Investigating the impact of switching to web in a longitudinal telephone survey: Potential effects on sample composition and attrition.

Shane Anthony Johnson

Master's Thesis for: MA Public Opinion and Survey Methodology

Director: Caroline Roberts

Expert: Boris Wernli

Session: Winter 2020

University of Lausanne – UNIL
Faculty of Social and Political Sciences
Institute of Social Sciences

University of Neuchâtel
University of Lucerne

Table of Contents

Abstract 5
Background 6
Literature Review 12
Rise of the web mode 12
Problems with web 14
Web mode effects 16
…on Measurement 16
… on selection 17
Longitudinal switch context 19
Devices and their effects 21
Attrition and its effects 23
Context of the study and Switzerland 25
Methods 26
Survey 26
Description 26
Questionnaires 27
Sample Composition 28
Data 31
Variables 31
Constructed variables 34
Analyses 35
Research Question 1: Would switching to web from telephone interviewing in the context of a longitudinal survey increase sample attrition? 35
Research Question 2: Would this attrition be selective among certain population subgroups? 36
Hypotheses 36
Results 38

Sample internet characteristics 38
Attrition risk characteristics 43
Mode preferences and stated willingness 45
Estimating attrition 51
Composition of attrition 62
Discussion 64
Internet characteristics 64
Mode preference 67
Stated Willingness 68
Predicting attrition factors 70
Estimating attrition 72
Limitations 73
Conclusion 76
References 79
Appendix 83

Abstract

This dissertation investigates the impact of switching the primary mode of data collection to web, from telephone, in the context of an ongoing longitudinal study. The growth of web surveying and the reduction in costs it represents for methodologists has driven an increase in longitudinal surveys switching to this mode mid-stream. The effects of which are important to establish in order to understand whether changes in time-series estimates are due to real change or just mode effects. This dissertation looks at the effects of switching mode on sample composition and attrition on a Swiss longitudinal study, the LIVES cohort study, which plans to switch to web in the near future. It finds that attrition is likely to be exacerbated due to switching to web, predominantly due to a sizeable proportion of respondents who prefer telephone interviewing, despite a largely technically competent panel. Additionally, it concludes that this attrition is not likely to be demographically selective. This importantly highlights that even in contexts of high internet competence, mode-related attrition may occur when switching to web, due to an entrenched preference for telephone interviewing.

Keywords: CATI, Web surveys, longitudinal surveys, mode switching, mode effects, attrition

Investigating the impact of switching to web in a longitudinal telephone survey: Potential effects on sample composition and attrition.

Background

Society, driven by technology, is "changing at an exponentially increasing rate" (Nicolaas et al., 2014, p.10). Researchers and survey methodologists need to find ways to keep up with the changing technological and social landscape, and, as such, increasingly so, they are turning to the web for data collection. This can be attributed to the need to respond to these societal and technological advances (Maslovskaya et al., 2019, p.327), but also the mounting need to reduce costs and increase the efficiency of data collection, thereby combining the benefits of "automation with those of self-administration" (Fricker et al., 2005, p.372). Nevertheless, while the prevalence of web-based surveys is rising due to the new opportunities its presents, the current state of knowledge is "not as advanced as necessary in times of rapid change" (Toepoel and Lugtig, 2015, p.155).

The web mode offers significant cost-advantages, and this is a major driver behind switching to web. It can help alleviate issues with coverage and nonresponse that traditional methods such as telephone data collection suffer with – such as the increase in mobile-only households and the associated decrease in response rates (Roberts and Vandenplas, 2017, pg.303) and number portability diluting geographical RDD telephone samples targeting efficiency (Horowitz et al., 2019, p.9). When mixed with other modes, it can potentially reduce the costs of mitigating nonresponse by offering alternative population coverage (Horowitz et al., 2019, p.131). There are also speed advantages to web questionnaires as the period of data collection can be shorter. Furthermore, it offers the respondent more liberty in

terms of when and where they complete the survey, potentially reducing the burden they face (Felderer et al., 2019, p.95).

Moreover, its self-administered nature means that the survey setting is more private, thereby negating the effect of any interviewer or social desirability biases on measurement error compared to interviewer-administered modes (Felderer et al., 2019, p.95). Relatedly, the rise of the web has meant a growing proportion of activities that are of interest to researchers occur in the digital environment, such a social media usage and smartphone activities; thereby creating a new territory from which data can potentially be harvested. The utilisation of the web mode of data collection, therefore, responds to both social evolution as well as methodological concerns. Tapping into the opportunities and territories that this technological milieu engenders is now vital to any industry, not least the survey industry.

The web mode also has its drawbacks, however. Accessibility issues mean that not all population subgroups have access to the internet and even when they do less have the competences required to use web-enabled devices efficiently. The 'unsupervised' nature of the web mode can be detrimental to data quality as respondents can misunderstand or inadequately reply to questions without interviewer prompts. Furthermore, there is a lack of an adequate email-address sampling frame (Toepoel and Lugtig, 2015). Not to mention the multifarious effects that the device used to complete the survey may have on data, as Toepoel and Lugtig (2015) profess: that "all web surveys should from now be thought of as mixed-device surveys" (p.155). These effects range from the fact that smaller and varying display sizes means that questions can appear in different ways, to the fact that different demographics own different devices (see de Bruijne and Wijnant, 2014; Maslovskaya et al, 2019) so such device-induced measurement error may be systematic in nature. Furthermore,

the web mode generally tends to have lower response rates (Nicolaas et al., 2014; Fricker et al., 2005; Felderer et al., 2019) which is concerning, however, web response rates are growing and starting to exceed the declining telephone response rates (Horowitz et al., 2019, p.131). Thus, the drawbacks above are increasingly less of a concern.

Switching or mixing modes can allow researchers to capitalise on the benefits of one mode and help negate the adverse effects of the former (Groves et al., 2004) thereby improving the efficiency of the survey. In this context, switching refers to a transition in the design of the survey that changes the mode of data collection from one mode to another; mixing refers to the introduction of a supplementary mode option, alongside the current mode. In the context of a longitudinal survey, which this dissertation will focus on, switching could potentially have severe effects on comparability between waves, as it is hard to isolate and then quantify effects on data quality explicitly linked to the mode change. Allum et al. (2018), state that any difference between responses in two modes is due to "differing degrees" of interplay between "two mechanisms" (p.46): Selection effects and Measurement effects. Selection effects are the result of different types of respondents systematically choosing to respond via different modes (or only being able to respond in certain modes) thereby the true values on variables of interest differ in different modes (Allum et al., 2018). This is most evident when, for example, different modes have different coverage problems. Measurement effects are caused by a variety of factors which means that the wrong value is recorded dependent on the mode used, i.e. the mode features cause differential measurement error (Allum et al., 2018). This could be due to both normative and cognitive considerations such as the presence (or not) of an interviewer or the aural or visual presentation of the questions leading to different cues and stimuli being used to formulate an answer (Vandenplas et al., 2017). Their inherent interaction means it is hard to disentangle the two

sets of effects, thereby, rendering it hard to ascertain the main cause of mode effects. Given that longitudinal studies investigate change, the fact that change could be due to mode effects, and not actual change, is problematic.

Furthermore, there are some specific concerns pertaining to switching modes in a longitudinal setting, namely attrition. Jäckle et al. (2015) highlight that response rates are more important than in a cross-sectional survey, in order to maintain the sample composition; hence there is a greater need to avoid nonresponse. Attrition of the sample members due to nonresponse can severely damage the viability of longitudinal studies, which by definition, rely on the responses from the same respondents at multiple time points. Secondly, they argue that longitudinal panel members' prior experience in another mode may be pertinent, as this familiarity "might improve the chances of response in the absence of interviewer explanation and persuasion" (p.58). This is relevant in switches from an interviewer-administered mode to a self-administered mode, where supervision is forgone for privacy and cost benefits.

Due to the potential risk to data quality, understanding the potential impact of switching to web in the context of an existing, ongoing longitudinal survey is of paramount importance to researchers. This dissertation will look at one longitudinal study which plans to switch mode from telephone to web in the near future and investigate what potential methodological and statistical implications could arise and how this would impact participation and attrition.

The Swiss LIVES Cohort study (LCS) will be used as a case study in this dissertation. The LCS is an annual survey following a cohort of young adults born between 1988 and 1997

who grew up in Switzerland. It was set up by the National Centre of Competence in Research LIVES and is conducted in collaboration with the FORS-led Swiss Household Panel (SHP). The LCS aims to "observe the transition into adulthood with a focus on the life course and vulnerability processes" (Spini et al., 2019, p.1) with an oversampling of second-generation immigrants, referred to as '*secondos'*, allowing for a more granular analysis of this often-under-represented subgroup. Due to cost considerations, the LCS, which is currently primarily CATI-based, is considering switching to being solely web-based as of the 2020-2021 wave. The proposed switch could have a variety of methodological and statistical implications on the continued utility of the study. This dissertation aims to dissect possible implications in terms of their potential effect on the sample composition and attrition levels.

This dissertation addresses the following two main research questions:

- RQ1: Would switching to web from telephone interviewing in the context of a longitudinal survey increase sample attrition?
- RQ2: If we find evidence of risk of attrition associated with switching to web, would this attrition be selective in terms of certain population subgroups?

In order to address RQ1, it will look at whether the switch could engender higher-than-average attrition, by considering a number of indicators of the risk of dropping out of the survey, if it becomes the principal mode of data collection (described later). For example, respondents' mode preferences or device access could be unsuited to web survey completion, say, if they prefer telephone interviewing and only have a smartphone. This question is pertinent as it could inform survey design decisions for upcoming waves, but more widely, it can contribute to the literature on mid-stream longitudinal data collection mode switches and their effects on attrition.

To address RQ2, this dissertation considers which subgroups are most susceptible to the risk of attrition, by looking at which subgroups are disinclined to respond via the web and seeing whether those liable to drop out are demographically similar. If there is evidence of selective attrition contributing bias to the sample's representativeness, this could be problematic for the utility of the study. This question is important as it gives an indication into what the LCS team can expect of the post-switch sample composition and whether it continues to meet the initial objectives, in relation to its oversampling of a vulnerable subpopulation, as well as, how to best navigate the transition to web. Furthermore, it can give wider insight into who is likely to drop out of panel surveys as a result of switches to web, if indeed any subgroup is more inclined to do so, especially in the cases of special populations.

Before presenting the methods used to address these questions, first the relevant background literature is reviewed, covering a broad range of topics pertinent to these research questions. Principally, research on the transition to web surveys, mode preferences, differences between web and telephone surveying and panel attrition will be covered. This review will help inform the research hypotheses.

Literature Review

The panel survey context is a unique situation in survey methodology. Where cross-sectional surveys need to be representative of the populations they are trying to study and equally need to obtain high response rates, panel surveys are "doubly challenged" to achieve these objectives over multiple time points (Sakshaug et al., 2019, p. 3). The difficulty of which is exacerbated by newer developments in the survey field, including device diversification and cost challenges. As such, contemplation and execution of a primary mode change mid-stream is a delicate endeavour that needs careful organisation.

Considerable evidence shows that the survey mode affects respondents' answers, even when questions are worded the same, as Dillman and Christian (2005) elucidate upon. The pertinence for panel surveys which aim to measure change between two different time points is clear. Discerning whether the change between two time points is attributable to the characteristic of interest or merely related to the mode change is the crux of the issue. Given that, increasingly so, longitudinal study mode changes are prevalent, it is important to dissect the issue and look at the motivations for mode switches.

Rise of the web mode

On a broader scale, the use of the web mode can be attributed to the need to respond to the advances in communication and technologies in our society (Maslovskaya et al., 2019, p.327). This is evident in the fact that web survey incidence has massively increased, especially since concerns over technological barriers and the digital divide have dissipated, and as internet access has increased (Roster et al., 2004, p.359). In her recent book, Zuboff (2019) posits that it is the cumulation of a decades-long process of "individualization" which

catalysed “second-modernity” needs for self-expression and self-determination, which in turn instigated the birth of the “burgeoning information apparatus” which surrounds us nowadays (p.36). Thus, the trend towards web data collection in the survey industry is part of a wider technological revolution of societal interaction, which has rendered virtual and web-based, a growing proportion of human communication. By reflecting the wider societal evolution, the move to the web of an increasing number of surveys is shown to be not just a methodological question but also a societal one.

Relatedly, the evolution of social communication towards the web has opened up new territory from which data can be harvested. The trend which has pushed a growing proportion of human interaction to the web, whether this is through social media or Fitbits, has, in essence, made reapable (or at least facilitated the reaping of) a large proportion of human activity. The ‘supply’ of this data has, in turn, boosted ‘demand’ for capturing it, which is met predominantly via web surveys, hence the rise in web survey incidence. Vital, therefore, is the capitalisation of this new digital milieu to the survey industry.

From a methodological perspective, the appeal of switching to the web mode can be attributed to a wide array of factors. Firstly, the cost and speed advantages of it. Cost benefits are one of the predominant driving forces behind the appeal of web data collection. Web-based surveys offer substantial data collection efficiencies and cost advantages (Roster et al., 2004, p.359). Revilla et al. (2014) and Schupp and Saßenroth (2015) both explain the “great potential for cost savings” (p.11) as marginal costs per web respondent are much lower in web than in face-to-face surveys and telephone surveys due to the elimination of interviewer costs (Allum et al., 2018). The elimination of interviewers can also improve data quality,

"due to the elimination of interviewer error and built-in checks that prohibit respondent errors" (Roster et al., 2004, p.359), further fuelling the web mode appeal.

Secondly, this rise has also been driven by the decline in the response rates of traditional methods such as landline telephones, which are being progressively supplanted by mobile phones, meaning that telephone interviewing is failing to provide adequate representative coverage of the population. Even at its peak, telephone interviewing had its limits in terms of coverage as it tended to underrepresent younger generations. The rise of the web has exacerbated this, while concurrently allaying fears of technology unevenness as internet access dissemination amongst diverse groups has increased – allowing web surveys to be increasingly representative of more general populations (Roster et al., 2004).

Problems with web

However, the web mode is not without disadvantages. In terms of representativeness, firstly, it assumes even accessibility across demographic categories. A major stopping block to the roll-out of web surveys historically has been the fact that not everyone had computer and internet access, which distorted the estimates derived from web surveys as they often missed out on those without access. Roster et al. (2004) found evidence supporting claims that internet surveys over-represent groups such as the youth, while under-representing ethnic minorities (p.364), but that this is likely to rescind in prevalence and can be corrected via weighting procedures in the meantime. Indeed, this divide has shrunk over time, and now internet penetration is widespread enough not to be a significant concern.

Also, causing representativeness issues is the lack of a suitable sampling frame. Couper (2008) claims "frames of internet users in a form suitable for sampling do not – and

likely will not – exist" (p.832) and this is another major drawback to web surveying – this lack of a universal, reliable sampling frame with high coverage. Although this issue can be alleviated by the rise of mobile devices which can be used for Random Digit Dialling (RDD) sampling (Toepoel and Lugtig, 2015, p.156), it can mean that sampling and recruitment have to be completed via another mode or the survey has to be completed in a non-probabilistic manner.

Secondly, it invokes issues with measurement. For example, the self-administered nature, while it can be positive in some respects, means that interpretation is left to the respondent, and if there is misunderstanding there is not an interviewer present to provide prompts. The result is that interpretation differs, and thus responses can be based on misunderstanding, this is particularly problematic when there are mixed modes, due to differential interpretation across modes. Device differences can exacerbate this. Also, the "unsupervised" nature of the web survey, may make participants more likely to adopt satisficing strategies (Allum et al., 2018; Fricker et al., 2005); however, the evidence on this is "rather mixed" (Allum et al., 2018, p.44). In any case, item and partial unit nonresponse have been found in such mixed-mode designs (Schupp and Saßenroth, 2015; Herzing, 2019) which is a significant concern. As is the higher break-off rate for mobile respondents that de Bruijne and Wijnant (2014b) claim is reported by several researchers.

Other issues include spam filters blocking participation requests, device diversification affecting responses (See *Device* section) and complications of mixing modes, including questionnaire unification strategies. Each issue has direct implications on the representativeness and measurement quality of web surveys; these implications are known as mode effects.

Web mode effects

Roster et al. (2004) state that there are some important differences between web and telephone responses. These differences, known as *mode effects* as they are the deviation from the true values caused by the mode, are critical and caution, therefore, needs to be exercised before assuming data equivalence between telephone and web surveys (Roster et al., 2004, p.371). There are two categories of mode effects, they are selection effects and measurement effects, and these two mechanisms are often hard to disentangle. Selection effects are when, due to the nature of the mode, different subgroups of respondents respond systematically in different modes, such that the estimates differ across modes. This is of particular concern when the primary data collection mode changes, as it is hard to identify and mitigate the divergence in estimates due to this effect – in part due to its interplay with the second component of mode effects, measurement effects. This is where due to the different nature of the mode such as its visual or aural nature, or presence of an interviewer, measurement error accrues systematically differently, such that across modes, estimates are skewed. When mode switches occur, it is hard to associate what proportion of the change in estimates is due to which of the two effects, though attempts have been made. Felderer et al. (2019) found that on the web mode, selection effects tend to be more pronounced, for example.

…on Measurement

Many factors contribute to measurement effects, especially in the answering process, as such the motivations for responding to a survey irrespective of mode or content and how respondents decide on a response, need to be outlined.

Tourangeau et al. (2000) propose a model of survey response that conceptualises how respondents arrive at an answer. For them, it is a four-step process of comprehension,

retrieval, judgement and response. Each step can provoke measurement error; for example, miscomprehension in the first stage could lead to the wrong answer being given. Such errors can be avoided by good question and questionnaire design and can be different depending on the mode of completion. Therefore, concerns about measurement error are salient, when switching from a "predominantly aural survey mode, the telephone, to a predominantly visual mode, the web" (Dillman and Smyth, 2007, p.S95) as this influences each stage in the response process. Dillman and Smyth (2007) attribute divergence between telephone and web responses to the fact that the survey modes rely on "fundamentally different types of communication" (p.S92). Web survey stimulus is, by definition, visual in nature, which provokes different cues and affects the way respondents interpret, understand, cognise and, eventually, answer questions. Thereby creating measurement effects.

Furthermore, another factor that can contribute towards measurement effects is that the mode may also affect the level of satisficing that respondents employ. This refers to the optimisation of the survey response process and the extent to which respondents 'shortcut' this – dependant on their ability or level of motivation. As different modes necessitate different levels of effort or induce different levels of motivation to make the effort, the effect of satisficing attempts differ between modes (Jäckle et al., 2010) and thus the response can differ due to the mode used. The display of the question on different modes can exacerbate this effect, for example, the grid display on question batteries in web surveys may render their similarity more salient, as Fricker et al. (2005, p.370) suggest.

… on selection

Of course, mode affects more than just measurement error. It can drive selection and nonresponse error too, as respondents tend to systematically differ in their propensity to

respond depending on their ability and willingness to respond via the offered modes. This is particularly concerning as web surveys tend to have lower response rates which increase the risk for selectivity and nonresponse bias, compared to the telephone survey (Felderer et al., 2019, p.95). Indeed, Felderer et al. (2019), found that web survey combined bias is larger than telephone survey bias, driven mainly by "comparatively larger nonresponse bias" (p.107). In a longitudinal setting, this can have important and salient consequences for attrition. This is of particular concern where, like in the LCS, there is a survey feature which aims to capture responses from 'vulnerable' subgroups, as Rothenbühler and Voorpostel (2016) found it is these groups that are most susceptible to attrition, in the SHP.

Moreover, response to web surveys can be affected by factors ranging from distrust of the internet to variation in competence and technology, according to Dillman and Smyth (2007, p.S91). Bosnjak et al. (2010), found that one's propensity to respond relies on "hedonic, affective, self-expressive, and trust-related factors" (p.357), that is to say, perceived enjoyment, attitudes towards participation, self-congruity and perceived trustworthiness (p.353). Further, Keusch (2015) summarises these 'external' factors and categorises them as either societal-level, sample-person-level or survey design level, with societal-level factors including culture, sample-person factors including gender and topic interest, and survey-design factors including incentive level, questionnaire length and prenotification (pp.186-189). Mode of completion also determines context and privacy of the completion environment thus the inherent privacy and lack of interviewer presence of the web mode might induce less bias for questions that are sensitive or susceptible to social desirability (Felderer et al., 2019, p.95; Fricker et al. 2005, p.374). While this example purports a positive consequence, it nonetheless highlights another way that telephone and

web answers can diverge, which is concerning in a longitudinal setting, where continuity and conformity is arguably more important.

Longitudinal switch context

Due to the growing appeal of web surveys, in part due to some of the reasons already mentioned, many longitudinal surveys are switching ongoing panel studies to web mid-stream creating potential disruption to time-series estimates and comparability between waves, as well as having potentially detrimental effects on attrition. Whether this by a wholesale change of mode (as envisioned by the LCS) or the introduction of a supplementary mode (as the LCS has experienced previously, by allowing web responses), major concerns arise about the level to which estimate changes, witnessed between pre- and post-switch, are attributable to actual change or mere mode effect distortion.

But also, switching involves, at its core, a change of contexts of where, when and how the survey could be filled out, Allum et al. (2018) refer to these contexts as environmental, temporal and internal contexts (p.46). For example, changing from CATI to a web survey means that participants may complete the survey on the go, in the early hours of the morning, and this freedom may engender a better internal context, i.e. a better mood. The effects of which may be far more widespread than those attributed to the mode change itself.

Allum et al. (2018) outline the psychological evidence that suggests that context effects may affect memory and recall accuracy due to the cue-driven nature of memory retrieval. In the context of longitudinal surveys, a mode switch changes the cues that have been available to respondents in previous waves, potentially having a detrimental effect on questions relying on memory retrieval. Their findings, however, suggest this effect is

minimal (Allum et al., 2018, p. 55). They do, however, raise the idea that those panel members who have prior experience of survey waves, can use their experience to "mentally reinstate" (p.46) the context of previous waves when confronted with the same questions in a different mode, thereby negating the effect of context in situations of mode switches.

Linking back to Tourangeau et al.'s (2000) model, Al Baghal and Kelley (2016) elucidate upon how context influences the process of survey response. They state that an answer to a question on subjective phenomena is derived from both chronically-accessible and temporarily-accessible information. Context effects occur when there is less reliance on chronically-accessible information, which by definition, is context-independent and more reliance on temporarily-accessible information which is context-dependent (Al Baghal and Kelley, 2016, p.145). Context could be found in the question order, previous survey experiences, survey topic or sponsors, or question position, amongst others as respondents actively try to seek context to answer questions. It is precisely because these context cues differ between modes and question types that mode divergence is a concern. This is particularly evident in questions on mode preference, which are often posed during waves prior to mode switches to help inform mode switch decisions, as these questions tend to rely on temporarily-accessible information and are thus subject to context effects. As such, mode preference questions tend to be an unstable measure of mode preference.

Altogether, switching represents a delicate procedure whereby careful planning and execution is required in order to adequately minimise potential mode effects, which can damage the veracity of time-series estimates and estimates of change over time.

Devices and their effects

Toepoel and Lugtig (2015) assert that "all web surveys should…be thought of as mixed-device surveys" (p.155) due to the fact the participants complete web surveys on a range of devices such as PCs, tablets and smartphones nowadays. This represents a new technological challenge as smartphones and tablets are increasingly rivalling, and rapidly overtaking, PCs in terms of internet access and survey completion (de Leeuw and Toepoel, 2018, p.52). The wait for internet access to become sufficiently pervasive has spawned a new access-related dilemma for survey methodologists – device type and ownership inconsistencies. This means that survey design needs to be reconceptualised as a "multi-device oriented" not merely "computer orientated" (de Leeuw and Toepoel, 2018, pg.52). Accommodation, therefore, needs to be made in survey design for mobile survey completion, in particular via a smartphone, as well as traditional desktop response. However, not only do these devices vary by screen size, but also data entry interface and owner characteristics meaning there is a question over the comparability of answers obtained from different devices (de Leeuw and Toepoel, 2018) in a similar way as between traditional modes.

The implications of this 'device divide' are manifold. There are many devices available to survey participants, but the most popular are smartphones, laptops, desktop computers and tablets. The users of each differ demographically because of access and competences. The interface and operating system of each differ, meaning questions and questionnaires can appear differently on different devices, potentially confounding answers. The environment within which respondents complete the survey can differ, depending on the device used – as mobile devices tend to be used in the presence of others or in locations where distractions are more numerous (Lynn and Kaminska, 2012). Relatedly, survey error and response motivations can differ between devices. The added complication of devices

within the web mode means that extra thought is needed when designing surveys, such as uni-mode design or specific mode optimisation (de Leeuw and Toepoel, 2018, p.54).

Maslovskaya et al. (2019) in a UK context, look at the different characteristics of respondents who respond via different modes. Their findings show that younger, female, employed people from smaller households are more likely to respond from mobile devices (smartphone/tablet) than older, male, unemployed people from larger households, who are more likely to use PCs or laptops (pp.342-343). These groups, they argue, could be motivated to respond to a greater extent by using advance communications highlighting that the survey is optimised for these devices (Maslovskaya, 2019, p.343). They (Maslovskaya, 2019) also found that marital status, children in the household, household income, number of cars and frequency of internet use are also highly related to device use (p.326). Age being a predictor of mobile response is corroborated by Jäckle et al., (2015); de Bruijne and Wijnant (2014); Watson and Wooden (2009); and Al Baghal and Kelley (2016). Gender, by Keusch (2015); de Bruijne and Wijnant (2014); and Watson and Wooden (2009). Employment status is also found to be relevant here by de Bruijne and Wijnant (2014); and Watson and Wooden (2009). Household size is also corroborated by de Bruijne and Wijnant (2014); Watson and Wooden (2009); and Schupp and Saßenroth (2015).

A certain emphasis has been put on smartphone response, which can be as high as 20-30% in certain surveys (Peterson et al., 2017, p.206). Indeed, smartphones are more frequently used than tablets (Revilla et al., 2014). Antoun et al. (2018) raise issues over the screen layout divergence and whether to employ app-based or browser-based surveys, the former can optimise to the screen size easier. They also look at question format and how simple questions and response options work best. These considerations allow for greater

readability on smartphones but also greater layout predictability amongst different devices. In terms of the participants, Herzing (2019) suggests that the inclusion of mobile devices can help recruit hard to reach groups such as "adolescents, refugees or migrants" (p.9) and it does improve age coverage (p.8). De Bruijne and Wijnant (2014) found that respondents under 35 were up to six times as likely to access surveys via smartphones. Given that the LCS is a panel of young people with immigrant backgrounds this is particularly pertinent, and suggests that the web mode, particularly accessed via mobile devices, seems to be suited to this panel.

A key topic relevant to this research is attrition, something that methodologists want to keep to a minimum, particularly in cases of mid-stream mode changes, due to the impact that measurement and selection effects can have on time-series estimates. The next section will look at this.

Attrition and its effects

Attrition is the decline of responding sample members over survey waves. It is a key concern for longitudinal survey methodologists, whose aim it is to minimise attrition in order to maintain the utility and viability of the panel. At the crux of the concern of attrition is selection bias, i.e. that attrition occurs in a non-random manner such that the representativeness of the sample is undermined. Watson and Wooden (2009) note that, at minimum, it has a detrimental effect on survey estimates accuracy, due to the higher nonresponse and potential selectivity, and that at sufficiently high levels, it can "threaten the viability of continuing a panel" (p.158), as the sample size reduces and the representativity worsens to the point where the sample is no longer sufficiently representative of the population.

Attrition's potential to impart bias to population estimates due to its non-random nature inducing selection bias (Watson and Wooden, 2009; Rothenbühler and Voorpostel, 2016) is a phenomenon feared by methodologists especially in mode switching contexts. If the mode change itself triggers attrition as respondents do not like the new mode, then this attrition is likely to not be random in nature and risks imparting selection bias rendering samples inadequate. This is particularly true when the mode switch is to web, as the demographics of web respondents compared others tends to be markedly different especially in terms of age (Al Baghal and Kelley, 2016, p.160), meaning the attrition tends to be non-random. Indeed, non-random attrition tends to be the norm (Rothenbühler and Voorpostel, 2016, p.221).

Rothenbühler and Voorpostel (2016) dissect non-random attrition into two categories: attrition that is selective on variables observed in the data and that which is selective on variables unobserved in the data. They go on to stipulate and find that attrition is often associated with the concept of vulnerability, that is those who have traits that "position them in low levels within the socioeconomic stratification" (Rothenbühler and Voorpostel, 2016, p.222), such as low level of education, foreign nationality, unemployment, poor health or being divorced. This may be because of difficulties in locating and contacting such people as they are more likely to move residence between waves; lower skills meaning that participating is relatively harder for such people; or lower perceived benefits of participation, as Rothenbühler and Voorpostel outline (2016, p.223). Nonetheless, their findings do suggest that vulnerability can provoke higher levels of attrition. This is a concern for the LCS which aims to track vulnerability processes over the transition to adulthood (Spini et al. 2019) if those who are most vulnerable refuse to participate then the study is weakened by selectivity bias.

Now the study and context of web surveying in Switzerland will be discussed.

Context of the study and Switzerland

The LIVES Cohort study, which will be used in this dissertation and which will be described later on is a longitudinal study of a specific age cohort in Switzerland with an oversampling of a special population – in this case, second-generation immigrants. The survey is looking to change its primary mode of data collection to the web. In the Swiss context, internet penetration is high, with over 90% of 16-74-year olds and 99.6% of Swiss adolescents using the internet on a regular basis (Herzing, 2019, p.10). Furthermore, 81% of 15-29-year olds, a similar age range to the LCS panel, went online with their mobile phone according to the Swiss Federal Statistical Office (Herzing, 2019, p.10). Therefore, Switzerland can be considered as having high internet penetration, which bodes well for the envisioned mode switch of the LCS. Indeed, Herzing states that mobile-only surveys are likely to be only useful for some target populations and while the LCS does not intend to become a mobile-only survey (but more widely web-only), it is an interesting assertion that web surveys can be used to target specific populations.

Methods

Survey

Description

The survey used to evaluate the impacts of a switch to web is the LIVES Cohort Study (LCS), set up by the Swiss National Centre of Competence in Research LIVES in collaboration with the FORS-led Swiss Household Panel (SHP). The LCS and the SHP resemble one another, except that the LCS is an individual-level panel study whereas the SHP is at the household level, asking everyone in a household. Furthermore, the aim of the LCS was to build an extensive sample of second-generation immigrants across Switzerland within a specific criterion. That is, that they were born between 1988 and 1997, schooled in Switzerland before their 10th birthday, resident in Switzerland on the 1st January 2013 and that both their parents were born abroad but arrived in Switzerland before their 18th birthday (Spini et al., 2019, p.402). Also, the parents had to have migrated from either Bosnia-Herzegovina, Croatia, Serbia, Kosovo, Macedonia, Montenegro, Italy, Portugal or Turkey.

The objective was that two-thirds of the sample were to be secondos, i.e. those that fulfil the criterion and one-third Swiss citizens from the same birth cohort. This, therefore, represents an oversampling of secondos which was the aim in order to build a sample which more fine-grain analysis of second-generation immigrant life patterns could be conducted, with particular emphasis on analysis vulnerability in the transition into adulthood (Spini et al., 2019, p.400). However, as we can see below the desired ratio was not recruited and has worsened over time.

Sampling was a multi-stage procedure, using a stratified simple random sample from the Swiss Federal Statistical Office population register, with a screening procedure to over-sample ‘secondos’. Then a multi-iteration controlled network sample, whereby those in the

networks of those selected from the population register, and then those in the networks of those in the networks of those from the original population register were screened and selected, with the aim of oversampling ‘secondos’ (Spini et al., 2019, p.402). As a result, the study is not representative of the Swiss population, and the weighting procedure is not standard.

The principal mode of interviewing is telephone, with earlier waves allowing face-to-face interviewing. Since wave three web interviewing has been permitted, only in cases of initial refusal. Due to funding structure changes which necessitate a reduction in the cost of running the LCS, from the eighth wave the mode will be switched to a web-based questionnaire only. It is the aim of this dissertation to investigate the potential ramifications of this switch.

Questionnaires

The questionnaires cover a broad range of topics from income to life satisfaction, religion to leisure activities. It is typically the same questionnaire as for the Swiss Household Panel, with a few added questions. The format for the questionnaire involves the base set of questions, which are repeated each year, complemented by a series of rotating modules which appear every few waves. For the sixth wave, which is the focus of this dissertation, these were ‘Religion and ‘Psychology’. Also, a special module was added to the LCS wave six on ‘Internet and device utilisation’ in order to give insight into panel members internet behaviour and device ownership, as well as mode preferences, to inform analysis on whether the web mode is suited to this cohort and thus the potential impact of the mode switch.

Sample Composition

As table 1 shows, 1,691 individuals were interviewed in the first wave, after the sampling procedure and there have since been six waves. In wave six, 784 individuals were interviewed, an overall attrition rate of over 53%. However, attrition has not been uniform over the waves; for example, in wave four, the retention rate was only 76.2% yet in wave five, this was 94.1%. The overall ratio of secondos has fallen, from representing 46.6% of the sample in the first wave to only 38.4% of the sample in the most recent wave. This is probably due to secondos no longer responding at a higher rate, rather than there being more non-secondos responding, due to the overall attrition rate falling.

In the first wave, a larger proportion of secondos were married, this gap has shrunk over the waves, presumably as non-secondos have increasingly become married. The proportion of men and women in the sample has remained somewhat similar over the waves, with only a slight divergence in the second and third waves when there was a higher proportion of men questioned. In terms of occupation, the ratio of the first wave, where almost 20% more secondos were in full-time work, has inversed itself, to the extent that in the sixth wave, almost 20% more non-secondos were in work – perhaps due to many non-secondos entering the workforce, who are of a greater number. Unemployment interestingly is equally balanced between secondos and non-secondos in the sixth wave, but this was not the case in the first wave where more secondos were out of work (61.7% compared to 38.3 of non-secondos).

Another clear trend is the decrease in the proportion of secondos in schooling, over each wave it has consistently fallen by 1-2% points from 42.5% to 36.8% - perhaps as more enter the workforce. However, this is not witnessed in the employment figures, as the

secondo ratio has fallen in this respect over the waves, perhaps, therefore, as secondos move in the workforce, they also stop responding and leave the panel. Furthermore, the amount of non-secondos with apprenticeship-level education has risen from 44.2% to 59.3%, over the same period, those with university-level education has not risen to the same extent.

Finally, the sample origin variable shows that those who were recruited via the network sample and are secondos haven fallen by nearly 10% over the waves to a much greater extent than those recruited from the main OFS sample, an interesting observation.

Table 1: Composition of the sample over the six waves across select demographics

	Wave 1		Wave 2		Wave 3		Wave 4		Wave 5		Wave 6	
Total N (retention %)	1691		1395 (82.5)		1187 (85.1)		904 (76.2)		851 (94.1)		784 (92.1)	
	Secondo (%)	Non-Secondo (%)	Secondo (%)	Non-Secondo (%)	Secondo (%)	Non-Secondo (%)	Secondo (%)	Non-Secondo (%)	Secondo (%)	Non-Secondo (%)	Secondo (%)	Non-Secondo (%)
Total	*46.6*	*53.4*	*42.9*	*57.1*	*42.3*	*57.7*	*40.6*	*59.4*	*40.5*	*59.5*	*38.4*	*61.6*
Cohort:												
1988-1993	*42.9*	*57.1*	*41.0*	*59.0*	*39.8*	*60.2*	*37.2*	*62.8*	*35.9*	*64.1*	*34.7*	*65.3*
1994-1999	*51.8*	*48.2*	*46.4*	*53.6*	*46.5*	*53.5*	*46.0*	*54.0*	*47.2*	*52.8*	*41.5*	*58.5*
Sex:												
Male	*46.6*	*53.4*	*44.1*	*55.9*	*43.5*	*56.5*	*40.2*	*59.8*	*41.0*	*59.0*	*38.0*	*62.0*
Female	*46.8*	*53.2*	*42.3*	*57.7*	*41.5*	*58.5*	*41.2*	*58.8*	*40.1*	*59.9*	*38.8*	*61.2*
Civil Status:												
Single/Divorced/Separated	*45.8*	*54.2*	*42.5*	*57.5*	*41.4*	*58.6*	*39.6*	*60.4*	*38.5*	*61.5*	*36.2*	*63.8*
Married/Partnership	*79.6*	*20.4*	*71.9*	*28.1*	*74.4*	*25.6*	*64.3*	*35.7*	*67.2*	*32.8*	*65.5*	*34.5*
Occupational Status:												
Full time work (37h min)	*58.1*	*41.9*	*50.2*	*49.8*	*47.4*	*52.6*	*42.7*	*57.3*	*40.8*	*59.2*	*40.2*	*59.8*
Part time work (36h and less)	*45.4*	*54.6*	*35.8*	*64.1*	*37.5*	*62.5*	*31.9*	*68.1*	*44.3*	*55.7*	*40.2*	*59.8*
In school or training	*42.5*	*57.5*	*41.5*	*58.5*	*40.9*	*59.1*	*39.9*	*60.1*	*37.2*	*62.8*	*34.8*	*65.2*
Unemployed	*61.7*	*38.3*	*55.2*	*44.8*	*45.5*	*54.5*	*54.5*	*45.5*	*64.7*	*35.3*	*50.0*	*50.0*
Other	*51.0*	*49.0*	*42.1*	*57.9*	*45.5*	*54.5*	*56.5*	*43.5*	*61.3*	*38.7*	*48.3*	*51.7*
Education Level:												
University/Academic high school	*32.0*	*68.0*	*32.9*	*67.1*	*31.5*	*68.5*	*30.7*	*69.3*	*34.2*	*65.8*	*33.0*	*67.0*
Bachelor/Maturity	*40.0*	*60.0*	*40.3*	*59.7*	*37.5*	*62.5*	*39.5*	*60.5*	*39.7*	*60.3*	*36.9*	*63.1*
Apprenticeship	*55.8*	*44.2*	*43.7*	*56.3*	*46.6*	*53.4*	*45.0*	*55.0*	*40.3*	*59.7*	*40.7*	*59.3*
Compulsory school/Elementary vocational training	*46.5*	*53.5*	*44.6*	*55.4*	*44.9*	*55.1*	*42.6*	*57.4*	*50.4*	*49.6*	*52.3*	*47.7*
Other	*47.1*	*52.9*	*46.5*	*53.5*	*47.2*	*52.8*	*41.9*	*58.1*	*41.9*	*58.1*	*39.3*	*60.7*
Sample origin:												
OFS sample	*26.7*	*73.3*	*29.6*	*70.4*	*27.0*	*73.0*	*22.1*	*77.9*	*21.4*	*78.6*	*27.1*	*72.9*
Network sample	*48.3*	*51.7*	*44.4*	*55.6*	*43.9*	*56.1*	*42.4*	*57.6*	*42.3*	*57.7*	*39.5*	*60.5*

Data

Data from the sixth wave of the LCS will be used principally. With a particular focus on the special module that was added to the questionnaire, asking respondents about their internet and device usage as well as mode preferences. Data from previous waves will be used to provide information on participation patterns across the waves as well as the evolution of the composition of the sample, for example, the '*secondo'* oversampling and general attrition levels. The results of which will be analysed to see the potential repercussions for the seventh wave and to give insight for the eighth wave which will be the first compiled under the web mode of data collection and wholly FORS-managed.

Variables

Apart from socio-demographic variables, such as sex, age, civil status, occupation and education, below is a list of other variables relevant to the analyses:

Variable Code	Description	
P18A251	Willingness to answer: Internet	If next year we asked you to complete a questionnaire on the internet, how likely is it that you would complete the questionnaire? Using a scale where 0 represents something you definitely would not do and 10 means something you definitely would do. 0-10
P18A252	Willingness to answer: Telephone	And if next year we approach you by telephone, how likely is it that you would complete the interview on the telephone? Using a scale where 0 represents something you definitely would do and 10 represents something you definitely would do. 0-10
P18A253	Willingness to answer: Smartphone	How willing would you be to complete an online questionnaire on your mobile phone for our Study?

		Very willing; Somewhat willing; A little willing; Not at all willing.
P18A254	Willingness to answer: Tablet	How willing would you be to complete an online questionnaire on your tablet for our Study? Very willing; Somewhat willing; A little willing; Not at all willing.
P18A250	Mode preference	Thinking about these two ways in which we could ask you to take part in the future, including telephone and a questionnaire via the internet, which one would you most prefer? A telephone interview; An internet questionnaire; No preference
P18A257-P18A262	Internet use: Locations	Where do you regularly use the internet? At home; At work; School, college, university; Travelling; In a public place (shop, café, restaurant, park etc.); Other location.
P18A212	Internet use: Frequency	How often do you use the Internet for your personal use? Every day; Several times a week; Several times a month; Once a month; Less than once a month; Never use; No access at home, work or elsewhere.
P18A213-P18A220	Internet use: Devices	Which of the following devices do you use to connect to the internet? Desktop computer; Laptop; Smartphone; Tablet; Feature Phone/Non-touchscreen mobile phone; E-book reader (e.g. Kindle); Smartwatch; Other.
P18A222-P18A234	Smartphone activities	Do you use your smartphone for the following activities? Browsing websites; Writing or reading an email (During leisure); Taking photos; Looking at content of social media websites/apps (e.g. looking at text, images, videos on Facebook, Twitter, Instagram); Posting context to social media websites/apps (e.g. posting text, images videos on Facebook, Twitter and Instagram); Making purchases (e.g. booking train tickets, buying clothes,

		ordering food); Online banking (e.g. checking account balance, transferring money); Installing new apps (e.g. from iTunes, Google Play store); Using GPS/location-aware apps (e.g. Google Maps, Foursquare, Yelp); Connecting to other electronic devices via Bluetooth (e.g. smartwatches, bathroom scales); Playing games; Streaming videos or Music); Other.
P18MODES	Mode of completion	CAWI; CATI. (Constructed)
P18A255	CAWI device used (If answering via web)	Could you tell us, what type of device did you use to answer the survey? If you used more than one device, please select all that apply. Desktop computer; Laptop; Large tablet (23cm or greater); Small tablet (Smaller than 23cm); Large-screen smartphone (13cm or greater); Standard-screen smartphone (Smaller than 13cm); Feature phone (Non-touchscreen devices); Other device.
P18D168	Secondo status	Yes; No. (Constructed)
INT1801	Interviewer: Respondents attitude	In general, what was the respondent's attitude toward the interview? Friendly and cooperative; Cooperative but not particularly interested; Impatient and restless; Hostile
INT1803	Interviewer: Case difficulty	How difficult was this case to get? Very difficult; Somewhat difficult; Somewhat easy.
INT1804	Interviewer: Participation next wave	Do you expect this respondent to participate in the next wave? Absolutely; Probably yes; May be; No

Constructed variables

For analysis purposes, some of the raw variables and answer categories were modified, and new variables were constructed from combinations of others. These are detailed in the below table:

Variable	Description
Net willingness	If willingness via telephone (0-10) minus willingness via the internet (0-10) is more than 0 then "more willing telephone". If willingness via the internet (0-10) minus willingness via telephone (0-10) is more than 0 then "more willing internet". Otherwise, if equal "Equally willing both".
Device category	If smartphone and no other device = "Smartphone-only". If tablet and no other device = "Tablet-only". If laptop and no other device = "Laptop-only" etc.
Mobile devices: Smartphone and or tablet only	If respondents use only a smartphone or only smartphone and tablet then 1, otherwise 0.
Age categorised	Under 25's = 0 Over 25's =1
Interviewer: Difficult case (INT1803)	Merged "Somewhat difficult" and "Difficult"
Sample origin	OFS; Network sample. Indicates whether the participant was recruited directly from the OFS sampling frame or the subsequent networking iterations.
Internet use: Frequency	For the logistic regression, the variable was dichotomised to 'Everyday' vs 'All other options'.
Loyalty	A variable counting how many waves the participant has participated in using data from Wave one to six. Hence, value between 1 and 6.
Occupation	For the models and testing, this variable was dichotomised. As such it became: 'In schooling or other' vs 'In work'.

Number of smartphone activities	A variable created to count the number of smartphone activities used, from 1 to 14.
Education	This was modified for the models to 'Maturité or lower', 'University level' and 'Apprenticeship/Vocational'
High political Interest	Created from the raw political interest variable as a proxy for social involvement. 5 or above = 1, 4 or lower = 0.
Interviewer: Attitude	Modified for models to 'Good attitude', 'Cooperative but not interested', 'Bad attitude'.

Analyses

Research Question 1: Would switching to web from telephone interviewing in the context of a longitudinal survey increase sample attrition?

For this question, descriptive analysis will be undertaken to look at the sample distribution in terms of internet use characteristics, mode preferences and stated willingness, to identify whether there is a subgroup of respondents who would not be suited towards responding via the web in the future. Hypotheses will be constructed in order to define who falls in this attrition-risk group. These hypotheses are detailed below. If the results show that there is a large proportion that falls into these at-risk categories, and thus are assumed to be highly likely to leave the sample in the case of a switch to web, we can assume that attrition will occur. The magnitude of which will be compared to current attrition rates to see whether this constituted an increase in sample attrition.

Pertinent here is the implicit fact that this aims to analyse hypothetical attrition by using stated willingness and preferences as a proxy for future participation. Haan et al. (2019) confirm that "A higher intention to use a device is strongly related to actually responding on that device" (p.526) and Wenz et al. (2019) link intention to actual behaviour (p.2), so there

is theoretical grounding to make this assumption. However, it is hard to estimate if unsuited willingness or preference will translate seamlessly into future nonparticipation.

Research Question 2: Would this attrition be selective among certain population subgroups?

Once we have identified those at risk of attrition due to the change of mode of data collection, analysis will be undertaken in order to ascertain the demographics of this group to see whether any of the attrition is selective in nature or whether it is a random.

Logistic regression models will try to predict the factors significant to having a greater stated net overall willingness of telephone in order to determine what demographic and internet use characteristics make one more likely to be more willing to respond via the telephone. If any demographic factors are found to be significant, we can state that the attrition is likely to be selective in nature.

Estimation of the attrition rate and composition can also be compiled using the predicted probabilities from the logistic models. This should give an indication into who is most at risk of leaving the sample, as those with the highest probability of preferring telephone interviewing are assumed to be most at risk of attrition. Descriptive analysis can then be completed on this estimate of attrition to see which demographic groups are most represented in this group, to give an indication of the selectivity of the future attrition rate.

Hypotheses

In order to address the research questions, it is first necessary to build some hypotheses in order to test them via the analyses and see if they hold true. The hypotheses are

aimed at determining who is at risk of leaving the panel, i.e. no longer responding, due to the mode switch.

H1) Those without internet access would be at high risk of leaving the panel

H2) Those who own only mobile devices would be susceptible to attrition.

H3) Those who do not use the internet frequently are at risk of leaving the panel.

H4) Those who express a preference for telephone interviewing are likely to leave the sample.

H5) Those who access the internet at work only are at risk of leaving the panel.

H6) Those who access the internet in public places only are at risk of leaving the panel.

H7) Those who access the internet from places of education only would be likely to leave the panel.

Results

Presented below are the results of the analyses, structured into the topics of the hypotheses. It starts by looking at the internet characteristics of the sample, before moving on to mode preference and stated willingness and finishing with estimating the predicted attrition rates and selectivity.

Sample internet characteristics

Firstly, table 2 shows the sample uses the internet and devices, split by their secondo status. This is important as it shows the breakdown of devices, internet use frequency and location of internet use, which can be used to determine the level to which the sample is competent and comfortable with internet tasks, which is relevant to whether the web mode is suited to this sample.

This sample overwhelmingly uses the internet every day, with only between 4.8-6% of respondents using it less frequently; the 0.2% of non-secondos who seem not to have internet access needs to be investigated. Moreover, in terms of locations, it is the home which is the most used location for regular internet access, followed by public places, work, travel, then places of education. These are not exclusive categories, and indeed the majority use a combination of these locations, important therefore is to see if there are any combinations which could induce context effects via environmental distractions, for example, this will be explored later on.

Moreover, smartphone use is ubiquitous, and there are no differences between secondos and non-secondos. On the other hand, the other major mobile device, the tablet, has relatively low usage, with only 35% of secondos and 32% non-secondos stating they use one to connect to

the internet. Desktops are the second-least used device with just under 60% of both groups using them. Finally, laptops are used by 86.7% of panel members, the same amount for both groups. In terms of device, members of this panel connect to the internet most with smartphones, which is logical as their portability and mobile data capabilities means that they can be used in most locations, whether these conditions are conducive to survey completion remains to be seen.

Table 2:

Internet use characteristics by secondo status

	Secondos (%)	Non-secondos (%)
N	**301**	**483**
Device Ownership:		
Smartphone	*100*	*99.6*
Tablet	*35.2*	*31.9*
Desktop	*59.5*	*58.5*
Laptop	*86.7*	*86.7*
Internet Use Frequency		
Everyday	*94.0*	*95.0*
Several times a week or less	*6.0*	*4.8*
Never/No Access	*0.0*	*0.2*
Locations		
Home	*99.7*	*99.4*
Work	*72.1*	*70.6*
Place of Education	*42.9*	*48.9*
Travelling	*59.8*	*62.5*
Public Place	*74.8*	*74.3*
Other	*21.6*	*20.7*

It is worth noting that this sample is a cohort where the maximum age is 32 years old, hence the findings of high smartphone penetration, regular internet use in multiple locations should not come as a surprise. Overall, it seems safe to assume that this sample seems adequately experienced in internet use.

Confirming this assumption is table 3, which looks at the smartphone activities that different subgroups engage in. It shows that the average sample member engages in 10.31 smartphone activities, reinforcing the assumption of a technologically competent sample. The frequencies of the activities show that photos and video/music are the two most engaged in smartphone activity. Along with websites, emailing, GPS, Apps and looking at social media content, over 90% of the sample engages in these activities. Games and Online banking represent the two least engaged in activities for almost every subgroup. Also, apart from the unemployed and those with only a compulsory level of education, over 90% every other subgroup engages in at least one of these activities on a daily basis.

Table 3

Smartphone activities by selected variables

Variables	N	Activity frequency		Mean number of activities (out of 13)	Smartphone activities												
		Everyday (%)	Less frequently (%)		Websites (%)	Email (%)	Photos (%)	Looking Social Media (%)	Posting Social Media (%)	Purchases (%)	Online Banking (%)	Apps (%)	GPS (%)	Bluetooth (%)	Games (%)	Videos/ Music (%)	Other (%)
Total	**784**	**752**	**32**	**10.31**	**745**	**739**	**760**	**729**	**598**	**562**	**518**	**732**	**735**	**607**	**386**	**754**	**216**
Age:																	
Under 25	357	*95.8*	*4.2*	10.24	*93.8*	*94.1*	*96.4*	*94.7*	*80.4*	*66.9*	*63.9*	*93.3*	*95.8*	*75.9*	*47.6*	*96.4*	*25.2*
25 and over	427	*93.9*	*6.1*	10.36	*96.0*	*94.4*	*97.4*	*91.6*	*72.8*	*75.6*	*67.9*	*93.4*	*92.0*	*78.7*	*50.6*	*96.0*	*29.5*
Sex:																	
Male	374	*94.6*	*5.4*	10.28	*94.9*	*92.8*	*95.2*	*90.4*	*69.8*	*67.9*	*64.2*	*93.9*	*92.5*	*84.0*	*55.6*	*96.0*	*31.3*
Female	410	*94.9*	*5.1*	10.33	*95.1*	*95.6*	*98.5*	*95.4*	*82.2*	*75.1*	*67.8*	*92.9*	*94.9*	*71.5*	*42.4*	*96.3*	*24.1*
Secondo:																	
Yes	301	*94.3*	*5.7*	10.56	*96.0*	*96.3*	*98.7*	*94.4*	*82.1*	*79.0*	*72.1*	*91.7*	*93.4*	*79.7*	*51.5*	*95.7*	*26.3*
No	483	*95.0*	*5.0*	10.15	*94.4*	*93.0*	*96.1*	*92.1*	*72.7*	*67.3*	*62.3*	*94.4*	*94.0*	*76.2*	*47.8*	*96.5*	*28.4*
Civil Status:																	
Single, never married	726	*94.7*	*5.3*	10.31	*94.9*	*94.2*	*96.7*	*93.3*	*76.4*	*71.1*	*65.4*	*93.3*	*94.4*	*77.7*	*49.3*	*96.6*	*27.7*
Married	58	*94.8*	*5.2*	10.29	*96.6*	*94.8*	*100*	*89.7*	*74.1*	*79.3*	*74.1*	*94.8*	*86.2*	*74.1*	*48.3*	*91.4*	*25.9*

Variables	N	Activity frequency		Mean number of activities (out of 13)	Smartphone activities												
		Everyday (%)	Less frequently (%)		Websites (%)	Email (%)	Photos (%)	Looking Social Media (%)	Posting Social Media (%)	Purchases (%)	Online Banking (%)	Apps (%)	GPS (%)	Bluetooth (%)	Games (%)	Videos/ Music (%)	Other (%)
Occupational Status:																	
Full time work (37h min)	311	*93.8*	*6.2*	10.54	*94.8*	*95.2*	*98.4*	*94.2*	*76.8*	*78.4*	*75.2*	*93.5*	*93.5*	*80.6*	*50.3*	*96.8*	*30.0*
Part time work 36h and less)	92	*95.7*	*4.3*	10.17	*95.7*	*97.8*	*97.8*	*92.4*	*77.2*	*76.1*	*63.0*	*91.3*	*92.4*	*76.1*	*46.7*	*95.7*	*15.2*
In school or training	333	*94.9*	*5.1*	10.08	*94.9*	*92.8*	*95.5*	*91.6*	*73.9*	*64.3*	*59.8*	*93.7*	*94.3*	*75.1*	*47.7*	*96.1*	*28.5*
Unemployed	18	*88.8*	*11.2*	10.89	*100*	*94.4*	*100*	*94.4*	*88.9*	*88.9*	*55.6*	*94.4*	*100*	*77.8*	*66.7*	*94.4*	*66.7*
Other	30	*96.7*	*3.3*	10.45	*93.3*	*90.0*	*93.3*	*96.7*	*86.7*	*60.0*	*60.0*	*93.3*	*90.0*	*73.3*	*53.3*	*93.3*	*26.7*
Education Level:																	
University/Academic high school	200	*97.5*	*2.5*	10.41	*97.0*	*98.0*	*98.0*	*93.0*	*72.0*	*78.5*	*67.0*	*96.5*	*95.5*	*77.0*	*42.0*	*98.5*	*27.5*
Bachelor/Maturity	236	*95.3*	*4.7*	10.17	*96.2*	*93.6*	*95.8*	*91.1*	*75.8*	*64.0*	*64.8*	*93.6*	*96.2*	*77.1*	*48.3*	*94.9*	*25.0*
Apprenticeship	199	*93.9*	*6.1*	10.24	*92.9*	*93.9*	*96.5*	*92.9*	*76.8*	*71.2*	*68.7*	*91.9*	*91.4*	*76.8*	*53.0*	*96.0*	*26.8*
Compulsory school/ Elementary vocational training	65	*84.6*	*15.4*	10.45	*96.9*	*89.2*	*96.9*	*98.5*	*81.5*	*72.3*	*53.8*	*95.4*	*71.1*	*80.0*	*63.1*	*93.8*	*32.3*
Other	83	*95.2*	*4.8*	10.54	*92.8*	*94.0*	*100*	*96.4*	*84.3*	*79.5*	*72.3*	*89.2*	*69.9*	*80.7*	*50.6*	*98.8*	*30.1*
Sample origin:																	
OFS sample	70	*94.3*	*5.7*	10.33	*95.7*	*91.4*	*98.6*	*95.7*	*70.0*	*67.1*	*65.7*	*98.6*	*95.7*	*77.1*	*55.7*	*95.7*	*25.7*
Network sample	714	*94.4*	*5.6*	10.31	*95.0*	*94.5*	*96.8*	*92.7*	*76.9*	*72.1*	*66.1*	*92.9*	*93.6*	*77.5*	*48.6*	*96.2*	*27.7*

Although the sample generally uses their smartphones for multiple activities, some differences in the types of activities engaged in can be observed across subgroups, for example, purchases are significantly associated to age ($\chi^2 = 7.11$, df = 1, $p < 0.01$), sex ($\chi^2 = 4.00$, df = 1, $p<0.05$), though not as much as secondo status ($\chi^2 = 10.61$, df = 1, $p<0.01$), occupation ($\chi^2 = 14.56$, df = 1, $p<0.001$) and education level ($\chi^2 = 10.75$, df = 2, $p<0.01$), so we can reject the null hypothesis and state that purchasing on one's smartphone is significantly associated to all these demographic variables. Emailing on a smartphone is significantly associated with one's level of education ($\chi^2 = 8.08$, df = 2, $p<0.05$). Taking photos is associated with sex ($\chi^2 = 4.02$, df = 1, $p<0.05$) as is looking at social media content ($\chi^2 = 5.02$, df = 1, $p<0.05$). As we can see, despite the ubiquity of the smartphone, there is some nuance in the utilisation of smartphones, albeit it is small. The associations between certain activities and certain demographics, notably sex, illustrates that there is some inter-group differentiation in terms of these activities, showing that although the mean figures show a competent sample in terms of smartphone activities, this is not uniform across all the demographics, and so some subgroups may be less competent than others. This can affect one's willingness to respond to a survey via a smartphone, something which will be looked at later.

Attrition risk characteristics

Table 4 looks at the 'attrition risk' characteristics as identified from the hypotheses detailed previously. As we can see, 5.1% of respondents, nearly 60% of whom are non-secondos, use the internet less than every day, compared to the vast majority of the panel who connect to the internet on a daily basis. If the data collection mode switches to web, these people may be less likely or less able to respond. Device-wise there are 28 who own just a smartphone and no other device, thus their choice of participation via the web mode is limited

to a smartphone, which as the literature suggests may not be suited to survey completion. A limited choice is also the same for the two people who only have a laptop. Furthermore, there are 12 respondents who only have a smartphone or tablet, while this is a less restricted choice of completion device, the smaller screen size may affect response willingness (as we will see later). Moreover, these devices rely on different interfaces and input methods which could skew results even if participation is garnered. In terms of the location, the hypotheses detailed that at work, in public or at a place of education, are the three locations which could have adverse effects on participation, due to environmental context distractions, for example.

Table 4

Attrition-risk characteristics by secondo status

	Secondos (%)	Non-secondos (%)	Total (%)
N	301	483	784
Internet use frequency:			
Less than everyday	17(42.5)	23 (57.5)	40 (5.1)
Never Use	0 (0.0)	1 (100.0)	1 (0.1)
Device:			
Smartphone only	13 (46.4)	15 (53.6)	28 (3.6)
Laptop only	0 (0.0)	2 (100.0)	2 (0.3)
Tablet only	0 (0.0)	0 (0.0)	0 (0.0)
Smartphone & Tablet only	7 (58.3)	5 (41.7)	12 (1.5)
Location:			
At work only	0	0	0
In education only	0	0	0
In public place only	0	0	0
Mode preference:			
Prefers Telephone	85 (36.3)	149 (63.7)	234 (29.8)
Interviewer Assessment:			
Likely to respond next wave: No	5 (83.3)	1 (16.7)	6 (0.8)
Difficult Case: Yes	3 (60.0)	2 (40.0)	5 (0.6)

However, a good sign is that no respondents use these locations solely. Indeed, the majority of respondents use the internet in a multitude of locations, predominantly including at home. Location, in this context, therefore, does not seem to be a concern in terms of inducing attrition if the mode was web. It shall thus be removed from the rest of the analysis as an independent variable.

Those who prefer telephone interviews and, thus, are deemed to be at risk of not responding to a web survey represent 29.8% of the overall sample. Of this group, 63.7% are non-secondos, who make up 61.6% of the sample; thus, there does not seem to be a disproportionate preference dependent on secondo status. Indeed, these two variables are not statistically associated ($\chi^2 = 0.64$, df = 2, p-value = 0.73). Finally, the interviewer adjudged variables show a total of 11 respondents, five of whom were difficult to recruit and six who are adjudged not to respond next time. Nevertheless, there is no significant association to suggest a selectivity issue within these variables.

Mode preferences and stated willingness

Table 5 below dissects the demographic breakdown of the sample's mode preferences. Interestingly there are not many significant associations, however, the mode of completion is significantly associated to mode preference ($\chi^2 = 25.91$, df = 2, $p<0.001$), as the literature has suggested, the mode of completion can distort questions on mode preferences due to context effects which will be discussed later. Moreover, there is a strong statistical association between preferring internet questionnaire and one's willingness to participate via a smartphone ($\chi^2 = 48.52$, df = 3, $p<0.001$) and via a tablet ($\chi^2 = 15.89$, df = 3, $p<0.01$), though to a lesser extent. This is to be expected though as those two devices are means for an

Table 5

Mode preference over socio-demographics

Variables	All	An internet questionnaire (%)	A telephone interview (%)	No preference (%)
Total	**784 (100.0)**	**446 (56.9)**	**234 (29.9)**	**101 (13.3)**
Age:				
Under 25	***357 (45.5)***	*208 (58.3)*	*104 (29.1)*	*45 (12.6)*
25 and over	***427 (54.5)***	*238 (55.7)*	*130 (30.4)*	*59 (13.8)*
Sex:				
Male	***374 (47.7)***	*199 (53.2)*	*124 (33.2)*	*51 (13.6)*
Female	***410 (52.3)***	*247 (60.2)*	*110 (26.8)*	*53(12.9)*
Secondo:				
Yes	***301 (38.4)***	*176 (58.5)*	*85 (28.2)*	*40 (13.3)*
No	***483 (61.6)***	*270 (55.9)*	*149 (30.9)*	*64 (13.3)*
Civil Status:				
Single, never married	***726 (92.6)***	*412 (56.8)*	*214 (29.5)*	*100 (13.8)*
Married	***58 (7.4)***	*34 (58.6)*	*20 (34.5)*	*4 (6.9)*
Occupational Status:				
Full time work (37h min)	***311 (39.7)***	*170 (54.7)*	*102 (32.8)*	*39 (12.5)*
Part time work 36h or less)	***92 (11.7)***	*59 (64.1)*	*22 (23.9)*	*11 (12.0)*
In school or training	***333 (42.5)***	*197 (59.2)*	*94 (28.2)*	*41 (12.3)*
Unemployed	***18 (2.3)***	*7 (38.9)*	*4 (22.2)*	*6 (33.3)*
Other	***30 (3.8)***	*13 (43.3)*	*11 (36.7)*	*5 (16.7)*
Education Level:				
Bachelor/Maturity	***236 (30.1)***	*134 (56.8)*	*71 (30.1)*	*31 (13.1)*
University/Academic high school	***200 (25.5)***	*116 (58.0)*	*53 (26.5)*	*31 (15.5)*
Apprenticeship	***199 (25.4)***	*107 (53.8)*	*69 (34.7)*	*23 (11.6)*
Compulsory school/ Elementary vocational training	***65 (8.3)***	*45 (69.2)*	*15 (23.1)*	*5 (7.6)*
Other	***83 (10.6)***	*44 (53.0)*	*26 (31.3)*	*14 (16.9)*
Sample origin:				
OFS sample	***70 (8.9)***	*44 (62.9)*	*21 (30.0)*	*5 (7.1)*

Variables	All	An internet questionnaire (%)	A telephone interview (%)	No preference (%)
Network sample	***714 (91.1)***	*402 (56.3)*	*213 (29.8)*	*99 (13.9)*
Mode of completion: Telephone	***720 (92.0)***	*391 (54.3)*	*231 (32.1)*	*97 (13.5)*
Web	***64 (8.2)***	*55 (85.9)*	*3 (4.7)*	*4(6.3)*
Mean mode willingness: Internet Questionnaire*	**8.40**	9.16	6.67	9.18
Telephone Interview*	**8.49**	7.53	10.00	9.31
Device willingness: Willing Smartphone**	***579 (73.9)***	*371 (64.1)*	*132 (22.8)*	*77 (13.3)*
Willing Tablet**	***179 (22.9)***	*118 (65.9)*	*38 (21.2)*	*23 (12.9)*

Note: * 11-point scale; ** Categories “Very Willing” and “Somewhat Willing” combined.

internet questionnaire. All other variables are not significant, but their relationship is nonetheless of interest. For example, women prefer internet questionnaires by 7% points, and men prefer telephone by almost the same margin. Secondos and non-secondos are rather similar with non-secondos slightly preferring telephone more; Married respondents to a greater extent prefer telephone interviewing. Part-time workers on average much prefer an internet questionnaire with over 64% of them preferring this mode; however, full-time workers are below the average. A similar finding for those with only a compulsory level of education, with 69.2% of them preferring an internet questionnaire, compared to 58% with the highest education level. Unemployed people are the group with the most ‘no preference’ answers at 33.3%.

Another indicator of mode preference is the participants’ stated mean willingness to complete the next wave of the survey by internet or telephone. As expected, it correlates well with the mode preference question as the highest willingness occurs in the most preferred

mode. Of note is the fact that those who prefer a telephone interview are less willing to participate via an internet questionnaire, than those who prefer an internet questionnaire are via a telephone interview. Furthermore, those who answered “no preference” in terms of their mode preference are slightly more willing to participate via a telephone interview than via the internet. Participants were also asked how willing they would be to participate in the next wave via a smartphone and tablet. The results show that those having high willingness via either device prefer an internet questionnaire by 64.1% and 65.9% respectively, with high smartphone willingness preferring telephone interviews slightly more than those with high tablet willingness. Overall, all subgroups prefer an internet questionnaire but differ in terms of their magnitude of telephone preference, and this telephone preference is what is key to ascertaining attrition in the case of switching to the web mode.

Table 6 takes the two mode willingness variables and constructs a sole variable, net willingness, from them. It leaves us with three groups, those who are more willing via telephone than web; those who are more willing via internet than telephone and those who are equally willing via both modes. Those who are equally willing via the two modes are the largest group at 45.6% of the sample, indicating a sample that is open to both modes. The second-largest group is those who are more willing via web, at 30.4%, compared to the final group at 23.9%, for those more willing via telephone. Excluding those who have equal willingness, those who are more willing via web make up 56% of the sample which bodes well for the envisioned switch, yet that leaves 44% who have an overall net willingness of telephone over web, and hence, could potentially be at risk of dropping out of the study in the event of a mode switch.

This pattern persists over age categories and sex, though only 20.5% of women are more willing via telephone one of the smallest percentages for this category. They too are more likely to have equal mode willingness as 48.5% of women fall into this group. Secondos are slightly more willing via the web mode than average, and nearly two percentage points more than non-secondos, suggesting that the envisioned switch will not damage the secondo oversampling ratio detrimentally.

Despite overall being more likely to be ‘more willing via web than telephone’, it is those in full-time work who are most ‘more willing via telephone than web’ compared to other occupational categories, with 26.1% compared to just 17.4% of those who work part-time in the same regard. Both full-time workers and part-time workers are similar in terms of being ‘more willing via web’, yet they differ greatly in the equal willingness category by almost 9%. Overall though, all occupational groups are more willing via web than telephone, except the ‘other’ category. Moreover, being ‘more willing via web’ is most prevalent at 33.6% with those in school or training, the highest percentage; with only 23.1% of this group being more willing via telephone than web.

In terms of education, all categories are more willing via web than telephone, but the difference is only 0.5% for apprentices. This divergence is starker for the highest level of education, University-level, at 9.5% and the starkest for the lowest education level, compulsory schooling, at a near 17% difference, suggesting a non-linear relationship between education and net willingness. In terms of being more willing via web, it is university-level education that tops the education categories, closely followed by compulsory schooling level, with apprentices who are the least willing by web. Although the differences were not

Table 6

Net willingness over socio-demographic variables

Variables	All	More willing via telephone (%)	More willing via web (%)	Equally willing via web or telephone (%)
Total	**784 (100.0)**	**187 (23.9)**	**238 (30.4)**	**359 (45.6)**
Age:				
Under 25	**357 (45.5)**	80 (22.4)	108 (30.3)	169 (47.3)
25 and over	**427 (54.5)**	107 (25.1)	130 (30.4)	190 (44.5)
Sex:				
Male	**374 (47.7)**	103 (27.6)	111 (29.7)	160 (42.8)
Female	**410 (52.3)**	84 (20.5)	127 (31.0)	199 (48.5)
Secondo:				
Yes	**301 (38.4)**	71 (23.6)	94 (31.2)	136 (45.2)
No	**483 (61.6)**	116 (24.0)	144 (29.8)	223 (46.2)
Civil Status:				
Single, never married	**726 (92.6)**	174 (24.0)	221 (30.5)	329 (45.4)
Married	**58 (7.4)**	13 (22.4)	17 (29.3)	28 (48.3)
Occupational Status:				
Full time work (37h min)	**311 (39.7)**	81 (26.1)	88 (28.4)	142 (45.7)
Part time work 36h and less)	**92 (11.7)**	16 (17.4)	26 (28.3)	50 (54.4)
In school or training	**333 (42.5)**	77 (23.1)	112 (33.6)	144 (43.2)
Unemployed	**18 (2.3)**	4 (22.2)	5 (27.8)	9 (50.0)
Other	**30 (3.8)**	9 (30.0)	7 (23.3)	14 (46.7)
Education Level:				
Bachelor/Maturity	**236 (30.1)**	53 (22.5)	70 (29.7)	113 (47.9)
University/Academic high school	**200 (25.5)**	49 (24.5)	68 (34.0)	83 (41.5)
Apprenticeship	**199 (25.4)**	54 (27.1)	55 (27.6)	90 (45.2)
Compulsory school/ Elementary vocational training	**65 (8.3)**	11 (16.9)	22 (33.8)	32 (49.2)
Other	**84 (10.7)**	20 (23.8)	23 (27.4)	41 (48.8)

Sample origin:				
OFS sample	**70 (8.9)**	17 (24.3)	22 (31.4)	31 (44.3)
Network sample	**714 (91.1)**	170 (23.8)	216 (30.3)	328 (45.9)
Attrition risk characteristics				
Never use internet	**1 (0.13)**	0	1 (100.0)	0
Internet use: Less than everyday	**40 (5.1)**	11 (27.5)	11 (27.5)	18 (45.0)
Device: Smartphone only	**28 (3.6)**	7 (25.0)	6 (21.4)	15 (53.6)
Device: Tablet only	**0 (0.0)**	0 (0.0)	0 (0.0)	0 (0.0)
Device: Laptop only	**2 (0.26)**	0	1 (50.0)	1 (50.0)
Device: Smartphone & Tablet only	**12 (1.5)**	2 (16.7)	4 (33.3)	6 (50.0)

significant (perhaps due to small sample sizes) this pattern of results would suggest a slightly increased risk of attrition for those with lower levels of education (excluding those with compulsory level education) – if being unwilling to complete the survey by web translate into nonparticipation.

Overall, descriptive analysis shows that some small attrition can be attributed to lack of internet use, lack of a suited non-mobile device, and preferring the telephone mode over web. Thus, it can be assumed that attrition will be slightly more than the average rate of 14% because of these factors, unique to only the wave after the switch.

Estimating attrition

Given that participants who stated they would be more willing to participate in the study by telephone than by internet are most likely to be at risk of dropping out in the event of a mode switch, the next analyses aim to model being more willing via telephone and preferring

telephone to ascertain the magnitude and selectivity of said attrition. This is important to address the second research question, which looks at whether the attrition caused by the switch to the web mode will be selective in nature. To do so, the parameter coefficients of a logistic regression model were estimated to see which characteristics of participants are predictive of telephone preference or stated willingness for telephone interviewing. In the following models, those who are already using the web mode are excluded they are not of a concern for this research.

Model 1 uses the net willingness variable which has been dichotomised to being more willing via telephone versus the other categories, allowing us to see which factors are relevant to falling into this group. Overall the model has a high Nagelkerke R_2 figure at 49%. The model χ^2 is significant, indicating the model is better than chance at predicting the dependent variable.

Preferring telephone is highly significant, with a p-value of less than 0.001, confirming that these two concepts, willingness and preference, are highly interconnected. This is an important finding in itself as it confirms that analysis on preference can be a proxy for willingness, yet, the strength of the correlation is unknown between preference and stated willingness and actual survey behaviour, though the literature suggests it is strong for device intention (Haan et al., 2019, p.526) and actual behaviour (Wenz et al., 2019, p.2). The odds ratio shows that preferring telephone makes someone 28.4 times more likely to be more willing to participate via telephone than web. Also significant in model 1, though to a lesser extent is, interestingly, using one's smartphone for looking at social media and the variable

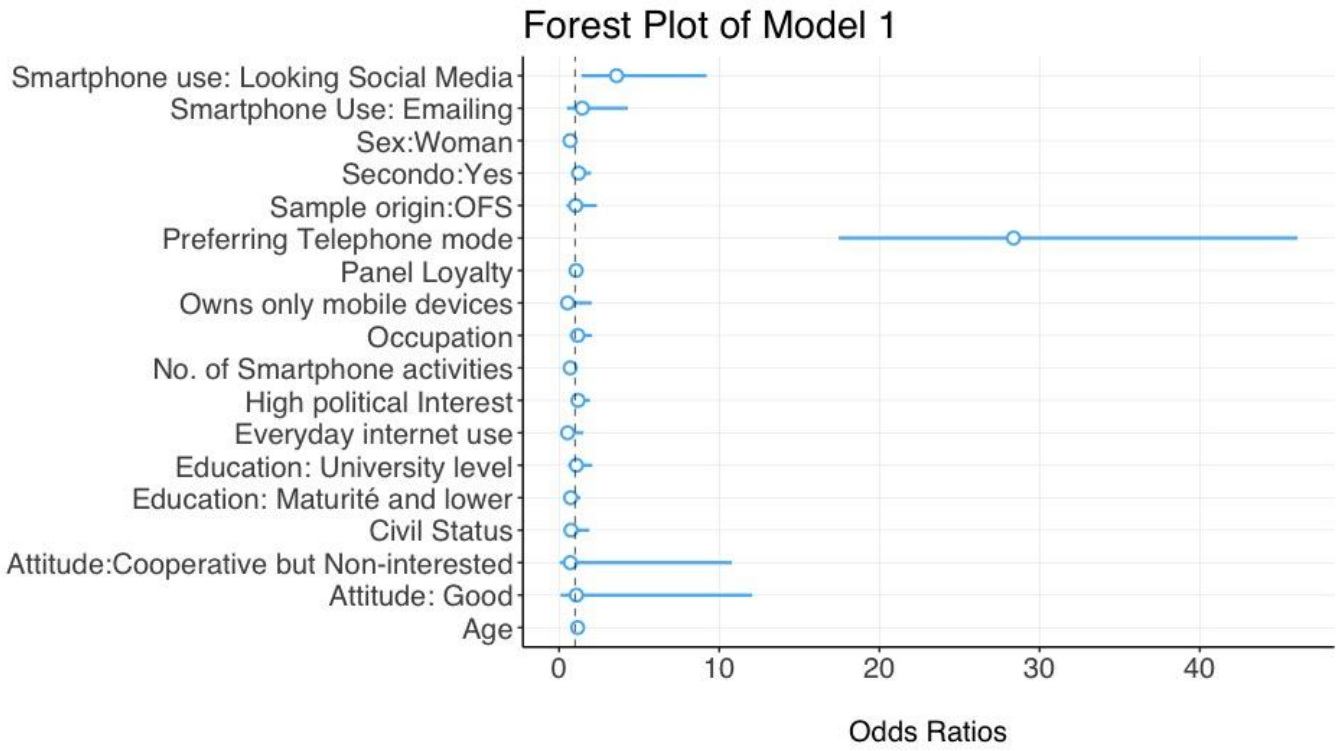
Model 1: More willing to participate via telephone interviewing than via an internet questionnaire

	B (SE)		95% confidence intervals for Odds ratio		
			Lower	Odds ratio	Upper
Constant	-3.31 (2.38)				
Mode Preference: Telephone (ref: Other)	3.35 (0.24)	***	17.96	28.37	46.19
Internet use frequency: Every day (ref: Several times a week or less)	-0.62 (0.51)		0.20	0.54	1.5
Age	0.05 (0.05)		0.96	1.06	1.17
Sex: Female (ref: Male)	-0.37 (0.23)		0.44	0.69	1.08
Secondo status: Yes	0.21 (0.24)		0.77	1.23	1.96
Number of waves participated in (1-6)	0.07 (0.12)		0.85	1.07	1.36
Uses smartphone for emailing: Yes	3.70 (0.51)		0.54	1.45	4.08
Use smartphone for looking at social media: Yes	1.28 (0.49)	**	1.40	3.60	9.49
Civil Status: Married (ref: Single)	-0.29 (0.47)		0.29	0.75	1.85
Mobile devices: Smartphone and/or tablet only	-0.62 (0.52)		0.19	0.54	1.46
Occupation: Schooling or other (ref: In-work)	0.16 (0.27)		0.69	1.17	1.98
Sample Origin: OFS (ref: Network sample)	0.04 (0.39)		0.48	1.04	2.20
Number of smartphone activities (0-11)	-0.19 (0.07)	**	0.72	0.82	0.95
Education: Maturité or lower (ref: Apprenticeship/Vocational)	-0.32 (0.29)		0.41	0.72	1.27
Education: University Level (ref: Apprenticeship/Vocational)	0.08 (0.33)		0.57	1.08	2.06
Political Interest: High	0.18 (0.24)		0.75	1.19	1.90
Interviewer adjudged attitude of respondent:			0.03	0.71	33.98
Cooperative but not interested (ref: Bad attitude)	-0.34 (1.83)				
Interviewer adjudged attitude of respondent: Good attitude (ref: Bad attitude)	0.08 (1.67)		0.06	1.08	41.34

Note. R_2 = .35 (Hosmer-Lemeshow), .33 (Cox-Snell), .49 (Nagelkerke). Model χ^2 = 289.59***

p < 0 '***' p < 0.001 '**' p < 0.01 'p < *' 0.05 p < '.' 0.1

Figure 1: Forest plot of model 1 odds ratios



that counts the number of smartphone activities someone does. Using a smartphone to look at social media means that one is 3.6 times more likely to be more willing to participate via telephone than web, whereas for every extra smartphone activity one does, one is 1.2 times less likely to be more willing to participate via telephone than web. The latter is logical as we would assume that someone who uses their smartphone for more activities would be more competent in web-related activities and thus more willing to participate via the web mode (and less via telephone).

Model 1 is largely driven by one extremely significant coefficient – ‘preferring telephone’; the effect of which is visible in the above forest plot. As such, it would be useful to see the same model without the telephone preference variable due to potential multicollinearity; that way, it could be seen if any other coefficients are significant.

Model 2 has the same dependent variable and the same independent variables as model 1 except the 'preferring telephone' variable which was extremely significant and aims to see if any other variables are significant but were masked by the telephone preference variable. It's Nagelkerke R^2 figure is much lower than model 1 at 4%, highlighting the explanatory power of mode preference and the relative lack of variance explanation that the independent variables in model 2 provide. The model χ^2 is also non-significant in this case, indicating that this is a weaker model than model 1.

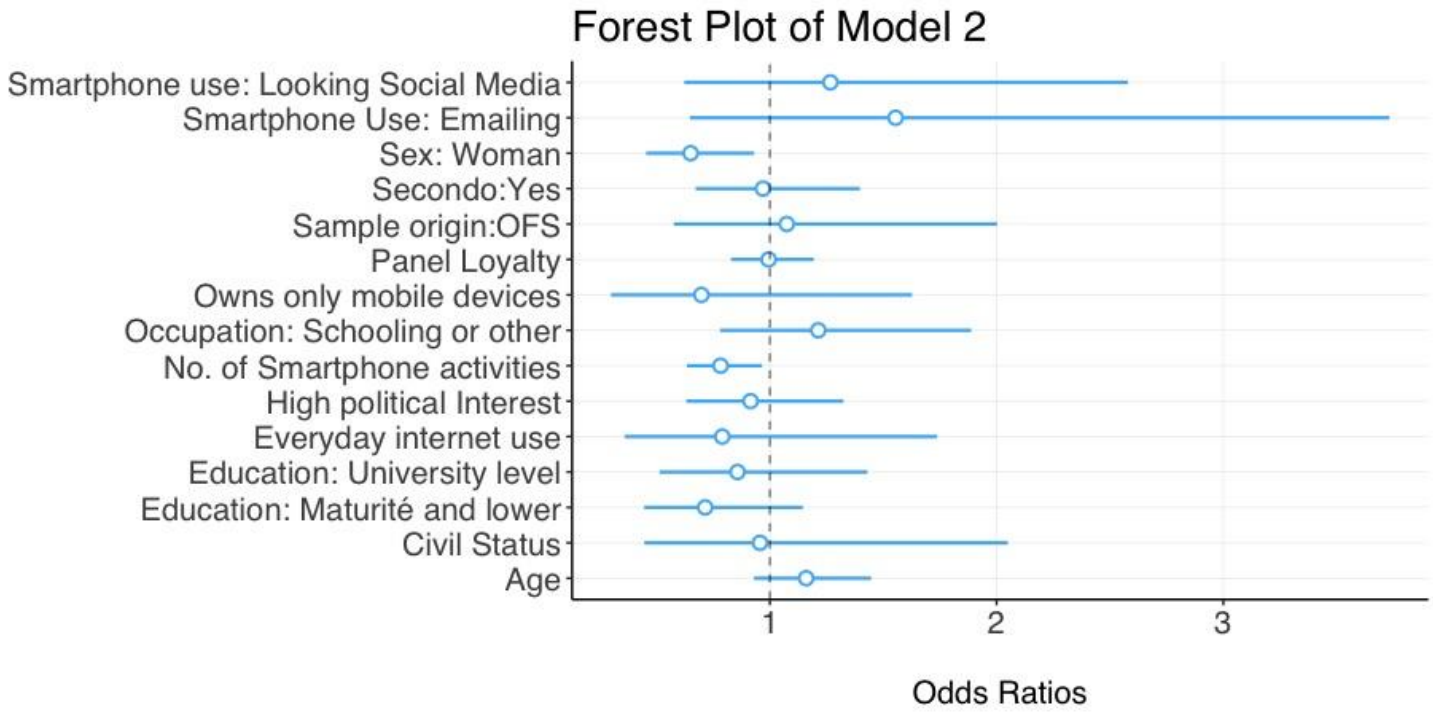
Only two variables are significant, the number of smartphone activities (the same as in model 1 though it is less significant in this case) and sex. This means that being a woman decreases one's likelihood to be more willing to respond via telephone than web by nearly 1.5 times and that for each extra smartphone activity done, one is 1.14 less likely to be more willing to respond via telephone than web. The finding on smartphone activities is roughly the same between the same model, but the finding on sex is hidden by telephone preference in the first model, suggesting some multicollinearity, i.e. some of the variance explained by sex in model 2 is explained by mode preference in model 1. Also, looking at social media is no longer significant in model 2.

Model 2: More willing to participate via telephone interviewing than via an internet questionnaire (Without mode preference as an Independent variable)

	B (SE)		95% confidence intervals for Odds ratio		
			Lower	Odds ratio	Upper
Constant	1.44 (1.76)				
Internet use frequency: Every day (ref: Several times a week or less)	-0.24 (0.39)		0.38	0.79	1.75
Age	0.05 (0.04)		0.98	1.06	1.14
Sex: Female (ref: Male)	-0.43 (0.18)	*	0.46	0.65	0.92
Secondo status: Yes	-0.03 (0.18)		0.67	0.97	1.39
Number of waves participated in (1-6)	-0.01 (0.10)		0.83	0.99	1.20
Uses smartphone for emailing: Yes	0.44 (0.42)		0.71	1.55	3.68
Use smartphone for looking at social media: Yes	0.24 (0.38)		0.61	1.27	2.71
Civil Status: Married (ref: Single)	-0.04 (0.37)		0.45	0.96	1.95
Mobile devices: Smartphone and/or tablet only	-0.36 (0.41)		0.29	0.70	1.52
Occupation: Schooling or other (ref: In-work)	0.19 (0.21)		0.80	1.21	1.84
Sample Origin: OFS (ref: Network sample)	0.07 (0.31)		0.58	1.07	1.92
Number of smartphone activities (0-11)	-0.13 (0.06)	*	0.79	0.88	0.98
Education: Maturité or lower (ref: Apprenticeship/Vocational)	-0.34 (0.23)		0.46	0.71	1.12
Education: University Level (ref: Apprenticeship/Vocational)	-0.15 (0.25)		0.52	0.86	1.41
Political Interest: High	-0.09 (0.18)		0.64	0.91	1.31
Interviewer adjudged attitude of respondent: Cooperative but not interested (ref: Bad attitude)	-0.02 (1.32)		0.08	0.97	23.75
Interviewer adjudged attitude of respondent: Good attitude (ref: Bad attitude)	0.39 (1.20)		0.17	1.47	31.34

Note. R_2 = .02 (Hosmer-Lemeshow), .03 (Cox-Snell), .04 (Nagelkerke). Model χ^2 = 18.72
p < 0 '***' p < 0.001 '**' p < 0.01 'p < *' 0.05 p < '.' 0.1

Figure 2: Forest plot of model 2 odds ratios.



Model 3 looks at the factors which contribute to preferring telephone as it has been shown to be a significant factor in the previous models. Understanding this could shed light on who is likely to no longer participate in the study if web becomes the mode of data collection, assuming that those who prefer telephone are likely to attrite. Overall, here the Nagelkerke R^2 is low, at 5%, so the model can only explain 5% of the variance, and thus it seems other unincluded factors can explain it to a greater extent. This weakness is reflected in the fact that the model χ^2 is non-significant.

The significant contributory factors in this model are ‘being female’ and using your smartphone to look at social media. Both of which make one less likely to prefer telephone interviewing. Being female makes one 1.4 times less likely to prefer telephone interviewing and using a smartphone to look at social media makes one 2.24 times less likely to do so. Also, being highly politically interested (above five on an 11-point scale) is almost significant

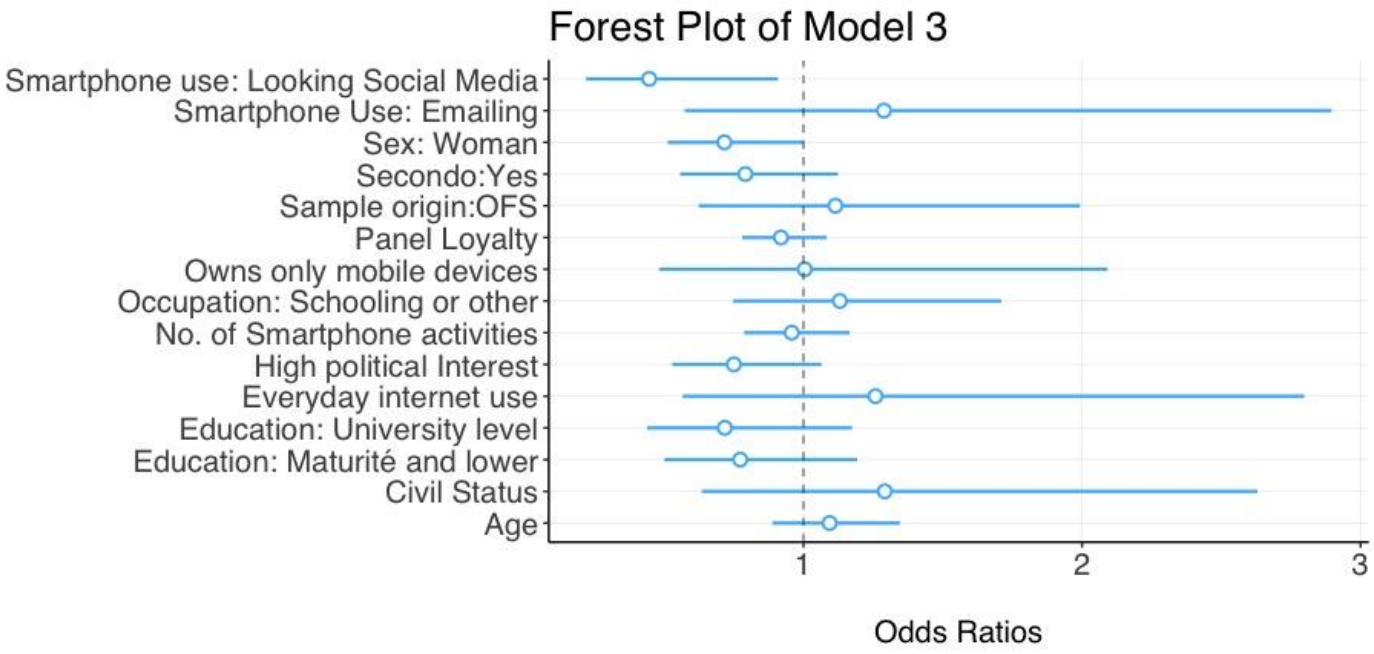
Model 3: Telephone preference

	B (SE)		95% confidence intervals for Odds ratio		
			Lower	Odds ratio	Upper
Constant	-0.69 (1.71)				
Internet use frequency: Every day (ref: Several times a week or less)	0.23 (0.39)		0.60	1.26	2.82
Age	0.03 (0.04)		0.96	1.03	1.11
Sex: Female (ref: Male)	-0.33 (0.17)	*	0.52	0.72	0.99
Secondo status: Yes	-0.23 (0.17)		0.56	0.79	1.11
Number of waves participated in (1-6)	-0.09 (0.09)		0.77	0.91	1.09
Uses smartphone for emailing: Yes	0.25 (0.39)		0.61	1.29	2.86
Use smartphone for looking at social media: Yes	-0.81 (0.35)	*	0.22	0.45	0.89
Civil Status: Married (ref: Single)	0.26 (0.34)		0.65	1.29	2.52
Mobile devices: Smartphone and/or tablet only	0.004 (0.36)		0.48	1.00	2.02
Occupation: Schooling or other (ref: In-work)	0.12 (0.20)		0.76	1.13	1.67
Sample Origin: OFS (ref: Network sample)	0.11 (0.29)		0.62	1.11	1.94
Number of smartphone activities (0-11)	-0.02 (0.05)		0.88	0.98	1.08
Education: Maturité or lower (ref: Apprenticeship/Vocational)	-0.26 (0.21)		0.51	0.77	1.17
Education: University Level (ref: Apprenticeship/Vocational)	-0.33 (0.24)		0.44	0.72	1.15
Political Interest: High	-0.29 (0.17)	.	0.53	0.75	1.05
Interviewer adjudged attitude of respondent: Cooperative but not interested (ref: Bad attitude)	0.34 (1.32)		0.13	1.40	34.43
Interviewer adjudged attitude of respondent: Good attitude (ref: Bad attitude)	0.68 (1.22)		0.23	1.97	42.85

Note. R_2 = .03 (Hosmer-Lemeshow), .03 (Cox-Snell), .05 (Nagelkerke). Model χ^2 = 24.05
p < 0 '***' p < 0.001 '**' p < 0.01 'p < *' 0.05 p < '.' 0.1

but the p-value is greater 0.05, yet this suggests being more political interested would make one less likely to prefer telephone interviewing (1.33 times). The forest plot below illustrates the odds ratios.

Figure 3: Forest plot of model 3 odds ratios.



Respondents were also asked how willing they would be to complete the survey on their smartphone; model 4 looks at this case of smartphone willingness. This could be useful to understand the willingness of those that have only a smartphone, or for future waves where smartphone participation may increase. Smartphone completion is a growing phenomenon in web surveys, even in unintended circumstances (De Bruijne and Wijnant, 2014) and so it is of use to understand the factors motivating, or disincentivising this. As such, the same independent variables as model 1 were inserted into this model with the dependent variable being high smartphone willingness, and the results are below.

The Nagelkerke R^2 for this model is 20%, a relatively good model compared to the previous models; this means that the independent variables explain 20% of the variance in the dependent variable, in this case, smartphone willingness. The model χ^2 is highly significant in this model, suggesting it is better than chance at predicting smartphone willingness. Overall, 74.34% of

Model 4: Smartphone willingness

	B (SE)		95% confidence intervals for Odds ratio		
			Lower	Odds ratio	Upper
Constant	-1.30 (1.76)				
Mode Preference: Telephone (ref: Other)	-1.38 (0.19)	***	0.17	0.27	0.37
Internet use frequency: Every day (ref: Several times a week or less)	0.35 (0.41)		0.61	1.42	3.15
Age	-0.03 (0.04)		0.89	0.97	1.05
Sex: Female (ref: Male)	0.36 (0.19)	.	0.99	1.43	2.09
Secondo status: Yes	0.09 (0.20)		0.74	1.09	1.61
Number of waves participated in (1-6)	-0.05 (0.10)		0.77	0.95	1.15
Uses smartphone for emailing: Yes	1.45 (0.40)	***	1.98	4.27	9.43
Use smartphone for looking at social media: Yes	-0.34 (0.40)		0.32	0.71	1.53
Civil Status: Married (ref: Single)	0.09 (0.38)		0.44	0.91	1.98
Mobile devices: Smartphone and/or tablet only	0.15 (0.43)		0.52	1.17	2.83
Occupation: Schooling or other (ref: In-work)	-0.05 (0.22)		0.61	0.94	1.47
Sample Origin: OFS (ref: Network sample)	0.37 (0.35)		0.75	1.45	2.95
Number of smartphone activities (0-11)	0.14 (0.06)	*	1.02	1.14	1.28
Education: Maturité or lower (ref: Apprenticeship/Vocational)	0.32 (0.24)		0.86	1.38	2.23
Education: University Level (ref: Apprenticeship/Vocational)	-0.28 (0.27)		0.45	0.76	1.28
Political Interest: High	-0.07 (0.20)		0.64	0.94	1.38
Interviewer adjudged attitude of respondent: Cooperative but not interested (ref: Bad attitude)	0.17 (1.22)		0.10	1.19	14.00
Interviewer adjudged attitude of respondent: Good attitude (ref: Bad attitude)	1.10 (1.12)		0.29	3.01	29.62

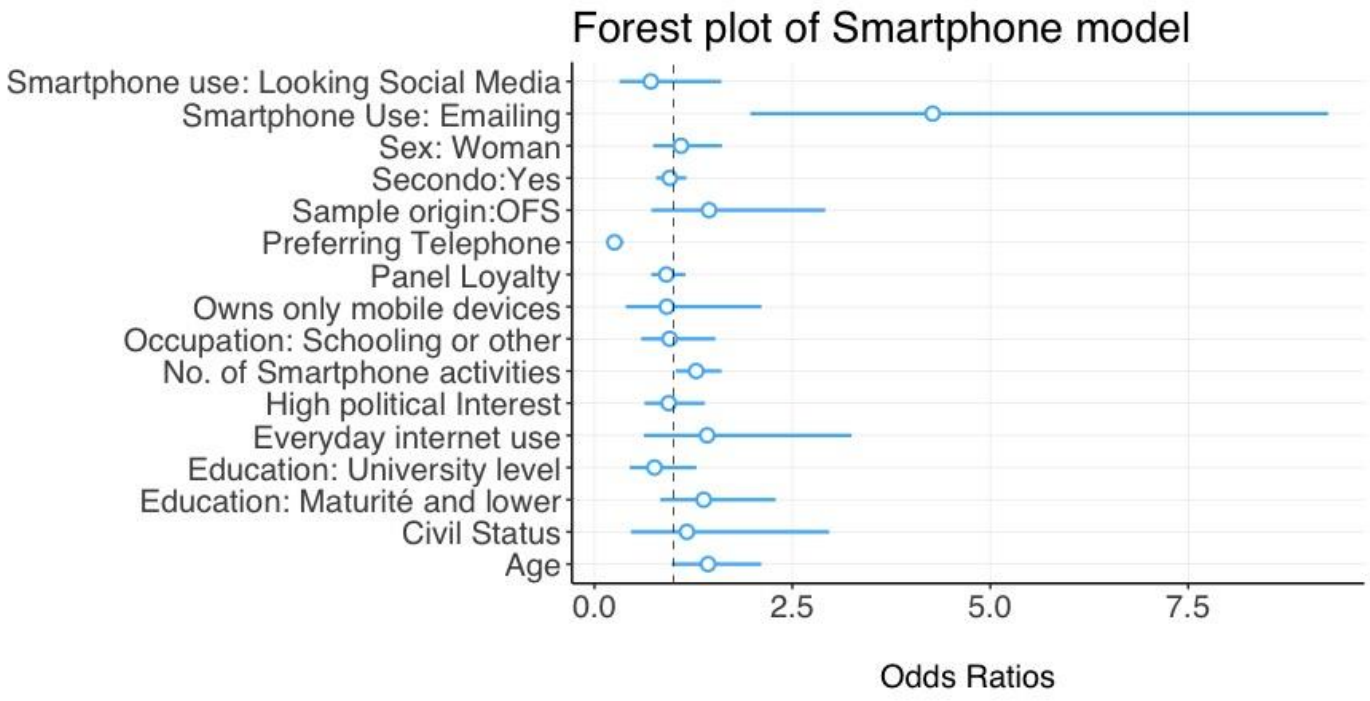
Note. R_2 = .13 (Hosmer-Lemeshow), .13 (Cox-Snell), .20 (Nagelkerke). Model χ^2 = 102.70 ***

p < 0 '***' p < 0.001 '**' p < 0.01 'p < *' 0.05 p < '.' 0.1

the sample (who are not already using the web mode option) are deemed to have high smartphone suggesting it is better than chance at predicting smartphone willingness. Overall,

74.34% of the sample (who are not already using the web mode option) are deemed to have high smartphone willingness. The results show that a lot of variables are significant to this concept. Firstly, preferring telephone is extremely significant (p-value <0.001) in a negative way, suggesting that preferring telephone makes one 3.7 times less likely to be willing to participate via a smartphone, this is to be expected, as by definition these respondents are less keen on web responding. Secondly, and also as highly significant is using a smartphone for emailing purposes, the odds ratio indicates that it makes one 4.27 times more likely to be willing to respond via telephone and the p-value is less than 0.001. Thirdly, the number of activities undertaken on a smartphone is also significant here, though, to a lesser extent (p-value less than 0.05), it indicates that for each extra activity undertaken one is 1.14 times more likely to be willing to respond via smartphone. Therefore, we can draw the conclusion that those who are more competent in smartphone activities, especially emailing and who prefer responding via the web would be best suited to smartphone responding.

Figure 4: Forest plot of smartphone model



Composition of attrition

As a way of surmising the socio-demographic breakdown of those who are at the highest risk of attrition, a variable was created whereby those with a predicted probability in the 9th decile or above of the predicted probability distributions of model 1 and model 2 were considered to be at high risk of attrition (N=119). Table 7 looks at the socio-demographics of these people compared to the overall sample to illustrate any issues where selective attrition could arise.

Table 7: High risk of attrition group demographics.

Variables	Overall (%)	Attrition risk group (%)	Difference (%)
Total (N)	**784**	**119**	--
Age:			
Under 25	*45.5*	*47.9*	*+2.4*
25 and over	*54.5*	*52.1*	*-2.4*
Sex:			
Male	*47.7*	*49.6*	*+1.9*
Female	*52.3*	*50.4*	*-1.9*
Secondo:			
Yes	*38.4*	*35.3*	*-3.1*
No	*61.6*	*64.7*	*+3.1*
Civil Status:			
Single, never married	*92.6*	*94.1*	*+1.5*
Married	*7.4*	*5.9*	*-1.5*
Occupational Status:			
Full time work (37h min)	*39.7*	*38.7*	*-1*
Part time work (36h or less)	*11.7*	*10.1*	*-1.6*
In school or training	*42.5*	*44.5*	*+2*
Unemployed	*2.3*	*1.6*	*-0.7*
Other	*3.8*	*5.1*	*+1.3*
Education Level:			
Bachelor/Maturity	*30.1*	*31.1*	*+1*
University/Academic high school	*25.5*	*26.1*	*+0.6*
Apprenticeship	*25.3*	*29.4*	*+4.1*
Compulsory school/ Elementary vocational training	*8.3*	*5.9*	*-2.3*
Other	*10.6*	*7.5*	*-3.1*
Sample origin:			
OFS sample	*8.9*	*10.1*	*+2.2*
Network sample	*91.1*	*89.9*	*-2.2*

Though there are no significant associations between these variables and being in the attrition risk group, there are some notable differences. The most marked difference is apprentices, with them being 4.1% more numerous in the attrition-risk group than the overall sample average. Elsewhere, under 25s are 2.4% more numerous in the attrition-risk group, men are also at a higher risk of attrition by 1.9%, so too are non-secondos (3.1%) and singletons (1.5%). Of the major occupational statuses, it is only ‘in school or training’ that has a higher than average percentage in the attrition-risk group. In terms of education, apart from apprenticeships, the only concern is that there are 1% more than average in the attrition risk group for bachelor/maturity level education, apart from University-level which is 0.6% high than average, the other categories are below average. Finally, the sample origin shows that those in the OFS sample are disproportionately more likely to be at risk of attrition.

Discussion

This research aimed to determine whether there would be any increased sample attrition due to the proposed mode switch from telephone interviewing to an internet questionnaire. Then if so, whether this attrition would be selective in nature. The literature suggests that switching to the web mode only could potentially induce mode effects due to different coverage issues and willingness rates of the web mode, as well as, measurement issues caused by the different nature of web surveys, namely the self-administered nature and visual presentation. This dissertation looked at the former, known collectively as selection effects and while measurement effects can play a role here, the interest is predominantly on attrition and selectivity caused by the mode switch.

The first research question aims to ascertain whether there will be mode-attributable increased sample attrition. Firstly, in order to determine if it increased, the attrition rate over time was analysed. The study has witnessed an average of 14% of attrition per wave since its inception and no longer reflects the intended target population recruited in the first wave. In particular, the ratio between secondos and non-secondos, which was a key pillar of the sample design has deteriorated.

Internet characteristics

Hypotheses on internet characteristics were set based on situations were web surveying was perceived not to be suited. In terms of internet access, the results show there is no one without internet access (With just one individual who does not use the internet for personal purposes), this bodes well for the suitability of the sample with web surveying. Therefore, H1 is disproved as internet access issues is not likely to cause attrition.

Those connecting to the internet via mobile devices only are of a relatively small proportion. It was hypothesised that these individuals would be at higher risk of leaving the sample due to the lack of device choice potentially increasing response burden. In particular, use of the smartphone only can increase the burden and thus disincentivise participation for a variety of reasons. Firstly, the reduced screen size means that responding to questions is harder, especially when responding to batteries. Also, the display of questions may not be uniform between devices such that comparison is hindered – a concern for methodologists. Secondly, respondents may associate smartphones with other activities of shorter time or more leisurely appeal. They, therefore, would prefer not to respond to a survey via it; this is reflected in the fact that some of those who are willing to participate via the web are not via a smartphone. Thirdly, their portable nature means that they are often used in environments where distractions are more frequent, and they become subject to context effects. If one is not fully focused on the survey administration, then satisficing techniques may be employed in order to quicken the process. De Bruijne and Wijnant (2014) support this by suggesting that it could be that surveys are associated with a longer task, thus more suited to conditions similar to computer tasks such as sitting down and in a stationary position (p.13)

The analyses found that there is a small proportion of participants who use only a smartphone, only a laptop or only both of these devices, as such, these could be at risk of leaving the sample. H2 is, therefore, not fully rejected as device access may engender a small amount of attrition. Though survey administration is entirely possible via these devices, so it is not certain they will leave (See appendix for the devices already being used for web response in the LCS).

The finding that the internet is used overwhelmingly on an everyday basis by this sample bodes well for future participation rates. Though there is one individual who claims they never access the internet, this respondent does use the internet for numerous activities, owns a PC, laptop, smartphone and tablet and prefers an internet questionnaire; therefore, it seems their response is an error. There is a small proportion who use the internet less than every day, though there is no evidence to suggest that they will leave the sample, which would require a once-a-year internet utilisation, because of this. H3 can, therefore, be rejected.

H5, H6 and H7 can also be rejected as the analyses found that no respondent uses the internet solely at work, in public places or in places of education; therefore, no attrition is expected to be caused because of location issues. Though respondents use the internet in these locations, they are not the sole location, so this bodes well as the three locations are, by definition, prone to context effects such as environmental distractions.

Analysis of the internet characteristics and usage of the current sample leads us to conclude that this sample is relatively competent technologically. The majority have a combination of devices, both fixed and mobile, and use mobile devices, in particular, smartphones, on a daily basis. This is to be expected of a cohort where the maximum age is 32 and so many have grown up in the technological milieu that Zuboff (2019) describes previously. Pertinent also was the fact that in terms of these internet characteristics, secondos and non-secondos are broadly similar; thus, the oversampling ratio is not threatened by attrition caused by lack of competence or usage.

The analysis in terms of smartphone activities attempted to moderate this finding to see if there was some nuance in smartphone competence, which is a good indicator of overall internet competence. While there are some significant associations between individual activities and some demographics, notably sex and secondo status, showing that different subgroups use smartphones for different purposes, no real conclusions can be drawn from this. However, as we have seen, some activities were found to be relevant in the later logistic models.

Mode preference

The concept of mode preference was analysed as this could be an indicator of potential future attrition if the mode preference is not suited to the proposed new primary mode. The literature has found that this measure is highly influenced by the mode of completion. This is because people often do not have preconceived views on what mode of survey data collection they prefer and thus their response is primarily based on temporarily-accessible information and socially desirable response patterns, such that the respondent answers the mode that they are currently replying in (Al Baghal and Kelley, 2016). Indeed, in this sample mode of competition and mode preference was found to be significantly associated. Nonetheless, in the case of the LCS where the majority of the respondents participate via CATI, over half claimed their mode preference was an internet questionnaire. Thus, this effect which is likely to inflate the CATI preference figure is not a significant concern here as the non-primary mode is the preferred mode.

Nearly 57% prefer an internet questionnaire, compared with just under 30% who prefer telephone interviewing – the current primary mode. The fact that twice as many respondents prefer the web mode than the current mode might be seen as boding well for a

switch to web, but the fact that 30% do not prefer the web mode is not a good sign. As such further analysis, including stated willingness, is required to determine the severity of this. In terms of subgroup differences with regard to mode preference, findings show that all subgroups analysed prefer web surveys over telephone interviewing, but the groups differ in terms of the amount they prefer each mode. For example, women prefer an internet questionnaire 7% more than men, non-secondos prefer telephone more than secondos, as well as married individuals compared to singletons.

Conversely, part-time workers prefer internet surveys more than any other occupational status. Thus, although all groups prefer an internet questionnaire, the groups that prefer telephone interviewing to a greater extent may be more prone to attrition, namely, men and married individuals. As such, H4 cannot be rejected, as it seems that this could be a potentially large cause of future attrition. It is, so far, the largest group that is unsuited to a proposed switch to web-only surveying. As mentioned, stated willingness might give further indications into the preferences of the sample, as it is a concept highly related to mode preference, but is numeric in nature.

Stated Willingness

A key finding is that when looking at mean willingness for internet surveys and telephone interviewing across the mode preference categories, we find that those who prefer telephone interviewing are less willing to participate via the web than those who prefer an internet survey are via telephone. In other words, those who prefer internet surveys are more willing to participate via the non-preferred mode. Furthermore, those who answered ‘no preference’ to the mode preference question are slightly more willing via telephone (9.31 vs 9.18). This indicates that despite the headline mode preference figures suggests a large

preference for web, this sample may be more suited to telephone interviewing as those who prefer web and those who have no preference are more willing via telephone, compared to how willing those who prefer telephone are towards web. However, this only considers the willingness means for the mode preference variable, which we know can be unreliable. Looking at the willingness itself may give a clearer image.

Therefore, the net willingness variable, constructed from subtracting the two mode willingness scores, was created. It shows that overall, more respondents are more willing via web that telephone than being more willing via telephone than web. However, the largest proportion has equal willingness via both modes, indicating that the sample is largely tolerant of both modes. This is a more reliable measure as it combines two different numeric variables and gives a sense of 'net' willingness which should indicate which mode is more likely to be used given the choice. It correlates well with mode preference and gives a more detailed image of willingness. For example, of those who said their mode preference was an internet questionnaire, 17 individuals are more willing via telephone than web, and 209 individuals are equally willing via both modes. The same applies to those who expressed a preference for telephone interviewing; four are more willing via web than telephone, and 79 are equally willing via both modes. So, while the two concepts correlate in that the majority of those who preferred web interviewing are more willing via web than telephone and vice versa, the net willingness variable shows a more detailed picture. Despite this, the proportion of the sample who are more willing via telephone that web is a cause for concern, these are likely to attrite if the primary mode of data collection changes to web-only and amounts to 187 people. If all of these panel members were to not respond in the next wave, then this would be a dramatic increase in the attrition rate. H4 cannot be rejected then, and it seems that mode preference

and the related concept of stated willingness could be factors which engender higher levels of attrition which is attributable to the mode switch.

Looking at subgroups, the results show that women, secondos, those in school or training, with university or compulsory school level of education are more than averagely likely to be more willing to participate via web than telephone. Compared with men, full-time workers, those over 25-year-old and apprentices are more likely than average to be more willing via telephone than web, who are at risk of attrition, yet these groups are overall more willing via web than telephone. Apprentices need particular care as they tend to have the smallest difference between the number of them who are more willing via web than telephone and those who are more willing via telephone than web. These findings are in line with the findings on mode preference and illustrate that there are group differences despite the fact that all subgroups are overall more willing to participate via web than telephone. A fact that could engender some selective attrition, however, no demographic variable analysed was found to be associated with mode preference or net willingness, suggesting a random element to willingness. Furthermore, this gives us no evidence that any attrition caused by a lack of willingness towards the web mode would be selective in nature.

Predicting attrition factors

Delving further, in order to ascertain who is at most risk of attrition, logistic regression modelled the factors contributing to whether someone was more willing via telephone than web. Results show that having a mode preference of telephone is, as expected, highly related to this measure, but so too is the number of smartphone activities one does, as well as, whether someone uses a smartphone to consult social media. When removing mode preference which is extremely significant, sex becomes significant as well. The finding that

consulting social media on a smartphone is interesting; it perhaps reinforces the idea that someone people associate the web for more leisurely activities such as social media and not with survey completion. The fact that sex is significant here is worrying, as it can suggest attrition can be selective according to sex. In other words, women are more willing to participate via web surveys and men via telephone interviewing. Therefore, we can state that attrition is likely to be more prevalent amongst men, in the case of a switch to web, if stated willingness translates into actual attrition. Overall, despite the potential multicollinearity between preferring telephone and being more willing via telephone than web, model 1 is a stronger model, yet for the purposes of understanding the most potential attrition risk factors, model 2 is also necessary as it gives more of these factors.

Due to the exceptional predictive ability of 'preferring telephone' to being more willing via telephone than web, a separate model was compiled in order to see what factors are relevant to this variable. Model 3 looked at this and found that sex and using a smartphone to look at social media were significant here. Also, political interest is almost significant, which is a good indicator of social involvement and in turn, survey cooperation (Rothenbühler and Voorpostel, 2016). Sex being relevant indicates that men prefer telephone interviewing more. Using a smartphone to look at social media makes one more likely to prefer telephone interviewing. This is in line with the findings on net willingness; it seems that using a smartphone to consult social media disincentivises someone to participate in surveys via the web – the literature does not offer an explanation for this. An explanation could be that these people associate smartphones, and to a broader extent, internet use, with leisure activities and not long survey administration, and so they prefer to be telephoned for survey completion, as de Bruijne and Wijnant (2014) propose. High political interest is almost significant, and its odds ratio indicates that having a higher political interest makes

one less likely to prefer the telephone, though the lack of significance means we cannot be sure this relationship is positive or negative. All in all, the fact that there are no significant, positive predictors of having a preference for telephone interview bodes well as there is no significant demographic predictor which predicates membership of this group which is at-risk of attrition.

In terms of the smartphone willingness model, which importantly looks at the factors relevant to participating via a smartphone – a growing phenomenon in web surveys, and something the LCS is witnessing (See appendix), there are few key points. Mode preference is expectedly significant, as smartphone willingness is, by definition, willingness for a web survey. The number of smartphone activities is again significant, and it seems this is a good proxy for smartphone competence throughout the models, suggesting that being willing to complete a web survey via a smartphone necessitates a degree of smartphone competence. Significant also is 'emailing on a smartphone' a potential explanation for this is that smartphone web survey participation is often garnered through an email link or the fact that emailing is a proxy for smartphone competence, like the number of smartphone activities variable. Sex is also almost relevant to smartphone willingness, with women being more likely to be willing. Altogether, smartphone willingness factors are broadly similar to overall web willingness, though the model is not as strong.

Estimating attrition

Finally, the identification of those most at risk of attrition by looking at those who are above the 9th decile in the predicted probability distributions of model 1 and 2 gives us a hypothetical idea of who is likely to leave the sample and what predicted attrition could resemble. The results indicate that under 25's, Men, Non-secondos, Singletons, those in

education and those who originate from the initial OFS sample are more numerous in this future-attrition group. These groups are, therefore, most at risk of attrition, and we would expect the distributions of these variables to change the most in the event of a switch to web. However, there are no significant associations between these variables and being in the attrition risk group. Sex comes the closest and is almost significant, suggesting this might be the most affected demographic by a mode switch.

Limitations

The analysis was based on some broad assumptions due to the hypothetical nature of the research questions. By aiming to ascertain future attrition rates and selectivity with the results of the previous wave, certain assumptions were needed to be a proxy for future nonparticipation. In this case unsuitability towards to web mode was seen as an indicator of future nonparticipation. In particular, a fundamental assumption was that preferring another mode over the web mode would mean that one would more at risk of leaving the sample post-switch. This is a fair assumption as the respondent burden is already high and completing it in a non-preferred mode is likely to increase this burden. However, it remains hard to determine the extent to which this state of unsuitability will translate into attrition. Though some literature finds a strong link between stated willingness and actual action (Haan et al. 2019, p.526), hypothetical measures are always subject to context effects (Wenz et al., 2019, p.2). This link could not be tested within this sample; therefore, the caveat persists that predicted risk of attrition may not lead to actual attrition. In the future, post-switch data may be able to confirm or reject these findings and test the link between stated willingness in this wave and actual action post-switch.

Furthermore, the study has already been subject to attrition over the previous waves, rendering the sample size relatively small for such an analysis on hypothetical future participation rates. The statistical power is thus diminished and may explain why only a small number of variables were found to be significant. If stronger evidence is sought, as to the effect of mid-stream mode switches on longitudinal surveys, then using a study with a greater sample size is advisable.

Additionally, the question on mode preference, stated willingness and internet characteristics have only been asked in the sixth wave of the LCS, giving us data on these variables at only one time point. Continued questioning on these topics may give a better idea as to participants' characteristics on these topics, as well as a direction of travel. With more data like this, this dissertation could have discerned deeper trends indicating web mode suitability.

--

The results of the analyses have many elements that are important for the future of the LCS pre- and post-switch. The planned switch of the primary data collection mode to web-only is likely to exacerbate the attrition rate, according to the reviewed literature, but the results show a more convoluted picture. The critical question is whether this future, hypothetical attrition rate will be higher than average and whether this can be attributed to the mode switch.

Altogether the sample are technologically competent and have the necessary access and devices to respond via an internet questionnaire – this does not seem to pose a problem. The main issue concerns the proportion of respondents who prefer telephone interviewing

and have an overall willingness for telephone interviewing. These people are most at risk of leaving the sample by future nonparticipation. These represent a greater proportion than average attrition, so if they all leave the sample, then attrition would be higher than average and predominantly attributable to the mode switch. However, it is difficult to say to what extent these respondents would have naturally left the sample if there was not a mode switch, or how many of those who are at risk of attrition would actually leave the sample. In part due to the fact that stated preference is often distorted towards the mode of competition, i.e. the results for telephone is likely to be inflated. The best guess of attrition is shown in table 7, and although this is based on imperfect models, the fact that there are not any significant associations is a good sign that any attrition caused by mode preference is not likely to be selective.

Conclusion

The growing appeal of web surveys, driven predominantly by societal changes and cost benefits, has meant that, increasingly so, social research surveys are employing this mode. In the case of ongoing longitudinal studies this has a number of implications, not least a disruption to time series estimates, but also this can have repercussions on the sample composition due to attrition. In the rare case of a wholesale, midstream primary mode switch, as is the case for LIVES Cohort study, which plans to switch to web-only from CATI, measurement and selection effects can distort estimates. This dissertation focused on the selection effects and their potential repercussions on the attrition rate.

It found that the LCS sample is largely competent in terms internet access, frequency of use and devices, with everyone having access, the majority using the internet on a daily basis and the majority connecting to the internet through multiple devices. Furthermore, the locations where the internet is connected to poses no issues in terms of context effects. Thus, there is not likely to be any attrition caused by a lack of technical competence.

However, in terms of mode preference, despite a majority preferring an internet questionnaire, there is a sizeable proportion of current respondents who cite a preference towards telephone interviewing. It is this proportion that is at risk of leaving the sample via future nonparticipation. Analyses using stated net willingness of telephone and preferring telephone show that engaging in a greater number of smartphone activities and being female makes one less likely to prefer the telephone, and therefore, men and those who are less competent with a smartphone are more susceptible to future nonparticipation.

Despite this, this is based on the assumption that not preferring the primary mode of data collection would lead a participant to leave the panel, this is not a given, and it is hard to say to what extent this would translate into future nonparticipation. The literature also suggests that mode preference is often distorted by the mode of completion, so in this case, the proportion of those who claim to prefer telephone is inflated because they answered via telephone. Nonetheless, estimation tells us that men, under 25's, non-secondos, singletons and apprentices are more likely to be future non-responders, but there are no significant associations.

At a wider level, the results from the LCS reinforce previous findings that internet penetration is high, smartphones are pervasive and general technological competence is high in Switzerland, albeit the sample are all under 35. General levels of preference show a majority prefer web interviewing, but there is still a sizeable proportion that prefers telephone interviewing for completing surveys. So, web surveys could still face nonresponse issues (i.e. lower responses rates), not because of access or competence issues but the fact that some people still prefer to be questioned via the telephone. Further research could investigate the extent to which preference and stated willingness translate into future participation in this context.

In conclusion, it is likely that there will be post-switch sample attrition in the LCS caused by the mode switch, at potentially higher rates than the average attrition rate of 14% due to the sizeable proportion who state they prefer and are more willing to participate via, telephone interviewing compared with an internet survey. Though there a higher likelihood of men leaving the sample, as they prefer telephone at higher rates, there is no conclusive

evidence that this attrition will be selective in nature and so sample composition will likely remain similar to its current state.

Recommendations

- Continue to ask questions on internet characteristic and mode preferences up until the proposed mode switch.
- Ensure post-switch questionnaires are operational on small-screen devices and browsers (See appendix for current device and browser use for web respondents)
- Consider shortening or chunking questionnaires, especially if smartphone participation increases, as they are associated with shorter leisurely activities.

References

Al Baghal, T., Kelley, J., 2016. The stability of mode preferences: implications for tailoring in longitudinal surveys. Methods, data, analyses: a journal for quantitative methods and survey methodology (mda) 10, 143–166.

Allum, N., Conrad, F., Wenz, A., 2018. Consequences of mid-stream mode-switching in a panel survey. 1 12, 43–58. https://doi.org/10.18148/srm/2018.v12i1.6779

Antoun, C., Katz, J., Argueta, J., Wang, L., 2018. Design heuristics for effective smartphone questionnaires. Social Science Computer Review 36, 557–574.

Bosnjak, M., Metzger, G., Gräf, L., 2010. Understanding the Willingness to Participate in Mobile Surveys: Exploring the Role of Utilitarian, Affective, Hedonic, Social, Self-Expressive, and Trust-Related Factors. Social Science Computer Review 28, 350–370. https://doi.org/10.1177/0894439309353395

Couper, M.P., Miller, P.V., 2008. Web survey methods: Introduction. Public Opinion Quarterly 72, 831–835.

De Bruijne, M., Wijnant, A., 2014. Mobile response in web panels. Social Science Computer Review 32, 728–742.

de Leeuw, E.D., Toepoel, V., 2018. Mixed-Mode and Mixed-Device Surveys, in: The Palgrave Handbook of Survey Research. Springer, pp. 51–61.

Dillman, D.A., Christian, L.M., 2005. Survey mode as a source of instability in responses across surveys. Field methods 17, 30–52.

Dillman, D.A., Smyth, J.D., 2007. Design Effects in the Transition to Web-Based Surveys. American Journal of Preventive Medicine, Critical Issues in eHealth Research 32, S90–S96. https://doi.org/10.1016/j.amepre.2007.03.008

Felderer, B., Kirchner, A., Kreuter, F., 2019. The Effect of Survey Mode on Data Quality: Disentangling Nonresponse and Measurement Error Bias. Journal of Official Statistics 35, 93–115. https://doi.org/10.2478/jos-2019-0005

Fricker, Scott., Galesic, Mirta., Tourangeau, R., Yan, T., 2005. An Experimental Comparison of Web and Telephone Surveys. Public Opinion Quarterly 69, 370–392. https://doi.org/10.1093/poq/nfi027

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R. (Eds.), 2004. Survey methodology, Wiley series in survey methodology. J. Wiley, Hoboken, NJ.

Haan, M., Lugtig, P., Toepoel, V., 2019. Can we predict device use? An investigation into mobile device use in surveys. International Journal of Social Research Methodology 1–15.

Herzing, J. M. E. (2019). Mobile web surveys. FORS Guide No. 01, Version 1.0. Lausanne: Swiss Centre of Expertise in the Social Sciences FORS. Doi:10.24449/FG-2019-00001

Horwitz, R., Lesser, V., Marken, S., Mathiowetz, G. N., Opsomer, J., Steiger, W. D., & Sterrett, W. D. (2010). Report of the AAPOR Task Force on Transitions from Telephone Surveys to Self-Administered and Mixed-Mode Surveys. Retrieved from AAPOR website: https://www.aapor.org/getattachment/Education-Resources/Reports/AAPOR-Mixed-Mode-Task-Force-Executive-Summary-FINAL.pdf.aspx

Jäckle, A., Lynn, P., & Burton, J., 2015. Going Online with A Face-to-Face Household Panel: Effects of a Mixed Mode Design on Item and Unit Non-Response. Survey Research Methods 9, 57–70.

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

Keusch, F., 2015. Why do people participate in Web surveys? Applying survey participation theory to Internet survey data collection. Management review quarterly 65, 183–216.

Lynn, P., Kaminska, O., 2012. The impact of mobile phones on survey measurement error. Public Opinion Quarterly 77, 586–605.

Maslovskaya, O., Durrant, G.B., Smith, P.W., Hanson, T., Villar, A., 2019. What are the Characteristics of Respondents using Different Devices in Mixed-device Online Surveys? Evidence from Six UK Surveys. International Statistical Review.

Nicolaas, G., Calderwood, L., Lynn, P., Roberts, C., 2014. Web Surveys for the General Population: How, why and when?

Peterson, G., Griffin, J., LaFrance, J., Li, J., 2017. Smartphone Participation in Web Surveys, in: Total Survey Error in Practice. pp. 203–233.

Revilla, M., Toninelli, D., Ochoa, C., Loewe, G., 2016. Do online access panels need to adapt surveys for mobile devices? Internet Research 26, 1209–1227.

Roberts, C., Vandenplas, C., 2017. Estimating Components of Mean Squared Error to Evaluate the Benefits of Mixing Data Collection Modes. Journal of Official Statistics 33, 303–334. https://doi.org/10.1515/jos-2017-0016

Roster, C.A., Rogers, R.D., Albaum, G., Klein, D., 2004. A Comparison of Response Characteristics from Web and Telephone Surveys. International Journal of Market Research 46, 359–373. https://doi.org/10.1177/147078530404600301

Rothenbühler, M., Voorpostel, M., 2016. Attrition in the Swiss Household Panel: are vulnerable groups more affected than others?, in: Surveying Human Vulnerabilities across the Life Course. Springer, Cham, pp. 223–244.

Sakshaug, J.W., Hülle, S., Schmucker, A., Liebig, S., 2019. Panel Survey Recruitment with or Without Interviewers? Implications for Nonresponse, Panel Consent, and Total Recruitment Bias. Journal of Survey Statistics and Methodology smz012. https://doi.org/10.1093/jssam/smz012

Schupp, J., Saßenroth, D., 2015. Changing from CAPI to CAWI in an ongoing household panel–experiences from the German Socio-Economic Panel (SOEP). Presented at the 6th Conference of the European Survey Research Association.

Spini, D., Morselli, D., Elcheroth, G., Gauthier, J.-A., Le Goff, J.-M., Dasoki, N., Tillmann, R., Rossignon, F., 2019. The LIVES-FORS cohort survey: A longitudinal diversified sample of young adults who have grown up in Switzerland. Longitudinal and Life Course Studies 10, 399–410.

Toepoel, V., Lugtig, P., 2015. Online surveys are mixed-device surveys. Issues associated with the use of different (mobile) devices in web surveys. methods, data, analyses 9, 8.

Tourangeau, R., Rips, L.J., Rasinski, K., 2000. The psychology of survey response. Cambridge University Press.

Vandenplas, C., Loosveldt, G., Vannieuwenhuyze, J.T.A., 2017. Assessing the Use of Mode Preference as a Covariate for the Estimation of Measurement Effects between Modes. A Sequential Mixed Mode Experiment. methods data, 24 Pages. https://doi.org/10.12758/MDA.2016.011

Watson, N., Wooden, M., 2009. Identifying factors affecting longitudinal survey response. Methodology of longitudinal surveys 1, 157–182.

Wenz, A., Jäckle, A., Couper, M.P., 2019. Willingness to use mobile technologies for data collection in a probability household panel. Survey Research Methods 13, 1–22. https://doi.org/10.18148/srm/2019.v1i1.7298

Zuboff, S., 2019. The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile Books.

Appendix

Figure 5: Modes used in Wave 6 of the LCS.

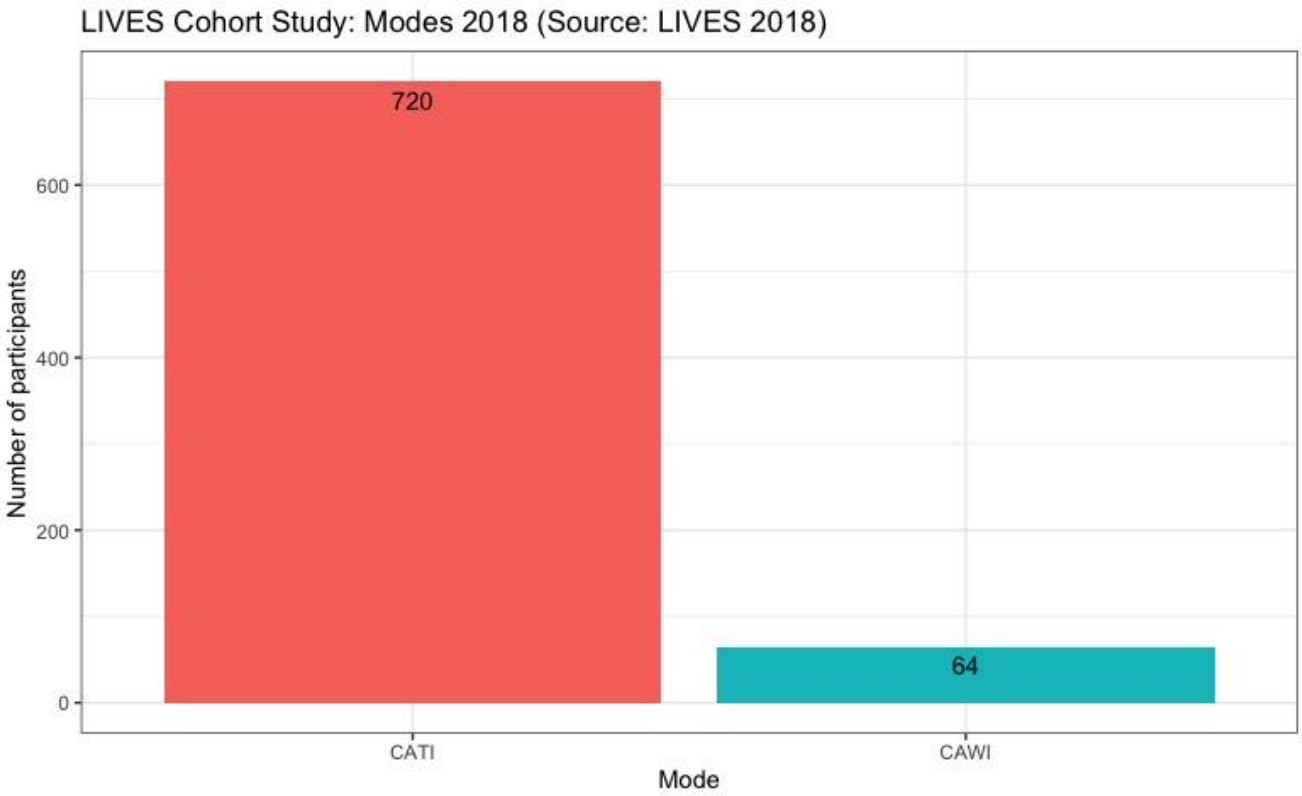


Table 8: Devices used for web responses in wave 6 of the LCS

Device	Frequency
Desktop Computer	11
Laptop	31
Large tablet (23cm or greater)	2
Small tablet (Smaller than 13cm)	0
Large screen smartphone (13cm or greater)	6
Standard screen smartphone (Smaller than 13cm)	12
Unknown computer device	1
Unknown mobile device	1
Total	**64**

Table 9: Browsers used for web responses in wave 6 of the LCS

Browser	Frequency
Firefox	10
Chrome	22
Internet Explorer	1
Safari	21
Edge	4
Samsung Browser	5
Opera	1
Total	**64**

Declaration of Authorship

I hereby certify that the thesis I am submitting is entirely my own original work except where otherwise indicated. Any use of the works of any other author is properly acknowledged. I am aware of the University of Lausanne's regulations concerning plagiarism, including the regulations concerning disciplinary actions that may result from it.

Date: 9th January 2020 Signature: