The Survey Statistician, 2023, Vol. 87, 12-26

Ask the Experts

Mixed mode official surveys. Current status and near future[1]

Barry Schouten[2], Jan van den Brakel[3], Deirdre Giesen[4(a)], Annemieke Luiten[4(b)] and Vivian Meertens[4(c)]

[2] Statistics Netherlands and Utrecht University, jg.schouten@cbs.nl
[3] Statistics Netherlands and Maastricht University, ja.vandenbrakel@cbs.nl
[4] Statistics Netherlands, [(a)]d.giesen@cbs.nl, [(b)]a.luiten@cbs.nl & [(c)]v.meertens@cbs.nl

Abstract

Recently, Schouten et al. (2022) synthesized literature and research on both design and analysis of mixed-mode surveys in the context of official statistics. The COVID-19 pandemic has pointed very clearly at the benefits and drawbacks of mixed-mode survey designs, especially when comparability over time is key. A brief account is given of the main themes in design and analysis of mixed-mode surveys, which may be seen as starting points for National Statistical Institutes trying to make their designs more robust.

Keywords: Online/web, telephone, face-to-face, mode effects.

1 Introduction

The rise of online surveys over the past two decades has kindled a strong interest in mixed-mode surveys. Web questionnaires are cheap and fast, but have a lower response rate than traditional interviewer-based or mailed-out paper questionnaire modes. Combining web surveys with another survey mode like telephone or face-to-face interviewing is thus an obvious way of boosting low web response rates. In 2018, a survey was held among European National Statistical Institutes (NSIs) on their (mixed mode) practices. At that time, all but one country used mixed-mode designs for some or all of their surveys, and about half said web surveys were one of the design modes (Murgia, Lo Conte and Gravem, 2018). The COVID-19 crisis of 2020 and 2021 gave a strong impulse in NSIs' thinking about web and mixed-mode surveys, as face-to-face interviews were no longer possible. The new situation gave rise to a new inventory among European NSIs as to their experiences during the COVID crisis (Beck et al., 2022). The inventory showed that mixed-mode expanded rapidly, although face-to-face interviewing was mostly replaced with telephone interviewing, while the level

[1] This article is based to a large extent on the book written by the authors: Schouten, B., Van den Brakel, J., Buelens, B, Giesen, D., Luiten, A. and Meertens, V. (2022): Mixed-Mode Official Surveys. Design and Analysis. Boca Raton: CRC Press.

of web surveys remained stable. It appeared far more complicated to introduce new modes, than to expand the use of modes that were already familiar. Although mixing modes allows more flexibility in unforeseen circumstances, not all countries envisage continuation of the new mixed mode designs after COVID. Flexibility comes with added complexity, both in designing questionnaires, fieldwork logistics, adjustment and analysis. On the other hand, rising fieldwork costs and coverage problems in single mode surveys force others to embrace mixed mode solutions, and include web into the mix. If countries do consider the continuation of the new mode design, it is primarily for reasons of cost-effectiveness (Beck et al., 2022).

The use of mixed-mode designs is not new, but the current emphasis on the effects on analysis and estimation is. The effects of modes on non-response and measurements errors have long been known and have been the subject of research for decades. See for example Groves (1989) and Toepoel et al. (2020) for an older and a more recent discussion of mode effects: the combined effects of mode specific measurement error and mode-specific selection error. The web mode has completely different characteristics than the interviewer-based telephone and face-to-face modes. In addition, the web mode is diverse, as many devices can access the Internet. Desktops, laptops, tablets, smartphones, smart TV's and even some wearables can go online. In recent years, survey methodology conferences such as AAPOR and ESRA held many sessions solely on mixed-mode survey design and its impact on survey errors, survey costs, and survey estimates.

In this article, we give a short summary of recent insights into some of the most challenging aspects of mixed mode data collection. Section 2 describes data collection and questionnaire design. Section 3 describes how to test the impact of mixed mode surveys and how to adjust for mode effects. Finally, in section 4, we take a look at the future and the possibilities and challenges of the use of smart devices like smartphones, wearables and other sensor systems.

2 The design of mixed mode surveys

When designing a survey strategy, there are several aspects to consider: which modes to choose, in which sequence to apply them, and what communication strategy to apply. Underlying are considerations of costs, the prevention of nonresponse and selection bias, and the optimisation of data quality. In addition, mixed-mode questionnaire design plays a very important role in preventing mode-specific measurement errors.

To date, all four modes (mailed-out paper questionnaires, web, telephone and face-to-face interviewing) may appear, even in a single survey. Schouten et al. (2022) distinguish six features of survey modes: intimacy, interaction, assistance, speed and pace, presentation (aural or visual) and timing. These features impact the survey taking experience and differ to a large extent between the modes. The largest difference comes from interviewer versus self-administered modes, which show disparity on all mode features. In addition, modes may also introduce specific answering behaviours and response styles. For example, interviewer modes, through their higher intimacy, interaction and assistance, may help avoid errors, but may on the other hand introduce socially desirable answering. Bais et al. (2017) studied mode as a source of differential measurement error. They use a scoring of survey item characteristics, to predict measurement error risk. The score for one of Statistics Netherlands' surveys was such that a large-scale experiment was set up to estimate mode-specific measurement biases relative to face-to-face (Schouten et al., 2013).

Schouten et al. (2022) show that various modes and mode combinations may exist within one and the same NSI. A specific mixed mode survey design will be determined by aspects like the complexity of the subject matter (that may warrant interviewer assistance), the survey length (long surveys are less suitable for web and telephone), whether or not measurement differences are to be expected (e.g., for sensitive or attitude questions), and whether or not longitudinal measurements are foreseen (investing in the relation with respondents pays off if they are to be interviewed several times). Different surveys within one NSI may warrant different designs as a result of these considerations.

The choice of a mixed-mode design also depends on the motives for mixing them: minimizing total survey error, reducing costs, or increasing the speed of data collection. If speed is the main consideration, one could offer all modes at the same time, in a so called concurrent design. If costs are the main consideration, one should employ a consecutive design, starting with the cheapest mode (web), and only introducing other modes after a period of time. In inventories among European NSIs in 2013 (Blanke and Luiten, 2014) and 2018 (Murgia et al., 2019), costs appeared to be the main driver for the introduction of web surveys. For fieldwork costs, this is evident. Web interviews may be 50 to 100 time less expensive than face-to-face interviews. However, introducing web may have negative cost implications for other aspects of fieldwork, like longer travel times for the remaining sample persons, or higher interviewer turnover because of the more difficult job.

The sequence of offering modes and the implication on response rates and costs has been the subject of many studies. Medway and Fulton (2012) showed in a meta-analysis, that offering sample persons a choice of paper or web questionnaires leads to slightly lower response rates than just offering a paper questionnaire. In addition, this kind of concurrent mixed mode design limits the possibility of saving costs, as a large majority of persons will choose the paper option. Careful design could counteract this tendency. Biemer et al. (2018) show that offering persons an incentive to choose the web option substantially increases the web uptake and is even costs effective. Whether or not concurrent designs have different implications for response rates than sequential designs is unclear (Patrick et al., 2018). There are studies showing that the one leads to higher response rates, but there are also studies showing the opposite result.

The inventory by Murgia et al. (2019) showed that web surveys have been introduced in a large number of surveys in most European countries. They occur in four kinds of designs:

1. As the first mode in a sequential design. Nonrespondents are followed up in other modes. This design offers opportunities for large costs savings, but runs the risk of lowering response rates.

2. As a concurrent mode, offered at the same time as other options. This design is most often used when the other option is a paper questionnaire, but combinations between web and telephone and/or face-to-face interviewing exist as well. Web as concurrent mode was the most commonly used mixed mode design in 2013 in Europe, and still was in 2018.

3. Web is offered to nonrespondents of earlier telephone or face-to-face approaches. This design is chosen when the aim is to improve response rates over the initial mode. This design is used but only rarely.

4. Web is offered in subsequent waves, when respondents in a previous wave can indicate their willingness to fill in a web questionnaire, and give their email address.

What mode or what sequence of modes is offered does not need to be uniform for all sample persons. Adaptive or responsive survey designs target different approaches to specific groups. A simple version would be to offer the web survey only to younger persons. More advanced versions use insight into response and measurement errors per group per mode to fine-tune mode allocation. Calinescu and Schouten (2016) illustrate an adaptive mixed-mode design for the Dutch Labour Force Survey, where mode switches depend on response probabilities of different subsamples.

2.1 Communication strategy

Most NSIs, including Statistics Netherlands, do not have access to registries of telephone numbers or email addresses. That means that, even if sample persons are invited to participate in a web survey, the most common means of letting them know of the request is by letter. There is an ever-increasing host of research into the best way of 'web-push' designs (Dillman, 2017). Elements of this communication strategy are how to describe the task of logging on to the web, how to remind the respondent and how often, how to convey trustworthiness, and if and how to offer incentives. Dillman, Smyth and Christian (2014) developed guidelines to increase web response rates:

Use multiple contact modes to increase the likelihood that invitations are received and heeded by sample persons.

Make contact by a mode different from the response mode to increase trust that the survey is legitimate and useful. So, even if email addresses or cell phone numbers are available, it is good practice to also send an invitation letter alongside emails or text messages.

Send a small cash incentive with an initial postal mail contact to increase trust in the survey. Dillman, Smyth and Christian (2014) also suggest a second cash incentive in a later contact. However, differential incentives between early and later respondents may be an (ethical) issue here.

Try to obtain contact information for more than one survey mode.

Statistics Netherlands has found that offering incentives when inviting people to a web questionnaire is a cost-effective way of increasing web response rates. Depending on the specific survey design, both unconditional incentives (5€ gift card included in the invitation letter), conditional incentives (10 or even 20€ gift cards upon completion of the survey) or lottery incentives will be effective. See Schouten et al. (2022) for an overview of the communication and incentive experiments that were performed by Statistics Netherlands, and their effect on response rates and sample composition.

The COVID crisis that pressed NSIs to rethink their data collection strategy, also invited rethinking the communication strategy. Beck et al. (2022) describe that NSIs made changes in their contact channels, for example by including a QR code in the invitation letter, or by using e-mails and text messages. The latter channels are especially used for reminders and additional information. In some cases, national legislation was adapted to give NSIs the right to access telephone numbers and email addresses that were available in other public administrations (like the tax authority). But getting the number or the email address from the respondent, by either requesting it in the invitation letter, or visiting (without entering) the household was also frequently employed.

2.2 Questionnaire design

Administering questionnaires in different modes runs the risk of introducing measurement differences. The way that respondents understand and answer questions can be affected by characteristics like the presence or absence of an interviewer, whether questions are heard or read, and the way questions are visually presented on a screen. Mixed mode questionnaire design is an important tool in preventing or reducing mode-specific measurement effects.

Ideally, the properties of the questionnaire would play a role in the design of the survey. For example, if it is identified in the conceptual phase that mainly sensitive questions will be asked, it is wise not to mix interview modes with non-interviewer modes. An alternative could be to mix the non-interviewer mode with face-to-face, where the sensitive questions can be completed by the respondents themselves (CASI, Computer Assisted Self Interviewing).

Mixed mode questionnaires can be designed with a focus on minimizing measurement differences between modes by presenting a unified stimulus in all the modes used. This is called the unified mode approach (e.g., Dillman, Smyth and Christian. 2014, Dillman and Edwards, 2016). Alternatively, the ‘best practices approach’ focusses on minimizing error within each mode, even if that means using different stimuli in different modes.

Tourangeau, Conrad and Couper (2013) and Tourangeau (2017) argue that the best practice approach will best estimate the ‘true value’, and is thus preferable if an overall point estimate is the main goal of a survey. On the other hand, if comparison between groups is important, the unified approach is better suited. As mode distribution in the response can change over time, comparability over time is an additional reason to prefer unified mode design.

In practice, comparisons between groups and over time are paramount for most surveys. Therefore, mixed mode questionnaire design generally focuses on minimizing measurement differences between modes, i.e. the unified mode approach. Also from a practical and costs efficiency point of view it is preferable to design questionnaires for various modes as identical as possible. Mind, however, that some differences between modes must of necessity exist, even in a unified approach (e.g. “press enter to continue”).

Schouten et al. (2022) advise to not just add a new mode like web to an existing survey, but to do a total redesign of the questionnaire if a mode is added. They also advise that mixed mode requirements should be considered in each stage of the questionnaire design process, e.g. the pre-testing of the questionnaire and the evaluation of the field work. If web is part of the mode mix, and the web questionnaire may be accessed by smartphone, consider designing the questionnaire for the smartphone. The next section is on this topic.

2.3 Smartphone questionnaires

If one of the modes is web, mixed mode questionnaire design will also entail mixed device design. Increasingly, respondents will access the web with a smartphone. This poses additional challenges to the design as it forces the developer to design for the smallest device. Figure 1 depicts how in continuous Statistics Netherlands’ web surveys, the percentage of smartphone logins has increased to 35%, while both tablets and PC logins are on the decline.

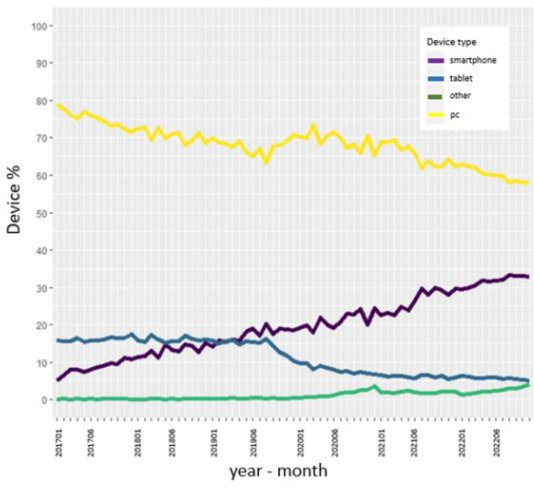

Figure 1. Device use (smartphone, tablets, PC, other) from 2017 to 2022 in percentage of number of first login attempts.

Note: The ‘other’ devices are mostly new smartphone types but some respondents do login with smart watches

Smartphones can impact the questionnaire experience, as the size of the screen is much smaller, potentially influencing how much text is visible and how easy it is to read the text. In addition, while respondents can use their keyboard on a PC, they need to use the touch screen for the smartphone. Open answers may be far more difficult to type in as a result. Respondents may also be less focused on the task, as the chance of distractions while answering the questions is highly probable (Couper, Antoun and Mavletova, 2017). However, well-designed questionnaires can prevent device-specific measurement errors (e.g., Tourangeau et al., 2017).

The European MIMOD – Mixed Mode Designs for Social Surveys – project (Schouten et al., 2018, Schouten et al., 2022) developed a list of indicators to assess if questionnaires are 'smartphone fit', or if device-related measurement risks can be expected. The first indicator counts the question length, presence of instructions, horizontal answer scales or matrix questions and the number of answer categories, to determine if the question will fit on a small screen. The second indicator evaluates the necessity of screen navigation, for example as a result of open questions and items with many answer categories. The third indicator is about response burden: the number of questions, the number of household members who need to respond, and 'underwater' interaction with databases, e.g., for the classification of occupation. The fourth is on task complexity. Persons may need to look up information in their administration, a task that may be incompatible with the location in which the respondent fills in the questionnaire. The final indicator is on perceived enjoyment or burden. These indicators may help making surveys more smartphone friendly or may even suggest a smartphone first design in a mixed mode context. Less enjoyable surveys may be extra susceptible to break-off or sub-optimal respondent effort if the smartphone does not facilitate a smooth respondent experience (Gravem et al., 2019).

Adapting questionnaires for mixed-device data collection is still a relatively new field. This list of indicators will be further developed and adjusted as more research into device effects become available.

3 The design of mixed mode surveys

3.1 Field tests

In survey methodology literature many references to experimental studies on improving the quality or efficiency of survey processes can be found. For example studies to compare the effect of different questionnaire designs, data collection modes or approach strategies on the main outcomes of a sample survey, with the purpose of reducing measurement errors, nonresponse bias or improving response rates. Many experiments conducted in this context are small scaled or conducted with specific groups. The value of empirical research into survey methods is strengthened as conclusions can be generalized to populations larger than the sample that is included in the experiment. This can be achieved by selecting experimental units randomly from a larger target population and naturally leads to randomized experiments embedded in probability samples, Fienberg and Tanur (1987, 1988, 1989). This enables the generalization of conclusions observed in an experiment to larger target populations and is particularly important if experiments are conducted to improve survey methods or to obtain quantitative insight in different sources of non-sampling errors in survey research. From this point of view they complement the insights obtained with small scale cognitive and usability testing strategies.

In an embedded experiment a probability sample is drawn from a finite target population, which is randomly divided into two or more subsamples according to a randomized experiment. In survey literature, such experiments are also referred to as split-ballot designs or interpenetrating subsampling, and date back to Mahalanobis (1946), but see also Cochran (1977, section 13.15), Hartley and Rao (1978). Designing and planning a randomized experiment embedded in a probability sample requires a careful preparation and planning of the fieldwork of the survey organization. Besides the standard decisions on the minimum required sample size to observe pre-specified treatment effects that should at least result in a rejection of the null hypothesis at pre-specified levels of significance and power, many considerations and decisions on how to implement and conduct the field work of the experiment have to be taken. Chapter 7 in Schouten et al. (2022) provides a discussion of the aspects of designing large scaled field tests. See also the edited volume by Lavrakas et al. (2019) for an extensive overview of experimental methods in survey research.

An important issue in the analysis of such experiments is to find the right inferential framework. The standard literature on design and analysis of experiments applies model-based inference procedures for the analysis of experiments. The observations obtained in the experiment are assumed to be the realization of a linear model. To test hypotheses about treatment effects, F-tests are derived under the assumption of normally and independently distributed observations. Inference in survey sampling is traditionally design-based, which implies that the statistical inference is based on the stochastic structure of the sampling design. Approximately design-unbiased parameter and variance estimators are derived under the concept of repeatedly drawing samples from a finite population according to the same sampling design, keeping all other population parameters fixed. If experiments are embedded in probability samples with the purpose to generalize conclusions to larger target populations, a design-based inference framework might be more appropriate than the model-based approach traditionally used in the analysis of randomized experiments. Van den Brakel and Renssen (2005) and Van den Brakel (2008, 2013) developed such a design-based framework for the analysis of a large range of randomized experimental designs embedded in probability samples.

3.2 Discontinuities and survey redesigns

Implementing a mixed-mode design in a repeatedly conducted survey, generally has systematic effects on the sample estimates. Survey samples contain besides sampling errors different sources of non-sampling errors that have a systematic effect on the outcomes of a survey. As long as the survey process is kept constant, this bias component is not directly visible. If, however, one or more components of the survey process are modified, these non-sampling errors are changed. Modifications in the survey process therefore generally have systematic effects on the survey estimates, sometimes called discontinuities, which disturb comparability with figures published in the past. To avoid confounding real period-to-period change in the parameters of interest with change due to alteration in the survey process, it is important to quantify the effect of changing the field work strategies and data collection modes.

Van den Brakel, Zhang and Tam (2020) propose a framework of statistical methods to measure discontinuities, depending on the type of change in the survey process. In the case of changing field work strategies or data collection methods, it can be expected that the micro data are not consistent under the old and new approaches. In such situations, a straightforward and reliable approach to quantify discontinuities is to collect data under the old and new approach alongside of each other at the same time. This is referred to as parallel data collection. Redesign of long-standing surveys like e.g. the US Current Population Survey and the US National Crime Victimization Survey are accompanied with a parallel run, (Dippo, Kostanich and Polivka, 1994, and Kindermann and Lynch, 1997). A parallel run is preferably designed as an embedded randomized experiment. The design-based inference approach, discussed at the end of Subsection 3.1, is particularly appropriate for the analysis of such parallel runs as it tests hypotheses about treatment effects on the sample estimates for the finite population parameters of the survey, accounting for both the sample design and the experimental design.

A strong point of parallel data collection is the low risk level to the regular publications during the changeover to the new design. Through a well-designed experiment the risk of failing to detect a discontinuity is minimized since the design of an experiment gives full control over the minimum detectable difference at a pre-specified significance and power level. A disadvantage of a parallel run is that extra cost is required for additional data collection.

If, due to lack of budget or fieldwork capacity, the new survey process is implemented without parallel data collection, discontinuities can be quantified using time series methods. Van den Brakel, Zhang and Tam (2020) discuss the use of structural time series models (STM). Under this approach, an appropriate STM is proposed for the population parameter of interest, for example with an appropriate model for the trend, the seasonal component and the unexplained population variation. This model is extended with a component that models the sampling error and a level intervention that changes from zero to one at the moment that the new mixed-mode design is implemented in the survey process. An estimate of the regression coefficient of the level intervention, usually obtained with the Kalman filter, can be interpreted as an approximation of the discontinuity. See for example Durbin and Koopman (2012) for details of STM and the estimation via the Kalman filter.

A major advantage of the time series approach is that no additional data collection is required, which makes this approach very cost-effective. Skipping a period of parallel data collection and relying on a time series model to estimate discontinuities also has several disadvantages and risks. During the period directly after the changeover, estimates for the discontinuities are prone to large revisions, each time new observations under the new survey become available. As a consequence, revisions must be accepted and final estimates are not timely. Another important drawback of this approach is that real developments and estimates for the discontinuities are confounded if the real evolution of the population parameter deviates from the assumed time series model. This situation can occur, for example if the changeover of the survey coincides with the start of the Global Financial Crisis in 2009 or the corona crisis in 2020. Finally, under the time series modelling approach there is no control over the precision and size of the minimum observable differences, resulting in an increased risk of failing to detect relevant discontinuities.

There might be budget or capacity to conduct a parallel run with a limited sample size, for example large enough to obtain sufficiently precise direct estimates for discontinuities at the national level, but not at the level of the planned domains of the regular survey. In that case small area estimation techniques (Rao and Molina, 2015) might be considered to obtain sufficiently precise discontinuity estimates at domain levels. Van den Brakel and Boonstra (2021) proposed a hierarchical Bayesian bivariate level model to combine direct domain estimates obtained from the old and new design for estimating domain discontinuities.

3.3 Re-interview designs to disentangle and adjust for mode effects

Survey methodology offers three options to make data collection strategies robust against mode effects. The first option is to reduce or minimize measurement effects by good questionnaire design, see, e.g., Dillman et al. (2014) and Schouten et al. (2022), Chapter 6. The second option is to avoid mode effects by adapting the choice of survey modes to the sample units using adaptive mixed-mode survey designs, see, e.g., Schouten, Peytchev and Wagner (2017) and Schouten et al. (2022), Chapter 11. The third option is to adjust mode effects by some form of weighting or matching afterwards, which will be explained in Subsection 3.4. Mode effects are the net result of multiple sources of non-sampling errors in the data collection phase of a survey process and can be divided in mode-specific selection effects and mode-specific measurement effects. In order to decide which of the aforementioned options need most attention and are most promising, it is crucial that mode effects are decomposed in relative selection bias and relative measurement bias.

Selection and measurement effects are typically strongly confounded when survey outcomes obtained under different modes are compared. Separation of selection effects from measurement effects in empirical studies requires carefully designed experiments in combination with weighting or regression based inference methods to control for selection effects, see, e.g., Jäckle, Roberts and Lynn (2010). As an alternative, Vannieuwenhuyze, Loosveldt and Molenberghs (2010, 2012) propose to disentangle measurement and selection effects using instrumental variables; they use a single-mode comparative survey that is assumed to be equally representative of the population and that has the same measurement effects. In Vannieuwenhuyze, Loosveldt and Molenberghs (2013) they compare this method to the generally applied backdoor method that assumes ignorable mode selection given a set of auxiliary variables. Biemer (2001) proposed a test-retest study and assumed a latent class model to separate selection bias and measurement bias in face-to-face and telephone modes. All these methods require auxiliary information that is strongly related with the target variables to separate mode-dependent selection effects from mode-dependent measurement bias. Available auxiliary information generally concerns standard socio-demographic variables, which are only weakly related to the target variables of a survey.

Schouten et al. (2013) proposed an experimental design to disentangle mode effects into mode-specific coverage effects, mode-specific nonresponse effects and mode-specific measurement effects using four modes, i.e. face-to-face interviewing (CAPI), telephone interviewing (CATI), web questionnaires and mailed-out paper questionnaires. The experiment, illustrated in Figure 2, consists of two waves. First a probability sample is drawn from the target population. In the first wave of the experiment, the sample units are randomly assigned to one of the four survey modes. The full sample, i.e. the respondents and the non-respondents of the first wave, is approached once more in the second wave using the benchmark mode, which is the face-to-face mode in this example. Based on the observations obtained in this experiment, estimates for mode-specific coverage effects, mode-specific nonresponse effects and mode-specific measurement effects are derived in Schouten et al. (2013).

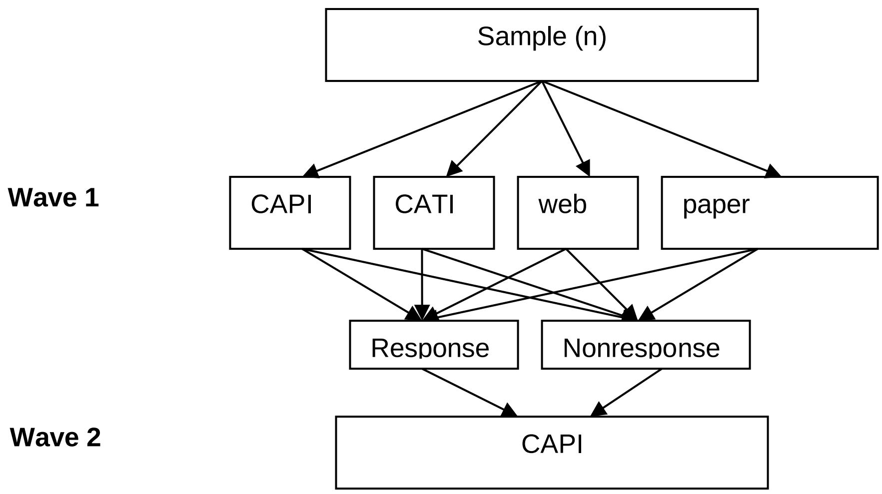

Figure 2: Repeated measurement design

Klausch et al. (2017) explored an estimation approach to adjust for measurement bias in sequential mixed-mode designs. The method is based on a re-interview approach. The idea behind this is to collect auxiliary information that is strongly correlated with benchmark variables to correct as best as possible for mode-dependent selection bias. Consider a sequential mixed mode design that starts with a data collection mode, say m_1, in the first round. In a sequential mixed mode design, the nonresponse of mode m_1 has a follow up using another mode, say m_2. To obtain strong auxiliary information for measurement bias adjustment, a random part of the response under mode m_1 is re-interviewed in the follow up using mode m_2. Estimators that adjust for mode-dependent measurement bias based on the re-interview data, are proposed by Klausch et al. (2017) and Schouten et al. (2022), Chapter 8.

3.4 Weighting methods for mixed-mode surveys

Survey sample estimates are generally based on calibration or generalized regression estimators, see, e.g., Särndal, Swensson and Wretman (1992). These techniques take unequal inclusion probabilities and selective nonresponse into account, and are aimed at correcting selection bias that could otherwise be present. Researchers and analysts working with mixed-mode data should not proceed carelessly as if analyzing single mode survey responses. When comparing survey estimates for population subgroups, observed differences could partially be the result of the fact that the groups were measured using different combinations of data collection modes. In a similar way, period-to-period changes in a repeated survey can partially be the result of different compositions of the respondents over the modes in both periods. When applying standard weighting methods to mixed-mode mode data, the resulting estimators consist of weighted combinations of observations collected through different modes. As a result, these traditional estimators contain components of differential measurement errors.

Regression, imputation and prediction techniques that attempt to remove measurement bias from the observations are developed by Suzer-Gurtekin, Heeringa and Vaillant (2012), Kolenikov and Kennedy (2014) and Park, Kim and Park (2016). Instead of correcting for mode dependent measurement bias, Buelens and Van den Brakel (2015, 2017) attempt to balance the response to neutralize adverse effects of varying response mode compositions. This is achieved by extending the calibration model with one or more components that assume a fixed distribution of the population over the data collection modes. In this way the measurement bias between domain estimates or between different survey editions is kept constant, such that differences are not affected by fluctuating measurement bias. Pfeffermann (2017) described a unified approach to handle inference from non-representative samples. Section 8.1 of that article suggests an extension from the essentially Bayesian approach to include measurement error arising in mixed-mode designs. The target of inference is the posterior probability distribution of the variable of interest, which is free from measurement bias, and corrected for selection effects.

4 The future of mixed mode surveys

The handbook Mixed-Mode Official Surveys (Schouten et al., 2022) was completed just before the COVID-19 pandemic. In its final chapter, Future of mixed-mode surveys, it is proclaiming that the face-to-face mode, while becoming more and more expensive, still is an important and stable mode in terms of coverage and response. Although only temporarily, the pandemic proved the authors wrong and pointed at the necessity of employing multiple modes. Or more specifically, at the necessity not to be dependent on one single mode. However, even with multiple modes, the impact of strong interventions such as COVID-19 is not to be underestimated. Interviewer modes, in particular face-to-face, are added because they bring in new types of respondents. Their addition does lead to shifts in statistics, and consequently, their omission to breaks in time series. Schouten et al. (2022) devote considerable attention to discussion of methodology for the estimation of such shifts accounting, if possible, for confounding with mode-specific measurement biases. The pandemic, thus, taught us that NSIs should have at least a basic understanding of the shifts in statistics that come from the inclusion of modes. This understanding must be updated on a periodic basis and ideally in such a way that it allows for some generalization across survey topics.

Surveys will undoubtedly continue to be very important tools in the production of official statistics for policy-making purposes. Surveys provide coherence and specificity in collected information and control over definitions and implementation. Given the increasing diversity in communication channels, there is also no doubt that surveys will keep using different modes to contact target populations and to administer data collection. But what will these surveys look like?

A clear trend that affects surveys in general, and mixed-mode surveys in particular, is the steady, gradual decline in response rates (Luiten, Hox and De Leeuw, 2020). It is not possible to attribute this trend to less willingness to participate in surveys alone; declining response rates have led to increasing costs per respondent which in turn have induced cost reductions through different mixes of modes. In other words, there is an interaction between the type of survey design and the response rates. Nonetheless, the prospect for future survey response rates is relatively grim. This prospect has already been forcing survey designers to tailor and adapt design features, and the survey mode is the most powerful of features. It is likely that a range of communication channels will remain to be needed in order to attract and recruit a diverse set of respondents, but in an adaptive, cost-effective fashion.

Face-to-face becomes less and less attractive from the cost perspective. To date, face-to-face often is a mode that is added in sequence to cheaper modes. This means the relatively hard cases are sent to interviewers, leading to lower interviewer response rates and thus higher costs per respondent. Furthermore, due to the unpredictable topographical variation in interviewer workloads, clustering of addresses within interviewers is complicated and face-to-face travel times may increase. These two consequences of mixed-mode surveys make face-to-face even more expensive, and may be the start of a downward spiraling of face-to-face. When face-to-face becomes more expensive then it may only be affordable for certain surveys or certain underrepresented population subgroups. This then leads to a smaller number of interviewers as workloads get smaller. Smaller numbers of interviewers lead to larger interviewer travel distances and travel costs, which in turn make face-to-face again more expensive. At Statistics Netherlands, this downward trend in face-to-face is already visible, despite being the most powerful mode still to recruit respondents.

The online mode is likely to undergo further changes as new and more powerful devices may emerge. Currently, the small screen sizes of mobile devices are still a barrier and a challenge in designing surveys and user-friendly user interfaces. It is foreseeable, however, that the drawback of small screens may be overcome by clever design, new navigation options or speech-to-text and other forms of additional communication. If so, these developments will force questionnaire designers to rethink how they construct question-answer processes. Although, universally, survey target populations have become more and more active online, there is likely to remain some digital divide that forces the same designers to align the potential of devices with more traditional modes.

Finally, it seems a natural avenue that surveys will more and more make use of existing data and of other means of collecting data besides asking questions. Existing data may be added in the estimation stage to strengthen time series. Existing data may also be added by respondents themselves as a form of data donation. Other forms of measurements, in particular data coming from mobile device sensors and other sensor systems, offer powerful options to further enrich survey data. Such surveys have been termed smart surveys. They are promising from the respondent perspective when they are not intrusive or burdensome, when respondents can access and control data and/or when respondents can learn about themselves. Nonetheless, smart surveys are likely to have a mixed-mode design as response rates to web invitation is often too low for official statistics. Consequently, such surveys require a renewed look at total error frameworks and demand for extension of existing methodology and development of new methodology. One of most crucial choices when including new types of data is between active (with explicit respondent involvement) and passive data collection (without a strong involvement of respondents). Making these choices, surveys move towards big data and big data move to surveys, creating hybrid forms of data collection in which the respondent still is a key person.

References

Bais, F., Schouten, B., Lugtig, P., Toepoel, V., Arends-Tóth, J., Douhou, S., Kieruj, N., Morren, M. and Vis, C. (2017) Can Survey Item Characteristics Relevant to Measurement Error Be Coded Reliably? A Case Study on Eleven Dutch General Population Surveys. *Sociological Methods and Research*, **48**, 1-33.

Beck F., Brilhault G., Burg T., De Vitiis C., Fekete-Nagy P., Lamei N., Mújdricza F., O'Callaghan F., O'Riordan F., Romano M. C., Sillard P., Smukavec A., Stare M. and Vereczkei Z. (2022) *Position Paper on Mixed-Mode Surveys*. Statistical Working Papers, Eurostat.

Biemer, P.P. (2001) Nonresponse Bias and Measurement Bias in a Comparison of Face-To-Face and Telephone Interviewing. *Journal of Official Statistics*, **17**, 295-320.

Biemer, P.P., Murphy, J., Zimmer, S., Berry, C., Deng, G. and Lewis, K. (2018) Using Bonus Monetary Incentives to Encourage Web Response in Mixed-Mode Household Surveys. *Journal of Survey Statistics and Methodology*, **6**, 240–261. https://doi.org/10.1093/jssam/smx015.

Blanke, K. and Luiten, A. (2014) *Query on Data Collection for Social Surveys*. ESSnet Project Data Collection for Social Surveys Using Multiple Modes. http://ec.europa.eu/eurostat/cros/system/files/Query_report_DCSS.pdf_en.

Buelens, B. and Van den Brakel, J.A. (2015) Measurement Error Calibration in Mixed-Mode Sample Surveys. *Sociological Methods & Research*, **44**, 391-426.

Buelens, B. and Van den Brakel, J.A. (2017) Comparing Two Inferential Approaches to Handling Measurement Error in Mixed-Mode Surveys. *Journal of Official Statistics*, **33**, 513-531.

Calinescu, M. and Schouten, B. (2016) Adaptive Survey Designs for Nonresponse and Measurement Error in Multi-Purpose Surveys. *Survey Research Methods*, **10**, 35-47.

Cochran, W.G. (1977) *Sampling Theory*. Wiley and Sons, New York.

Couper, M.P., Antoun, C. and Mavletova, A. (2017) Mobile Web Surveys: A Total Survey Error Perspective. In: *Total Survey Error in Practice* (Eds. (eds. P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg and B. West), John Wiley and Sons, New York, 133-154.

Dillman, D.A. (2017) The Promise and Challenges of Pushing Respondents to the Web in Mixed-Mode Surveys. *Survey Methodology*, **43**.

Dillman, D.A. and Edwards, M.L. (2016) Designing a Mixed-Mode Survey. In: *The SAGE Handbook of Survey Methodology* (eds. C. Wild, D. Joye, T.W. Smith and Y-C. Fu), Sage, London, 255-269.

Dillman, D.A., Smyth, J.D. and Christian, L.M. (2014) *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailor Design Method* (4th edition), Wiley and Sons, New York.

Dippo, C.S., Kostanich, D.L. and Polivka, A.E. (1994) Effects of Methodological Change in the Current Population Survey. In: *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 260-262.

Durbin, J. and Koopman, S.J. (2012) *Time Series Analysis by State Space Methods* (second edition). Oxford University Press, Oxford.

Fienberg, S.E. and Tanur, J.M. (1987) Experimental and Sampling Structures: Parallels Diverging and Meeting. *International Statistical Review*, **55**, 75-96.

Fienberg, S.E. and Tanur, J.M. (1988) From the Inside Out and the Outside In: Combining Experimental and Sampling Structures. *Canadian Journal of Statistics*, **16**, 135-151.

Fienberg, S.E. and Tanur, J.M. (1989) Combining Cognitive and Statistical Approaches to Survey Design. Science, **243**, 1017-1022.

Gravem, D., Meertens, V. Luiten, A., Giesen, D., Berg. N., Schouten, B. and Bakker, J. (2019) *Smartphone Fitness of ESS surveys. Case studies of the ICT Survey and the LFS*. ESSnet MIMOD Deliverable 4 Work package 5, https://www.istat.it/en/research-activity/international-research-activity/essnet-and-grants.

Groves, R.M. (1989). *Survey Errors and Survey Costs*. John Wiley & Sons.

Hartley, H.O. and Rao, J.N.K. (1978) Estimation of Nonsampling Variance Components in Sample Surveys. In: *Survey Sampling and Measurement* (ed. N.K. Namboodiri). Academic Press, New York, 35-43.

Jäckle, A., Roberts, C. and Lynn, P., (2010) Assessing the Effect of Data Collection Mode on Measurement. *International Statistical Review*, **78**, 3–20.

Kindermann, C. and Lynch, J. (1997) *Effects of the Redesign on Victimization Estimates*. Technical report, US Department of Justice, Bureau of Justice Statistics, http://www.ojp.usdoj.gov/bjs/abstract/erve.htm.

Klausch, T., Schouten, B. Buelens, B. and Van den Brakel, J.A. (2017) Adjusting Measurement Bias in Sequential Mixed-Mode Surveys Using Re-Interview Data. *Journal of Survey Statistics and Methodology*, **5**, 409-432.

Kolenikov, S. and Kennedy, C. (2014) Evaluating Three Approaches to Statistically Adjust for Mode Effects. *Journal of Survey Statistics and Methodology*, **2**, 126-158.

Lavrakas, P.L., Traugott, M.W., Kennedey, C., Holbrook, A.L., De Leeuw, E.D. and West. B.T. (2019) *Experimental Methods in Survey Research*. Wiley and Sons, New York.

Luiten, A., Hox, J. and De Leeuw, E. (2020) Survey Nonresponse Trends and Fieldwork Effort in the 21st Century: Results of an International Study across Countries and Surveys. *Journal of Official Statistics*, **36**, 469-487.

Mahalanobis, P.C. (1946) Recent Experiments in Statistical Sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, **109**, 325-370.

Medway, R.L. and Fulton, J. (2012) When More Gets You Less: A Meta-Analysis of the Effect of Concurrent Web Options on Mail Survey Response Rates, *Public Opinion Quarterly*, **76**, 733–746.

Murgia, M., LoConte, M., Coppola, L, Frattarola, D., Fratoni, A., Luiten, A. and Schouten, B. (2019) *Mixed-Mode Strategies for Social Surveys: How to Best Combine Data Collection Modes*. ESSnet MIMOD Deliverable, available at *Istat.it - Projects funded by Eurostat: Essnet and Grants*.

Murgia, M., Lo Conte, N. and Gravem. D., (2018) *Report on MIMOD Survey on the State of the Art of Mixed Mode for EU Social Surveys*. ESSnet MIMOD, WP1 - Deliverable 1, ISTAT, https://www.istat.it/en/research-activity/international-research-activity/essnet-and-grants.

Park, S., Kim, J. K. and Park, S. (2016) An Imputation Approach for Handling Mixed-Mode Surveys. *The Annals of Applied Statistics*, **10**, 1063-1085.

Patrick, M.E., Couper, M.A., Laetz, V.B., Schulenberg, J.E., O'Malley, P.M., Johnston, L.D. and Miech, R.A., (2018) A Sequential Mixed-Mode Experiment in the U.S. National Monitoring the Future Study. *Journal of Survey Statistics and Methodology*, **6**, 72–97.

Pfeffermann, D. (2017) Bayes-Based Non-Bayesian Inference on Finite Populations from Non-Representative Samples: A Unified Approach. *Calcutta Statistical Association Bulletin*, **69**, 35-63.

Rao, J.N.K. and Molina, I. (2015) *Small Area Estimation*. Wiley and Sons, New York.

Schouten, B., Blanke, K., Gravem, D., Luiten, A., Meertens, V. and Paulus, O. (2018) *Assessment of Fitness of ESS Surveys for Smartphones WP5: Challenges for Phone and Tablet Respondents within CAWI*. Deliverable 5.1 of Mixed Mode Designs for Social Surveys – MIMOD. https://www.istat.it/en/research-activity/international-research-activity/essnet-and-grants.

Schouten, B., Van den Brakel, J., Buelens, B, Giesen, D., Luiten, A. and Meertens, V. (2022) *Mixed-Mode Official Surveys. Design and Analysis*. CRC Press, Boca Raton.

Schouten, B., Van den Brakel, J.A., Buelens, B. and Klausch, T. (2013) Disentangling Mode-Specific Selection Bias and Measurement Bias in Social Surveys. *Social Science Research*, **42**, 1555-1570.

Schouten, B., Peytchev, A. and Wagner, J. (2017) *Adaptive Survey Design*. Chapman and Hall, New York.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992) *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Suzer-Gurtekin, Z.T., Heeringa, S. and Valliant, R. (2012) Investigating the Bias of Alternative Statistical Inference Methods in Sequential Mixed-Mode Surveys. *Proceedings of the JSM, Section on Survey Research Methods*, 4711-4725.

Toepoel, V., De Leeuw, E.D., Hox, J.J., Atkinson, P., Delamont, S., Cernat, A., Sakshaug, J.W. and Williams, R.A. (2020) *Single- and Mixed-mode Survey Data Collection*. SAGE publications.

Tourangeau, R. (2017) Mixing Modes. Trade-offs among Coverage, Nonresponse and Measurement Error. In: *Total Survey Error in Practice* (eds. P.P. Biemer, E.D. De Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, N.C. Tucker and B.T. West), Wiley, New York, 115-132.

Tourangeau, R. Conrad, F.G. and Couper, M.P. (2013) *The Science of Web Surveys*. Oxford University Press, New York.

Van den Brakel, J.A. (2008) Design-Based Analysis of Experiments with Applications in the Dutch Labour Force Survey. *Journal of the Royal Statistical Society, Series A*, **171**, 581-613.

Van den Brakel, J.A. (2013) Design Based Analysis of Factorial Designs Embedded in Probability Samples. *Survey Methodology*, **39**, 323-349.

Van den Brakel, J.A. and H.J. Boonstra (2021) Hierarchical Bayesian Bivariate Fay-Herriot Model for Estimating Domain Discontinuities. *Survey Methodology*, **47**, 151-189.

Van den Brakel, J.A. and Renssen, R.H. (2005) Analysis of Experiments Embedded in Complex Sampling Designs. *Survey Methodology*, **31**, 23-40.

Van den Brakel, J.A., Zhang, X. and Tam, S.M. (2020) Measuring Discontinuities in Time Series Obtained with Repeated Sample Surveys. *International Statistical Review*, **88**, 155-175.

Vannieuwenhuyze, J. and Loosveldt, G. (2013) Evaluating Relative Mode Effects in Mixed-mode Surveys: Three Methods to Disentangle Selection and Measurement Effects. *Sociological Methods and Research*, **42**, 82–104.

Vannieuwenhuyze, J., Loosveldt, G. and Molenberghs, G. (2010) A Method for Evaluating Mode Effects in Mixed-Mode Surveys. *Public Opinion Quarterly*, **74**, 1027–1045.

Vannieuwenhuyze, J., Loosveldt, G. and Molenberghs, G. (2012) A Method to Evaluate Mode Effects on the Mean and Variance of a Continuous Variable in Mixed-Mode Surveys. *International Statistical Review*, **80**, 306–322.