Current methodologies to deal with mode effects and mode bias in mixed-mode designs

Version: final (rev. 1)

Date: 20 June 2018

Authors: Bart Buelens[1], Jan A. van den Brakel [1], Barry Schouten [1]

[1] CBS (Statistics Netherlands)

Funding: European Commission, Eurostat, Directorate F: Social Statistics

Grant agreement number: 07112.2017.010-2017.786

ESSnet MIMOD – MIxed MOde Designs for social surveys

Work Package 2: Mode bias/mode effects and adjustment for mode-effects

Deliverable 1: A report containing an overview on current methodologies adopted at the ESS NSIs to deal with mode bias/mode effects in mixed-mode designs.

Summary

This report is the first deliverable of Work Package 2 (WP2) of the ESSnet project on mixed mode designs for social surveys, MIMOD. WP2 addresses mode effects, their assessment, and estimation strategies that adjust for unwanted mode effects. The report contains an overview of approaches to mode effect assessment and adjustment found in the literature. A section reporting on results from a survey held among statistical agencies in ESS countries about their mixed-mode experiences and activities is included as well.

Combining and comparing data collection modes has happened since the onset of sample surveying itself, in the first half of the twentieth century. It is the emergence of the internet and the possibility and appeal of web interviewing that sparked a renewed interest in mixed-mode designs and associated mode effects. In current mixed-mode designs, web interviewing is almost always one of the data collection modes. The literature review in this report concentrates on articles published since 2005. Many present-day articles on mode effects and adjustments cite some seminal articles from that year, including de Leeuw (2005), Voogt and Saris (2005), Dillman and Christian (2005), and Fricker et al. (2005).

Mode assessment studies are sometimes limited to quantifying the total mode effect, but are more insightful when they separate the total effect into selection and measurement components. Selection effects are caused by the selection mechanism of a mixed-mode survey design which results in the partitioning of the sample into respondents and non-respondents. Selection effects are a combination of coverage and non-response effects. Measurement effects are caused by specifics of the modes employed in the survey and affect the recorded responses to the survey questions. They arise from the same respondent potentially giving different answers to the same questions in different modes. Experimental designs specifically aimed at separating mode effects into selection and measurement effects are preferable, but costly, and hence less common. Such designs include parallel, independent surveys, embedded experiments, and re-interview studies. Some authors report methods to separate mode effects in observational studies, usually relying on socio-demographic covariates that explain the selection mechanism. Using such covariates, approaches like reweighting, regression, or sample matching have been applied.

Adjustment methods are not as commonly encountered in the literature as assessment methods. Adjustment techniques are aimed at correcting survey estimates for bias induced by one or several modes, or by the specific combination of several modes. Adjustment for bias

requires the presence of a definition – or choice – of reference mode or design that serves as a benchmark, since bias of some design is only meaningful with respect to some other design. Adjustment techniques that have appeared in the literature include reweighting and calibration approaches, imputation, and prediction approaches. The latter are generally considered in the potential-outcomes framework, predicting so-called counterfactuals: answers that respondents would have given had they responded through some mode other than the one they actually responded through. Reliable adjustment methods require mode effects to be separable into selection and measurement effects, which is most successful in experimental designs. One approach to achieve this is to have an embedded experiment within a mixed-mode survey design. Klausch et al. (2017) propose such a design based on re-interviews, which will be further explored in the context of WP2, with a report to follow in deliverable 2.

The MIMOD survey held among statistical agencies in ESS countries confirms the picture arising from the literature review, that mode effect assessments are more widespread than mode effect adjustments. While about two thirds of countries report to have undertaken activities related to mode effect assessments, only one third says to have taken measures to adjust for mode effects. Somewhat remarkably, only just less than half of the countries report to have plans for future mode effect assessment or adjustment studies.

Table of contents

Summary	2
Table of contents	4
1. Introduction	5
2. Mixed-mode designs	7
3. Mode effects	8
4. Assessment of mode effects	10
5. Adjustment methods	14
6. MIMOD survey results	17
7. Conclusions	20
References	23

1. Introduction

This report is the first deliverable of Work Package 2 (WP2) of the ESSnet MIMOD. WP2 addresses mode effects in mixed-mode survey designs. The results of this work package are expected to provide all countries in the ESS with an updated overview about methodological solutions and strategies to improve the quality of estimates produced in mixed-mode surveys.

Well-designed mixed-mode surveys may reduce costs and non-sampling errors (coverage, nonresponse, and measurement errors). However, possible mode selection effects (resulting from errors of nonobservation), and mode measurement effects (resulting from observation errors) can affect the survey results due to the use of different data collection modes. Mode effects need to be properly assessed and adjusted in order to ensure accurate estimates.

This report provides an overview of methods for this purpose. It contains a review of recent literature on methodologies to assess or to adjust for mode effects, highlighting solutions to common problems in the area of mixed-mode data collection.

Comparisons between surveys conducted using different modes are available in the literature almost from the time when sample surveying became common practice. It seems to be the emergence of web technology that has instigated renewed interest in research into the effects of using different modes of data collection. The year 2005 appears to mark the onset of this latest wave of interest, with particular attention to the combined use of multiple modes in the same survey. In that year, some often-cited articles were published. De Leeuw (2005) lists advantages and pitfalls of mixing modes. Voogt and Saris (2005) discuss the trade-off between improved selection and possibly hampered measurements in mixed-mode surveys. Dillman and Christian (2005) recognize the issue of differential measurement effects between modes and suggest preventing this issue through the design of questionnaires that prevent this phenomenon from occurring. Fricker et al. (2005) conducted an experiment to compare web and telephone surveys.

Since then, many articles have been published. The present report does not aim to provide an exhaustive inventory of these publications. Rather, it aims at reporting on significant contributions, and at providing an overview of the field by providing examples from the literature. For each example chosen in this report, there are sometimes other similar examples that were not chosen but that are equally relevant.

The reader is expected to be familiar with basic concepts of survey sampling in general and mixed-mode designs in particular. Section 2 briefly introduces the concept of mixed-mode survey designs, and section 3 defines the concept of mode effects as understood in the present report. An overview of studies assessing mode effects is given in section 4. Literature on approaches to adjust survey estimates for mode effects are presented in section 5. Results from a survey held among statistical institutes in ESS countries are summarised in section 6. Conclusions are drawn in section 7.

2. Mixed-mode designs

In this report we consider survey designs in which more than one mode of data collection is used. Ordinarily, data collection modes include telephone, face-to-face, postal and internet interviewing. More generally, a data collection mode is a communication medium. When different strategies are employed all using the same data collection mode, most of the methods discussed in this report could be of use as well, although they do not have our primary attention. Different strategies could be to use fewer reminders in some population subgroups than in others, to work with incentives in some population subgroups, or to vary contact times between subgroups in telephone surveys. If the same communication medium is used in those settings, they are not considered as mixed-mode surveys in our context. Rather, we would consider these to be adaptive survey designs (Schouten et al., 2017).

Mixed-mode designs can employ multiple data collection modes in different ways. A first classification of mixed-mode designs can be made regarding the choice of modes: does the agency conducting the survey assigns sample units to mode groups, or can the sample units choose the mode through which they respond to the survey? A second classification can be made based on a distinction between designs in which each respondent can only respond through a single mode (assigned or chosen), and designs in which different modes are offered to the same respondents. Mixed-mode designs in which multiple modes are used simultaneously are known as concurrent designs. In contrast, sequential designs use one mode first and then re-approach non-respondents using a different mode; combinations with more than two modes are possible.

All mixed-mode surveys, regardless of their precise design, result in a bipartition of the sample into respondents and non-respondents. The respondents have provided answers to the survey questions, and not all of them did so through the same data collection mode. This phenomenon can give rise to mode effects, discussed in the next section.

3. Mode effects

The term mode effect is used differently in different contexts, and in its most general form refers to effects that are due to the use of one mode compared to another, or a combination of modes to a single mode, or to a different combination of the same or other modes.

Effects of this kind manifest themselves in the survey outcomes, typically estimates of population means and totals. Mode effects are related to bias and variance of the estimators of the survey variables.

In principle, an effect such as bias could be defined with respect to the true, unobserved quantity. This approach, however, has no practical use since the unobserved quantity remains unknown at all times. Once it is observed, mode effects come into play. Therefore mode effects are usually evaluated relative to some benchmark mode, which is sometimes regarded as the gold standard, but it does not need to be; it could just as well be the data collection mode that has always been used in the past, for example.

In the present report two kinds of mode effects are distinguished. First, selection effects are caused by the selection mechanism of a mixed-mode survey design which results in the partitioning of the sample into respondents and non-respondents. Selection effects are a combination of coverage and non-response effects. Second, measurement effects are caused by specifics of the modes employed in the survey and affect the recorded responses to the survey questions. They arise from the same respondent potentially giving different answers to the same questions in different modes. Sometimes measurement effects are referred to as measurement bias, or as pure mode effects.

Often, only a joint mode effect can be observed, which is the combined effect of selection and measurement effects. Unless in experimental designs, selection and measurement effects are generally confounded and are difficult to separate.

In an earlier ESSnet project, on Data Collection for Social Surveys, Körner (2014) produced a report on the definition, identification and analysis of mode effects.

The present report takes a somewhat different approach on the topic. While the definitions given by Körner (2014) largely correspond to those in the present report, the identification of mode effects by Körner (2014) is not an integral part of the present report. Identification of mode effects refers to describing and explaining why different modes may exhibit relative selection and measurement effects, the latter being the primary focus of chapter 3 in Körner

(2014). For example, reasons for mode effects can be found in the different stages of the cognitive process a respondent goes through when confronted with a questionnaire (Tourangeau et al., 2000), and how data collection modes affect the cognitive processes in these stages. The presence or absence of an interviewer, the speed of the interview, computer literacy, the perceived confidentiality, and the type of question are all elements that can cause mode effects. Recent work by Kim et al. (2018) studies straightlining answering behaviour in different modes and the measurement effects this can induce in survey estimates. Many reasons can be conjectured or shown to cause mode effects. In this report we do not elaborate further on these reasons. Here, focus is on the assessment of mode effects (section 4), and on adjustment methods (section 5).

An important point the present authors wish to stress is that mode effects are not necessarily bad. Mode effects, when present, can either improve or worsen the quality of survey estimates. An obvious improvement that could be had from mode effects is a less selective sample of respondents in a mixed-mode survey compared to a single-mode survey. In this case a selection effect may be present, which manifests itself as a difference in survey estimates. Researchers can study the representativity and may come to the conclusion that the mixed-mode survey is to be preferred, and that the mode effect introduced is an improvement compared to the former survey design, the single-mode survey. Generally, mode dependent selection effects indicate a difference in representativity of the response collected through a mixed-mode design and a benchmark design. If the difference is such that the mixed-mode response is less selective, the selection effect corresponds to an improvement in survey estimates.

Measurement effects in mixed-mode designs are generally not desirable. Such effects typically arise when different modes have different associated biasing effects: they do not measure the target quantity at the same level, or with the same precision. Since mixed-mode designs produce responses using a combination of modes, the individual responses may become incomparable, as they are not all measured using the same measurement instrument (data collection mode in this setting).

4. Assessment of mode effects

Assessment of mode effects is carried out by studies into effects of using one or several data collection modes in comparison with some reference or benchmark design, characterised by the use of another or several other modes. Jäckle et al. (2010) identify issues in assessing mode effects, outlining that confounding of selection and measurement is a key problem. They stress that an assessment of mode effects should not only consider their presence, but also their direction, size, and significance.

A key distinction in mode assessment studies is whether studies employ experimental designs or non-experimental designs. Experimental designs include, among others, embedded experiments, split sample designs and repeated measurement designs. Non-experimental designs are observational studies and are generally based on mixed-mode surveys that are conducted not with the primary aim of mode assessments. In such settings, weighting or regression-based inference methods to control for selection effects can be applied, see, for example, Jäckle et al. (2010).

Some assessment methods extend naturally to adjustment techniques, hence, adjustment methods usually incorporate – implicitly or explicitly – an assessment of mode effects. Mode adjustment methods are discussed in the next section. The present section focusses on research in which the assessment of mode effects dominates.

Relatively early work was done by Biemer (2001), who applied an interview re-interview approach analysed with a latent class model to disentangle selection and measurement bias in face-to-face and telephone data collection modes.

Vannieuwenhuyze et al. (2010) and Vannieuwenhuyze and Loosveldt (2013) outline some analyses that could be conducted comparing a mixed-mode with a single-mode design with the purpose of unravelling confounded selection and measurement effects due to survey mode. They state that successful separation of selection and measurement effects is only possible under strong assumptions, or when specific data are available, such as observed variables that are insensitive to the survey mode.

Buelens and van den Brakel (2010) compare different versions of a sequential mixed-mode design in which the last stage consists of face-to-face interviewing. Simply by omitting the face-to-face respondents from the analysis they study the effect of including this data collection mode in the mix of modes applied in their survey. The found effect is an overall mode effect, not disentangled into selection and measurement effects.

Experiments embedded in probability samples are useful to estimate relative differences between data collection modes. The sample design of the survey provides a framework to design efficient randomized experiments. For details see Fienberg and Tanur (1987, 1988), Van den Brakel and Renssen (2005), and Van den Brakel (2008). Van den Brakel and Renssen (2005), Van den Brakel (2008, 2013) developed design-based inference procedures for the analysis of embedded experiments that account for the sample design as well as the superimposition of the applied experimental design on the sampling design. Examples of experiments aimed to assess differences between data collection modes are included in Van den Brakel (2008). In such experiments, a probability sample drawn from a finite target population is divided randomly into two or more subsamples, each of which is assigned to a treatment, in this case a data collection mode. Hypotheses about differences between estimated population means and totals can be tested using Wald or *t*-statistics.

Lugtig et al. (2011) propose to use a propensity score matching approach where respondents from one mode are matched to respondents from a different mode. The difference in survey estimates on the matched samples is taken to be the measurement effect. This is based on the assumption that the covariates in the propensity score models explain the selection fully. The authors find measurement effects between telephone and web modes.

Capacci et al. (2018) apply a similar propensity score matching technique in which respondents from independent surveys are matched, and then compared. The surveys under consideration use different survey modes, hence, differences observed after matching are said to be measurement effects. This finding is conditional on the same assumptions as Lugtig et al. (2011).

Schouten et al. (2013) conducted a large scale experiment consisting of two waves. The first wave was a split sample design in which sample units were randomly assigned to one of four modes. The second wave was a follow-up, repeated measurement in a single mode. This design allowed the decomposition of mode effects into selection effects and measurement effects. This study was based on a crime victimisation survey and found important effects in key survey variables.

Based on the same experiment, Klausch et al. (2015) considered the split samples and their follow-ups in a disjoint fashion and compared selection effects of the resulting mixed-mode designs. Total bias, and bias resulting from selection and measurement in such an approach is discussed in Klausch et al. (2017). As in Schouten et al. (2013), the key to success in these analyses is the experimental design allowing for fully explaining selection effects.

Wagner et al. (2014) conducted an experiment in which the sequence of modes in a mixed-mode survey was varied. They focussed on response rates, possibly indicative of selection effects. They did not include any substantial analysis of measurement effects.

Vandenplas et al. (2016) investigate whether mode preference could be a helpful covariate to explain selection effects in mixed-mode designs. This research is based on the conjecture that sample units have higher response probabilities when approached in their favourite mode, and that they give better answers in that case too – better: in the sense of answers that are closer to the truth than answers they would have given in another mode. Typically, known socio-demographic variables are used to explain selection effects. Adding a question about mode preference to the questionnaire delivers an additional covariate that could be used when separating selection from measurement effects. The results are mixed, insofar that participation in web is better explained, but at the same time the decomposition of the mode effects into the different components gives counterintuitive results.

Roberts and Vandenplas (2017) discuss an experiment in which components of mean square error are obtained due to selection and measurement effects in mixed-mode designs. They conclude that mixing modes reduces bias in general, but that the relative contribution to the total survey error from different sources varies with the survey design. In addition, these general results vary with the type of variable that is measured as well.

The fact that mode effects may be different for different variables in a mixed-mode survey is also investigated by Klausch et al. (2013), who study measurement effects of attitudinal rating scale questions in a mixed-mode experiment. In such experimental settings, selection effects can be conditioned on, and analysis of measurement effects is possible. Important measurement differences seem to exist between interviewer and non-interviewer modes.

In contrast, Sarracino et al. (2017) find in their study that measurements of subjective variables do not suffer from significant measurement differences in web compared to telephone interviews.

Cernat (2015) studies mode effects in a longitudinal study, using a quasi-experimental design consisting of random allocation to mixed-mode or single mode, within an existing panel study. The results are analysed using models that take advantage of the longitudinal character of the data, such as quasi-simplex models, which are a type of structural equation models where survey responses to earlier waves are included in the models.

Reliability was studied, and no differences were found. In a follow-up paper, Cernat et al. (2016) study mode effects in the same quasi-experimental design using a panel where respondents were allocated randomly to one of three modes. They use latent class models to study mode effects and find effects predominantly between interviewer administered modes (face-to-face and telephone) and the non-interviewer mode (web).

The common theme and challenge in these articles is the decomposition of the total mode effect into contributions originating from selection and from measurement. Experimental designs allow controlling for selection effects, and hence the unbiased assessment of measurement differences between modes. Observational studies require covariates that explain the selection mechanisms. If available, differences between mode groups are attributed to measurement differences, conditional on the covariates. Validating this assumption can be achieved when variables are available that are observed without error, potentially available from data sources other than the survey.

When the available covariates do not fully explain the selection mechanism, the decomposition of the total mode effect into selection and measurement effects may be incorrect. Since both effects could have either positive or negative signs independently of each other, it is not necessarily the case that an underestimation of the selection effect corresponds to an overestimation of the measurement effect; both effects could be equal in absolute value with opposite signs, in which the total mode effect would be zero. If both effects work in the same direction, though, an underestimation of the magnitude of the selection effect corresponds to an overestimation of the magnitude of the measurement effect.

5. Adjustment methods

Mode adjustment methods are methods that adjust survey estimates obtained from mixed-mode designs to correct for mode effects induced by the use of multiple modes of data collection.

In survey sample research, adjustment for selection effects due to coverage and nonresponse problems is typically conducted. Methods commonly used for this purpose include weighting, calibration and regression methods (Bethlehem et al., 2011). If mixed-mode designs result in adverse selections of respondents, these common methods can be applied in the same way as they are used to correct for selection effects due to coverage or nonresponse issues. In this respect, mixed-mode designs are not unlike single mode designs in which selection effects are corrected for in order to remove or reduce bias in survey estimates.

The adjustment methods discussed in this section are aimed at handling measurement effects, possibly in combination with the familiar selection effect adjustment methods. Common estimation methods in single mode survey sampling do not handle measurement issues. These adjustment methods are specific to mixed-mode designs, which is the likely reason why they are not studied very extensively yet. This explains at the same time why literature on adjustment methods is still somewhat limited.

Adjustment to survey estimates in mixed-mode surveys are warranted and desirable when the point or variance estimates are biased compared to estimates from some benchmark design. Of course, it is assumed that the adjusted estimates are better – in mean square error sense – than the unadjusted survey estimates.

Suzer-Gurtekin et al. (2012) presented some early results on estimation methods in the context of mixed-mode designs, expanding upon this work in her PhD thesis (Suzer-Gurtekin, 2013). In this work, mixed-mode measurements are regarded as treatments in a causal modelling framework of counterfactuals (Rubin, 2005), with potential outcomes defined as answers that would be given to survey questions through a mode that was not actually used for the respondent. Potential outcomes are obtained through regression modelling. Overall survey estimators of means and totals are proposed to be combinations of real answers and of potential outcomes. Uncertainty resulting from models to predict the counterfactuals adds to the total variance. Depending on the choice of benchmark, different mixes of counterfactuals can be produced; it is suggested to seek a mix that minimises the mean square error.

Kolenikov and Kennedy (2014) compare the regression modelling approach with multiple imputation of non-observed answers, framing the problem rather as an imputation and missing data problem. In addition they studied a third approach, an imputation technique based on an econometric framework of implied utilities in logistic regression modelling. The multiple imputation method came out as the preferential technique.

Park et al. (2016) too propose an imputation approach to impute unobserved observations with counterfactuals. They propose to use fractional imputation and obtain variance estimators using Taylor linearization. They present a limited real-world application in addition to a simulation study.

A recent application is discussed by Fessler et al. (2018) where they extend the potential outcomes approach to distributional characteristics other than means and totals, to estimate measures of income inequality in Austria.

Another approach to mode adjustment is reweighting of the survey response. Buelens and van den Brakel (2015) address a situation where the composition of the survey response varies between population subgroups such as regions or age classes, or between editions of a survey in the case of regularly repeated surveys. Such variations hamper the comparability of survey estimates as the measurement effect in subgroups or editions is not constant due to the variability in the mode compositions. Their solution is to apply a calibration correction by reweighting the survey response to fixed mode distributions. This method can be applied to non-experimental data assumed that there are no confounding variables that are not accounted for.

Buelens and van den Brakel (2017) compare their mode calibration method with the potential outcomes approach and regression modelling (Suzer-Gurtekin, 2013). They discuss parallels and differences of the two methods and give circumstances in which both methods are equivalent. They provide an example from the Labour Force Survey in the Netherlands and find that in this specific case no adjustments due to imbalances in mode distributions are required.

Vannieuwenhuyze et al. (2014) propose covariate adjustments to correct for mode effects. While such methods are common to correct for selection effects, they propose to apply these methods to correct for measurement effects. Covariates must then be chosen not so that they explain selection differences between modes, but rather so that they explain measurement differences between modes. Which covariates can be used for this purpose remains an ad-hoc choice.

Mariano and Elliot (2017) propose a Bayesian hierarchical model from the field of Item Response Theory (IRT). Applying such model in the context of mixed-mode surveys assumes the existence of latent traits that are measured through different data collection modes. They successfully applied this method to a randomized survey experiment.

Pfeffermann (2017) described a unified approach to handle inference from non-representative samples. Section 8.1 of this 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. This discussion is largely conceptual and has not been applied or simulated.

Klausch et al. (2017) propose an experimental design attached to an observational study, in which some respondents of one mode are re-interviewed in another mode. This allows for estimation of the measurement effect, and consequently, adjustment of the survey estimates to a benchmark mode. They compare the performance of different estimators and conclude that an inverse regression estimator based on a linear measurement error performs best. This is conditional on the error model being the true model. This design can be attractive for practitioners as the experimental component does not interact with the regular survey procedures.

6. MIMOD survey results

In the context of this MIMOD ESSnet project a survey was held among statistical offices in ESSnet countries. The survey contained questions on data collection strategies, questionnaire design, use of smartphones and tablets, methods to deal with mode effects, and case management systems. Here, we report on answers received to the questions in the section on methods to deal with mode effects. It is anticipated that a wider appraisal of the survey results will appear elsewhere at a later stage.

Responses to the MIMOD survey were received from all 31 countries in the survey: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Norway, Poland, Portugal, Romania, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, The Netherlands, United Kingdom.

Table 1 summarizes reported activities undertaken by the agencies to assess mode effects in social surveys. Each country could report multiple activities. One third of the countries did not conduct any assessments of mode effects in their social surveys, as can be seen from the last row of the table. Out of the activities undertaken by the agencies who did conduct some assessments, pre-testing or experiments with questionnaire designs are most common. Other often conducted assessments include pre-testing or experiments with sensitive or core questions, conducting pilot surveys, comparing distributions in socio-demographic or target variables, comparing various quality indicators, and parallel runs of different data collection strategies.

It cannot be seen in this table how many activities were undertaken in a combined fashion by individual agencies. Analysis of the results learns that most agencies who report at least some activity did actually undertake several activities. Countries reporting only a single activity are exceptions.

Table 2 lists measures taken by agencies to adjust for mode effects. Reporting multiple measures was allowed. From the last row it is seen that two thirds of the agencies have not taken any measures so far. The minority of countries that have taken measures did so predominantly by applying weighting corrections. Only a few countries applied calibration or correction adjustments. One country reported having applied another method, but did not provide details.

Finally, 14 of the 31 countries report to have future plans for research into mode effect assessment and/or adjustment methods. Most of these plans focus on assessment and to a lesser extent on adjustment. The plans for assessments are often quite rigorous in that they involve pilot studies, experimental designs or parallel execution of different strategies. Some agencies anticipate the need for mode effect adjustments, but none report to have plans for research into adjustment strategies specifically. The plans involve mostly empirical and applied research.

Table 1. Activities undertaken by 31 ESS countries to assess mode effects in mixed-mode designs. Each country could report multiple activities.

Activity undertaken	**Percentage of countries**
Pre-tests, experiments on questionnaire design	48 %
Pilot surveys	42 %
Differences in distributions of socio-demographic or target variables	39 %
Pre-tests, experiments on sensitive or core questions	35 %
Differences in quality indicators (e.g. total or item non response rates, break-off rates, reliability indicator, failure rates of consistency rules, …)	35 %
Previous and new data collection strategies running simultaneously (independent sampling)	32 %
Separating selection, nonresponse and measurement effects	26 %
Calculation of representativeness indicators of various designs	23 %
Pre-tests, experiments on split sample approach	19 %
Subsampling of groups receiving different data collection strategies (e.g. control group)	19 %
Pre-tests, experiments on the use of different devices (smartphones, tablets, …)	19 %
Re-interview studies	6 %
Other types of pre-tests and/or experiments	3 %
Other activities	6 %
No activity conducted in recent years	*32 %*

Table 2. Measures taken by 31 ESS countries to adjust for mode effects in mixed-mode designs. Each country could report multiple measures.

Measure taken	**Percentage of countries**
Weight adjustments	29 %
Calibration to fixed mode distributions	13 %
Estimate measurement errors and correct responses to a benchmark mode	10 %
Other	3 %
No measure taken	*61 %*

7. Conclusions

The ESS country experiences reported in the MIMOD survey reflect our findings in the literature reviews on methods for mode effect assessment and adjustment. Both reported activities and published literature on mode effect assessments are more widespread than on mode effect adjustment techniques. Sometimes assessment of mode effects may be sufficient, but when detected, some effects may need to be corrected for, in particular measurement effects.

While in our view a distinction between selection effects on the one hand and measurement effects on the other is essential to make, this is not always done in the literature on mode effect assessments. An important reason is that it is difficult to separate selection from measurement effects, but easy to assess their combined effect. The main difficulty is the confounding of selection and measurement effects in observational studies. The two effects can be separated in experimental studies, but these are rather rare because of the associated costs.

Appropriate adjustment methods, however, require the separation of selection and measurement effects in order to correct each, potentially by different types of approaches. Adjustment methods are aimed at correcting survey estimates for undesired mode effects, typically bias resulting from measurement effects. Literature on adjustment methods is scarce. Reweighting approaches seek to correct through applying adjustments to the usual survey weights, whereas prediction methods attempt to predict so-called potential outcomes: answers that respondents would have given had the survey been conducted in another mode. When this is done at the item level, imputation techniques can be used.

Assessments as well as adjustments are most sensibly conducted in a comparative manner, by comparing a mixed-mode design with a single mode design, or with another multimode design. In assessment studies, the representativity of the response, the response rate, and distributional socio-demographic characteristics of the respondents can be studied to gain insight into the selection mechanism of a mixed-mode design. Generally it is of course desirable that the response collected through a mixed-mode design is better in some way: less selective and/or higher than for example through a single-mode design. In this sense, selection effects are desirable and could reduce selection bias of survey estimates. Adjustments for selection effects in mixed-mode designs are no different from adjustments in single-mode designs, and are generally needed because of selective coverage and nonresponse.

The adjustments we refer to in the context of mixed-mode designs are necessitated by the potential occurrence of measurement effects. Measurement effects arise when respondents give different answers to the same questions in different modes. As a result, comparability of population subgroups who responded through different data collection modes may be compromised. Assessment of measurement effects may show that there are systematic differences between measurements obtained through one mode compared to a different mode. When applying adjustments, the researcher must choose a reference design as the benchmark, since true measurement errors with reference to some unknown underlying construct are impossible to recover. The benchmark design can consist of a single data collection mode, or of a mix of several modes where the proportion of each mode in the mix is fixed at a specific level. Measurements that deviate from the benchmark design are said to suffer from measurement effects and are in need of adjustments to remove the bias with respect to the benchmark. Adjustments can be applied in several ways. One approach is through reweighting to enforce the mixing proportions of the mixed-mode design to those of the benchmark design. This method is not applicable if the benchmark design is a single-mode survey. In such cases one could use an imputation approach where counterfactuals are imputed: predictions of measurements that would have been obtained had the data been collected through a different mode. Alternatively, systematic measurement differences between two modes could be estimated at aggregated levels, and subsequently used in an additive correction, for example.

Assessment and adjustment strategies are most reliable and hinge less on assumptions when conducted in experimental settings. In such cases selection and measurement effects can be separated, which is important specifically in adjustment approaches. Separation of selection from measurement effects generally proceeds by explaining the selection using some covariates, and attributing remaining differences to measurement. Hence, when separating the effects is not completely successful, selection effects are not fully explained, and as a result estimated measurement effects are biased.

Since separating selection from measurement effects are a prerequisite for successful mode effect assessments and adjustments in mixed-mode designs, a promising line of future research is the development of mixed-mode designs that allow for this, for example through embedded experiments. An example of such a design consists of conducting re-interviews through a second mode for a subset of respondents who already responded through a first mode (Klausch et al. 2017). Within WP2 of this ESSnet project, this design will be studied further: a cost-benefit analysis will be conducted. Alternative designs that allow for separating measurement and selection effects, and for which suitable mode adjustment estimators can be defined, are expected to appear and would deliver a very valuable contribution to the practical usability and theoretical validity of mixed-mode sample surveys.

References

Bethlehem, J., Cobben, F., & Schouten, B. (2011). Handbook of nonresponse in household surveys (Vol. 568). John Wiley & Sons.

Biemer, P. (2001). Nonresponse bias and measurement bias in a comparison of face to face and telephone interviewing. Journal of Official Statistics 17:295-320.

Buelens, B., & van den Brakel, J. (2010). On the necessity to include personal interviewing in mixed-mode surveys. Survey Practice, 3(5).

Buelens, B., & van den Brakel, J. A. (2015). Measurement error calibration in mixed-mode sample surveys. Sociological Methods & Research, 44(3), 391-426.

Buelens, B., and J. van den Brakel (2017). "Comparing two inferential approaches to handling measurement error in mixed-mode surveys," Journal of Official Statistics, 33(2), 513-531.

Capacci, S., Mazzocchi, M., & Brasini, S. (2018). Estimation of unobservable selection effects in on-line surveys through propensity score matching: An application to public acceptance of healthy eating policies. PloS one, 13(4), e0196020.

Cernat, A. (2015). The impact of mixing modes on reliability in longitudinal studies. Sociological Methods & Research, 44(3), 427-457.

Cernat, A., Couper, M. P., & 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.

De Leeuw, D. (2005). To mix or not to mix data collection modes in surveys. Journal of official statistics, 21(2), 233.

Dillman, D. A., & Christian, L. M. (2005). Survey mode as a source of instability in responses across surveys. Field methods, 17(1), 30-52.

Fessler, P., Kasy, M., & Lindner, P. (2018). Survey mode effects on measured income inequality. The Journal of Economic Inequality, 1-19.

Fienberg, S.E., and J.M. Tanur (1987). Experimental and sampling structures: parallels diverging and meeting. International Statistical Review, 55, 75-96.

Fienberg, S.E., and J.M. Tanur (1988). From the inside out and the outside in: combining experimental and sampling structures. Canadian Journal of Statistics, 16, 135-151.

Fricker, S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An experimental comparison of web and telephone surveys. Public Opinion Quarterly, 69(3), 370-392.

Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the effect of data collection mode on measurement. International Statistical Review, 78(1), 3-20.

Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D. P. (2018). Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail–Web Mixed-Mode Surveys. Social Science Computer Review, 0894439317752406.

Klausch, T., Hox, J. J., & Schouten, B. (2013). Measurement effects of survey mode on the equivalence of attitudinal rating scale questions. Sociological Methods & Research, 42(3), 227-263.

Klausch, T., J. Hox, and B. Schouten (2015), "Selection Error in Single- and Mixed Mode Surveys of the Dutch General Population," Journal of the Royal Statistical Society: Series A (Statistics in Society), 178(4), 945–961.

Klausch, T., Schouten, B., Buelens, B., & Van Den Brakel, J. (2017). Adjusting measurement bias in sequential mixed-mode surveys using re-interview data. Journal of Survey Statistics and Methodology, 5(4), 409-432.

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

Kolenikov, S., & Kennedy, C. (2014). Evaluating three approaches to statistically adjust for mode effects. Journal of survey statistics and methodology, 2(2), 126-158.

Körner, T. (2014). Report on the definition, identification and analysis of mode effects. Deliverable for work package III of the ESSnet on Data Collection for Social Surveys using Multiple Modes. Discussion paper, Statistics Germany.

Lugtig, P., Lensvelt-Mulders, G. J., Frerichs, R., & Greven, A. (2011). Estimating nonresponse bias and mode effects in a mixed-mode survey. International Journal of Market Research, 53(5), 669-686.

Mariano, L. T., & Elliott, M. N. (2017). An Item Response Theory Approach to Estimating Survey Mode Effects: Analysis of Data from a Randomized Mode Experiment. Journal of Survey Statistics and Methodology, 5(2), 233-253.

Park, S., Kim, J. K., & Park, S. (2016). An imputation approach for handling mixed-mode surveys. The Annals of Applied Statistics, 10(2), 1063-1085.

Pfeffermann, D. (2017). Bayes-based Non-Bayesian Inference on Finite Populations from Non-representative Samples: A Unified Approach. Calcutta Statistical Association Bulletin, 69(1), 35-63.

Roberts, C., & Vandenplas, C. (2017). Estimating Components of Mean Squared Error to Evaluate the Benefits of Mixing Data Collection Modes. Journal of Official Statistics, 33(2), 303-334.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association, 100(469), 322-331.

Sarracino, F., Riillo, C. F. A., & Mikucka, M. (2017). Comparability of web and telephone survey modes for the measurement of subjective well-being. In Survey Research Methods (Vol. 11, No. 2, pp. 141-169).

Schouten, B., van den Brakel, J., Buelens, B., van der Laan, J., & Klausch, T. (2013). Disentangling mode-specific selection and measurement bias in social surveys. Social Science Research, 42(6), 1555-1570.

Schouten, B., Peytchev, A., Wagner, J. (2017). Adaptive survey design. Chapman and Hall/CRC.

Suzer-Gurtekin, Z. T., Heeringa, S., & Vaillant, R. (2012). Investigating the Bias of Alternative Statistical Inference Methods in Sequential Mixed-Mode Surveys. Proceedings of the JSM, Section on Survey Research Methods, 4711-2.

Suzer-Gurtekin, Z.T. (2013). Investigating the Bias Properties of Alternative Statistical Inference Methods in Mixed-Mode Surveys. Ph.D. thesis, University of Michigan.

Tourangeau, R., L. Rips, and K. Rasinski (2000): The psychology of survey response. Cambridge: Cambridge University Press.

Tourangeau, R. (2017). Mixing Modes. In: Biemer, P. P., de Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L. E., ... & West, B. T. (Eds.), Total survey error in practice. John Wiley & Sons, pp. 115-132.

Van den Brakel, J.A. (2008). Design-based analysis of embedded experiments with applications in the Dutch Labour Force Survey. Journal of the Royal Statistical Society, Series A, 171, pp. 581-613.

Van den Brakel, J.A. (2013). Design based analysis of factorial designs embedded in probability samples. Survey Methodology, 39, pp. 323-349.

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

Vandenplas, C., Loosveldt, G., & Vannieuwenhuyze, J. T. (2016). Assessing the use of mode preference as a covariate for the estimation of measurement effects between modes: a sequential mixed mode experiment. Methods, data, analyses: a journal for quantitative methods and survey methodology, 10(2), 119-142.

Vannieuwenhuyze, J., G. Loosveldt, and G. Molenberghs. 2010. A method for evaluating mode effects in mixed-mode surveys. Public Opinion Quarterly 74:1027–1045.

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

Vannieuwenhuyze, J.T., Loosveldt, G., Molenberghs, G. (2014) Evaluating mode effects in mixed-mode survey data using covariate adjustment models. J. Off. Stat. 30(1), 1–21

Voogt, R. J., & Saris, W. E. (2005). Mixed mode designs: Finding the balance between nonresponse bias and mode effects. Journal of official statistics, 21(3), 367.

Wagner, J., Arrieta, J., Guyer, H., & Ofstedal, M. B. (2014). Does sequence matter in multimode surveys: Results from an experiment. Field methods, 26(2), 141-155.