**Dealing with mode effects**

Orietta Luzi, Istat, [luzi@istat.it](mailto:luzi@istat.it)

Bart Buelens, CBS, [b.buelens@cbs.nl](mailto:b.buelens@cbs.nl)

**Abstract**

*Workpackage 2 of the European project MIMOD (MIxed MOde Designs for social surveys) is devoted to the evaluation of methods for dealing with mode effects in surveys using mixed-mode data collection. 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. The paper reports some outcomes of the workpackage activities in the first half of 2018, including some results of the MIMOD query and main evidences from literature update. The overall strategy which will be adopted in order to assess mode effect estimation (and adjustment) in selected ESS surveys will be also delineated.*

**Keywords:** Mixed-mode, mode effects, social surveys

**1. Introduction**

As known, the use of different modes of data collection in a same survey produces advantages in terms of increasing response rates and coverage of target populations, and reduction of surveys costs and respondents burden. However, mixed-mode strategies originate the so-called *mode effects* (selection and measurement effects), which may highly affect estimates accuracy. Furthermore, selection effect and measurement differences across modes are usually confounded. For an overview of issues on mixed-mode surveys and mode effects, see for example De Leeuw (2005). Mode effects need to be properly taken in to account at both the survey design and the estimation phases in order to reduce their biasing effects on target parameters and to ensure accurate estimates. In particular: 1) measurement errors may be prevented at the design phase by using proper data collection tools and by adopting the most appropriate survey design; 2) selection errors (and in some cases measurement errors) need to be “treated” by using suitable estimation strategies.

In the MIxed MOde Designs for social surveys - MIMOD project, Work package 1 studies the topic of mode organization (adaptive/responsive, concurrent/sequential mixed-mode designs,…) so that modes are optimally allocated to survey’s units to minimize measurement bias, costs and nonresponse rate. On the other hand, Work package 4 looks at the impact on data quality/respondent burden of the questionnaire design elements (in particular, unimode vs. mode-specific questionnaires).

The purpose of Work package 2 (WP2 hereafter) is to look into ways to cope with mode effects (e.g. weighting, imputation, other methods and models), and to analyse differences in the final sample composition based on different modes across time, countries and survey types. More specifically, the activities within WP2 aim to:

1. provide an updated overview on methodologies for mode effect assessment and adjustment, particularly those currently used in the ESS;
2. evaluate the suitability of selected statistical approaches and methods to deal with selection errors and to adjust for measurement errors based on practical applications and statistical analyses.

Istat and CBS are the partner countries which are collaborating to carry out the above activities, supported by the French Statistical Office (INSEE). Based on the results of these activities, a set of general, practical, evidence-based guidelines for the use of the selected methodologies in ESS surveys will be elaborated.

In the first months of 2018, WP2 activities have been focused on objective a). The main output delivered is a report containing a discussion of assumptions, advantages and disadvantages of the theoretically associated to the use of various approaches. The literature update has been complemented by the results of the survey conducted in the MIMOD project (see Signore, 2018).

As for objective b), in the first project’s period the analysis of selected methods for mode effect estimation and adjustment has been deepened, and the main elements of experimental applications to be carried out have been defined. Methods evaluated within WP2 include reinterview designs and causal inference models. As known, all methodologies dealing with mode effects require auxiliary variables (e.g. from registers, administrative archives, etc.), validation data, or repeated measurements for the same respondents in different modes. The latter situation can be achieved by reinterview designs: a power-cost analysis of reinterview designs for decomposing mode effects into selection and measurement will be performed within WP2. Reinterview designs require a consistent investment, and their application may only be viable under certain conditions and assumptions: these conditions and assumptions will be evaluated in WP2 for selected ESS surveys. On the other hand, a number of models (such as propensity score matching, multiple imputation, methods based on R-indicators, etc.) can be used at the estimation phase for disentangling confounded effects (Vannieuwenhuyze et al., 2012) and for obtaining unbiased estimates of population parameters. Some of these approaches will be evaluated on current mixed-mode social surveys in WP2.

The paper is structured as follows. Section 2 contains a summary of activities carried out to update the state of the art about methods and approaches dealing with mode effects in mixed-mode designs. In Section 3, the ongoing activities aimed at the application to ESS social surveys of selected methods to manage mode effects are summarized. Conclusions and perspectives are reported in Section 4.

**2. The state of the art**

Over the last years, a series of papers has become available in the literature, explicitly discussing the estimation of mode effect components in mixed-mode surveys. This literature needs to be summarized, compared and judged in terms of its applicability (advantages and possible risks) in the ESS context. Actually, findings from literature review (section 2.1), complemented by some results of the MIMOD query (section 2.2), provide an updated overview on existing solutions to common problems in this area.

*2.1. Update of recent literature*

Comparisons between surveys conducted using different data collection 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 multi-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.

Two types of mode effects are distinguished. First, selection effects are caused by the selection mechanism of a multi-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. Mode effects, when present, can either improve or worsen the quality of survey estimates. An obvious improvement to be had from mode effects is a less selective sample of respondents in a multi-mode survey compared to a single-mode survey. Measurement effects in multi-mode designs are generally not desirable. Such effects typically arise when different modes have differential bias: they do not measure the target quantity at the same level, or with the same precision.

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. The WP2 report will list approaches that have been published in the literature. Experimental designs specifically aimed at studying mode effects are preferable, but costly, and hence less common. These include parallel, independent surveys, or reinterview 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 or sample matching have been reported. References to specific articles will be provided in the actual WP2 report.

Less common than mode assessments are mode adjustments. Adjustment techniques are aimed at correcting survey estimates for bias induced by one or several modes, or by the specific combination of several modes. Adjustments for bias require 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. Faithful 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, which will be further explored in the context of WP2, see section 3.1 below.

*2.2. Selected results from the MIMOD query*

The query carried out within the project includes a specific section investigating activities and methods used in recent years in ESS countries to assess and/or adjust for mode effects in mixed-mode surveys.

A limited number of specific questions have been formulated on this topic. In addition, given the complexity of the subject, respondents have been asked to supply specialized documentation either by providing links to on-line papers, or by directly uploading documents on the MIMOD project website. This allowed us to analyse more in depth the applications designed and the methodologies used in the different national contexts, and to acquire useful information for the update of literature review.

It has to be underlined that one of the main problems in the definition of this part of the questionnaire has been the identification of proper questions and their answer categories, and the phrasing of the wording. As for the latter, due to the vast literature and the lack of a common glossary on the topic, we wanted to minimize the risk of misunderstanding with respect to the information content actually required.

In this section, only some main quantitative results are reported for length restrictions. Results relate to the 31 responding countries (100% response rate).

Concerning the question “Has your organization conducted any activity to assess mode effects for any social survey in recent years”, it resulted that the most used approaches are: *Pre-tests and/or experiments on questionnaire design* (48.4%), *Pilot surveys* (42%), *Differences in distributions of socio-demographic or target variables* (about 39%). Frequently used approaches are also *Pre-tests and/or experiments on sensitive or core questions* (35.5% of countries), *Differences in quality indicators (e.g. total or item non response rates, break-off rates, reliability indicator, failure rates of consistency rules, …)* (35.5%), and *Previous and new data collection strategies running simultaneously (independent sampling)* (about 32%). Approaches used with lower frequency are *Separating selection, nonresponse and measurement effects* (25.8%) and *Calculation of representativeness indicators of various designs* (22.6%). One of the less used methods are *Reinterview studies* (6.4%).

Concerning the question “What measures have been taken by your organisation to adjust for mode effects”, most responding countries did not take any measure (61.3%). The most used approach is *Weight adjustment if distributions differ over modes* (25.8%). A lower percentage of countries adopt either *Calibration to fixed mode distributions* (about 13%) or *Estimate measurement errors and correct responses to a benchmark mode* (about 10%).

Concerning plans for future research into mode effect assessment/adjustment for any social survey, about half of the countries (54.8%) declare not to have any plans. The plans of the other countries involve assessment studies more than research on adjustment methods.

**3. The planned experimental studies**

*3.1. Reinterview studies*

Since the separation of selection from measurement effects is a prerequisite for successful mode effect adjustments in multi-mode designs, a promising line of 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 reinterviews through a second mode for a subset of respondents who already responded through a first mode (Klausch et al. 2017). The reinterview designs under consideration are mixed-mode sample survey designs in which a portion of respondents is interviewed twice. The purpose of the reinterviews is to estimate the extent of mode-dependent measurement bias, and to use this estimate in an adjustment technique.

In WP2, a cost-benefit analysis will be conducted to study optimality of the allocation of interview capacity and budget to either nonresponse follow-up, or to reinterviews. Ideally, without constraints, all respondents and non-respondents of a first mode would be followed up for interviewing through a second mode. For the respondents of the first mode, these would be reinterviews. In order to not inflate costs, it should be possible to balance the interview capacity of the second mode between follow-up of non-respondents on the one hand, and reinterviews of the respondents of the first mode on the other hand.

The criterion that will be used in the optimisation strategy is mean square error (MSE) of the adjusted survey estimates. Klausch et al. (2017) considered a range of estimators and found that a regression estimator in which a measurement error model is embedded performs best. Since the error model is inverted to predict the adjusted measurements, this optimal estimator is referred to as an inverse regression estimator. Variance estimates are obtained through bootstrapping. It will be this estimator, and its associated variance estimates that will be used in the study to be conducted in WP2.

*3.2. Methods to deal with mode effects at the estimation stage*

In last years, Istat started studying and implementing different approaches which have been proposed in literature for treating mode bias/mode effects: methods based on R-indicators to study the representativeness of the samples (Schouten *et al.*, 2011), method for testing the assumption of equivalence of measurement across modes (Hox *et al.,* 2015); methods based on Propensity Score approach (Lee and Valliant, 2008), methods based on Multiple Imputations which allow to overcome the non-ignorability of the mode selection assumption (Park *et al*., 2016, Suzer-Gurtekin, et al., 2012), methods based on Calibration (GREG) (Buelens and Van den Brakel, 2013), aiming at defining a robust estimation procedure with respect to variations in the distribution of respondents to different techniques (see De Vitiis et al., 2018).

In the context of WP2, practical applications of selected methods on current Istat mixed-mode social surveys are planned:

* Multipurpose survey on households - Citizens and leisure time. It collects information about recreational and cultural activities in free time, such as sports, reading, cinema, music, the Internet, social relations. A sample of about 24.000 households is annually selected through a two stage sample design. A mixed technique using sequential Web-PAPI is adopted;
* Multipurpose survey on households - Aspects of Daily Life. It collects information about many topics related to the quality of life of people. A sample of about 24.000 households is annually selected through a two stage sample design. A mixed technique using sequential Web-PAPI is adopted: for year 2017 an experimental design was set up, splitting the sample of households in two independent samples: one sample (consisting of 2/3 of total sample units) was subjected to a sequential Web-PAPI data collection strategy, the other one was interviewed in PAPI mode.

Auxiliary information (socio-demographic and economic variables) from registers and administrative sources is used. In particular, information from the Italian “Archimede” Project (Integrated Archive of Economic and Demographic Microdata) is exploited.

The choice of methods for the analysis and treatment of mode-effects to be implemented will face some critical issues: 1) missing values on covariates, so that method for incomplete data analysis should be considered to avoid the issue of extending the results to the incomplete part of the sample, and 2) total nonresponse resulting from the selection process and which cannot be addressed with usual methods.

Experimental studies aim to identify feasible methodologies to evaluate the measurement error, even if the ideal goal, especially for multipurpose surveys, is to define a weighting adjustment. What we are studying at present is the possibility of adjusting for the selection effect: we know it is a desirable effect, but in some cases it introduces in the sample of respondents a sort of imbalance that seems to have to be corrected or, at least, accounted for in the calibration phase. The ultimate goal of the analysis is the comparison of the estimates of some target parameters obtained using alternative approaches based on different assumptions, with the aim to obtain an assessment of these assumptions as well. Based on the performed applications, main advantages and risks theoretically associated to the use of the analysed methods will be discussed. Furthermore, a general “protocol” for analysing and treating mode effects in sequential and concurrent strategies and general recommendations to prevent and limit mode effect /mode bias will be released.

**4. Conclusions and future work**

Within WP2 of the MIMOD project methods and approaches to deal with mode effects in mixed-mode surveys in the ESS are investigated. In the first half of 2018, activities have been focused on an update of recent literature on the topic, on the results of the MIMOD query, and on the planning of experimental applications of selected methods to assess and adjust for mode effects. By the end of the year, a final report containing a general discussion and guidelines about the methodological solutions which can be adopted to deal with mode-effects in mixed-mode designs will be delivered.

The results of WP2 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 social surveys involving the web. However, given that only two countries are involved in WP2, the results of the performed analyses will allow to delineate quite general guidelines on possible risks and advantages of combining modes of data collection in this context. For the same reason, the WP2 results could not cover all the specific situations of each single country and application context. Suitable modes of collaboration could be identified in the future to proceed with developments in this area, e.g. through a network of countries interested in continuing the discussion on methodological issues.

**5. References**

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

de Leeuw, E. (2005), To mix or not to Mix Data Collection Modes in Surveys, Journal of Official Statistics, 21(2), pp. 233-55.

De Vitiis, C., Guandalini, A., Inglese F. and Terribili M. (2018), Assessing and adjusting bias deriving from mode effect in mixed mode social surveys, European Conference on Quality in Official Statistics (Q2018), Kraków, Poland. 26-29 June.

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

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

Hox J., De Leeuw E. D. and Klausch T. (2015), Mixed Mode Research: Issues in Design and Analysis, International Conference on total survey error: improving quality in the era of big-data, Baltimora, September 2015.

Klausch T., Hox J. and Schouten B. (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), pp. 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.

Lee S. and Valliant R. (2008), Post-Survey Weighting Methods Using Propensity Scores: A Review, chapter 8 in Advances in Telephone Survey Methodology, John Wiley.

Park S., Kim J. K. and Park S. (2016), An imputation approach for handling mixed-mode surveys, The annals of Applied Statistics. Vol. 10, No. 2, pp. 1063-1085.

Schouten B., Shlomo N. and Skinner C. (2011), Indicators for Monitoring and Improving Representativity of Response, Journal of Official Statistics 27: pp. 231–253.

Signore, M. (2018), Improving the quality of multi-mode data collection. The European MIMOD project, European Conference on Quality in Official Statistics (Q2018), Kraków, Poland. 26-29 June.

Suzer-Gurteki Z. T., Heeringa S. G. and Valliant R. (2012), Investigating the bias of alternative statistical inference methods in sequential mixed-mode surveys, Section on Survey Research Methods – JSM 2012.

Vannieuwenhuyze J. and Loosveldt G. (2012), Evaluating Relative Mode Effects in Mixed-Mode Surveys: Three Methods to Disentangle Selection and Measurement Effects, Sociological Methods & Research, 42, pp. 82–104.

Buelens B. and van den Brakel J. (2013), Measurement error calibration in mixed-mode sample surveys. Discussion paper. Statistics Netherlands. The Hague/Heerlen, 2013.

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.