GenPopWeb2: The utility of probability-based online surveys – literature review

Oriol J. Bosch[1], Olga Maslovskaya[2]

[1]*Department of Methodology, The London School of Economics and Political Science*

[2]*Department of Social Statistics and Demography, School of Economic Social and Political Science, University of Southampton*

26 May 2023

Introduction

In recent years the web has become one of the most common modes of data collection in surveys (Macer and Wilson, 2014; Peterson et al., 2017). Although its hegemony is mostly linked to the popularity of nonprobability online surveys (Callegaro et al., 2014), recent years have also seen an increase in the interest and popularity of probability-based online surveys for the general population (e.g., Blom et al., 2016; Bosnjak et al., 2013; De Vos, 2010; Revilla et al., 2016). The COVID-19 pandemic has further accelerated this trend, given the ability of online surveys to distribute and collect data from self-administered questionnaires in a fast and cost-effective manner (e.g., Mannheim Corona Study (Blom et al., 2020)). The European Social Survey (ESS), which historically used face-to-face mode of data collection, has recently announced a plan to transition from face-to-face interviews to a 'web first self-completion' design[1].

Probability-based online surveys can be defined as surveys which use an online mode to administer the questionnaire to all or most[2] of the sampled participants selected using a probability-based sampling approach. As all other approaches to conducting survey data collection, they come with opportunities and challenges. For instance, while their self-administered nature can help increase the measurement quality of the data collected by reducing the social desirability bias (Kreuter et al., 2008) and cancelling interviewer effects (West and Blom, 2017), response rates have been found to be lower than for other modes of data collection (Daikeler et al., 2020; Manfreda et al., 2008; Shih and Fan, 2008). In addition, probability-based online surveys are not homogeneous, varying in key design characteristics such as in the strategies used to recruit participants or in ways how they deal with sampled units without access to the internet, which can also have an impact on data quality. As a result, researchers and survey practitioners looking to transition to online probability-based surveys need to understand how these factors can affect the quality of their data, as well as what best design options are available for them.

To aid researchers in this transition, this paper reviews the literature on probability-based online surveys to 1) provide an up-to-date compilation of their general utility and 2) compare the various dimensions in which they can vary from a data quality perspective. Our aim is to provide a guide that enables researchers to better understand the specific errors associated with

[1] https://www.europeansocialsurvey.org/about/singlenew.html?a=/about/news/essnews0130.html

[2] Surveys which use other modes of data collection together with the online mode are mixed-mode designs. In this review we consider that surveys that use the online mode as the main mode of data collection can still be considered as mainly online.

different design choices. To achieve this, we use the Total Error Survey (TSE) framework (Groves et al., 2009) to present the differences between probability-based online surveys and other modes of data collection, as well as between different approaches used for online surveys, for each error source. The TSE framework allows to identify and estimate potential errors and the effects of those on estimates and how to minimise them (Biemer, 2010).

The remainder of this review is organised as follows. First, we present the methodological procedure followed and provide a more detailed definition of the dimensions in which probability-based online surveys can vary. Second, we present the results of the literature review for each error source investigated. Finally, we discuss the practical implications of this review and recommend ways in which researchers can use it to improve the way they design an implement probability-based online survey.

Methodological procedure

To conduct the literature and structure the results accordingly, we looked for evidence of the data quality of probability-based online surveys. This term encompasses a range of factors, including accuracy, credibility, comparability, usability, relevance, accessibility, timeliness, completeness, and coherence. To guide our analysis, we adopted the TSE framework by Groves et al. (2009), which conceptualises data quality in terms of accuracy. This framework identifies two groups of errors: errors of representation, and errors of measurement. Errors of representation occur when eligible members of the population of interest are not measured, leading to selection bias. These include coverage, sampling, nonresponse, and adjustment errors. In contrast, errors of measurement occur when the concept of interest that researchers want to measure differs from the processed measure collected. These include specification (or validity), measurement, and processing errors.

From this list of errors, we conducted an in-depth search for publications mentioning these error sources for probability-based online surveys in academic journals and book chapters. We searched four major electronic reference databases: Scholar, WebSM, JSTOR and Web of Science. Besides, we used a snowballing technique to inspect the reference lists of the selected publications. To avoid the "publication bias" problem, we included unpublished and non-peer-reviewed research (e.g., reports, presentations) obtained through contacts of the GenPopWeb2 network[3]. To select publications for our final review, we only selected studies which used probability-based sampling strategies. The focus was on general population

[3] https://www.ncrm.ac.uk/research/genpopweb2/

surveys, however, in some cases we used results from studies of specific populations (see Appendix A for a summery on literature used for this review). In addition, we only included online surveys, and excluded mixed-mode surveys even if one mode of data collection was online. The exception from this rule was the inclusion of online surveys with a mixed-mode component when offline options were provided: 1) to offliners (e.g., American Trends Panel), 2) to offliners and as an option for nonrespondents in mainly online panels (e.g., NatCen panel) and 3) for offline panels transitioning part of their panellist to the online mode (e.g., Understanding Society's Innovation Panel).

Furthermore, the goal of this literature review is also to compare different aspects of the data quality of probability-based online surveys. We consider that four main dimensions are relevant when discussing the data quality of probability-based online surveys, not denying that other dimensions might still be relevant for specific error sources (see Figure 1 for a visual representation): 1) the one-time or panel nature of the survey, 2) recruitment strategy, 3) treatment of offliners and 4) questionnaire optimisation for mobile devices or "mobile-first" survey design approach. Section 2.1 further discusses these characteristics. While reviewing the literature, we specifically searched for empirical evidence and/or theoretical arguments that assess if differences on data quality were found across dimensions or not. Finally, in relation to the empirical evidence, it is often difficult to extract general conclusions since studies differ on the quality indicators used. For instance, a wide range of measurement quality indicators are considered in the literature, such as different response style behaviours (e.g., selection of extreme and middle options in attitudinal questions), satisficing bias or different measures of reliability and validity. We considered the different types of indicators available for each error source.

After conducting an exhaustive search of both published and unpublished research, including sources obtained through contacts of the GenPopWeb2 network, a total of 255 publications were identified and reviewed for potential inclusion in our study. Of these, 176 were found through database searching, while 79 were obtained through other sources. We utilised 92 of these sources, and a list containing summary information for each source can be obtained on request from the authors.

Distinguishing dimensions of survey design

In this section, we discuss the four dimensions related to the design of probability-based online surveys that we consider relevant: survey type, recruitment strategy, treatment of offliners, and mobile device optimisation.

The first dimension is the distinction between **one-time cross-sectional surveys and panels**. *One-time cross-sectional surveys* are unique surveys conducted at a specific point in time (e.g., UK Active Lives Survey[4] or the Fundamental Rights Survey[5]), while *panels* involve a sample of individuals who agree to periodically complete surveys via the internet. Panels can be used for longitudinal surveys to measure change over time or for conducting repeated cross-sectional surveys across time (e.g., Kantar Public Voice).

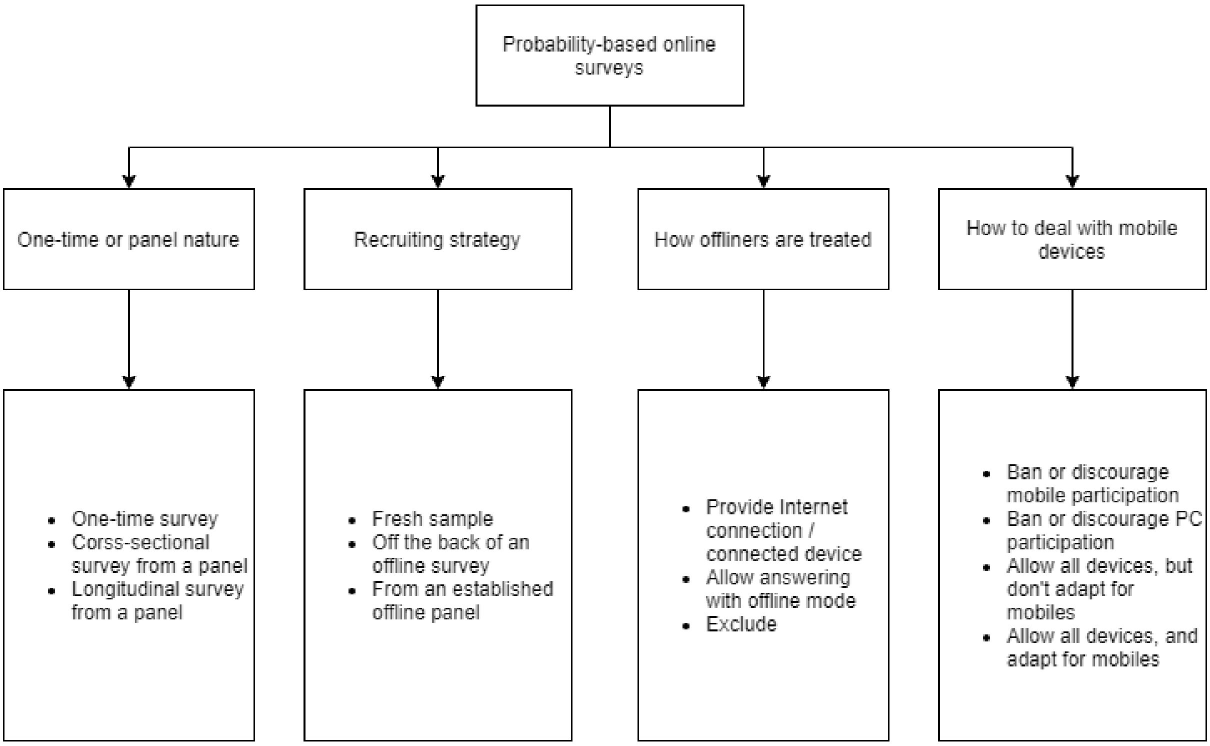

Figure 1. The four dimensions of probability-based online surveys

The second dimension concerns the **recruiting strategy** used to build the sample. Some sampling frames for the general population do not provide enough information to contact units directly online. Therefore, probability-based online surveys often rely on offline sampling frames. In the UK, Address-Based Sampling (ABS) is commonly used, with the Postcode Address File (PAF) or AddressBase as the sampling frame. However, to participate in an online

[4] https://www.sportengland.org/know-your-audience/data/active-lives?section=methodology#adultsurvey
[5] https://fra.europa.eu/en/publication/2020/fundamental-rights-survey-trust#TabPubMethodologyQA2

survey using offline frames, sampled units must have access to an URL or QR code to be able to access the survey questionnaire. To facilitate this, push-to-web strategies are used. These strategies use an offline contact to request responses over the internet without offering the option of answering with offline modes until later stages (Dillman, 2017) or sometimes never. There are different push-to-web strategies, applied to different types of surveys, with varying implications for data quality. These strategies include (a) *contacting a fresh sample using an offline mode*, where a fresh sample is drawn from an offline frame like PAF, and sampled units are contacted using offline modes (mail) (e.g., Kantar Public Voice; Community Life Survey); (b) *recruiting on the back of an existing face-to-face or telephone survey*, where participants are asked to participate in an online survey or join an online panel at the end of an offline survey, and those who give their consent and e-mail addresses are sent a survey or several surveys on a regular basis (e.g., CROss-National Online Survey (CRONOS) panel using the ESS Round 8, NatCen Panel recruiting from the British Social Attitudes (BSA) survey); and (c) *using already established probability-based panels to push participants to answer an online questionnaire*, build a panel, or transition part or all of the offline panel to the online mode of data collection (e.g., Understanding Society's COVID-19 study (Burton et al., 2020)).

The third dimension is related to the issue of **how offliners are treated**. Offliners are those who do not have access to the internet[6] and, hence, cannot answer online surveys if no alternative option is provided. An option is to *directly exclude those from the study*, for instance, because no extra resources are available or it is believed that the impact of excluding them is expected to be low (e.g., Norwegian Citizen Panel (NCP)). Alternatively, *internet connection and/or a connected device can be provided*, only to offliners (e.g., German Internet Panel (GIP) or electronic questionnaire device (EQD) in ESS) or to all participants, e.g., mobile devices connected using 3G/4G connection (ELIPSS), to ensure that all sampled individuals can participate regardless of their online status. Finally, another option is *to allow offliners to participate using an offline mode* (e.g., postal (GESIS panel) or telephone (NatCen Panel)). When offline respondents are offered to participate in a different mode, the survey stops being exclusively online. However, the high costs associated with providing offliners with connected

[6] The definition of who *has access* to the internet is not that clear. In the current context, individuals have access to the internet through different devices (laptops, mobile devices), with different types of connections (personal Wi-Fi, 4G, public Wi-Fi) and in different contexts (work, shared, personal) (Revilla et al., 2016). Besides, individuals might have access to the internet at home, or might be provided internet, but that does not mean they have the *ability to* respond online.

devices have made the approach of offering a different mode a considerably more popular alternative.

The fourth and final dimension is **the presence and optimisation of mobile devices**. Online surveys are multidevice by nature, given the various types of devices individuals can use to access the internet. However, this poses challenges to the quality and comparability of collected data, specifically related to some characteristics of mobile devices such as smaller screens and touchscreen keyboards. Researchers have taken different approaches over time to address these challenges, such as *banning or discouraging the use of mobile devices* or *allowing their use* while either optimising the questionnaire design for mobile devices or not. Although not common for probability-based surveys, it is common for some non-probability surveys to be mobile-only (i.e., allowing survey completion only using mobile devices but not desktops/laptops), especially when innovative data collection strategies are implemented (see Bosch et al. (2019); Revilla et al. (2020) for details).

Literature Review

In the following subsections, we present our results for each error component, with Appendix A providing descriptive information on the literature we used for the review. For error components in which the size and/or causes of error varied across dimensions, we further separated our results into different subsections, each addressing a specific varying dimension. In contexts where the errors were the same across all dimensions or insufficient evidence was available, we present a general discussion of the errors specific to probability-based online surveys and draw comparisons to offline alternatives.

Coverage Error

Coverage error in probability-based online surveys occurs when a sampling frame used to select a sample differs from the target population. In general, probability-based online surveys for the general population rely on offline frames to sample and contact individuals. In the UK, to achieve a good population coverage the PAF or AddressBase sampling frames are commonly used. The main limitation of the frames is that they are not individual-level sampling frames and they do not usually have names, characteristic of individuals, phone numbers or email addresses. They allow households to be contacted by mail, and then the respondents within households can be directed to an online questionnaire. For high coverage frames like the PAF or AddressBase, coverage errors are typically minimal. However, evidence suggests that address-based frames have higher risk of introducing coverage bias for certain variables compared to individual-level sampling frames or ABS augmented with personal information (Amaya et al., 2018; Kölln et al., 2019).

Specific coverage issue for some general population probability-based online-only surveys is the under-coverage of individuals with no internet access, as they have a zero chance of participating in the survey. This issue is specific to online surveys that do not take steps to include offliners, such as providing individuals with internet access or offering alternative offline answering modes. Furthermore, additional coverage errors can also be introduced depending on how mobile devices are handled by a specific survey, for example, in the past some surveys would not allow respondents to complete questionnaires using mobile devices.

Treatment of offliners

Coverage errors can be introduced if offliners are excluded from participation by design, and their characteristics are systematically different from those who can complete surveys online. Previous research has explored potential coverage errors by comparing the characteristics of survey respondents with and without access to the internet (Couper et al., 2018). Offliners have been found to be older (Blom et al., 2015; Bosnjak et al., 2013; Jessop et al., 2016; Leenheer and Scherpenzeel, 2013; Revilla et al., 2016), more likely to live in smaller households (Blom et al., 2015; Leenheer and Scherpenzeel, 2013), and in less urban areas (Blom et al., 2015; Leenheer and Scherpenzeel, 2013; Revilla et al., 2016). Additionally, they are more likely to be non-Western immigrants (Leenheer and Scherpenzeel, 2013), retirees, single, widowers, and divorcees (Revilla et al., 2016), and not be in paid work or education (Jessop et al., 2016). Women are also more likely to be offliners (Blom et al., 2015; Bosnjak et al., 2013; Jessop et al., 2016), although some studies disagreed with this finding (Revilla et al., 2016).

Offliners have also been found to differ from onliners in personality traits, political views, and interests. Specifically, offliners tend to be less open and extraverted, more conscientious (Bosnjak et al., 2013), less politically interested (Blom et al., 2015), and more likely to have voted for Brexit, prioritise government spending on areas other than education, and vote for conservative parties (Jessop et al., 2016).

Despite these differences, Eckman (2016) reported that the under-coverage of offline households in the LISS panel did not introduce bias into different multivariate estimates. Similarly, Bach et al. (2023) demonstrated that, despite sociodemographic differences between offliners and onliners in the GIP, conclusions drawn from univariate and multivariate substantive analyses are only marginally affected by excluding offliners, if at all.

Mobile devices

As the number of people who exclusively access the internet through mobile devices continues to grow, excluding these "mobile-only" users from online surveys may introduce coverage errors. In the US, for instance, 10% of adults had only a smartphone data plan back in 2015 (Smith, 2015), while in Europe, 7% of households had mobile-only internet access around the same time (European Comission, 2014). If online surveys prohibit mobile participation, the risk of coverage errors emerges if mobile-only users are systematically different from those who participate via desktops/laptops.

While no direct studies have assessed the differences between mobile-only users and the general population, researchers have found some distinguishing characteristics of mobile respondents. For example, mobile device respondents tend to be younger (Cook, 2014; de Bruijne and Wijnant, 2014; Gummer et al., 2019; Lambert and Miller, 2015; Lugtig et al., 2016; Wells et al., 2014), female (Cook, 2014; de Bruijne and Oudejans, 2015; de Bruijne and Wijnant, 2014; Gummer et al., 2019; Lambert and Miller, 2015; Wells et al., 2014), and from higher income groups (Toepoel and Lugtig, 2014a), while living in larger households (Wells et al., 2014), and being more progressive (de Bruijne and Wijnant, 2014). Conversely, mobile-only respondents are less likely to have higher education degrees and to be married, Hispanic or African American (Lugtig et al., 2016).

In an opposite scenario, restricting survey participation to only those who use mobile devices and excluding "Personal Computer (PC) only" individuals could also introduce coverage errors. Though limited research exists on the impact of such an approach, Antoun et al. (2019) found average absolute coverage deviation of 1.5 percentage points when surveys only allowed mobile participation, indicating that such a strategy is likely to introduce (small) coverage errors.

Sampling errors

Sampling error is defined as the errors that arise because of inclusion of a subset of the population of interest rather than the entire population. The non-observation of all population members introduces deviation from the achieved sample statistics and the same statistic obtained from the target population (true population value). Several factors can contribute to sampling errors, for instance, the size of the sample, the sample design (e.g., stratification/clustering), and unequal selection probabilities.

When an online probability-based survey selects a sample from offline frames, in the same way as it would be done for offline surveys, this should produce similar small sampling errors. However, as mentioned earlier, in some countries like the UK, samples are drawn using ABS sampling frames (e.g., PAF), which do not include names of individuals. When using face-to-face recruitment, interviewers can use well-defined methods to randomly select individuals within the same address (e.g., Kish grid). When invitations are sent by mail, the selection of individuals within addresses can only be done by the address residents, which must follow quasi-random protocols. In this case, a recipient of the invitation letter is instructed on how to select one or various household members to act as respondent(s) (Nicolaas et al., 2014). Household could be asked to select one adult in the household with the next or most recent birthday (e.g., Battaglia et al., 2008; Olson and Smyth, 2017). In a large proportion of cases errors are made when using these methods (Park and Humphrey, 2014; Villar, 2013; Williams, 2015), which can lead to sampling errors and selection biases. An alternative to this is sampling all members of the address, using incentives to achieve full completion. However, this can increase fraud and fake interviews (Murphy et al., 2016; Williams, 2015). The summary of different methods of selection of individuals within households can be found in Nicolaas' report (2022).

The differences between approaches will mostly be determined by the approach used to invite participants to join the online survey/panel. For instance, we would expect probability-based online panels recruiting their panellists for fresh samples with mail invitations to present a higher risk of introducing sampling errors than those recruiting on the back of telephone or face-to-face surveys, since the latter ones do not need to rely on quasi-random protocols. However, to the best of our knowledge, specific empirical research comparing the size of sampling errors between different approaches has not been yet conducted.

Nonresponse errors

Nonresponse errors occur when data cannot be collected for all sampled members. Nonresponse can occur at the unit level, where no information is available for any measure for a given unit, or at the item level, where information is not available for a specific variable(s) for a given unit. When nonresponse occurs, estimates are drawn based only on the subset of the sample with available answers, potentially introducing biases if this subset is systematically different from the target population.

The magnitude of nonresponse bias can vary between probability-based online surveys and offline alternatives, such as face-to-face or telephone surveys, for various reasons. For instance, the self-administered nature of online surveys may affect respondents' willingness and ability to complete the survey. Shih et al. (2008), Manfreda et al. (2008) and Daikeler et al. (2020) report that online surveys suffer from lower response rate than other modes of data collection. These surveys can also have higher nonresponse bias than offline and interview-administered modes (e.g., Felderer et al., 2019). Goodman et al. (2022) demonstrate that online mode of data is associated with the increased response rate when compared to telephone-only approach and improved item nonresponse for sensitive questions such as number of units of alcohol consumed but with increased item nonresponse on questions about pay and other financial matters.

One-time survey or panel and recruitment strategy

Nonresponse can be introduced at different stages of the survey process. Hence, the number and types of stages between the moment when a unit is sampled and asked to participate in a specific survey and the end of data collection might be associated with the proportion of people not responding, and the associated bias. The one-time or panel nature of the survey, as well as the recruitment strategy followed determine these stages. Therefore, we discuss these two dimensions together. The basic nonresponse process is the one observed for *one-time survey* recruited using a *fresh sample*. In this case, nonresponse can be introduced as: 1) units cannot be contacted, and if contacted, 2) units refuse to cooperate when contacted or 3) are not able to do so (e.g., due to language barriers) (Groves et al., 2009). Any change in the nature or recruitment strategy of the panel will introduce new steps and, hence, new opportunities for the participants not to provide data.

Panel surveys. In a panel context, nonresponse will be a combination of initial participants not joining the panel, attrition in the panel, and wave-specific nonresponse (Lynn and Lugtig, 2017). Both the initial panel nonresponse and the wave-specific nonresponse can be further subdivided into separate events, such as noncontact, refusal to participate in the recruitment interview, refusal to become a panel member or physical or intellectual inability to participate (Lugtig et al., 2014). It is to be expected, therefore, that panels will present higher nonresponse errors than one-time surveys, especially in later waves. Past research has indeed investigated the potential biases that each of these stages might introduce to surveys. First, individuals joining probability-based online panels have been found to be younger (Blom et al., 2015; Hoogendoorn and Daalmans, 2009; Lugtig et al., 2014), more educated (Bosnjak et al., 2018; DiSogra et al., 2009), with higher incomes (Bosnjak et al., 2018; Hoogendoorn and Daalmans, 2009), and more likely to live in smaller households (Blom et al., 2015) and be homeowners (Lugtig et al., 2014) than those deciding not to join. In addition, they have been found to be less likely to come from single-individual households, extremely urban areas

(Lugtig et al., 2014) and cell-phone-only households (DiSogra et al., 2009). Second, panellists abandoning the panel (attrition) have also been found to differ from those remaining in the panel. Specifically, individuals with children (Kruse et al., 2009), less politically interested (Frankel and Hillygus, 2014), non-White (Frankel and Hillygus, 2014; Kruse et al., 2009), younger (Frankel and Hillygus, 2014; Lugtig, 2014; Skjervheim et al., 2020) and less educated (Frankel and Hillygus, 2014; Lugtig, 2014) are more likely to discontinue participation in the panel.

Recruiting on the back of an existing face-to-face or telephone survey. When obtaining a sample on the back of another survey, the survey nonresponse of that survey is carried to the new survey or panel. For instance, for probability-based online panels recruited on the back of another survey, unit nonresponse is the cumulative result of previous survey nonresponse, initial nonresponse (not joining the panel), attrition, and the wave-specific nonresponse. Previous research has found that most of the selection bias from probability-based online panels recruited on the back of a survey come from the initial offline nonresponse (Jessop et al., 2016; Pew Research Center, 2015). More specifically, it has been reported that those participating to the initial survey of both the American Trends Panel (recruited in 2014 by telephone on the back of Political Polarization and Typology Survey (PPTS)) and the NatCen panel (recruited via face-to-face interviews on the back of British Social Attitudes survey (BSA)) were more likely to be older and more educated (Jessop et al., 2016; Pew Research Center, 2015). Specifically, while those participating in the PPTS were also found to be more likely to be White, married and politically engaged (Pew Research Center, 2015), in the BSA they were women, and people in managerial and professional occupations, living in a single-person household, and owning their own home (Jessop et al., 2016).

Furthermore, past research has found that those accepting to join a panel after answering an offline survey are more likely to be more educated (Bosnjak et al., 2013; Jessop et al., 2016; Pew Research Center, 2015), younger (Bosnjak et al., 2013; Jessop et al., 2016; Maslovskaya and Lugtig, 2022), active in politics (Pew Research Center, 2015), and less conscientious (Bosnjak et al., 2013). In some cases, these differences moderated some of the biases of the base offline surveys (e.g., age), whereas other biases were introduced or exacerbated (e.g., education, political affiliation or income) (Jessop et al., 2016).

In terms of attrition, Bosnjak et al. (2013) found that individuals with children were more likely to discontinue participation in the panel, while older panellists were less likely to do so. For the CRONOS panel, Maslovskaya and Lugtig (2022) demonstrated that older respondents and those with lower level for education were progressively getting more underrepresented in the panel in Estonia and Slovenia but not in the UK, where young people (18-29) became one of the most underrepresented groups at the end of the panel data collection. Interestingly, Jessop et al. (2016) reported that attrition contributed little to the overall selection bias of the NatCen Panel.

Using an already established offline panel to push participants to answer an online survey. When using already established panels, previously accumulated biases can be brought forward into the new survey. Additionally, asking panellists that were already used to answer in an offline mode to do it online might increase unwillingness to participate.

In an experiment exploring the effect of transitioning from a face-to-face single-mode design to a mixed-mode design with the online mode used as the primary mode in the Understanding Society's Innovation Panel, Jäckle et al. (2015) found that several demographic groups were less likely to participate when offered to respond in mixed-mode design when compared to face-to-face single-mode: men, White, living in rural location, web users, those for whom email address was available, age 21-30, single respondents with children, couples with children and 2+ unrelated adults with children, individuals who reported that they would not do a survey by web. These differences in the likelihood to participate and in sample composition reduced in subsequent waves. After two more waves the likelihood to participate in mixed-mode design was identical compared to the face-to-face single-mode design with only minimal differences in sample composition (Bianchi et al., 2017).

Treatment of offliners

Differences in unit nonresponse can arise depending on how offliners are dealt with within surveys. As discussed earlier, surveys can provide offliners with an internet connection and/or a connected device or allow them to answer with an offline mode.

Research has explored the effectiveness of providing offliners with an internet connection and/or a device. Leenheer and Scherpenzeel (2013) found that in the LISS panel, recruited panellists who accepted the internet connection and/or device constituted a representative random sample of offline households in terms of ethnical background, region, and household composition. However, the sample was not completely representative in terms of age, with older age groups without internet being underrepresented in the panel. For the GIP panel, the inclusion of offliners made it more representative in terms of age, education, and household size (Bloom et al., 2015).

Wave nonresponse and attrition can undermine the efforts of providing offliners with a connection and/or a device if these have a higher likelihood of not answering and/or attriting.

However, Cornesse and Schaurer (2021) found no differences in survey completion across waves for offliners and onliners in the GIP. Moreover, the inclusion of offliners reduced selection bias for all waves. For the LISS panel, offline households were found to be more loyal and less likely to leave the panel (Leenheer and Scherpenzeel, 2013). According to Lugtig (2014), giving panellists a device with a connection was a strong predictor of not dropping out. Contrary to these results, Frankel and Hillygus (2014) reported the higher likelihood of abandoning the American National Election Study (ANES) for those panellists who were provided with a connected device. Regardless of the likelihood to attrite, the absolute bias of the LISS panel decreased over time, suggesting that selectivity in voting behaviour for attrition classes corrected for some of the bias introduced during the panel recruitment process (Lugtig, 2014).

For those surveys offering offliners an offline option, similar findings were reported. Bosnjak et al. (2018) found that allowing offliners and those not willing to answer online to use an offline mode made the GESIS panel more representative in terms of education than if only online units were allowed to become panellists. Similarly, Benzeval et al. (2021) demonstrated that allowing non-regular internet users to answer through the phone reduced the bias for the Understanding Society's Covid-19 study. However, Jessop et al. (2016) reported little effect of adding offliners, mostly balancing the educational profile of the sample but exacerbating existing issues with the age profile of participants. In terms of wave nonresponse and attrition of offliners, Jessop (2017) demonstrated that offliners participate in regular panel waves at a lower rate than online members. Although Cornesse and Schaurer (2021) also found that offliners experience a significantly lower survey completion rate than online participants across waves, their inclusion still reduced selection bias at later waves.

Mobile devices

The presence and optimisation of questionnaire for mobile devices can impact nonresponse bias. At the unit level, unit nonresponse can be introduced if mobile devices are not allowed and some participants are highly unwilling to participate using PCs. If those unwilling to switch to a PC significantly differ from those willing to switch (Peterson et al., 2017), results could be biased. In terms of item nonresponse, literature demonstrates no differences between mobile and PC respondents with respect to the distribution of item nonresponses (Lee et al., 2019; Toepoel and Lugtig, 2014a; Wells et al., 2013).

Optimising survey questionnaires to mobile devices reduces the nonresponse rates compared to those not optimised (Couper et al., 2017, Horowitz, 2016; McClain and Crawford, 2013; Sarraf et al., 2015), but not enough to entirely cancel the differences when compared to PC respondents (Antoun, 2015; Toepoel and Lugtig, 2014b; Wells et al., 2013). Additionally, optimisation of questionnaire generally reduces item nonresponse (McGeeney and Marlar, 2013; Sarraf et al., 2015), although some research has found these differences to be nonsignificant (McClain and Crawford, 2013).

All in all, the differential nonresponse between PC and mobile surveys has been associated with an average absolute deviation of 1.0 percentage points (Antoun et al., 2019), which in most cases can be considered as close to negligible.

The “mobile-first” approach to survey design, when the questionnaire is designed with the small screens in mind, is the current best practice and the situations when some devices are not allowed for survey completion now represent the past trends and, therefore, there is no risk that these designs have negative impacts on the current surveys.

Adjustment errors

When modelling and producing survey estimates, researchers can employ weighting and/or imputation strategies to deal with missing data. However, deficiencies in nonresponse error and coverage error weighting adjustments and imputation approaches for item missingngness can introduce adjustment errors. To apply weighting and imputation adjustments, auxiliary data are needed. Auxiliary data can be defined as a set of variables measured in the survey and for which information on the distributions of these variables in the population, the frame, or the nonrespondents is available (Scherpenzeel and Bethlehem, 2011). Examples of auxiliary data can be sampling frame data (e.g., geographic region), various paradata such as interviewer observations (e.g., type of housing unit) and linked micro-geographic data (e.g., purchasing power in the area) (Cornesse, 2020). The amount of meaningful and high-quality auxiliary data determines how well adjustment techniques can reduce selection and nonresponse biases (Kreuter et al., 2010; Little and Vartivarian, 2005; Olson, 2013). The availability and richness of this information depends on different aspects, such as the country (i.e., in some countries population registers used as sampling frames, they contain rich individual-level information but in others, no such information is available), the mode of data collection (i.e., in face-to-face interviewers can collect observations from the dwelling, the area and in address-based frames units addresses can be linked to micro-geographic data) or even the budget (i.e., micro-geographic data are expensive). The lack of interviewer available for recruitment and/or the interviewing stage make probability-based online surveys different from offline options where interviewer is present. In terms of recruitment, for probability-based online panels different modes have been used to recruit sampled units such as face-to-face, telephone or mail, which allow to collect different types of auxiliary data. Those surveys which recruit on the back of another survey might be able to obtain more auxiliary information available from the original surveys which would help to adjust for, for instance, nonresponse.

When adjustment errors are considered, the main variations within probability-based online surveys can be observed between one-time surveys or panels and between recruitment strategies.

One-time survey or panel

For probability-based online panels, populations of interest can be dynamic. Eligibility status can change over time in ways in which are not always clear to researchers (Lynn and Kaminska, 2012), and which can make computing selection probabilities hard (Lavallee, 2007). Besides, nonresponse patterns can be more complex than for cross-sectional surveys. If attrition is not monotone (i.e., nonrespondents can participate in subsequent waves), weights change for every wave, with the complexity escalating quickly. Besides, keeping track of the eligibility status is a complex endeavour which can introduce errors. Non-identified transitions to ineligibility (e.g., died but research do not know it) can make weighing adjustments to overrepresent sample members who share covariate characteristics with those unidentified ineligibles (Lynn and Lugtig, 2017).

Recruitment strategies

Obtaining a sample for a one-time survey or a panel from an offline existing panel can have some advantages, compared to using a fresh sample. Using already existing surveys or panels means that some information about nonrespondents might be available (Benzeval et al., 2021). These set of nonresponse predictors vary depending on the survey or panel used but will most likely go beyond basic sociodemographic data. Previous research has shown that inverse probability/ response-propensity weights based on information from previous surveys or waves are highly effective in reducing the overall bias (Benzeval et al., 2021; Jessop et al., 2016), and work substantially better in comparison to when only bias adjustment based on calibration weights from demographic characteristics alone is used (Benzeval et al., 2021; Couper et al., 2007; Schonlau et al., 2009).

Specification errors

A specification error (also known as (in)validity) arises when the concept being measured differs from the concept of interest (Biemer, 2010). Hence, errors are introduced when researchers first define the concept that they want to study and then design the survey question(s) to measure this concept. Specification errors are not expected to vary between online and other modes of data collection.

Measurement errors

Measurement errors are produced when the value obtained from a sampled unit deviates from the true value that the measurement should have if no errors happened when collecting the data. Measurement errors are results of, for example, human memory limitations, interviewers' influence, deliberate falsification, or comprehension errors. In terms of measurement errors, there are some general differences between online surveys and offline surveys. First, an interviewer is not present in online surveys as it is a self-administered mode of data collection. Second, in online surveys questions are presented visually, with participants having to read questions on their screens, compared to face-to-face or telephone interviews, in which questions are read to respondents by interviewers. Finally, devices used to answer online surveys are heterogeneous, not only between broad groups (PC, tablet and smartphone), but also within groups (e.g., screen size). These characteristics can affect how questions are presented on screen. For instance, online surveys often optimise questionnaires for smaller screens by changing horizontal scales to vertical ones, and by transforming grids to item-by-item questions with a paging design (Revilla and Couper, 2017). These differences have consequences on measurement errors between online and other modes of data collection, especially interviewer-administered ones.

Previous research using probability-based surveys has found no significant difference in terms of nondifferentiation between face-to-face and online modes (Cernat and Revilla, 2020) or acquiescence bias (Cernat and Sakshaug, 2020; Heerwegh and Loosveldt, 2008), as well as no difference in terms of primacy effects between online and telephone of those switching modes in a panel study (Kocar and Biddle, 2020). For other indicators, results are mixed. Some studies have reported that online surveys suffer from a lower social desirability bias (Cernat et al., 2016; Lee et al., 2019; Tourangeau et al., 2013, Goodman et al., 2022), whereas others have found online mode to present a slightly higher bias (Cernat and Sakshaug, 2020). Furthermore, while Cernat et al. (2016) found that responses collected online had higher recency effects when compared to face-to-face mode of data collection, Cernat and Revilla (2020) showed that primacy effects were larger in the CRONOS panel when compared to ESS round 8 data.

Considering that different errors can offset each other, Revilla and Saris (2013) compared the measurement quality (defined as the product of the squared reliability and validity coefficients) of the same questions asked in the ESS and the LISS panel. They reported similar average quality between both modes. Similarly, Revilla (2013) found that there is no effect on measurement quality of answering in online or a face-to-face mode. Additionally, Felderer et al. (2019) showed that, while online answers to socially desirable questions (e.g., whether the participant received benefits) are biased when compared with administrative data, this bias is lower than the one observed using a telephone mode of data collection.

Measurement errors can vary across probability-based online surveys. These differences can be associated with 1) the longitudinal or cross-sectional nature of the panel; 2) how offliners are treated and 3) the proportions of participants answering using mobile devices.

One-time survey or panel

Compared with one-time surveys, participants' answering surveys in a panel might affect the quality of their answers, both in positive and negative ways. On one hand, panel conditioning can decrease the quality of answers to knowledge questions (Das et al., 2011; Toepoel et al., 2009), and increase the likelihood of participants' engaging in straightlining (Schonla and Toepoel, 2015). On the other hand, Struminskaya (2016) found that more experienced respondents provide more "don't know" answers in knowledge questions, which demonstrate both panel experience and honesty. Halpern-Manners et al. (2014), similarly, found that participants with previous experience answering about sensitive behaviours such as drink driving are more likely to report having engaged on those behaviours in subsequent waves.

Treatment of offliners

Probability-based online surveys offering offliners to respond using an offline mode can be impacted by mode effects, i.e., the situation when respondents' answers to survey questions differ depending on the mode of data collection rather than on real difference in responses. As mentioned earlier, online and interviewer-administered modes of data collection can present different measurement errors. Combining multiple modes that influence participants differently can compromise the accuracy of comparisons between respondents interviewed in different modes, and can even bias multivariate analyses if the selection into a specific mode is linked to other variables of interest.

Klausch et al. (2013) found significant differences in the measurement properties (threshold bias, systematic bias, and random error) of self-administered (online and postal) and interview-administered (face-to-face and telephone) modes. These differences were identified across different topics, formats, and position of questions in the questionnaire, suggesting that mode effects might not be mitigable by optimising the design of the questionnaire. However, these differences were not observed between online and postal (self-administered) modes and face-to-face and telephone (interviewer-administered) modes. Therefore, if more than one mode is to be used, mode effects can be mostly prevented by not combining interview- and self-administered modes. Additionally, Hox et al. (2015), compared interviews administered using online or face-to-face mode of data collection in the same mixed-mode survey. They found (partial) measurement equivalence in most of the studied measures, i.e., the same construct is being measured across methods.

Mode effects can be especially problematic for longitudinal studies if participants can switch modes across waves. When mode effects are present, measures of change might reflect changes in the size of the measurement errors and not in the concept of interest.

Mobile devices

Mobile devices might present different measurement properties, potentially impacting the measurement quality of data collected in online surveys. Compared with participants responding to surveys using PCs, those responding on mobile devices have been found to provide less accurate answers (Antoun, 2015), shorter answers to open-ended questions (Struminskaya et al., 2015) and to be more susceptible to primacy effects (Lugtig and Toepoel, 2016a), acquiescence bias (Clement et al., 2020) and straightlining (Maslovskaya et al., 2020; Struminskaya et al., 2015). The evidence, however, is mixed. Some studies have found no evidence of differences between mobile devices and PCs for primacy effects (Erens et al., 2019; Toepoel and Lugtig, 2014b; Wells et al., 2014), straightlining (Erens et al., 2019), social desirability bias (Antoun, 2015), and satisficing behaviours (Antoun et al., 2017; Clement et al., 2020). Focusing only on participants switching devices across waves, Lugtig and Toepoel (2016b) and Struminskaya et al. (2015) found no relevant behavioural change in the way of answering the questionnaires, indicating that any supposedly negative behaviour associated with mobile devices should probably be attributed almost entirely to the characteristics of those choosing to participate with mobile devices rather than to a choice of device.

Bosch et al. (2018) used a Multitrait-Multimethod (MTMM) experiment included in the NCP and reported that when including smartphone respondents (with optimised design), the average measurement quality was reduced. This could indicate that the measurement quality of smartphone answers is lower. However, differences in measurement quality between all the responses and those from PCs and tablets were not statistically significant. This finding is in agreement with other studies which demonstrated that data coming from PCs and mobile devices present no meaningful differences in terms of reliability, validity (Tourangeau et al., 2017, 2018) and measurement errors (Antoun et al., 2019).

Nevertheless, differences between the size of measurement errors can be expected 1) across types of mobile devices and 2) between surveys which optimised questionnaires for mobile participation and those that did not do so. Wenz (2017) found that users of small smartphones are significantly more likely to provide shorter answers to open-ended questions as well as to be more prone to straightlining. Additionally, surveys optimised for mobiles lead to higher quality data (see Antoun et al. (2018)), specifically reducing straightlining (Borger and Funke, 2015; McClain and Crawford, 2013) and producing more correct answers to test items (Borger and Funke, 2015) when compared to non-optimised alternatives.

Processing errors

Processing errors arise during the data processing stage. For survey data, processing errors can be introduced during data entry, coding, editing, and variable transformations (Bosch and Revilla, 2022). In principle, probability-based online surveys should present a lower risk of introducing processing errors since all responses are recorded online. If designed properly, the risk of introducing errors during data entry stage should be lower when compared to interviewer-administered surveys. Additional benefit of online surveys is that routing is easier when compared to other modes of data collection, especially to postal surveys. In terms of other sources of processing error, these should not differ across modes. However, no empirical comparison has been conducted, to the best of our knowledge.

Discussion

The main aims of this report were to provide an up-to-date and comprehensive compilation of evidence exploring the utility of probability-based online surveys, as well as to compare the different dimensions in which probability-based online surveys vary from a data quality perspective. Overall, the evidence suggests that, for some error sources, differences should be expected between online and offline modes. More specifically, it appears that the larger nonresponse errors are observed in online surveys when compared to offline modes of data collection. The evidence also suggests that the size of the errors can vary across different design choices. To better comprehend the utility of probability-based online surveys, it is, therefore, important to understand how these vary across different design dimensions.

Main results

Based on our literature review, we can provide a summary of the results:

Coverage errors. Some specific design decision of probability-based online surveys can introduce coverage errors. If offliners are excluded from the sample, research has unanimously found that specific populations will be excluded, biasing the sample (e.g., older, less educated, less urban). However, some evidence seems to suggest that excluding offliners does not affect univariate and multivariate statistics (see Eckman, 2016; Bach, Cornesse and Daikeler, 2023). In addition, excluding either mobile-only or PC-only participants might increase bias, although little empirical evidence is available to date.

Sampling errors. In countries such as the UK, probability-based surveys rely on address-based sampling frames, which do not include names of individuals. In some instances, researchers might need to invite sampled units through mail letters. In these cases, the selection of individuals within addresses can only be done by the address residents, which must follow quasi-random protocols (e.g., most recent birthday). These quasi-random protocols present a higher likelihood of introducing selection biases. To avoid quasi-random protocols, all sample members from each address might be asked to participate, but this could increase fraud and fake interviews if incentives are offered as reward for survey participation.

Nonresponse errors. The nonresponse bias has been generally found to be higher for online surveys. This is driven by lower response rates and differential response propensities. The extent to which online surveys introduce nonresponse errors is highly moderated by the design choices made. First, one-time surveys and panels present different nonresponse processes, which can affect the size of their errors. Specifically, individuals joining online panels (e.g., younger, more educated, higher incomes) have been found to be significantly different than those who do not. In addition, those who attrit from panels have also been found to be significantly different than those who stay, although results are mixed in terms of their characteristics. Second, different recruitment strategies also have the potential of affecting nonresponse errors. When conducting a survey on the back of an offline cross-sectional survey or panel, most nonresponse errors come from the initial survey (e.g., older, more educated, women) or panel recruitment. However, the nonresponse associated with joining the panel/answering the survey can be moderated (e.g., age) or exacerbated by introducing more errors (e.g., education, political affiliation, income). Third, providing offliners with internet or allowing them to answer through an offline mode has been found to produce an improvement in the representativeness of online surveys. While those provided with internet connection or connected devices present a similar or even higher loyalty, those offered an offline option are more likely to not respond and/or attrite. Regardless of this, the positive effect on representativeness persists across waves for both approaches. Fourth, regarding the inclusion and optimisation for mobile devices, the existing evidence suggests a significantly higher unit and item nonresponse rate for mobile participants, which can be reduced by optimising the design of the questionnaires for mobile devices.

Adjustment errors. The lack of an interviewer for the recruitment and/or interviewing stages can reduce the amount of auxiliary data available to make adjustments compared to face-to-face surveys. Of special interest is the difference between using fresh samples or recruiting on the back of another offline survey or an offline panel. Weights produced using the information from previous surveys or waves have been found to be highly efficient in reducing the overall bias, substantially more than those computed for fresh samples (e.g., calibration weights from demographics).

Specification errors. There is no reason to expect differences in terms of specification errors, although no empirical research is available.

Measurement errors. Although there are good reasons to expect online surveys to present different measurement errors than offline surveys (e.g., absence of interviewer, visual presentation of questions), most research has demonstrated either no significant differences or the results are mixed. Indeed, the few studies exploring the measurement quality of online surveys compared to face-to-face ones through MTMM analyses have found no significant differences. When it comes to comparing measurement errors across dimensions, some differences have been found, although most appear to be small or inconsistent (presenting mixed results). First, panels can suffer from panel conditioning effects. Mixed results have been found for both positive and negative effects of panel conditioning. Second, allowing offliners to use offline modes can introduce unwanted mode effects within the survey sample, which might negatively affect estimates. The only studies available have found that differential unwanted mode effects exist between interviewer-administered surveys and online ones, but not between mail and online surveys. Third, the presence of mobile devices has also been considered as a potential danger to measurement quality. Nonetheless, regardless of the survey being optimised or not, online and offline surveys have been found to present similar measurement errors, when either comparing response quality indicators or directly exploring measurement quality. In addition, evidence suggests that optimisation of questionnaires further reduces the likelihood of some unwanted behaviours. And as mentioned earlier, the current best practice recommends "mobile-first" questionnaire design when questionnaires are designed with small screens of mobile phones in mind.

Processing errors. Processing errors are expected to be lower for online surveys than offline alternatives, especially than for postal surveys.

Limitations

There are several limitations that should be reported here. First, no quantitative approach was used for summarising the results (e.g., meta-analysis) because of the many different indicators explored across and within different sources of errors in this review. Second, although we have used several databases (Scholar, WebSM, JSTOR, Web of Science) and have tried to capture as many unpublished reports as possible, some published or unpublished results might have not been captured by our review. Third, we did not explore mixed-modes surveys, apart from the mentioned exceptions. Fourth, the four dimensions used to differentiate between probability-based online surveys might not be the only ones to have a substantial impact. These four dimensions were considered the most appropriate for this review. Fifth, it could be argued that for some of the sources of errors there was no need to exclude non-probability surveys (e.g., measurement errors). However, we wanted to avoid any potential effect that professional panellists could have on the reported results. Sixth, the different error sources, as well as the dimensions considered, can interact with each other (see Tourangeau (2020) for a discussion about how errors cumulate). For the sake of simplicity, we have not considered how errors might interact. Nevertheless, researchers should account for these interactions when designing and analysing probability-based online surveys. Finally, for some sources of error and dimensions there is not enough research published to draw solid conclusions.

Practical recommendations

Based on our results, we provide some practical recommendations. We only consider the impact on data quality, regardless of the costs:

1. Although the internet penetration has been growing during last years, providing alternatives to offliners instead of excluding them leads to less biased samples. To avoid potential mode effects, it is recommended to provide internet connection rather than allowing offliners to answer with offline modes. However, this might not always be economically feasible (especially for one-time surveys). In those cases, offering a paper questionnaire as an alternative to offliners is recommended as the likelihood of mode effects is lower in comparison to the situation when an interviewer-administered mode is offered.
2. Although allowing to participate with mobile devices might introduce nonresponse errors (mostly associated with break-offs and item nonresponse), these seem to be offset by the potential coverage errors of excluding the participants willing to take part using mobile devices. Therefore, we recommend allowing mobile devices in all survey which has been

the common practice in survey data collection for a long time now. Since optimising the survey design for mobile devices seems to reduce nonresponse and measurement errors, we also recommend doing so or using "mobile-first" approach to questionnaire design which is considered being the best practice for a number of years now.

3. Recruiting on the back of another cross-sectional offline survey or an offline panel might be a feasible alternative comparable or even a better one than recruiting a fresh sample of respondents. Existing research has demonstrated that little extra bias is introduced when recruiting the panellist this way. Besides, using the rich set of information from the base survey or panel used to obtain the sample can help designing better adjustment strategies while this is not available for fresh samples.
4. Some variables have been linked to nonresponse in most of the literature reviewed, regardless of the one-time or panel nature of surveys, or different recruiting strategies used. Better educated respondents have been found to be more likely to join online surveys/panels and/or to participate in all types of surveys. Literature also demonstrates that people with higher income are more likely to participate in surveys, whereas men and non-white/non-native English speakers are less likely to take part. Therefore, specific strategies and targeted procedures should be explored to tackle these differences in the likelihood of joining online surveys and probability-based online panels.

References

Amaya, A.; Zimmer, S.; Morton, K.; and Harter, R. (2018). "Does Undercoverage on the U.S. Address-based Sampling Frame Translate to Coverage Bias?" *Sociological Methods & Research,* 004912411878253. doi: 10.1177/0049124118782539

Antoun, C. (2015). "*Mobile Web Surveys: a First Look at Measurement, Nonresponse, and CoverageErrors*" (University of Michigan). Retrieved from https://deepblue.lib.umich.edu/bitstream/handle/2027.42/116722/antoun_1.pdf?sequence=1

Antoun, C.; Conrad, F. G.; Couper, M. P.; and West, B. T. (2019). "Simultaneous Estimation of Multiple Sources of Error in a Smartphone-Based Survey." *Journal of Survey Statistics and Methodology, 7*(1): 93–117. doi: 10.1093/jssam/smy002

Antoun, C.; Couper, M. P.; and Conrad, F. G. (2017). "Effects of Mobile versus PC Web on Survey Response Quality." *Public Opinion Quarterly, 81*(S1): 280–306. doi: 10.1093/poq/nfw088

Antoun, C.; Katz, J.; Argueta, J.; and Wang, L. (2018). "Design Heuristics for Effective Smartphone

Questionnaires." *Social Science Computer Review, 36*(5): 557–574. doi: 10.1177/0894439317727072

Bach, R. L.; Cornesse, C.; and Daikeler, J. (2023). "Equipping the Offline Population with Internet Access in an Online Panel: Does It Make a Difference?" *Journal of Survey Statistics and Methodology, 00*: 1–14. doi: 10.1093/JSSAM/SMAD003

Battaglia, M. P.; Link, M. W.; Frankel, M. R.; Osborn, L.; and Mokdad, A. H. (2008). "An Evaluation of Respondent Selection Methods for Household Mail Surveys." *Public Opinion Quarterly, 72*(3): 459–469. doi: 10.1093/poq/nfn026

Benzeval, M.; Burton, J.; Crossley, T. F.; Fisher, P.; Gardiner, C.; and Moore, J. (2021). "*High frequencyonline data collection in an annual household panel study: some evidence on bias prevention and bias adjustment.*"

Bianchi, A.; Biffignandi, S.; and Lynn, P. (2017). "Web-face-to-face mixed-mode design in a longitudinal survey: Effects on participation rates, sample composition, and costs." *Journal of Official Statistics, 33*(2): 385–408. doi: 10.1515/jos-2017-0019

Biemer, P. P. (2010). "Total survey error: Design, implementation, and evaluation." *Public Opinion Quarterly*. doi: 10.1093/poq/nfq058

Blom, A. G.; Bosnjak, M.; Cornilleau, A.; Cousteaux, A.-S.; Das, M.; Douhou, S.; and Krieger, U. (2016). "A Comparison of Four Probability-Based Online and Mixed-Mode Panels in Europe." *Social Science Computer Review, 34*(1): 8–25. doi: 10.1177/0894439315574825

Blom, A. G.; Cornesse, C.; Friedel, S.; Krieger, U.; Fikel, M.; Rettig, T.; … Reifenscheid, M. (2020). "High-Frequency and High-Quality Survey Data Collection: The Mannheim Corona Study." *Survey Research Methods, 14*(2): 171–178. doi: 10.18148/srm/2020.v14i2.7735

Blom, A. G.; Gathmann, C.; and Krieger, U. (2015). "Setting Up an Online Panel Representative of the General Population." *Field Methods, 27*(4): 391–408. doi: 10.1177/1525822X15574494

Borger, C.; and Funke, F. (2015). "Responsive questionnaire design for higher data quality in mobilesurveys." *GOR Conference*.

Bosch, O. J.; and Revilla, M. (2022). "When survey science met web tracking: Presenting an error framework for metered data." *Journal of the Royal Statistical Society: Series A (Statistics in Society), 185*(S2): S408–S436. doi: 10.1111/RSSA.12956

Bosch, O. J.; Revilla, M.; DeCastellarnau, A.; and Weber, W. (2018). "Measurement Reliability, Validity, and Quality of Slider Versus Radio Button Scales in an Online Probability-Based Panel in Norway." *Social Science Computer Review, 37*(1): 119–132. doi:

10.1177/0894439317750089

Bosch, O. J.; Revilla, M.; and Paura, E. (2019). “Answering mobile surveys with images: an exploration using a computer vision API.” *Social Science Computer Review, 37*(5): 669–683. doi: 10.1177/0894439318791515

Bosnjak, M.; Dannwolf, T.; Enderle, T.; Schaurer, I.; Struminskaya, B.; Tanner, A.; and Weyandt, K. W. (2018). “Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany.” *Social Science Computer Review*, *36*(1): 103–115. doi: 10.1177/0894439317697949

Bosnjak, M.; Haas, I.; Galesic, M.; Kaczmirek, L.; Bandilla, W.; and Couper, M. P. (2013). “Sample Composition Discrepancies in Different Stages of a Probability-based Online Panel.” *Field Methods*, *25*(4): 339–360. doi: 10.1177/1525822X12472951

Burton, J.; Lynn, P.; and Benzeval, M. (2020). “How Understanding Society: The UK Household longitudinal study adapted to the COVID-19 pandemic.” *Survey Research Methods*, *14*(2): 235–239. doi: 10.18148/srm/2020.v14i2.7746

Callegaro, M.; Baker, R.; Bethlehem, J. G.; Göritz, A. S.; Krosnick, J. A.; and Lavrakas, P. J. (2014). “Online panel research: History, concepts, applications and al look at the future.” In *Online Panel Research: A Data Quality Perspective* (pp. 1–22).

Cernat, A.; Couper, M. P.; and Ofstedal, M. B. (2016). “Estimation of Mode Effects in the Health and Retirement Study Using Measurement Models.” *Journal of Survey Statistics and Methodology*, *4*(4): 501–524. doi: 10.1093/jssam/smw021

Cernat, A.; and Revilla, M. (2020). “Moving from Face-to-Face to a Web Panel: Impacts on Measurement Quality.” *Journal of Survey Statistics and Methodology, 0*: 1–19. doi: 10.1093/jssam/smaa007

Cernat, A.; and Sakshaug, J. W. (2020). “The impact of mixed modes on multiple types of measurement error.” *Survey Research Methods, 14*(1): 79–91. doi: 10.18148/srm/2020.v14i1.7450

Clement, S. L.; Severin, M. C.; and Shamshiri-Petersen, D. (2020). “Device effects on survey response quality. A comparison of smartphone, tablet and PC responses on a cross sectional probability sample.” *Survey Methods: Insights from the Field*. doi: 10.13094/SMIF-2020-00020

Cook, W. A. (2014). “Is mobile a reliable platform for survey taking?” *Journal of Advertising Research, 54*(2): 141–148. doi: 10.2501/JAR-54-2-141-148

Cornesse, C. (2020). “The utility of auxiliary data for survey response modeling: Evidence from the

German Internet Panel." *Survey Methods: Insights from the Field*. doi: 10.13094/SMIF-2020-00008

Cornesse, C.; and Schaurer, I. (2021). "The Long-Term Impact of Different Offline Population Inclusion Strategies in Probability-Based Online Panels: Evidence From the German Internet Panel and the GESIS Panel." *Social Science Computer Review*, 089443932098413. doi: 10.1177/0894439320984131

Couper, M. P.; Antoun, C.; and Mavletova, A. (2017). "Mobile web surveys: a total survey error perspective." *Total Survey Error in Practice*. doi: 10.1002/9781119041702.ch7

Couper, M. P.; Gremel, G.; Axinn, W.; Guyer, H.; Wagner, J.; and West, B. T. (2018). "New options for national population surveys: The implications of internet and smartphone coverage." *Social Science Research, 73*: 221–235. doi: 10.1016/j.ssresearch.2018.03.008

Couper, M. P.; Kapteyn, A.; Schonlau, M.; and Winter, J. (2007). "Noncoverage and nonresponse in an Internet survey." *Social Science Research*, *36*(1): 131–148. doi: 10.1016/j.ssresearch.2005.10.002

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

Das, M.; Toepoel, V.; and van Soest, A. (2011). "Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys." *Sociological Methods & Research*, *40*(1): 32–56. doi: 10.1177/0049124110390765

de Bruijne, M.; and Oudejans, M. (2015). "Online surveys and the burden of mobile responding." In *Survey measurements: Techniques, data quality and sources of error* (pp. 130–145).

de Bruijne, M.; and Wijnant, A. (2014). "Mobile Response in Web Panels." *Social Science Computer Review*, *32*(6): 728–742. doi: 10.1177/0894439314525918

De Vos, K. (2010). "*Representativeness of the LISS-panel 2008, 2009, 2010*." Retrieved from https://www.lissdata.nl/sites/default/files/bestanden/Representativeness of the LISS panel 2008%2C 2009%2C 2010.pdf

Dillman, D. A. (2017). "The promise and challenge of pushingrespondents to the web in mixed-mode surveys." *Survey Methodology*, *43*(1): 3–30.

DiSogra, C.; Callegaro, M.; and Hendarwan, E. (2009). "Recruiting probability-based web panel members using an address-based sample frame: Results from a pilot study conducted by knowledge networks." *Joint Statistical Meetings*. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.553.3459&rep=rep1&type=pdf

Eckman, S. (2016). “Does the Inclusion of Non-Internet Households in a Web Panel Reduce Coverage Bias?” *Social Science Computer Review*, *34*(1): 41–58. doi: 10.1177/0894439315572985

Erens, B.; Collins, D.; Manacorda, T.; Gosling, J.; Mays, N.; Reid, D.; and Taylor, W. (2019). “Comparing data quality from personal computers and mobile devices in an online survey among professionals.” *Social Research Practice*, *7*: 15–26. Retrieved from https://researchonline.lshtm.ac.uk/id/eprint/4650983/

European Comission. (2014). “*E-communications and telecom single market household survey, special Eurobarometer 414*.” Retrieved from https://ec.europa.eu/digital-agenda/en/news/special-eurobarometer414-e-communications-household-survey

Felderer, B.; Kirchner, A.; and Kreuter, F. (2019, March 1). “The Effect of Survey Mode on Data Quality: Disentangling Nonresponse and Measurement Error Bias.” *Journal of Official Statistics*, Vol. 35, pp. 93–115. doi: 10.2478/jos-2019-0005

Frankel, L. L.; and Hillygus, D. S. (2014). “Looking beyond demographics: Panel attrition in the ANES and GSS.” *Political Analysis*, *22*(3): 336–353. doi: 10.1093/pan/mpt020

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

Groves, R. M.; Fowler, F. J.; Couper, M. P.; Lepkowski, J. M.; Singer, E.; and Tourangeau, R. (2009). “Survey Methodology, 2nd Edition.” *Wiley Series in Survey Methodology*.

Gummer, T.; Quoß, F.; and Roßmann, J. (2019). “Does Increasing Mobile Device Coverage Reduce Heterogeneity in Completing Web Surveys on Smartphones?” *Social Science Computer Review*, *37*(3): 371–384. doi: 10.1177/0894439318766836

Halpern-Manners, A.; Warren, J. R.; and Torche, F. (2014). “Panel Conditioning in a Longitudinal Study of Illicit Behaviors.” *Public Opinion Quarterly*, *78*(3): 565–590. doi: 10.1093/poq/nfu029

Heerwegh, D.; and Loosveldt, G. (2008). “Face-to-Face versus Web Surveying in a High-Internet-Coverage Population: Differences in Response Quality.” *Public Opinion Quarterly*, *72*(5): 836–846. doi: 10.1093/poq/nfn045

Hoogendoorn, A. W.; and Daalmans, J. (2009). “Nonresponse in the recruitment of an internet panel based on probability sampling." *Survey Research Methods*, *3*(2): 59–72. doi: 10.18148/srm/2009.v3i2.1551

Hox, J. J.; De Leeuw, E. D.; and Zijlmans, E. A. O. (2015). "Measurement equivalence in mixed mode surveys." *Frontiers in Psychology*, *6*(FEB): 87. doi: 10.3389/fpsyg.2015.00087

Jäckle, A.; Lynn, P.; and Burton, J. (2015). "Going online with a face-to-face household panel: Effects of a mixed mode design on item and unit non-response." *Survey Research Methods*, *9*(1): 57–70. doi: 10.18148/srm/2015.v9i1.5475

Jessop, C. (2017). "*Developing the NatCen Panel*." Retrieved from https://natcen.ac.uk/media/1484228/Developing-the-NatCen-Panel-V2.pdf

Jessop, C.; Wood, M.; and Marshall, L. (2016). "*Social and political attitudes of people on low incomes: Feasibility Study Part 3: Establishing a Panel*."

Klausch, T.; Hox, J. J.; and Schouten, B. (2013). "Measurement Effects of Survey Mode on the Equivalence of Attitudinal Rating Scale Questions." *Sociological Methods & Research*, *42*(3): 227–263. doi: 10.1177/0049124113500480

Kocar, S.; and Biddle, N. (2020). "*Panel mixed-mode effects: does switching modes in probability-based online panels influence measurement error?*" Retrieved from https://csrm.cass.anu.edu.au/sites/default/files/docs/2020/2/Panel_mixed_mod_effects_does_switching_modes_in_probability_based_online_panels_influence_measurement_error.pdf

Kölln, A.-K.; Ongena, Y. P.; and Aarts, K. (2019). "The Effects of Sampling Frame Designs on Nonresponse and Coverage Error: Evidence from the Netherlands." *Journal of Survey Statistics and Methodology*, *7*(3): 422–439. doi: 10.1093/jssam/smy016

Kreuter, F.; Olson, K.; Wagner, J.; Yan, T.; Ezzati-Rice, T. M.; Casas-Cordero, C.; … Raghunathan, T. E. (2010). "Using proxy measures and other correlates of survey outcomes to adjust for non-response: examples from multiple surveys." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(2): 389–407. doi: 10.1111/j.1467-985X.2009.00621.x

Kreuter, Frauke; Presser, S.; and Tourangeau, R. (2008). "Social Desireability Bias in CATI, IVR, and Web Surveys: the effects of modeand question sensibility." *Public Opinion Quarterly*, *72*(5): 847–865. doi: 10.1093/poq/nfn063

Kruse, Y.; Callegaro, M.; Dennis, J. M.; Subias, S.; Lawrence, M.; DiSogra, C.; and Tompson, T. (2009). "Panel conditioning and attrition in the AP-Yahoo! news election panel study." *64th Conference of the American Association for Public Opinion Research (AAPOR).*

Lambert, A. D.; and Miller, A. L. (2015). "Living with Smartphones: Does Completion Device Affect

Survey Responses?” *Research in Higher Education, 56*(2): 166–177. doi: 10.1007/s11162-014-9354-7

Lavallee, P. (2007). “*Indirect Sampling.*” Berlin, Germany: Springer.

Lee, H.; Kim, S.; Couper, M. P.; and Woo, Y. (2019). “Experimental Comparison of PC Web, Smartphone Web, and Telephone Surveys in the New Technology Era.” *Social Science Computer Review, 37*(2): 234–247. doi: 10.1177/0894439318756867

Leenheer, J.; and Scherpenzeel, A. C. (2013). "Does it pay off to include non-internet households in an internet panel?." *International Journal of Internet Science*, *8*(1).

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

Lugtig, P. (2014). “Panel Attrition.” *Sociological Methods & Research, 43*(4): 699–723. doi: 10.1177/0049124113520305

Lugtig, P.; Das, M.; and Scherpenzeel, A. C. (2014). “Nonresponse and attrition in a probability-based online panel for the general population.” In *Online Panel Research: A Data Quality Perspective* (pp. 135–153). John Wiley & Sons Ltd.

Lugtig, P.; and Toepoel, V. (2016a). “The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey.” *Social Science Computer Review, 34*(1): 78–94. doi: 10.1177/0894439315574248

Lugtig, P.; and Toepoel, V. (2016b). “The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey.” *Social Science Computer Review, 34*(1): 78–94. doi: 10.1177/0894439315574248

Lugtig, P.; Toepoel, V.; and Amin, A. (2016). “Mobile-only web survey respondents.” *Survey Practice, 9*(4). doi: https://doi.org/10.29115/SP-2016-0020

Lynn, P.; and Kaminska, O. (2012). “Combining refreshment or boost samples with an existing panel sample: challenges and solutions.” *International Workshop on Panel Survey Methods*.

Lynn, P.; and Lugtig, P. (2017). “Total survey error for longitudinal surveys.” In *Total survey error in practice* (pp. 279–298). Wiley & Sons.

Macer, T.; and Wilson, S. (2014). “*Confirmit 2013 annual MR software survey.*” Retrieved from https://www.meaning.uk.com/resources/reports/2013-Confirmit-MR-technology-survey.pdf

Manfreda, K. L.; Bosnjak, M.; Berzelak, J.; Haas, I.; and Vehovar, V. (2008). “Web Surveys versus other Survey Modes: A Meta-Analysis Comparing Response Rates.” *International Journal of*

Market Research, 50(1): 79–104. doi: 10.1177/147078530805000107

Maslovskaya, O.; and Lugtig, P. (2022). "Representativeness in six waves of CROss-National Online Survey (CRONOS) panel." *Journal of the Royal Statistical Society Series A: Statistics in Society, 185*(3), 851-871.

Maslovskaya, O.; Smith, P. W. F.; and Durrant, G. (2020). "*Do respondents using smartphones produce lower quality data? Evidence from the UK Understanding Society mixed-device survey.*" Retrieved from http://eprints.ncrm.ac.uk/4322/1/DataQuality_UnderstandingSociety_NCRMWorkingPaper.pdf

McClain, C.; and Crawford, S. D. (2013). "Grid formats, data quality, and mobile device use: Toward aquestionnaire design approach." *AAPOR Annual Conference*.

McGeeney, K.; and Marlar, J. (2013). "Mobile browser web surveys: Testing response rates, data quality,and best practices." *AAPOR Annual Conference*.

Murphy, J.; Biemer, P.; Stringer, C.; Thissen, R.; Day, O.; and Hsieh, Y. P. (2016). "Interviewer falsification: Current and best practices for prevention, detection, and mitigation." *Statistical Journal of the IAOS, 32*(3): 313–326. doi: 10.3233/SJI-161014

Nicolaas, G. (2022) Within-household Selection for Push-to-Web Surveys. GenPopWeb2 Report. https://www.ncrm.ac.uk/documents/Within%20household%20selection%20for%20push%20to%20web%20surveys.pdf.

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

Olson, K. (2013). "Paradata for Nonresponse Adjustment." *The ANNALS of the American Academy of Political and Social Science, 645*(1): 142–170. doi: 10.1177/0002716212459475

Olson, K.; and Smyth, J. D. (2017). "Within-Household Selection in Mail Surveys." *Public Opinion Quarterly, 81*(3): 688–713. doi: 10.1093/poq/nfx025

Park, A.; and Humphrey, A. (2014). "*Mixed-mode surveys of the general population - Results from the European Social Survey mixed-mode experiment.*" Retrieved from http://www.websm.org/uploadi/editor/1416295092Park_Humphrey_2014_Mixed_Mode_Survey_Of_The_General_Population.pdf

Peterson, G.; Griffin, J.; LaFrance, J.; and Li, J. (2017). "Smartphone Participation in Web Surveys." In *Total Survey Error in Practice*. doi: 10.1002/9781119041702.ch10

Pew Research Center. (2015). "*Building Pew Research Center's American Trends Panel.*" Retrieved from https://www.pewresearch.org/methods/2015/04/08/building-pew-research-centers-american-trends-panel/

Revilla, M. A. (2013). “Measurement invariance and quality of composite scores in a face-to-face and a web survey.” *Survey Research Methods*, *7*(1): 17–28. doi: 10.18148/SRM/2013.V7I1.5098

Revilla, M.; Cornilleau, A.; Cousteaux, A.-S.; Legleye, S.; and de Pedraza, P. (2016). “What Is the Gain in a Probability-Based Online Panel of Providing Internet Access to Sampling Units Who Previously Had No Access?” *Social Science Computer Review*, *34*(4): 479–496. doi: 10.1177/0894439315590206

Revilla, M.; and Couper, M. P. (2017). “Comparing Grids With Vertical and Horizontal Item-by-Item Formats for PCs and Smartphones.” *Social Science Computer Review*. doi: 10.1177/0894439317715626

Revilla, M.; Couper, M. P.; Bosch, O. J.; and Asensio, M. (2020). “Testing the Use of Voice Input in a Smartphone Web Survey.” *Social Science Computer Review*, *38*(2): 207–224. doi: 10.1177/0894439318810715

Revilla, M.; and Saris, W. E. (2013). “A comparison of the quality of questions in a face-to-face and a web survey.” *International Journal of Public Opinion*, *25*(2): 242–253.

Sarraf, S.; Brooks, J.; Cole, J.; and Wang, X. (2015). “What is the impact of smartphone optimization onlong surveys?” *APOR Annual Conference*.

Scherpenzeel, A. C.; and Bethlehem, J. G. (2011). “How representative are online panels? Problems of coverage and selection and possible solutions.” In *ocial and behavioral research and the Internet: Advances in applied methods and research strategies* (pp. 105–132). Routledge.

Schonla, M.; and Toepoel, V. (2015). “Straightlining in Web survey panels over time.” *Survey Research Methods*, *9*(2): 125–137. doi: 10.18148/srm/2015.v9i2.6128

Schonlau, M.; van Soest, A.; Kapteyn, A.; and Couper, M. (2009). “Selection Bias in Web Surveys and the Use of Propensity Scores.” *Sociological Methods & Research*, *37*(3): 291–318. doi: 10.1177/0049124108327128

Shih, T.-H.; and Xitao Fan. (2008). “Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis.” *Field Methods*, *20*(3): 249–271. doi: 10.1177/1525822X08317085

Skjervheim, O.; Høgestøl, A.; Bjørnebekk, O.; Eikrem, A.; and Wettergreen, J. (2020). “*Norwegian Citizen Panel 2020 - Nineteenth Wave Methodology report.*”

Smith, A. (2015). “U.S. Smartphone Use in 2015. Pew Research Center.” *Pew Research Center: Internet, Science & Tech*. Retrieved from http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/

Struminskaya, B. (2016). “Respondent Conditioning in Online Panel Surveys.” *Social Science*

Computer Review, 34(1): 95–115. doi: 10.1177/0894439315574022

Struminskaya, B.; Weyandt, K.; and Bosnjak, M. (2015). "The Effects of Questionnaire Completion Using Mobile Devices on Data Quality. Evidence from a Probability-based General Population Panel." *Methods, Data, Analyses, 9*(2): 261–292. doi: 10.12758/mda.2015.014

Toepoel, V.; Das, M.; and Van Soest, A. (2009). "Relating question type to panel conditioning: Comparing trained and fresh respondents." *Survey Research Methods, 3*(2): 73–80. doi: 10.18148/srm/2009.v3i2.874

Toepoel, V.; and Lugtig, P. (2014a). "What Happens if You Offer a Mobile Option to Your Web Panel? Evidence From a Probability-Based Panel of Internet Users." *Social Science Computer Review, 32*(4): 544–560. doi: 10.1177/0894439313510482

Toepoel, V.; and Lugtig, P. (2014b). "What Happens if You Offer a Mobile Option to Your Web Panel? Evidence From a Probability-Based Panel of Internet Users." *Social Science Computer Review, 32*(4): 544–560. doi: 10.1177/0894439313510482

Tourangeau, R. (2020). "How Errors Cumulate: Two Examples." *Journal of Survey Statistics and Methodology, 8*(3): 413–432. doi: 10.1093/jssam/smz019

Tourangeau, R.; Conrad, F. G.; and Couper, M. P. (2013). "The science of web surveys." In *Oxford University Press*. doi: 10.1093/acprof:oso/9780199747047.001.0001

Tourangeau, R.; Maitland, A.; Rivero, G.; Sun, H.; Williams, D.; and Yan, T. (2017). "Web Surveys by Smartphone and Tablets." *Public Opinion Quarterly, 81*(4): 896–929. doi: 10.1093/poq/nfx035

Tourangeau, R.; Sun, H.; Yan, T.; Maitland, A.; Rivero, G.; and Williams, D. (2018). "Web Surveys by Smartphones and Tablets." *Social Science Computer Review, 36*(5): 542–556. doi: 10.1177/0894439317719438

Villar, A. (2013). "Feasibility of Using Web to Survey at a Sample of Addresses: a UK ESS experiment." *NCRM Workshop: Web Surveys for the General Population: How, Why and When?* Retrieved from http://www.websm.org/uploadi/editor/1372075808Villar_2013_Feasibility_of_Using_Web_to_Survey.pdf

Wells, T.; Bailey, J. T.; and Link, M. W. (2013). "Filling the Void: Gaining a Better Understanding of Tablet-Based Surveys." *Survey Practice, 6*(1). doi: https://doi.org/10.29115/SP-2013-0002

Wells, T.; Bailey, J. T.; and Link, M. W. (2014). "Comparison of Smartphone and Online Computer Survey Administration." *Social Science Computer Review, 32*(2): 238–255. doi:

10.1177/0894439313505829

Wenz, A. (2017). "*Completing web surveys on mobile devices: Doesscreen size affect data quality?*" (No. 2017–05). Retrieved from https://www.econstor.eu/bitstream/10419/163548/1/890214735.pdf

West, B. T.; and Blom, A. G. (2017). "Explaining interviewer effects: A research synthesis." *Journal of Survey Statistics and Methodology, 5*(2): 175–211. doi: 10.1093/jssam/smw024

Williams, J. (2015). "*Community Life Survey - Investigating the feasibility of sampling all adults in the household.*"

Appendix A: Descriptive information of the literature used

Categories	Types of literature	Percentage of papers
Revision process	Peer-reviewed	64.1
	Not peer-reviewed	28.3
	Not applicable or unknown	7.6
Outlet	Journal	65.2
	Book or chapter	5.4
	Report	15.2
	Conference	13.0
	Dissertation	1.1
Country	UK	18.9
	USA	26.7
	Netherlands	24.4
	Germany	11.1
	Slovenia	2.2
	Estonia	2.2
	Norway	2.2
	Denmark	1.1
	Belgium	1.1
	France	1.1
	Europe (General)	2.2
	Australia	1.1
	South Korea	1.1
	Not applicable	8.9
Population	General	79.3
	Other	13.1
	Not applicable	7.6