors can be correlated with non-response and potentially influence survey responses and estimates of change in longitudinal analyses. This paper investigates the patterns by which people transition from one mode of interview to another in a nationally-representative, sequential mixed-mode (Web and face-to-face) longitudinal study. Using mixed-mode waves 5–10 of the Understanding Society Innovation Panel, we perform a latent class analysis on respondents and their mode switching behaviors. We identify five distinct classes of respondents: Face-to-face respondents, Web respondents, Face to face/late drop-offs, Single mode/early drop-offs and Switchers. Furthermore, we show that these classes differ with respect to respondent characteristics and significantly contribute to the prediction of future wave participation and mode of response, even after controlling for socio-demographic characteristics, interview mode at the previous wave, and previous non-response behavior. Practical implications of these results are discussed and possible strategies to use this information for targeting and correcting for non-response in longitudinal studies are proposed.

Keywords: mixed modes; longitudinal survey; latent class analysis; non-response; face-to-face survey; web survey

1 Introduction

Increasingly surveys are using mixed-mode designs to collect social science data. This involves conducting interviews using a combination of self-administered (e.g. Web) or interviewer-administered (e.g. face-to-face) modes mixed either concurrently or sequentially (de Leeuw, 2005). The use of mixed-mode designs is also common for longitudinal studies. For example, since wave 7 Understanding Society—The UK Household Longitudinal Study (UKHLS) has transitioned from a single-mode Computer Assisted Personal Interviewing (CAPI) design to a sequential mixed-mode (Web and CAPI) design (Bianchi, Biffignandi, & Lynn, 2017). The National Longitudinal Study of Adolescent to Adult Health (or Add Health) also recently transitioned from a single-mode CAPI to a mixed-mode design in wave 5 with Web/mail as primary modes and CAPI used for nonresponse follow-up (Biemer, Harris, Liao, Burke, & Halpern, 2021). Another example is the National Child Development Study (NCDS), which moved to a sequential mixed-mode design (Web-telephone) in the Age 55 survey from a previously used single-mode design (Brown, 2016). While such design decisions are typically driven by costs, there are still many unknowns regarding the impact of mixing modes on data quality. Longitudinal studies such as those highlighted above are mainly concerned with estimating change over time and how such changes are influenced by socio-demographic and other factors. The introduction of a sequential mixed-mode design in a longitudinal study adds an extra dimension that can influence both non-response and measurement error: the shift in the response mode used by respondents over time. Although we know that the mode of response can influence both survey participation (Lynn, 2013) and measurement (Cernat, 2014, 2015), there is a need for further research on the patterns of shifting between sequentially-offered modes in a longitudinal context and how they correlate with these dimensions.

Understanding the patterns that respondents shift from one mode of response to another in a sequential mixed-mode design can be valuable information for survey methodologists for at least three reasons. First, understanding the transitions between modes can better inform different strategies for targeting and other adaptive designs used in longitudinal studies (Lynn, 2017). This is especially important in household surveys where large cost-savings can occur when a significant proportion of household members respond in

Contact information: Alexandru Cernat, University of Manchester, Humanities Bridgeford Street, Manchester M13 9PL, (E-Mail: alexandru.cernat@manchester.ac.uk)

a cheaper mode, such as Web, as opposed to a potentially more expensive mode, such as face-to-face (Bianchi et al., 2017; Wagner, Arrieta, Guyer, & Ofstedal, 2014). Thus, understanding the way in which different groups switch between modes can inform targeting strategies that either push respondents to Web or decrease the chances of having to deploy more expensive modes to obtain a response. Second, the switching of modes can also be related to general non-response propensity. For example, a person who previously participated using the first offered mode in a sequential mixed-mode design, but failed to do so in a subsequent wave, triggering the deployment of a second mode, might indicate their reluctance to participate in this wave and/or further waves of the study. Thus, information about mode switching behaviors may be potentially valuable for correcting for non-response: the shifting mode patterns of respondents in previous waves can be included in non-response models, such as those used in weighting or multiple imputation (Kalton, 1986). And third, understanding mode shifts can also inform issues around measurement error and bias in longitudinal estimates of change. Although measurement error is not explicitly addressed in this study, previous research has shown that the introduction of a mixed-mode design can lead to a spurious increase in estimates of change in longitudinal surveys (Cernat, 2015). Thus, accounting for mode switch patterns can potentially help explain measurement mode effects and inform procedures for correcting for these effects.

Against this backdrop, the present study aims to better understand one of the key characteristics of a mixed-mode longitudinal design: the patterns by which people transition from one mode to another over time in a sequential mixed-mode study. Specifically, this paper investigates the patterns of changing modes by looking at several waves of the UKHLS Innovation Panel. Latent class analysis is used to identify the underlying patterns of change in modes over time. These patterns are used both as dependent variables, to understand who are the respondents in each class, and as independent variables, to predict future wave outcomes, including mode of response, whether a more expensive follow-up mode is needed (because of failure to obtain a response in the first offered mode), and non-response. Findings will inform the design and use of longitudinal sequential mixed-mode studies, such as the UKHLS and others.

2 Background

The shift from a single mode to a mixed-mode design, where a combination of interview modes is used to collect data, is one of the key challenges of contemporary longitudinal studies. The decision about which modes to use and how to implement them can have a big impact, influencing costs, non-response (and attrition), and measurement error (Bianchi et al., 2017; Biemer et al., 2021; Cernat, 2015; Converse, Wolfe, Huang, & Oswald, 2008; Dillman & Christian, 2005; Klausch, Schouten, & Hox, 2017; Lynn, 2013; Millar & Dillman, 2011; Sakshaug, Cernat, & Raghunathan, 2019, 4; Sakshaug, Yan, & Tourangeau, 2010; Wagner et al., 2017). Costs are one of the main driving forces behind the shift from single- to mixed-mode designs. Sequential mixed-mode designs, in particular, can be cost-effective when they start with a relatively inexpensive (usually self-administered) mode, such as Web, and follow up the remaining nonrespondents with a more expensive (usually interviewer-administered) mode, such as CAPI (Bianchi et al., 2017; Wagner et al., 2014). As such, there is a growing literature on "pushing" respondents to Web in order to save costs (Dillman, 2017; Dillman, Smyth, & Christian, 2014; Millar & Dillman, 2011). Sequential mixed-mode designs can also be advantageous for minimizing non-response as they offer a greater chance of reaching certain subgroups of the population that have low chances of being contacted or responding in a particular mode (de Leeuw, 2005; Lynn, 2013; Sakshaug et al., 2019, 4; Sakshaug & Eckman, 2017; Vannieuwenhuyze, 2014).

In the case of the UKHLS and its Web-CAPI sequential mixed-mode design there are at least four distinct mechanisms that can potentially influence whether or not respondents shift between these two modes over time. The first and second mechanisms relate to respondents who respond in the same mode in each wave ("mode stayers"), whereas the third and fourth mechanisms involve respondents who start in one mode and switch to the other mode over the course of the longitudinal study ("mode switchers"). The first mechanism relates to respondents who always complete the survey using the CAPI mode. These respondents are cooperative but unwilling to complete surveys online. They do not value the advantages of responding online (as opposed to CAPI) and may require further persuasion to make the switch to the Web mode. In the sense of leverage-saliency theory (Groves, Singer, & Corning, 2000), they do not place a high level of importance on the Web mode and would rather not participate than comply with the Web survey request. These individuals are likely to have negative attitudes towards the internet and computers and are not comfortable navigating a Web survey instrument (De Leeuw & Hox, 2011; Fan & Yan, 2010; Fuchs & Busse, 2009; Hoogendoorn & Daalmans, 2009; Kwak & Radler, 2002). Older individuals and those with low "digital affinity" tend to fall in this category and are more likely to complete the survey using an "offline" mode of data collection (Herzing & Blom, 2018).

The second mechanism relates to respondents who always complete the survey online. These respondents are cooperative and might participate in any mode offered at the beginning of the field period. On the other hand, they may be hard-to-reach and work during the day and/or evenings and therefore prefer the convenience of participating online. Moreover, they may not like interacting face-to-face or over the telephone and generally prefer the asynchronous communication channel that is inherent to the Web mode. Respondents in this group are likely to be younger and have higher digital affinity compared to CAPI-only respondents (Herzing & Blom, 2018).

Relevant to mode switchers, the third mechanism concerns respondents who bypass the Web mode initially and use the CAPI mode in the early waves and then switch to the Web mode in later waves. This switching behavior may be influenced by an increase in digital affinity and development of more positive attitudes towards using the internet over time (Herzing & Blom, 2018). It is conceivable that this group would be inclined to switch from CAPI to the less-expensive Web mode sooner if given more encouragement, training, or a larger incentive to switch. Thus, this group may be a viable target for "push-to-Web" strategies (Dillman, 2017; Millar & Dillman, 2011). The fourth mechanism relates to the reverse mode switch pattern, where respondents switch from the initial Web mode to the subsequently offered CAPI mode over time. Such behavior could be a sign that respondents' interest in the study is dwindling and that they are on the verge of attriting. This group would therefore be a prime target for attrition avoidance strategies (Bianchi & Biffignandi, 2017; Lynn, 2017).

However, what is currently unknown is the extent to which different mode switch patterns manifest in longitudinal studies and the characteristics of respondents underlying those patterns. Additionally, while it is important to know how mode switching occurs, it is also important to know when it occurs. This information is useful for budgetary planning of subsequent waves as well as the optimal timing of targeted intervention strategies to reduce attrition and entice respondents to respond in the Web mode sooner, minimizing the use of more expensive follow-up modes. Further, knowing respondents' past mode switching behaviors provides an additional source of auxiliary data that can potentially be exploited to improve weighting and imputation models to adjust for non-response bias. Although longitudinal studies possess large amounts of background information on respondents (in addition to their previous non-response behavior) that can be used in non-response adjustment models, additional information on mode switching patterns may improve the fit of these models even further. Such a finding would be a useful contribution to survey practice, where stronger auxiliary variables are highly sought after for purposes of non-response adjustment (Smith, 2011). Finally, knowing the patterns of mode switching and who belongs to each pattern will inform researchers on the possible extent that mixing modes might explain measurement error in estimates of change.

The proposed research in the present study is used to address most of these aspects by identifying the underlying patterns of mode changes in sequential mixed-mode (Web-CAPI) waves 5–9 of the UKHLS Innovation Panel. In this research, latent class analysis is used to find the latent patterns of mode change. This approach helps summarize complex patterns in classes or groups of respondents. Similar approaches have been used to identify classes of attriters in longitudinal surveys (Lemay, 2009; Lugtig, 2014). Unlike ad-hoc approaches, latent class analysis enables one to make inferences about parameters of interest (by producing standard errors and p-values) and has a number of fit indices that can be used to compare models. The approach is especially useful when we have multiple types of participation outcomes and waves making the inclusion of the original variables too complex.

The classes created using this approach are then used both as dependent variables, in order to understand the types of individuals and households that are found in each class, and as independent variables to see if they predict future participation, mode of interview, and deployment of the more expensive CAPI mode in wave 10 (a hold-out sample). Although mode switch patterns have important implications for both non-response and measurement error in longitudinal studies, as mentioned above, we focus solely on the issue of non-response and leave the topic of measurement error for future work. Specifically, we address the following research questions (RQ):

RQ1. What are the underlying classes of mode switching in a sequential mixed-mode (Web-CAPI) design?

The first part of the research will investigate the latent classes of respondents based on the modes of interview in which they participated in waves 5-9 of the Innovation Panel (the mixed-mode sample).

RQ2. What types of respondents and households are found in each mode switch class?

The second research question investigates the types of respondents and households that are more likely to be in each of the identified latent classes. The findings from this research question can be used to inform targeting or adaptive designs in the main UKHLS survey as well as other longitudinal studies. For example, if respondents that are currently under-represented tend to be in a group that shows more shifting from one mode to another then this would be a group that could potentially be targeted with extra incentives or effort to push them to the Web.

RQ3. Can the identified mode switch classes predict future wave outcomes: mode of interview, non-response, and whether an interviewer-administered (CAPI) follow-up mode is needed?

Lastly, we investigate whether the identified classes from waves 5–9 are also predictive of future outcomes for Wave 10 (which is a hold-out sample), including mode of response, the need for an interviewer-administered follow-up mode, and non-response. If these classes show predictive power (after controlling for socio-demographic predictors and previous non-response behavior) then they could be used to improve non-response models, such as those used for weighting, imputation, as well as survey planning and targeting strategies.

3 Data and Methods

3.1 Data source

The Understanding Society Innovation Panel is an annual longitudinal survey representative of the UK population (Jäckle, Al Baghal, Burton, Kaminska, & Lynn, 2018). The panel began data collection in 2008 and is mainly used for methodological research and to inform the main Understanding Society survey. The initial sample was based on postal codes and selected from a stratified sample based on government office regions, proportion of non-manual workers, and population density. Based on these strata, 120 sectors were selected with probability proportional to population size. Within each sector 23 addresses were selected for data collection, leading to a sample of 2,760 addresses. The survey interviewed all members of the household that were 16 years or older using face-to-face (CAPI) in waves 1 to 4, with the exception of wave 2 where a random sub-sample of respondents were included in a telephone-face-to-face sequential design. In wave 4, a further 960 addresses were added as a refreshment sample.

Starting with wave 5, a random two-thirds of the panel was allocated to a sequential mixed-mode (Web followed by face-to-face) survey. From wave 6 onward a "mop-up" stage was introduced at the end of the fieldwork using telephone. For the purposes of this paper, these few cases (ranging from 0–20 individuals with an average of 9.2 per wave) are removed. In the present analysis, we investigate respondents who were eligible to participate in the mixed-mode design in wave 5, which yields a total analytic sample size of 1,636 individuals after excluding youth respondents and other ineligible sample members. An advantage of investigating waves 5 to 9 is that data collected in the previous wave (wave 4) can be used to predict membership and assess respondent composition in each mode switching class. As the previous wave data were collected using a single-mode (face-to-face) design, there is no issue of mode effects confounding the predictive ability of the wave 4 variables.

In wave 1 of the Innovation Panel the household response rate was 59% while the conditional individual response rate was 88.9% (Jäckle et al., 2018). The household response rate for the refreshment sample in wave 4 was 54.5% while the conditional response rate was 81%. For waves 5 to 9 the household response rates conditional on previous wave participation for the mixed-mode sample were 76.7%, 83.8%, 80.4%, 86.5% and 86.1%, respectively. For full response rates and details of the sampling procedures, see (Jäckle et al., 2018). For our analysis we did not use any weighting.

3.2 Methods

To answer the three research questions, we use latent class analysis based on the mode of response in waves 5–9 of the Innovation Panel (the mixed-mode sample). This analysis enables us to identify the underlying classes, or clusters, that explain participation and modes of interview over time. The main formula for latent class analysis is (Collins & Lanza, 2010):

$$P(Y = y) = \sum_{c=1}^{C} \gamma_c \prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{j,r_j \vee c}^{I(y_j=r_j)}$$

where the mode of interview at each wave (Y) is decomposed into the different classes (γ_c represents the size of the classes) and the propensity to participate in a particular mode at each wave (ρ) is conditional on the class membership. In the formula c indexes the classes while j indexes the variables, and r_j indexes the response category.

The latent class approach has a number of advantages: 1) it summarizes the data in a principled way (e.g., using fit indices), 2) takes into account uncertainty in class membership and corrects for missing data (using Full Information Maximum Likelihood), 3) is flexible as it can be used with dichotomous or nominal dependent variables, and 4) can be included in structural models both as a dependent or independent variable.

To answer research question two, we use a multinomial regression model to predict class membership conditional on individual and household characteristics collected at wave 4. The three-step approach (Vermunt, 2010) is used as implemented in Mplus 8.2 (Asparouhov & Muthén, 2012) in order to treat the classes as fixed (i.e., they do not change due to the independent variables) while taking into account the uncertainty in class membership. Finally, for research question three we use the identified classes as independent variables in three different logistic regression models of survey participation in wave 10 (the out-of-sample data): the first one modeling the likelihood of participation in wave 10 of the Innovation Panel, the second modeling the mode of interview in wave 10, and the third modeling the likelihood that a face-to-face interview was attempted in wave 10. Each regression model includes a number of socio-demographic control variables from wave 9, an indicator of prior non-response behavior, and is fitted with and without the latent class variables to assess their performance in the models. Standard diagnostics are used to assess the extent to which model fit is improved by including the latent class information.

4 Results

4.1 Identifying class membership

We first describe the distributions of the possible response outcomes and the mode switches over time (Figure 1). This figure shows the five possible outcomes observed in waves

5–9 of the Innovation Panel: response by Web, response by face-to-face, refusal, other nonresponse (also includes non-contact) and not-issued (counts are provided in Appendix A1). It is apparent that most respondents answer by Web and this proportion is stable in time. Furthermore, it is evident that a large proportion of respondents switch from face-to-face to Web mode. This is especially true from wave 5 to wave 6. There is also a smaller degree of switching in the opposite direction, from Web to face-to-face, and this is especially small between waves 5 and 6 and between waves 8 and 9. We also observe an increase of cases that were not issued. These were typically due to non-response in prior waves and to the moving out of the scope of the study. This is accompanied by a decrease in other types of non-response.

Using the outcomes presented in Figure 1 over the five waves as dependent variables we estimate five different latent class models starting from a simpler one, with only two classes, to a more complex one, with six classes. Table 1 presents the fit indices for the different models. BIC (smaller is better), AIC (smaller is better), and Entropy (larger is better) as well as the class profiles are used to determine the number of classes. The three fit indices point towards the six-class model as the best fitting one (although Entropy only marginally so). We investigated the five and six class solutions to understand how they are different. The six-class solution is similar to the five class one but splits one of the classes in two and leads to two small classes (under 10%). Given the latent class separation (Collins & Lanza, 2010) and small class sizes we opted for the five class solution.

In order to understand the latent classes, we recreate Figure 1 by showing the mode switch patterns in the Innovation Panel but this time separately by each class (class membership being based on the class with the highest likelihood). Figure 2 shows these patterns by class together with class name and prevalence. The solution includes two "stayer" classes where respondents tend to consistently participate in the same mode. The largest class, which we call "Web respondents", includes around 35% of the sample and is comprised mainly of respondents who participate in all of the waves using the Web mode. The other stayer class is called "Face to face respondents", which includes respondents who tend to participate mainly using the face to face mode. They represent around 14% of the sample.

The solution includes also three classes defined by their movements between different modes and outcomes over time. For example, the "Switchers" class (22% of the sample) includes people that often change their status between Web, face to face and non-response. The last two classes include higher numbers of non-respondents and non-issued cases. The "Face to face/late drop-offs" (12% of the sample) includes respondents who tend to participate by face to face in waves 5 and 6 and then shift to non-response or Web by wave 9. The final class, "Single mode/early-drop-offs" (16.5% of the sample) includes people who tend to participate in one, or at most two, waves before they become non-respondents and are not issued anymore.

4.2 Predicting class membership

To better understand the composition of the predicted classes we use 19 variables measured in wave 4 to predict class membership (descriptive statistics are provided in Appendix A2). As mentioned in the methods section, the three-step approach is used. In this way, the class profiles are fixed and are not influenced by these independent variables. At the same time, the uncertainty around the class membership is accounted for. The fitted model estimates a multinomial regression in which the classes are the dependent variables and the wave four variables are the independent ones. For the full model results see Appendix A3 and, to facilitate interpretation, figures depicting the observed probabilities conditional on class membership can be seen in Appendix B1.

If we compare the two "stable" classes, "Web respondents" and "Face to face respondents" (reference), we observe some interesting patterns. Respondents that have a partner, a higher degree, use the internet daily and use a mobile, and stated that they prefer Web as a mode of survey participation in wave 4 are more likely to be in the Web class. These patterns paint a picture of a highly educated, tech savvy and willing participant. In contrast, respondents who lived in London or the North of the UK and required more than 6 telephone calls in wave 4 were less likely to belong to the Web class.

Comparing the "non-stable" classes to the "face-to-face respondents" class (reference), we observe that respondents with A level education, have missing data on the question about internet use in wave 4, and prefer answering by mail or Web are more likely to belong to the "Switchers" class. Respondents who are less likely to belong to the "Switchers" class tend to be those in the 56–75 age category and those who volunteer, indicating that the "Switchers" may be less cooperative with survey requests compared to respondents who tend to always respond via face-to-face. Respondents who received between 3–6 telephone calls in wave 4 were more likely to be in the "Face to face/late drop-offs" class whereas those living in London were less likely to be in this class. Finally, respondents with an A level education, older than 75, and declared Web or mail as their preferred mode were more likely to belong to the "Single mode/early drop-offs" class, and respondents who lived in London or the North of the UK were less likely to be in this class.

4.3 Using the classes to predict future outcomes

Now that we have a basic understanding of the five classes we investigate if they are useful in predicting future outcomes. This serves both as a validation for the classes and as a way to show how they could be used in non-response

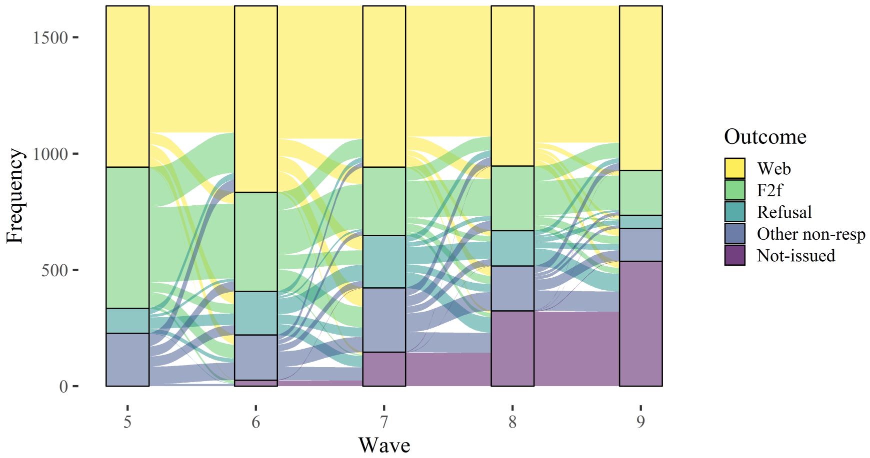

Figure 1. Visual representation (Sankey diagram) of mode proportions and non-interview outcomes in waves 5–9 of the Innovation Panel. The bars show the frequency of each outcome by wave. The lines between the wave bars represent the transitions from one outcome to another. Larger lines indicate more transitions.

Table 1
Fit indices of latent class models with 2 to 6 classes based on participation and mode of interview in waves 5–9 of the Innovation Panel.

Nr. classes	LL	Parameters	n	AIC	BIC	Entropy
2	−10957	39	1636	21991	22202	1.00
3	−10310	59	1636	20738	21057	0.92
4	−9883	79	1636	19924	20351	0.89
5	−9625	99	1636	19448	19983	0.89
6	−9491	119	1636	19221	19863	0.90

prediction models either to correct for non-response or for targeting certain subgroups and planning future waves.

To test the predictive power of the classes we model three outcomes from wave 10 of the Innovation Panel, which were not included in the latent class model (and thus can be used for validation). The three dependent variables are: 1) if sample members were non-respondents in wave 10; 2) for those that did respond in wave 10, in which mode they responded in, and 3) whether a face-to-face follow-up interview was attempted (because a Web response could not be obtained in the initial phase of data collection). These dependent variables are each used in two types of logistic regression models. In the first models (1) we include information from wave 9 and earlier to predict the two outcomes. This is similar to the normal procedure used to model non-response in the UKHLS (Jäckle et al., 2018). The model we consider includes the following covariates: education, age, gender, having a partner, living in an urban area, living in London or in the North, an indicator of non-response in any of the earlier waves 5–9, and responding by face-to-face in wave 9 (descriptive statistics are available in Appendix A4). The motivation for including socio-demographic and previous response behavior as control variables stems from their usage in the survey literature for modeling nonresponse/attrition (Kern, Weiss, & Kolb, 2019; Lugtig, 2014; Siegers, Belcheva, & Silbermann, 2019; Vandecasteele & Debels, 2007; Watson & Wooden, 2009), mode of response (Lynn, 2020; Roberts, Joye, & Ernst Stähli, 2016), and the fact that these variables are observed for the entire analysis sample and are not restricted to subgroups. In the second models, we extend the previous models to include the identified latent classes as independent variables with the "Web respondents" class as the reference group.[1] Here, we are interested to know whether the classes are significant predictors of the three outcomes conditional on the control variables, and how much extra variation they explain.

Table 2 presents the results for the six models (three outcomes for two types of models). The results show that the latent classes are statistically significant when added to the

[1]We investigated the multi-collinearity of the covariates in the models and all the VIF scores were below 10, indicating no signs of overly high correlations.

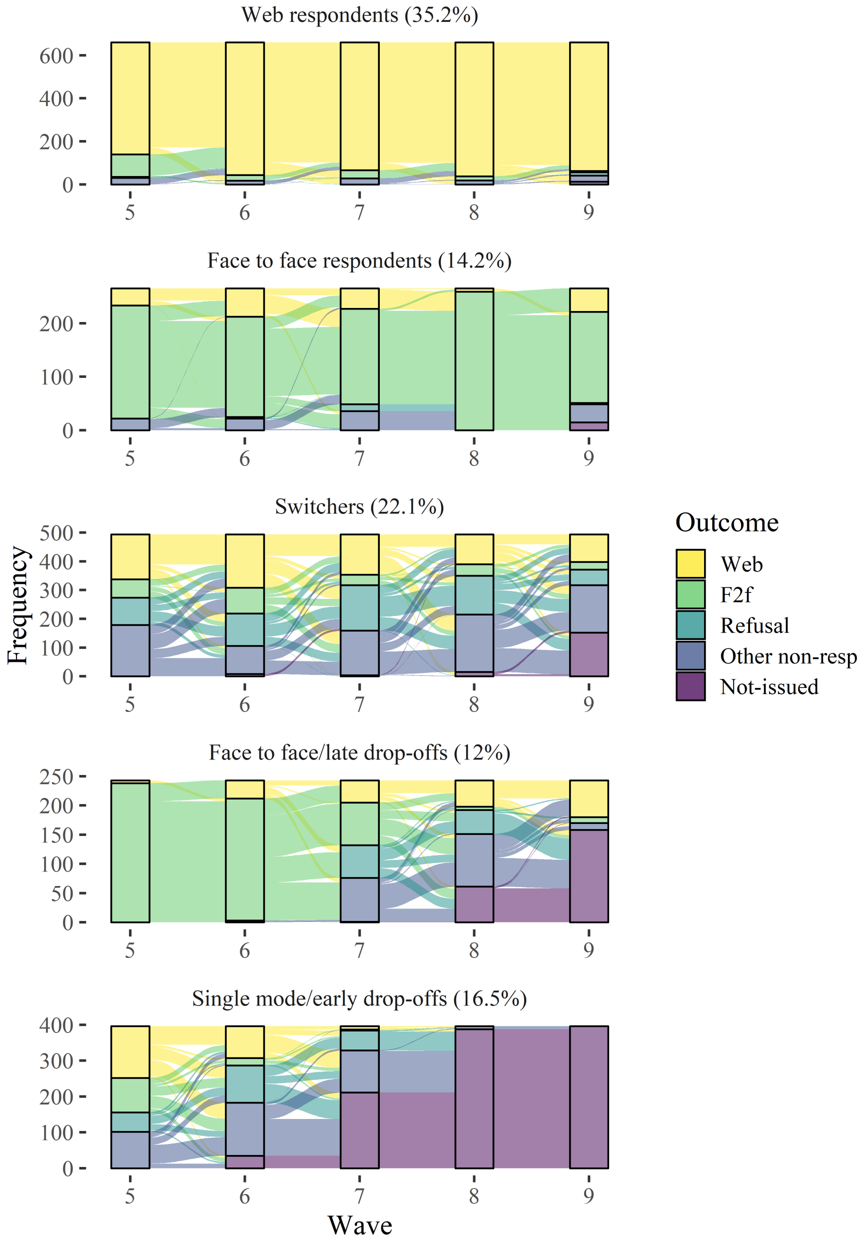

Figure 2. Visual representation (Sankey diagram) of mode proportions and non-interview outcomes in waves 5–9 of the Innovation Panel based on the predicted class membership. The bars show the frequency of each outcome by wave. The lines between the wave bars represent the transitions from one outcome to another. Larger lines indicate more transitions.

three models. The model fits also show overall improvements. The AIC, Area Under the Curve (AUC), and the Root Mean Square Error (RMSE) always show a modest improvement in the fit of the second model over the first model. The BIC shows the same pattern for two of the models, the model predicting mode of response and the likelihood of being offered the face-to-face mode, but not for the model predicting non-response in wave 10.

Table 2 highlights some of the differences between the classes. For example, respondents in the "Face to face/late drop-offs" class are more likely to be non-respondents in wave 10 compared to the "Web respondents" class. The other classes are not predictive of non-response. As expected, all other classes are more likely to participate in the face-to-face

Table 2
Logistic regression coefficients predicting non-response, mode of interview, and being offered the face-to-face mode in wave 10 of the Innovation Panel with no latent classes (1) and with latent classes (2) as covariates.

	Nonresponse		Face to face		Offered Face to face	
	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	−2.69***	−2.78***	−1.42**	−1.75***	−1.00**	−1.20***
	(0.52)	(0.53)	(0.44)	(0.46)	(0.35)	(0.36)
Edu: A level	0.56	0.52	−0.20	−0.21	0.09	0.04
	(0.36)	(0.37)	(0.32)	(0.33)	(0.25)	(0.26)
Edu: GCSE	0.65	0.66	0.18	0.08	0.34	0.31
	(0.36)	(0.36)	(0.30)	(0.31)	(0.25)	(0.25)
Edu: Other	0.47	0.37	0.25	0.16	0.39	0.29
	(0.53)	(0.54)	(0.40)	(0.40)	(0.34)	(0.34)
Age: 36-55	−0.43	−0.47	−0.48	−0.30	−0.44	−0.33
	(0.37)	(0.38)	(0.35)	(0.36)	(0.28)	(0.29)
Age: 56-75	−1.24**	−1.27**	−0.85*	−0.79*	−1.01***	−0.97**
	(0.44)	(0.44)	(0.37)	(0.37)	(0.30)	(0.31)
Age: >75	−1.57	−1.63	−0.03	0.07	−0.23	−0.22
	(1.08)	(1.08)	(0.53)	(0.55)	(0.46)	(0.47)
Female	−0.24	−0.31	−0.06	−0.03	−0.14	−0.17
	(0.28)	(0.28)	(0.24)	(0.25)	(0.20)	(0.20)
Partner	−0.06	0.06	−0.13	−0.10	−0.21	−0.16
	(0.34)	(0.35)	(0.27)	(0.28)	(0.23)	(0.23)
Urban	0.40	0.39	−0.46	−0.57*	−0.22	−0.32
	(0.35)	(0.36)	(0.26)	(0.28)	(0.22)	(0.23)
London	−0.43	−0.30	0.15	0.12	−0.19	−0.19
	(0.57)	(0.58)	(0.45)	(0.45)	(0.39)	(0.39)
North	−0.42	−0.38	−0.37	−0.51	−0.47*	−0.56**
	(0.29)	(0.30)	(0.26)	(0.27)	(0.21)	(0.21)
Wave 9 f2f	−0.53	−0.33	3.60***	2.38***	3.07***	2.09***
	(0.42)	(0.52)	(0.27)	(0.36)	(0.25)	(0.34)
Any non-response	1.35***	1.09**	0.36	−0.01	0.90***	0.56*
	(0.28)	(0.35)	(0.27)	(0.35)	(0.21)	(0.27)
Class: Face to face resp.	-	−0.30	-	1.90***	-	1.45***
		(0.58)	-	(0.38)	-	(0.34)
Class: Face to face/late drop-offs	-	1.29**	-	1.26*	-	1.37***
		(0.44)	-	(0.50)	-	(0.37)
Class: Switchers	-	0.36	-	1.27**	-	0.87*
		(0.43)	-	(0.47)	-	(0.35)
AIC	421.97	418.61	525.76	506.62	731.04	711.88
BIC	487.97	498.75	590.61	585.36	797.04	792.02
Log Likelihood	-196.99	-192.30	-248.88	-236.31	-351.52	-338.94
Deviance	393.97	384.61	497.76	472.62	703.04	677.88
Num. obs.	824	824	759	759	824	824
RMSE	0.69	0.68	0.81	0.79	0.92	0.91
AUC	0.77	0.78	0.86	0.89	0.83	0.85

Reference groups are respondents with: degree, under 36, male, no partner, rural, other area, Web in wave 9, continuous participant and Class: "Web respondents". The "Class: Single mode/early drop-offs" was excluded due to the small number of eligible respondents.
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

mode in wave 10 compared to the "Web respondents" class. The same pattern is evident when looking at the likelihood of being offered face-to-face in wave 10—all the classes are more likely to be offered the face to face follow-up mode compared to the "Web respondents" class.

5 Discussion

In this paper we investigated the mode switching patterns of respondents in a nationally-representative longitudinal mixed-mode (Web followed by face-to-face) survey. We used five waves (waves 5–9) of the Understanding Society Innovation Panel to identify latent classes based on survey participation and mode of data collection used.

Five classes of respondents were identified. Two of the classes are "stayers." These are respondents that have a high likelihood of participating and they tend to always answer in the same mode. They represented around half of the sample with around 35% almost always answering by Web and approximately 14% always answering by face-to-face. The remaining three classes were defined by their change in outcomes. For two of the classes the switching led to non-response and to not being issued anymore. These two classes, "Single mode/early drop-offs" and "Face to face/late drop-offs," are mainly distinguished by the mode of participation and the timing of the drop-off. They represented 12% and 16.5% of the sample, respectively. The last class, "Switchers," comprised around 22% of the sample. This is a group of respondents which tends to switch often between different modes and non-response outcomes.

Predicting the classes using socio-demographic information from wave 4 provided further insights about the types of respondents belonging to the classes. For instance, when compared against the "Face-to-face respondents" reference class, respondents who were partnered, highly educated, tech savvy, and indicated they prefer the Web mode were more likely to belong to the "Web respondents" class. Respondents with A level education, missing data on a question about internet use, and those who preferred the Web or mail survey modes were more likely to belong to the "Switchers" class, whereas older people and those who volunteered were less likely to be in the "Switchers" class (in favor of the "Face-to-face respondents" class). Older respondents and those who received many telephone contact attempts in wave 4 were more likely to belong to the "Single mode/early drop-offs" class and "Face to face/late drop-offs" class, respectively, as opposed to the "Face-to-face respondents" reference class.

Lastly, we investigated the predictive power of the five latent classes. Overall, the classes were significant predictors of three subsequent wave outcomes: survey non-response, mode of interview, and whether a face-to-face follow-up mode was deployed in wave 10 of the Innovation Panel. The latent classes led to an improvement for most of the fit indices we investigated, even after controlling for previous-wave interview mode, participation in earlier waves, and socio-demographic control variables. Compared to the "Web respondents" reference class, all classes were associated with being offered the face-to-face follow-up mode and participating in the face-to-face mode in wave 10. Only the "Face to face/late drop-offs" class was associated with non-response in wave 10.

Our findings demonstrate that latent class analysis might prove to be a useful tool for survey methodologists in understanding the patterns of mode switching in a longitudinal survey. Identifying such classes can provide a new and useful source of auxiliary information that may be considered as a complement to existing auxiliary data sources for non-response adjustment as well as targeting and early intervention strategies (Bianchi & Biffignandi, 2017; Lynn, 2017). In the case of the UKHLS Innovation Panel, respondents that are identified as mode stayers (either Web or face-to-face) probably require no intervention. They show a high likelihood of participation and a low likelihood of switching or non-response. The other three classes might present good avenues for targeting. For example, the "Switchers" class comprises participants who tend to fluctuate between different modes and response outcomes. They might benefit from a push-to Web or even an increase in their incentives to encourage longer-term participation. The "Single mode/early drop-offs" class and the "Face to face/late drop-offs" class seem to be in the process of disengaging from the study, with the former class developing more quickly than the latter class, and could also benefit from active interventions to increase retention.

As with all research, this study has some limitations. Firstly, this research was conducted on a single longitudinal study with a particular mixed-mode design (sequential Web-face-to-face) running in the UK. Although other longitudinal studies run similar Web-first mixed-mode designs, such as Next Steps (previously known as the Longitudinal Study of Young People in England Calderwood & Sanchez, 2016) and the National Child Development Study (Brown, 2016), further research is needed to determine whether our proposed latent class approach is useful in other survey contexts to identify distinct mode classes that are predictive of survey outcomes. Fortunately, the proposed approach can be easily implemented in any given mixed-mode longitudinal survey to identify mode classes. Another limitation is the use of only one hold-out wave (wave 10) to evaluate the predictive properties of the latent classes. In practice, it would be useful to know whether a fixed set of classes can reliably predict survey outcomes over multiple future waves of a study. This issue is particularly important for survey planning as budget and design decisions for future waves of longitudinal studies are often made well in advance (sometimes many years in advance) of actual data collection.

Future research should consider investigating the utility of this approach in improving correction for non-response, such as weighting. If the use of latent classes can improve predictions of participation (as we showed here) or make prediction models more parsimonious then they should be considered as auxiliary variables for non-response adjustment procedures (Smith, 2011), assuming that the latent classes are also correlated with the substantive variables of interest, which is a requirement for optimal adjustment (Kreuter & Olson, 2011; Little & Vartivarian, 2005). Finally, researchers should consider mode switching patterns and the use of latent classes in the context of studying measurement mode effects and adjusting for these effects (de Leeuw, 2005; Jäckle, Gaia, & Benzeval, 2017). Just as mode switching and latent classes are correlated with participation, mode of interview, and the deployment of certain modes, they may also be correlated with substantive responses and estimates of change reported by respondents over time. This possibility raises important measurement error issues in mixed-mode longitudinal designs which should be given high priority in future research. Additionally, future research should look into confirming such types of classes in other studies and using other clustering approaches to enhancing our understanding of what causes these patterns. Finally, the classes identified here could be used together with other experiments carried out in the Innovation panel in order to investigate potential avenues for targeting.

6 Acknowledgements

This research was funded by the Understanding Society Methods Fellowship (Nr. 120720).

References

Asparouhov, T., & Muthén, B. (2012). Auxiliary variables in mixture modeling: A 3-step approach using Mplus. Mplus Web Note No. 15; pp. 1–39). Retrieved from https://statmodel.com/examples/webnotes/AuxMixture%5C_submitted%5C_corrected%5C_webnote

Bianchi, A., & Biffignandi, S. (2017). Targeted letters: Effects on sample composition and item non-response. *Statistical Journal of the IAOS*, *33*(2), 459–467.

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.

Biemer, P. P., Harris, K. M., Liao, D., Burke, B. J., & Halpern, C. T. (2021). Modeling mode effects for a panel survey in transition. In A. Cernat & J. Sakshaug (Eds.), *Measurement error in longitudinal data*. Oxford: Oxford University Press.

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., London.

Calderwood, L., & Sanchez, C. (2016). Next steps (formerly known as the longitudinal study of young people in England). *Open Health Data*, *4*(1), e2. doi:10.5334/ohd.16

Cernat, A. (2014). The impact of mixing modes on reliability in longitudinal studies. *Sociological Methods & Research*, 1–31. doi:10.1177/0049124114553802

Cernat, A. (2015). Impact of mixed modes on measurement errors and estimates of change in panel data. *Survey Research Methods*, *9*(2), 83–99. doi:10.18148/srm/2015.v9i2.5851

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* (1st ed.). New Jersey: Wiley-Blackwell.

Converse, P. D., Wolfe, E. W., Huang, X., & Oswald, F. L. (2008). Response rates for mixed-mode surveys using mail and e-mail/web. *American Journal of Evaluation*, *29*(1), 99–107. doi:10.1177/1098214007313228

De Leeuw, E. D., & Hox, J. J. (2011). Internet surveys as part of a mixed-mode design. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the internet: Advances in applied methods and research strategies* (pp. 45–76). New York: Routledge/Taylor & Francis Group.

de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, *21*(2), 233–255.

Dillman, D. A. (2017). The promise and challenge of pushing respondents to the web in mixed-mode surveys. *Statistics Canada*, *43*(1), 3–30.

Dillman, D. A., & Christian, L. M. (2005). Survey mode as a source of instability in responses across surveys. *Field Methods*, *17*(1), 30–52. doi:10.1177/1525822X04269550

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method.* New Jersey: John Wiley & Sons.

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, *26*(2), 132–139. doi:10.1016/j.chb.2009.10.015

Fuchs, M., & Busse, B. (2009). The coverage bias of mobile web surveys across European countries. *International Journal of Internet Science*, *4*(1), 21–33.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: Description and an illustration. *The Public Opinion Quarterly*, *64*(3), 299–308.

Herzing, J. M., & Blom, A. G. (2018). The influence of a person's digital affinity on unit nonresponse and attrition in an online panel. *Social Science Computer Review*, *37*(3), 404–424.

Hoogendoorn, A., & Daalmans, J. (2009). Nonresponse in the recruitment of an internet panel based on probability sampling. *Survey Research Methods*, *3*, 59–72.

Jäckle, A., Al Baghal, T., Burton, J., Kaminska, O., & Lynn, P. (2018). *Understanding society – the UK household longitudinal study, innovation panel, waves 1–9, user manual.* Colchester: University of Essex.

Jäckle, A., Gaia, A., & Benzeval, M. (2017). *Mixing modes and measurement methods in longitudinal studies.* London, UK: UCL Institute of Education.

Kalton, G. (1986). Handling wave nonresponse in panel surveys. *Journal of Official Statistics*, *2*(3), 303–314.

Kern, C., Weiss, B., & Kolb, J.-P. (2019). A longitudinal framework for predicting nonresponse in panel surveys. ArXiv Preprint ArXiv:1909.13361.

Klausch, T., Schouten, B., & Hox, J. J. (2017). Evaluating bias of sequential mixed-mode designs against benchmark surveys. *Sociological Methods & Research*, *46*(3), 456–489.

Kreuter, F., & Olson, K. (2011). Multiple auxiliary variables in nonresponse adjustment. *Sociological Methods & Research*, *40*(2), 311–332.

Kwak, N., & Radler, B. (2002). A comparison between mail and web surveys: Response pattern, respondent profile, and data quality. *Journal of Official Statistics*, *18*(2), 257.

Lemay, M. (2009). Understanding the mechanism of panel attrition. University of Maryland. Retrieved from https://drum.lib.umd.edu/handle/1903/9631

Little, R. J., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Survey Methodology*, *31*(2), 161.

Lugtig, P. (2014). Panel attrition: Separating stayers, fast attriters, gradual attriters, and lurkers. *Sociological Methods & Research*, *43*(4), 699–723.

Lynn, P. (2013). Alternative sequential mixed-mode designs: Effects on attrition rates, attrition bias, and costs. *Journal of Survey Statistics and Methodology*, *1*(2), 183–205. doi:10.1093/jssam/smt015

Lynn, P. (2017). From standardised to targeted survey procedures for tackling non-response and attrition. *Survey Research Methods*, *11*(1), 93–103. doi:10.18148/srm/2017.v11i1.6734

Lynn, P. (2020). Evaluating push-to-web methodology for mixed-mode surveys using address-based samples. *Survey Research Methods*, *14*(1), 19–30.

Millar, M. M., & Dillman, D. A. (2011). Improving response to web and mixed-mode surveys. *Public Opinion Quarterly*, *75*(2), 249–269. doi:10.1093/poq/nfr003

Roberts, C., Joye, D., & Ernst Stähli, M. (2016). *Mixing modes of data collection in swiss social surveys: Methodological report of the LIVES-FORS mixed mode experiment.* LIVES Working Papers, 48, 1–42.

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*, 545–571. doi:10.1093/jssam/smy024

Sakshaug, J. W., & Eckman, S. (2017). Following up with nonrespondents via mode switch and shortened questionnaire in an economic survey: Evaluating nonresponse bias, measurement error bias, and total bias. *Journal of Survey Statistics and Methodology*, *5*(4), 454–479. doi:10.1093/jssam/smw039

Sakshaug, J. W., Yan, T., & Tourangeau, R. (2010). Nonresponse error, measurement error, and mode of data collection: Tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opinion Quarterly*, *74*(5), 907–933. doi:10.1093/poq/nfq057

Siegers, R., Belcheva, V., & Silbermann, T. (2019). *Data documentation: Documentation of sample sizes and panel attrition in the German Socio-Economic Panel (SOEP)(1984 until 2017).* SOEP Survey Papers, DIW-Berlin.

Smith, T. W. (2011). The report of the international workshop on using multi-level data from sample frames, auxiliary databases, paradata and related sources to detect and adjust for nonresponse bias in surveys. *International Journal of Public Opinion Research*, *23*(3), 389–402. doi:10.1093/ijpor/edr035

Vandecasteele, L., & Debels, A. (2007). Attrition in panel data: The effectiveness of weighting. *European Sociological Review*, *23*(1), 81–97.

Vannieuwenhuyze, J. (2014). On the relative advantage of mixed-mode versus single-mode surveys. *Survey Research Methods*, *8*(1), 31–42. doi:10.18148/srm/2014.v8i1.5500

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*(4), 450–469. doi:10.1093/pan/mpq025

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. doi:10.1177/1525822X13491863

Wagner, J., Schroeder, H. M., Piskorowski, A., Ursano, R. J., Stein, M. B., Heeringa, S. G., & Colpe, L. J. (2017). Timing the mode switch in a sequential mixed-mode survey: An experimental evaluation of the impact on final response rates, key estimates, and costs. *Social Science Computer Review*, *35*(2), 262–276. doi:10.1177/0894439316654611

Watson, N., & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In P. Lynn (Ed.),

Methodology of longitudinal surveys. doi:10.1002/978
0470743874.ch10

Appendix A
Tables

Table A1
Frequency of outcomes in waves 5–9 of the Innovation Panel based on an analysis sample size of 1,636 units.

Outcome	Wave 5	Wave 6	Wave 7	Wave 8	Wave 9
Web	694	802	694	689	708
F2f	607	426	294	278	193
Other non-response	227	194	277	193	142
Refusal	108	188	225	152	56
Not-issued	0	26	146	324	537

Table A2
Descriptive statistics of variables used in wave 4. The base sample consists of respondents eligible for the mixed-mode sample in waves 5–9

Variable	Freq.	%
Education		
Higher	371	29
A level	234	18.3
GCSE	287	22.5
Other	282	22.1
Missing	104	8.1
Age categorical		
Under 35	285	22.3
35-55	447	35
56-75	358	28
Over 75	72	5.6
Missing	116	9.1
Female		
No	543	42.5
Yes	637	49.8
Missing	98	7.7
Partner		
No	400	31.3
Yes	780	61
Missing	98	7.7
London		
No	1079	84.4
Yes	101	7.9
Missing	98	7.7
North		
No	692	54.1
Yes	488	38.2
Missing	98	7.7

Continued on next page

Continued from previous page

Variable	Freq.	%
Urban		
No	294	23
Yes	886	69.3
Missing	98	7.7
Refreshment		
No	779	61
Yes	401	31.4
Missing	98	7.7
In work		
No	36	2.8
Yes	702	54.9
Missing	540	42.3
Net use		
Daily	576	45.1
Sometimes	343	26.8
Never	170	13.3
Missing	189	14.8
Mobile use		
No	76	5.9
Yes	1104	86.4
Missing	98	7.7
Volunteers		
No	935	73.2
Yes	245	19.2
Missing	98	7.7
Donated		
No	373	29.2
Yes	807	63.1
Missing	98	7.7
Voted		
No	792	62
Yes	388	30.4
Missing	98	7.7
Mode preference		
Web	260	20.3
Face to face	650	50.9
Telephone	14	1.1
Mail	146	11.4
Other	110	8.6
Missing	98	7.7
Cooperative		
No	313	24.5
Yes	867	67.8
Missing	98	7.7
Not suspicious		
No	196	15.3

Continued on next page

Continued from previous page

Variable	Freq.	%
Yes	984	77
Missing	98	7.7
Mixed mode in wave 2		
No	666	52.1
Yes	514	40.2
Missing	98	7.7
Number of calls wave 4		
1-2 c	428	33.5
3-6 c	512	40.1
>6 c	240	18.8
Missing	98	7.7

Table A3

Results from the 3-step model predicting class membership using variables from wave 4. The "Face to face respondents" class is used as the reference category. Statistically significant predictors at the 0.05 level are in boldface type.

	Single mode/ early drop-offs		Face to face late drop-offs		Switchers		Web respondents	
Class	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Intercept	−0.48	0.80	0.15	0.88	−1.21	1.60	−1.99	0.76
Female	−0.13	0.24	0.39	0.27	−0.04	0.27	0.23	0.20
Having partner	−0.25	0.26	−0.26	0.30	0.07	0.31	0.64	0.23
Region (Ref. categ.: not in London or the North)								
London	−1.29	0.44	−1.39	0.60	−0.66	0.46	−0.98	0.36
North	−0.93	0.25	−0.54	0.29	−0.46	0.30	−0.54	0.22
Urban	0.39	0.29	0.60	0.33	0.04	0.30	0.26	0.23
Refreshment	0.31	0.37	−0.37	0.41	−0.08	0.40	0.16	0.30
Education (Ref. categ.: No education)								
Degree	0.40	0.39	0.14	0.39	0.39	0.43	0.65	0.30
A level	1.09	0.39	−0.15	0.44	1.18	0.42	0.67	0.32
GCSE	0.57	0.38	0.06	0.37	0.58	0.44	0.32	0.30
Age group (Ref. categ.: < 36)								
36-55	−0.52	0.35	−0.11	0.40	−0.13	0.37	0.43	0.29
56-75	−0.12	0.38	−0.54	0.45	−1.51	0.51	0.37	0.33
>75	1.89	0.67	1.17	0.81	−0.54	1.14	0.62	0.72
Internet use (Ref. categ.: none)								
Daily	0.56	0.41	−0.57	0.43	0.44	0.51	0.87	0.34
Sometimes	−0.05	0.40	−0.46	0.39	0.00	0.52	0.23	0.32
Missing	0.55	1.46	−3.08	1.89	3.55	1.52	0.12	1.39
Mobile use	−0.24	0.42	0.33	0.48	0.84	1.20	1.10	0.45
Employed	0.30	0.28	−0.34	0.31	−0.32	0.29	0.36	0.23
Volunteers	−0.51	0.29	−0.52	0.34	−1.04	0.45	−0.35	0.23
Donates	0.15	0.29	0.33	0.31	−0.09	0.31	0.29	0.25
Voted	0.05	0.26	−0.06	0.28	−0.10	0.31	−0.34	0.22
Mode preference (Ref. categ.: face to face or telephone)								
Web	1.53	0.37	0.39	0.46	1.61	0.40	1.25	0.32
Mail	1.15	0.37	0.18	0.43	1.10	0.41	0.49	0.33
Other	1.49	1.29	2.40	1.31	−0.15	1.22	1.88	1.23
Interviewer observations (dummies)								
Cooperative	−0.15	0.32	−0.65	0.35	−0.61	0.35	−0.27	0.28
Not suspicious	−0.06	0.40	0.43	0.44	0.71	0.59	0.38	0.35
Mixed mode wave 2	−0.32	0.34	−0.57	0.35	−0.66	0.37	−0.72	0.28
Number of calls (Ref. categ.: 1–2 calls)								
3–6	−0.25	0.29	0.66	0.30	0.21	0.34	0.04	0.24
> 6	−0.45	0.32	−0.52	0.39	0.09	0.36	−0.75	0.27

Table A4
Descriptive statistics of variables used in wave 9 to predict outcomes in wave 10. The base sample consists of respondents eligible for the mixed-mode sample in waves 5–9.

Variable	Freq.	Perc.
Education		
Higher	367	34.5
A level	246	23.1
GCSE	227	21.3
Other	94	8.8
Missing	130	12.2
Age categorical		
Under 35	237	22.3
35-55	338	31.8
56-75	309	29
Over 75	56	5.3
Missing	124	11.7
Female		
No	407	38.3
Yes	533	50.1
Missing	124	11.7
Partner		
No	324	30.5
Yes	614	57.7
Missing	126	11.8
Urban		
No	258	24.2
Yes	682	64.1
Missing	124	11.7
London		
No	876	82.3
Yes	64	6
Missing	124	11.7
North		
No	559	52.5
Yes	381	35.8
Missing	124	11.7
Face to face wave 9		
No	857	80.5
Yes	161	15.1
Missing	46	4.3
Any non-response		
No	653	61.4
Yes	411	38.6

Appendix B
Figures

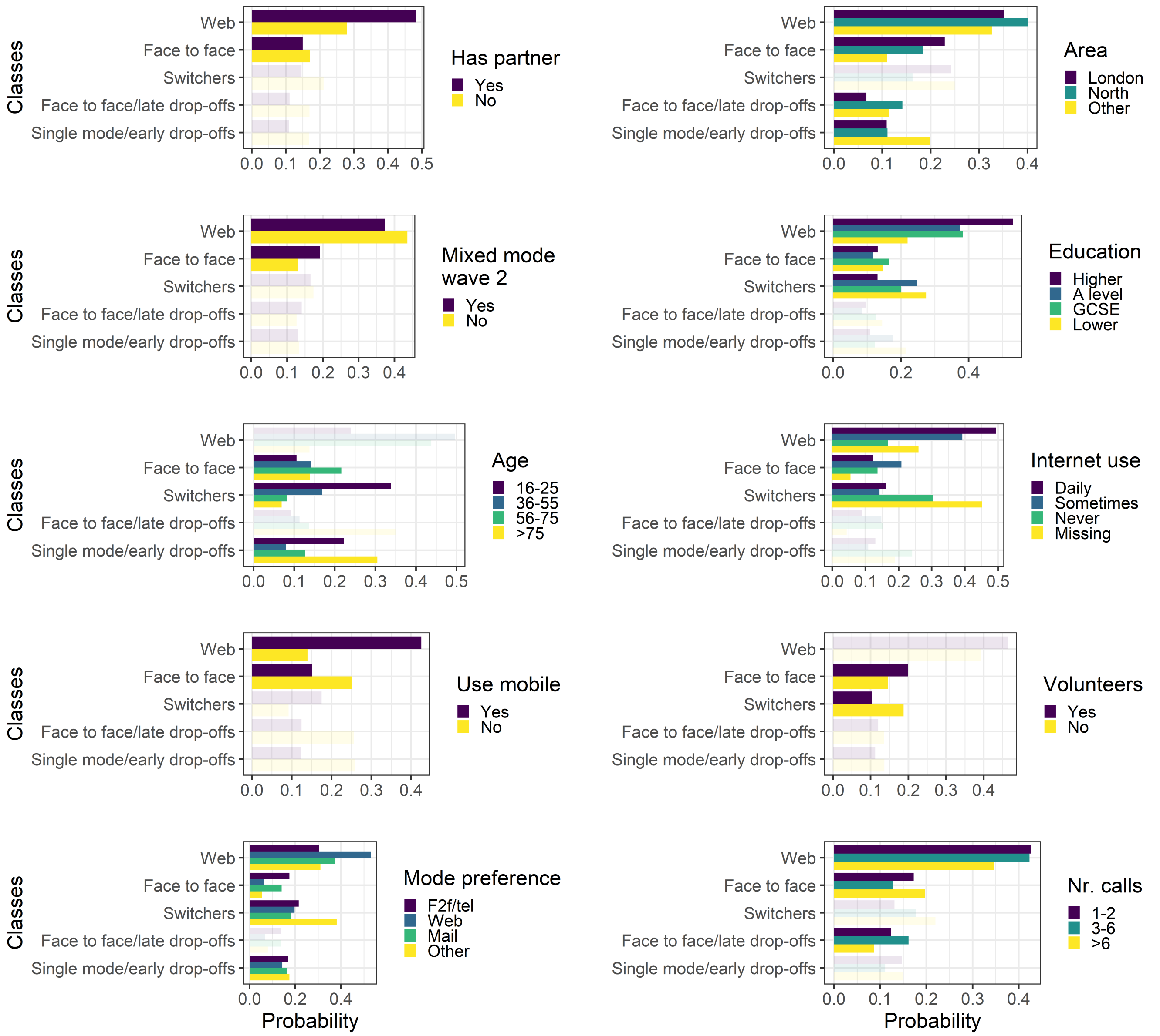

Figure B1. Observed distribution of variables in wave 4 conditional on class membership. Differences that are not statistically significant in model presented in Appendix 3 (at the 0.05 level) are made transparent.