ROMANIA

Reimbursable Advisory Services Agreement on
Romania Capacity Building for Statistics (P167217)

OUTPUT No. 6c

Report on advisory services provided to Recipient on the Methodology for assessing the effects of changing the data collection method, in the context of switching to multimodal methods (CAPI/CAWI), for surveys and statistics in the inter-census periods

October 2022

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

Project co-financed from the European Social Fund through the Operational Programme for Administrative Capacity 2014-2020

Disclaimer

This report is a product of the staff of the World Bank. The findings, interpretation, and conclusions expressed in this paper do not necessarily reflect the views of the Executive Directors of the World Bank or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work and does not assume responsibility for any errors, omissions, or discrepancies in the information, or liability with respect to the use of or failure to use the information, methods, processes, or conclusions set forth. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of the World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.
This report does not necessarily represent the position of the European Union or the Romanian Government.

This report has been delivered in October 2022 under the Reimbursable Advisory Services Agreement on Romania Capacity Building for Statistics (P167217) signed between the National Institute of Statistics in Romania and the International Bank for Reconstruction and Development on September 17, 2019. It contibutes to Output 6 under the above-mentioned agreement.

TABLE OF CONTENTS

List of Figures

List of Tables

List of Acronyms

CAPI	Computer-assisted personal interview
CATI	Computer-assisted telephone interviewing
CASI	Computer-assisted self-interviewing
CAWI	Internet-computer-assisted web interview
INS	National Institute of Statistics
MSE	Measures the amount of error
NSO	National Statistics Office
PAPI	Pencil and paper interviewing
PDCA	Plan-Do-Check-Act
RAS	Reimbursable Advisory Services
WB	World Bank

Executive summary

The purpose of this report is to provide indications and recommendations for assessing the effects of changing the data collection method, in the context of switching to multimodal methods (CAPI/CAWI), for surveys and statistics in the inter-census periods. Usually, the **multi-mode** surveys refer to different data collection modes used to administer different sections of the survey questionnaire to the same sample unit. An example of a multi-mode survey is the Household Budget Survey (HBS) that in some countries uses the following multi-mode approach for each sampled household: i) Self-recording (e.g., PAP mode) for consumption expenditures (diary) and ii) Personal interview (e.g., CAPI mode) to collect general information on the household and data on infrequent expenditures, saving, income. Strictly related to multi-mode surveys, there is the **mixed-mode data collection,** this term is used to identify the combined use of different data collection modes to administer the same survey questionnaire to all sample units. Multi-mode and mixed-mode are sometimes used as synonymous in an interchangeable manner.

This report is divided into four (4) sections. The first section describes the main reasons to move from a unimodal data collection to a mixed one, discussing the key factors that guide this choice, as well as organizational and infrastructural aspects that affect all the data collection system. Recommended good practices, as well as bad practices, are selected from previous experiences in European countries.

The second section proposes multi-mode and mixed mode designs, describing the different steps in which they can be organized. A specific attention has been devoted to adaptive mixed mode data collection design.

Since multi-mode and mixed mode data collection really affects all the data collection system, the third section reports key aspects to consider in communicating with the respondents, also reporting good practices and actions to avoid. In addition, the effects of multimode and mixed mode data collection on the IT and organizational infrastructure are provided. Finally, in the last section, the document describes the so-called mode effect, reporting suggestions, proposals, and issues in assessing, adjusting, and controlling the mode effect.

This document represents Output 6c under the Reimbursable Advisory Services (RAS) Agreement on Romania Capacity Building for Statistics (project No. P167217). The project is implemented by the National Institute of Statistics with support from the World Bank.

1. Reasons for multimodal methods

1.1. Switching from uni-mode to multimode

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 or multi-mode surveys in our context.

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 (see section 4).

Mail and face-to face surveys are the oldest recorded data collection modes. Therefore, it is not surprising that the earliest forms of mixed-mode designs combine face-to-face interviews with mail surveys. For example, in longitudinal or panel surveys, face-to-face interviews were used in the recruitment phase to maximize response and to administer base-line questionnaires to household members. In the next waves, data were then collected with less costly mail surveys. Almost the same approach is now in vogue for establishing “access” panels and Internet panels: telephone interviews for recruitment and far less costly web surveys for follow-up data collection.

The popularity of telephone surveys led to a new mixed-mode approach as mixes of face-to-face and telephone surveys were implemented. For instance, many Labour Force Surveys used telephone interviews in a quarterly panel design. In this mixed-mode design all first interviews were conducted in person, and the follow-up interviews were conducted by telephone.

The rapid growth of computer technology caused the next important change in data collection. Computer-assisted equivalents were developed for all major data collection methods with a generally positive effect on data quality and a potential for new applications. The greater efficiency and more effective case management of computer-assisted telephone interviewing (CATI) made this a powerful tool for the screening of potential respondents and for nonresponse follow-ups. A more recent form of computer-assisted self-interviewing has come about by means of the establishment of computerized household panels, where households are equipped with computers and software and questionnaires are sent electronically on a regular basis. CATI facilities are still necessary to recruit panel members and assist respondents with problems.

The latest development is the web or Internet survey. Internet or web surveys are very cost and time efficient, and this, together with the novelty value, have made them very popular in a short time. They have a great potential, but they also still have limitations (e.g., noncoverage, nonresponse). The rapidly growing interest in web surveys, their potential and limitations, gave a new impetus to mixed-mode designs.

1.2. Decision making

An optimal data collection method is defined as the best method, given the research question and given certain restrictions. The basic research question defines the population under study and the type of questions that should be asked. Survey ethics and privacy regulations may restrict the design, as may practical restrictions like available time and funds. When designing a survey, the goal is to optimize data collection procedures and reduce total survey error within the available time and budget. In other words, it is a question of finding the best affordable method, and sometimes the best affordable method is a mixed-mode design. Survey designers choose a mixed-mode approach because mixing modes gives an opportunity to compensate for the weaknesses of each individual mode at affordable cost.

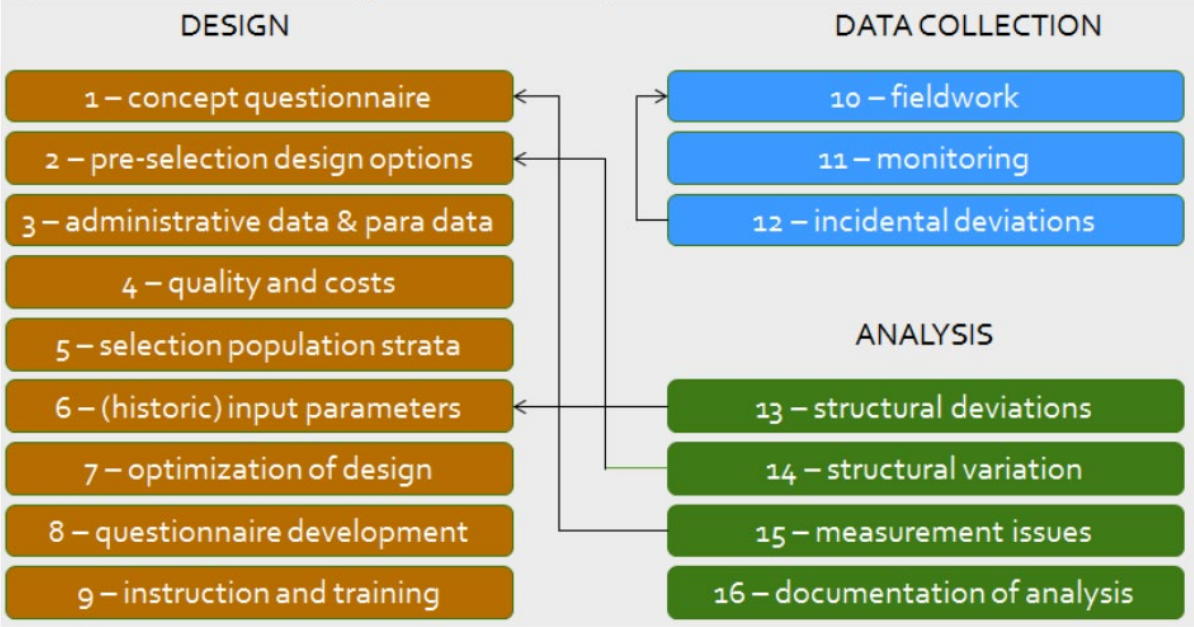
The most cost-effective method may not be optimal for a specific study. By combining this method with a second more expensive method we have less costs and less error than in a unimode approach. In mixed-mode designs there is an explicit trade-off between cost and errors, focusing on non-sampling errors – that is, frame or coverage error, nonresponse error and measurement error.

Actually, the survey mode is the design feature that has the strongest impact on survey **errors** and survey **costs**. In addition, modes often have their own logistics and infrastructure and combining modes in one design is challenging for survey case management, monitoring and analysis. Consequently, a discussion of mode choice and mode allocation is a discussion of virtually all aspects of survey design and analysis. References to general discussions of survey errors and costs are Groves (1989), De Leeuw, Hox and Dillman (2008), Bethlehem (2009), Groves et al (2009) and Dillman et al (2014).

Figure 1 depicts the plan-do-check-act (PDCA) cycle of survey design, where the “act” phase is represented as feedback arrows to the “plan” phase of survey design. Mode choice is part of steps 2 and 7 but influences also design steps 3, 4, 6, 8 and 9. Mode allocation through adaptive survey design depends on steps 3, 5 and 6 in static designs, and, additionally, on steps 11 and 12 in dynamic designs.

An important message from any PDCA cycle, as represented by figure 1, is that any design option is not fixed but change over time. Mode choice and mode allocation are no exceptions to this rule. Gradual changes in survey errors (population coverage, response rates, response representativeness, measurement) and survey costs (both fixed and variable), represented by steps 13, 14 and 15 alter the business case of modes in mode design. During data collection and from one wave to the other, response rates may decrease (and costs go up) and/or contributions of modes to the overall response may change, altering also the overall measurement characteristics and comparability. Good examples are the emergence of web as a survey mode, the more recent diversification of devices that allow access to web and the decreasing coverage of landline phones. Apart from the omnipresence of modes and preferences that respondents may have to use them, there is also the general tendency of decreasing response rates that has changed trade-offs in survey errors and costs. These changes, themselves, imply that mode choice and mode allocation cannot have absolute viewpoints.

Figure 1 - A sketch of the survey plan-do-check-act cycle



Nevertheless, it is possible to determine the strongest influences of modes on survey errors, costs, and operations.

1.2.1. Survey errors and mixed-mode design

From the survey error perspective, modes strongly influence: 1) population coverage, 2) survey response and 3) survey measurement.

Upcoming modes and vanishing modes, typically, have a lower **coverage** of the population. Coverage of modes is country-dependent, so that trade-offs become country-dependent as well. Survey response rates vary greatly between modes and could differ between online and face-to-face modes. As response rates do not have a direct relation to nonresponse bias, also contrasts/variation in response rates over relevant population subgroups need to be considered. Here, the picture is less clear, but modes do also affect variation, although less strongly as overall rates. Cornesse and Bosnjak (2018) concluded, for example, that web response rates vary more than mixed-mode response rates based on a meta-analysis.

Coverage error is one of the biggest threats to inference from telephone and web surveys. Although phone and Internet access is growing everywhere, the picture is diverse for sub-populations. Furthermore, those covered differ from those not covered, with the elderly, lower-educated, lower-income, and minorities less well-represented online. To compensate for coverage error in telephone and web surveys, mixed-mode strategies are now employed: for instance, one can use a mixed-mode or hybrid survey approach, in which all sampled units are contacted by means of a paper letter and given the choice to either use the Internet or request a paper questionnaire. Telephone and web hybrids have become increasingly popular as the development of special multi-mode CATI/CAWI software also indicates.

Obviously, **survey response** is conditional on mode coverage, so that part of the difference in level and variation of response rates is due to varying coverage. However, mixed-mode applications can reduce **nonresponse error**. Response rates have been declining over the years, in official statistics. To achieve higher response rates, while keeping the overall costs low, mixed-mode strategies are used, starting with the less costly method first. A prime example is

the American Community Survey, which is a mail survey with follow-up telephone interviews for nonrespondents, followed by face-to-face interviews for a subsample of the remaining nonrespondents. Telephone follow-ups appear to be effective in raising response and may even reduce nonresponse bias in mail. To reduce selective nonresponse, for instance to include ethnic groups, one can have a mail survey, with an explicit statement on the cover in several languages, urging respondents interested in completing a telephone survey to contact the survey centre where bilingual interviewers are available. Incentives, together with mail and telephone follow-ups, are employed to raise response rates.

Measurement error also varies between modes. Modes vary on several administration features, the most prominent being intimacy, interaction, assistance, presentation, speed/pace of the interview, and timing. The first three all relate to the presence of an interviewer: Interviewers create a form of intimacy that may affect answers for certain types of questions. Interviewers naturally lead to interaction, which gives the survey a face or voice and a spokesperson. Interaction can also be simulated in web surveys. The interaction may increase motivation and concentration. Interviewers also help respondents by navigating through the questionnaire and/or by explaining context.

Presentation of the modes refers to aural or visual questionnaires, which affect the cognitive effort and skills to answer questions. Modes influence the speed and pace of an interview. In self-administered modes respondents can choose their own speed and pace, allowing them to think more carefully but also less carefully. Finally, the timing of the interview is affected by the mode. Again, self-administered modes offer more freedom; respondents can choose the time and place. Measurement error follows response and coverage, and it is often hard to remove confounding of the errors without experimental designs.

One of the most consistent findings in mode comparisons is that self-administered forms of data collection perform better than interview-modes when **sensitive questions** are asked. Therefore, mixed-mode approaches using a paper self-administered form to elicit sensitive information in a face-to-face interview have been standard good practice for a long time. The more private computer-assisted self-administered forms led to more accurate reporting of socially undesirable attributes. This suggest using computer-assisted self-interviewing (CASI), where respondents answer most questions privately by directly entering the answers in the computer, and only a few non-threatening questions are asked by an interviewer.

A positive aspect is also that using mixed-mode strategies enabled some simplification of questions and questionnaires.

1.2.2. Survey Cost and mixed-mode design

From the survey **cost** perspective, modes vary, perhaps, even more than for survey errors. The differences in costs per respondent in web and face-to-face are very large; face-to-face may be as much as 50 to 100 times more expensive, depending on the implementation. The consequence is that sample sizes can be made much larger for cheaper modes and precision can be greater, despite the lower coverage and response rates. The difference in costs between modes arise mostly from fieldwork, especially travel costs, and not from design or analysis. From the survey operations perspective, modes are very different. Naturally, self-administered modes are operationally much easier as they do not require planning and management of interviewers.

Another main distinction is between computer-assisted and paper-assisted administration. Interviewer modes allow for both, and, naturally, web and paper modes allow only for one type of administration. One type of administration is not necessarily more complex than the other

when starting from scratch, but they are different. This points at the most complex part in mixed-mode designs, that of combining mode operations in a single fieldwork. This holds especially true for combinations of self-administered and interviewer assisted and computer-assisted and paper-assisted. All in all, mixing modes has been a means to improve survey costs, while not affecting survey errors too much, weighed against the increased complexity of the survey operations.

1.3. General advice, good and bad practices

Summarizing, the main advantages of mixed-mode data collection are the following ones:

- cost reduction
- improved coverage
- the ability to adapt to respondents' preferences and convenience
- increased response rate

In terms of disadvantages of mixed-mode data collection, we have:

- organizational complexity and investments required
- mode effects and measurement error
- questionnaire development costs
- if the data collection is not mixed in the proper way: poorer coverage and reduced response rates.

Particularly with online surveys, sometimes we may have some poorer interviewer motivation, since the "easiest" respondents participate to online surveys, leaving the more difficult cases to be followed up by interviewers.

Regarding bad practices, we can mention the following five:

1. Don't send too many reminders
2. Don't send reminders indiscriminately
3. Don't introduce questionnaire differences that can lead to bias and time series breaks
4. Don't underestimate case management and IT infrastructure costs
5. Don't make questionnaires too long.

General advice on good practices for online surveys:

1. Gradually introduce web sample to go mixed-mode, measure effects.
2. Login and completion assistance, Make login as easy as possible.
3. Questionnaire development: adapt questionnaires strongly, ensure smartphone compliance, modularize questionnaires.
4. Communication strategies: find the right communication strategy for each survey, add a brochure with web completion instructions to your cover letter, shorten e-mail notifications.
5. Data collection organization: use an online first approach, do CAWI breakoff reclaim directly in CATI, in CAPI/CATI designs use the same interviewer for the same respondent.
6. Technical aspects: monitor servers constantly, enable restart at breakoff points.

2. Multimodal methods for data collection and their effects on survey design

2.1. Types of mixed-mode designs and multi-mode system, multimode steps in data collection

There are many forms of mixed-mode designs and many ways of summarizing them. One can focus on the data collection and its objectives, on timing of interacting with respondents (i.e., contact phase, response phase, and follow-up phase) as organizing principle. These forms overviews can be integrated and expanded by including both actual data collection mixtures and mixtures of means of communication. It is important to realize that survey researchers communicate with sample members at different points in time and that they may use different modes of communication to do so. Prime examples are prenotifications, screening procedures and reminders. This goes beyond the mere data collection itself, and it is often used the term mixed- or **multi-mode system**. Sometimes, although the actual data collection is unimode, the data collection system is multi-mode with mail and telephone advance notifications and reminders.

In Figure 2 we give a systematic representation of different types of mixed-mode systems, the reasons to employ them and the potential effects on survey quality. This overview is ordered according to phase.

Figure 2 - Types of mixed-mode systems, rationales, and effects on survey data quality

Mixed-mode survey system	*Rationale for implementation*	*Effect on survey quality*
Contact phase		
Advance notification **in mode different from data collection mode**	▪ Correct sampling frame ▪ Raise response rate ▪ Enhance credibility/trust	▪ Reduce coverage and nonresponse error ▪ No threats to measurement (if data collection is uni-mode)
Recruitment / Screening / Selection **in mode different from data collection mode**	▪ Reduce cost ▪ Enhance efficiency ▪ Update / expand contact information for main mode	▪ Improved timeliness ▪ If pure screening no threats to measurement ▪ If screening plus first part data collection in other mode, then risk of potential mode effects on measurement
Response phase		
One sample, one time period, one questionnaire, but different sample persons, different modes	▪ Reduce costs ▪ Improve coverage ▪ Improve response	▪ Reduce coverage and nonresponse error ▪ Mode effects on measurement (confounded with subgroups)
***One sample, one time point,* but** different modes for different parts of questionnaire (for same person)	▪ Improve privacy of measurement ▪ Reduce social desirability bias	▪ Improved data quality, especially with very sensitive questions
***One sample, multiple time points,* but** same persons measured with different modes at different time points	▪ Reduce costs	▪ Measurement differences causing confounding of time effects and mode effects
Follow-up phase		
Reminder **in mode different from data collection mode**	▪ Raise response rate	▪ Reduce nonresponse error ▪ If pure reminder no threats to measurement ▪ If reminder plus part of data collection in other mode then risk of mode effects on measurement
Partly based on:	Dillman (2000)	Balden (2004)

2.2. Concurrent and consecutive multimode data collection

A different situation occurs when an advance notification is used to invite sample members to complete a questionnaire and where it is left to the respondent to choose a mode. There can be, for instance, a paper mail advance letter with an invitation to complete a web survey, but also offering the opportunity to ask for a paper questionnaire or by telephone via a toll-free number. This is a form of **concurrent multiple modes**: both modes are being implemented at the same time.

Concurrent mixed-mode designs apply when different data collection modes are in the field at the same time. Modes can be assigned in advance to sub-groups of sample units or sample units can choose the mode they prefer.

A procedure like this is often used to reduce coverage error, but as the data collection itself now is multi-mode other errors come into the picture. First, self-selection may cause differences in socio-demographic variables; secondly, the mode itself may cause measurement

differences. The researcher must decide which scenario is the best: multiple-mode with reduced coverage error at the price of increased measurement error or a uni-mode approach with a larger coverage error component. In web surveys, where the risk of coverage error is still high, researchers usually opt for the multi-mode approach and take the risk of mode effects on measurements. Since self-selection and mode effects are completely confounded in such designs, it is difficult to correct for mode effects.

One sample, one time period, one questionnaire: The first situation indicated in Figure 2 is where one mode of data collection is used for some respondents of a sample and another mode for others in the same sample in order to collect the same data. An example of a **concurrent mixed-mode design** for this situation is a paper mail survey with a web option. The aim is to reduce coverage bias and still complete the survey at reasonable costs. It is also assumed that giving a sample member a choice may reduce nonresponse, as some persons may express certain mode preference. Still, giving respondents a choice may be a good strategy, as it may create goodwill and reduce costs. In business surveys it is more common than in household surveys to allow respondents to choose a data collection method.

Far more common and effective are sequential multi-mode systems to reduce survey nonresponse. Usually, an inexpensive mode is used as the main mode for the whole sample and then a more expensive mode is used for the nonresponse follow-up to improve response rates. Strategies beginning with mail and telephone and followed-up with other methods, provide response rates as high as face-to-face, for half of the costs. Sequential mixed-mode surveys will increase response both for the general population, for different racial and ethnic groupings. There is also evidence that a sequential mixed-mode raises the response rates in establishment surveys.

Sequential mixed- or multi-mode data collection methods are effective in reducing nonresponse, but a coin has two sides and there is a potential for measurement error as the modes used may cause measurement difference

One sample, one time point, but different modes for different parts of the questionnaire*:* A second form of a mixed-mode is when different modes are used for a subset of questions in the questionnaire during a single data collection period. Usually, a mix of interview and self-administered forms is used to exploit the strong points of both methods. For instance, within an interview a self-administered form of data collection such as CASI is used for sensitive questions to reduce social desirability and enhance privacy as neither the interviewer nor any other person in the vicinity will know the answers given. This situation is basically positive and is not a case for concern

One sample, multiple time points: The third form is a longitudinal study or a panel in which the same respondents are surveyed at different time points, and different modes are used at those different time points. Here practical considerations and costs are the main reasons to use this multiple mode approach. One issue might be the availability of a good sampling frame with contact information. Sometimes addresses are available, but telephone numbers or e-mail addresses are not and have to be collected first; sometimes no frame is available and area probability sampling is the only option. This means that an initial contact should be made using a face-to-face method. Together with the greater flexibility of an interviewer to gain cooperation at the doorstep, a face-to-face interview is often the preferred choice for the base-line study of a panel. When possible, a less expensive method is used after the first wave to reduce costs. A combination of face-to-face interviews for the first wave and telephone surveys for the second is used for labor force surveys in several countries. Another example of a mixed-mode panel combines an initial face-to-face interview with mail surveys in subsequent waves.

Sometimes modes can even be switched back and forth. For instance, after an initial face-to-face survey, telephone and mail surveys are employed with an occasional face-to-face survey interspaced at crucial points. Internet panels, which are fashionable, are often formed after an initial telephone survey. In sequential mixed-mode studies time and mode effects are confounded, and it is difficult to decide if a changeover time is real or the result of a change of mode

2.3. How to best combine mode: deciding the modes sequence

Concurrent mixed-mode designs are characterized by a large variety of mode combinations, but in concurrent designs CAWI is used less frequently than CAPI, CATI or PAPI. This is probably a consequence of the organizational complexity of the concurrent design: the presence of an interviewer allows to better manage the different modes available on the field at the same time, especially when respondents' mode choice is allowed. The higher use of interviewer-assisted modes has an impact on costs, that could be contained in case the fieldwork is totally managed in-house. The fieldwork organization is, in fact, a further relevant aspect for these designs, both for the organization complexity and for costs.

In concurrent mixed-mode designs, the two strategies of mode choice, by respondents or by the INS, are possible. The choice of mode by respondents is harder to manage, but it is used to reduce response burden and encourage participation. On the other side, the decision of the INS on which mode to assign is less complicated but requires the availability of auxiliary information.

Mode choice and mode allocation are at the core of survey design and analysis. Since survey modes have an impact on both representation errors and measurement errors, mode choice and mode allocation are a complex endeavour that demands for explicit objectives and risks assessment. Although, survey designers acknowledge the important role of mode-specific measurement, mode choice and allocation rarely involve explicit criteria for data quality in practice. Comparability and equivalence in respondent answers are usually restricted to the questionnaire design stage. However, in recommendations for mode choice and allocation, mode-specific measurement biases must be accounted for, even after careful questionnaire design.

The choice of a mixed-mode survey design - which modes to use and in which sequence - strongly depends on the purposes of the survey design and on its constraints. Maximizing data quality, that is minimizing the total survey error, is in general the main aim to be reached, considering constrains like costs, timing, response burden, etc. (Istat, 2018). Vice versa the aim could be to minimize costs under the constraint of maintaining the same level of accuracy in terms of total survey error. Modes have a different impact on quality and costs; in mixed-mode strategies these differences depend on modes sequence. In terms of impact of the sequences, there are disparate results in the existing literature (see Wagner et al. 2014). It is, therefore, impossible to provide indications on the best sequence that meets all the requirements, national backgrounds and specific survey features. However, it is possible to define several steps to follow, when deciding which mode strategy to use for a specific survey. These steps are listed below:

1. Determine and prioritize aims by estimating the quality-cost trade-off
2. Identify risks
3. Determine the candidate modes
4. Evaluate different mixed-mode design options:
   a) Define the mode sequence/administration
   b) Define which devices respondents may use to participate if CAWI mode is offered

c) Design and test the questionnaire to reduce mode-specific measurement effect
d) Evaluate the complexity of the logistics and operations, i.e., case management and human resources
e) Evaluate the possibility of enhancing the response rate and/or population coverage (incentives, mode choice)
f) Design the communication strategy: advance letter, reminders, interviewer instructions and training and survey website
g) Test the mixed-mode design(s)

5. Implement the mixed-mode design and collect data
6. Compute estimates
7. Assess the results in terms of the defined aims
8. Document

The eight steps make up the decision process of the entire survey design. Results of tests (step 4g) and of assessment (step 7) might activate a 're-design' step as part of the PDCA cycle (step 2 of figure 1). Here, the need to prioritize aims is highlighted, since different priorities influence the choice of mode design options (step 4). Step 2 is about the identification of risks. The main risks when using mixed-mode strategies, are break in time series, mode and selection effects and budget overruns. Once priority of aims is established and risks are identified, then the potential modes to adopt can be determined. It is obvious that mode adequacy for the survey topics is another ingredient to be taken into account in step 3 (besides costs and errors), as well as the availability of contact information (telephone numbers, postal addresses, e-mails).

The evaluation of the mode design options, step 4, is made of different sub-steps. The first sub-step (4a) is about the choice of the sequence of mode administration. Literature is poor about experiments on large scale that provide indications on which mode combinations and sequences to use (concurrent versus sequential). A comparison among different mode administration can be found in Mauz et al. (2018). They describe the results of an experiment where a sequential and a concurrent design are evaluated in terms of costs, response rates and estimates for the main indicators. Results are, anyway, country and survey specific. Wagner et al. (2014) report the results of several studies testing the influence on costs and response rates of different sequences of modes. All studies agree on the fact that using the cheapest mode as the first one positively affects costs but show controversial positions about the influence that the mode sequence has on the response rate.

It is not possible to define the 'optimal survey design', since it is country and survey specific as literature on this topic shows, therefore, only general rules that are based on the priorities defined in step 1, are provided. Specifically:

a) If cost reduction has the highest priority, a sequential design where the cheapest mode is offered first can be adopted. However, cost reduction needs to be evaluated overall. If, for instance, CAWI as first method reduces fieldwork costs, the mixed-mode strategy as whole may not. In fact, higher costs may come from the increased complexity of the organization, questionnaires implementation and testing, monitoring, higher travelling costs for face-to-face interviewers, etc. Therefore, it is important to keep all these factors under control to reach the aim of containing costs.
b) If the aim is to contain non-response and coverage errors, then the concurrent mix of modes that allow for a higher response rate and/or coverage is advisable; a sequential design with most "effective" mode offered first can be an alternative. Generally, the modes that guarantee this aim are interviewer administered.

c) Strategies in between, like party sequential-partly concurrent designs, can help in balancing costs and errors. An example can be a self-administered mode (cheapest) that is offered first and left open as an option, when an interviewer-administered mode is added to improve coverage and response.

As to sequential designs, it is important to take into account the sequence of modes, because it could have an impact on the response rate: if a mode with an expected lower response rate is offered first in a sequence, this might increase resistance to later attempts in a different and more effective mode(s). As to step 4d (evaluate the complexity of organization based on CMS and human resources), it is important to mention that without a unique IT platform, questionnaire development and testing can be difficult and burdensome. Similarly, monitoring would increase in complexity especially for concurrent designs.

For what concerns the **human resources**, it is necessary:

i) to predict the interviewers' workload and costs, especially travel expenses, and
ii) to define a payment system that does not influence the response rate per mode, when self-administered and interviewer-administered modes are combined.

All this always holds, but it is even more delicate/important in case the fieldwork is outsourced (totally or partially). Sub-step 4e) is about the possibility of enhancing response rates and/or coverage using incentives and/or by giving respondents the chance of choosing the mode they prefer among those offered. As to incentives, literature suggests that, in general, incentives may increase survey participation: monetary and especially cash are more effective than non-monetary incentives and unconditional incentives are more effective than conditional ones. Concern has been raised that incentives might reduce data quality, but there appears to be little empirical evidence supporting this. On the contrary, some studies indicate that incentives may influence respondents to put more effort into completing the questionnaire (see Olsen et alt., 2012, for a review). However, offering incentives might raise costs. The choice of using them or not depends, therefore, on the priority of aims and on the budget available. A possible way for not affecting budget too much, is to offer incentives only to certain sub-groups of respondents, that are in general those 'harder' to involve in the survey.

The mode choice from respondents could be another way of enhancing response rate. However, literature converges on the fact that doing so does not appear to increase response rate (see Couper, 2011, for a review). Besides, offering this choice could add an additional danger of confounding mode effects and measurement effects with self-selection bias (De Leeuw, 2005). Anyway, many INSs use this option quite frequently for their concurrent designs, most likely because it reduces respondents' burden and better prepares them to do the interview.

Attention should be paid to the potential increase of the **operational complexity** that mode choice might induce. This is particularly true for concurrent mixed-mode designs. To limit this danger, an "in between" strategy may be used, that is to manage the choice of mode concurrently, but then to perform data collection sequentially. To do this, the choice of the preferred mode can be proposed to sample units in the invitation letter, in which different, sequential, time periods of data collection for each mode are also announced.

Sub-step 4f) is about the communication strategy. Decisions on communication strategy depend on survey modes and their sequence (steps 3 and 4a). As far as the decision process is concerned, particular attention should be given to the advance letter to increase respondents' participation and eventually to push them toward a specific mode (e.g., push to web strategy). Besides number and timing of reminders, as well as the communication means, should be decided carefully, paying attention to contain costs and harassment of respondents. Helpful indications can arise from testing. In general, all decisions taken from step 3 to 4f should be

tested in order to establish whether they are in line with the expectations. Evaluation of mode and selection effects should be part of the analysis.

Step 5 follows with the implementation of the mixed-mode design and data collection. During this phase, two viewpoints must be taken into account: that of respondents and that of interviewers (if they are involved). As to the first it is important to always facilitate respondents' tasks by setting a contact centre not only to provide respondents with technical and thematic assistance, but also to assure them about the confidentiality of the data they provide. For what concerns interviewers, training and motivation are essential. This is particularly true in mixed-mode surveys using web as first mode, where it can happen that the "easiest" respondents participate to the online questionnaire and the "most difficult" cases are left to interviewers. Steps 6 to 8 conclude the survey process. Step 7 about the assessment has the purpose to evaluate the design performance in terms of data quality and, therefore, to provide indications for future survey editions.

2.4. Adaptive mixed-mode survey design

Adapting the allocation of modes to different subgroups is a next step to mixing modes. It offers more flexibility in making trade-offs between survey errors and survey costs; a non-adaptive, i.e., uniform, design is just a special case of an adaptive design. It also demands for (even) more flexibility in survey operations. In the literature, there is a distinction between adaptive and responsive survey designs. Here, we will avoid the distinction and simply refer to adaptive survey design. Survey designs can be measured along two dimensions: the extent to which a sample is stratified and treated differently, and the extent to which design decisions are postponed to the actual data collection. If a survey design does not stratify the sample and treats strata differently before or during data collection, then we do not call the design adaptive. So, the first dimension needs to be present to a minimal extent. There are five levels of adaptive survey designs:

0. Uniform design, i.e., no stratification in any way
1. The sample is stratified at the moment a sample is drawn based on administrative data, frame data or data from previous surveys/waves in a panel
2. The sample is stratified based also on paradata that will become available during data collection, but the treatment allocation is decided beforehand. In other words, at the start of data collection it is known what treatment a sample unit will get for each possible realization of paradata
3. Like 2, but the treatment allocations are not (fully) known at the start of data collection and are decided upon at pre-defined phase points during data collection. In other words, even if paradata is available, it is not known in advance what treatments are employed. However, the set of possible treatments is specified beforehand
4. Like 3, but also the set of treatments itself is not yet specified at the start of data collection.

It must be clear that each level introduces complexity. Level 1 is usually called "static" and levels 2 to 4 are "dynamic". Levels 3 and 4 are sometimes referred to as "responsive". In steps 2 to 4, paradata are included. Examples are automated process data such as number and timing of reminders, calls and visits, type of online device used and whether a break-off occurred. Other examples are interviewer observations on the sampled persons and/or their dwellings and neighbourhoods.

To date, adaptive survey designs have focused almost entirely on survey response. There are a few exceptions in the literature where also measurement error is considered (Calinescu and

Schouten 2015 and 2016). For adaptive survey designs that focus on mode, measurement error is, however, an important component of quality.

So how to make design decisions in adaptive mixed-mode surveys? Given that these designs inherit all the complexity of mixed-mode surveys, there are again no absolute viewpoints. However, guidelines have been developed. The four main ingredients to adaptive survey design are 1) quality-cost objectives, 2) stratification of the population/sample, 3) design features to vary, and 4) an optimization strategy. These ingredients correspond to steps 4, 5, 2 and 7 in figure 1. In the context of mixed-mode surveys, ingredient 3, design features, is the choice of modes and sequences of modes. Wagner and West (2016) developed a checklist for adaptive survey design, in which the four ingredients come up. The checklist consists of the following steps, which are further elaborated, as in the previous subsection:

1. Identify priorities
2. Identify major risks:
    a. Consider risk of incomparability in time
    b. Consider risk of incomparability between subgroups
    c. Consider risk of budget overrun and heavy interviewer workloads in follow-up modes
3. Define quality and cost indicators:
    a. Consider nonresponse indicators
    b. Consider measurement error indicators
    c. Consider cost indicators
4. Define decision rules from:
    a. Trial-and-error
    b. Case prioritization
    c. Quota
    d. Mathematical optimization
5. Modify the survey design and monitor the outcomes:
    a. Develop a dashboard for survey errors
    b. Develop a dashboard for survey costs
6. Compute estimates
7. Document

The first three steps concern quality and cost indicators, the fourth and fifth step concern the design features and optimization strategy, and the last two steps involve future changes and replication. The choice of strata is implicitly included in step 3. The first two steps are critical and imperative; without a consensus on the priorities and risks, it will be impossible to set explicit quality and cost criteria, which in turn are necessary to make decisions. Priorities have been discussed above, but risks have not. The major quality risks in mixed-mode surveys are incomparability in time and incomparability between population subgroups. To most users, comparability is as important as accuracy of survey statistics.

Mode-specific biases in selection and measurement may create such mode effects in time or between subgroups. Adaptive designs, therefore, must account for mode effects in time series. In practice this may imply that a constraint is added on the size of mode effects caused by modes relative to a benchmark. Alternatively, minimization of method effects due to modes could be the objective itself, subject to constraints on budget and precision. The major cost risks are budget overruns and/or heavy workloads on interviewers. In fact, dynamic adaptive survey designs, in particular the levels 3 and 4 designs, mainly arose in order to manage costs during data collection. These designs keep a constant watch on effort and costs. In adaptive designs, it means that costs are included as constraints, usually in a conservative way.

Interviewer workloads are another imminent risk in sequential mixed-mode designs where interviewer modes follow self-administered modes. Due to sampling variation in the first mode(s), the workload becomes to some extent unpredictable.

Adaptive survey design typically uses dashboards to monitor data collection, e.g., Kreuter (2013). Quality indicators to make decisions focus on nonresponse and employ estimated response propensities in relevant subgroups. Examples are CV (Coefficient of variation) and R-indicators (Schouten, Cobben and Bethlehem 2009) that transform response propensity variation to a top-down metrics. It would go beyond the scope of this paper to give an account of indicators. Chapter 10 of Schouten, Peytchev and Wagner (2017) provides a detailed discussion and explains the links between indicators. As costs for one survey are often hard to separate from other surveys, dashboards often consider indirect indicators, such as number of calls, visits, reminders, and so on. Schouten and Shlomo (2017) distinguish different optimization strategies and decision rules:

- Trial-and-error: There is no explicit set of quality and cost indicators, and mode allocations would be based on expert knowledge and historic survey data. An example is Luiten and Schouten (2013), where web, paper and telephone are allocated to different subgroups. The advantage of this approach is its simplicity. The disadvantage is its unpredictable outcome and subjective nature
- Case prioritization: Response propensities are estimated and sorted in ascending order. The lowest response propensities are allocated first to follow-up modes. This approach will be demonstrated in the HS/EHIS case study. The advantage of this approach is the link to response propensities. The disadvantage is the risk of allocating effort to unsuccessful follow-up and the lack of an explicit quality criterion
- Stopping rules based on quota: Follow-up in strata is based on quota, say 50% or 60% stratum response rates. When thresholds are met follow-up is stopped. Implicitly, the approach attempts to obtain equal stratum response rates. The advantage is again its relative simplicity. The disadvantage is the unpredictable fieldwork effort. This approach is simulated by Lundquist and Särndal (2013) for the Swedish EU-SILC
- Mathematical optimization: The most advanced but also demanding approach is to formulate an optimization problem in which mode allocation probabilities are decision variables. The optimization problem chooses a quality or cost indicator as objective function and optimizes subject to constraints on other indicators. The advantage of the approach is it transparency and link to indicators. The disadvantage is the requirement to estimate all components in quality and cost functions.

Adaptation of mode allocation is completely dependent on the possibility to link relevant auxiliary data, either from frame data, linked administrative data or paradata. In settings where frame data and/or linked administrative data are rich, such as in many Nordic countries, stratification may even require parsimony in choosing strata. In settings where there are barely any frame or administrative data, it may simply not be possible to employ static designs. In the US, this has been the case for a long time, and focus has been on dynamic designs, therefore.

Paradata is less setting- or country-dependent than frame data and administrative data but has limitations when the survey mode is the main design feature. One reason is that self-administered modes offer very little paradata for non-respondents; thus, giving little opportunity to adapt in follow-up waves. Furthermore, paradata are mode-specific, so that designs that have concurrent mode elements may offer different paradata for use in follow-up modes.

Given auxiliary data, the main question is how to stratify the population. Basically, there are three approaches. The first is to model nonresponse and choose strata that explain nonresponse best. The second is to model one or a few main survey variables and choose strata that explain these variables. The third is to model costs and choose strata that are most heterogeneous in costs. The approaches may also be combined. At the Central Agency for Statistics Netherlands (CBS), the practice is to pre-select auxiliary variables from the weighting model, i.e., that relate to survey outcome variables, use this set to model nonresponse and then form strata.

3. Implementing and managing the multimodal data collection

3.1. Survey communication in multimode

Survey communication comprises all contacts with respondents, including the contact phase, response phase, and follow-up phase. Different communication tools and modes are generally used for pre-notifications, invitations and reminders. Thus, the mixed-mode concept can be extended to the survey communication strategy.

Communication strategies are a fundamental part of the survey design and are, in general, planned to be compliant with aims and constrains of the data collection design. They are, therefore, a component/factor to be considered in the decision process aimed at defining the 'optimal data collection' strategy. In this paragraph some 'ingredients' of the communication strategies that might be used in the decision process. Incentives are a way to increase survey participation. Whether they are effective or not should be evaluated survey by survey.

Communication strategies are devoted to informing respondents about a survey in order to increase their participation. There are various channels to reach individuals (direct to respondents, advance letters, reminders to non-respondents and to break-offs, or to a vaster audience, institutional website, advertising campaigns on TV/radio stations/social networks.

3.2. Pre-contact and multiple modes

In the pre-contact phase, mixed-mode systems are used for prenotification and recruitment. A classic example is the use of paper advance letters in telephone surveys. This mix is chosen because it is easier to establish legitimacy and trust through an official letter that has a letterhead, contact information and a signature, than through a mere voice over the phone. That advance letters indeed reduce nonresponse in telephone survey. In establishment surveys the opposite mix – a telephone pre-contact before a mail or web survey – has been found to be effective. Business surveys face different methodological issues than household surveys, and a telephone conversation is far more efficient than a letter in getting past gatekeepers and in identifying the targeted most knowledgeable respondent in the establishment. As the actual data collection in these cases is unimodal, the mixed-mode system has no implication for measurement error at all but will reduce nonresponse error: a win –win situation.

Advance letters and reminders can be sent in different ways: using the "traditional" contact means (postal or through interviewers) or "digital" ones (e-mail or SMS). Usually the advance notification/invitation is a paper letter while reminders, that have in general shorter contents, can be sent by e-mails or by SMS. Text messages are still not frequent, mainly because of non-availability of mobile phones numbers; moreover, some INSs prefer to use them only for smartphone designed surveys. The different tools (paper letters, e-mails, text messages, etc.) have a different impact on the survey costs and on the scheduling of shipments; the latter, in turn, depends on the duration of the fieldwork. Moreover, the number and timing of reminders may also depend on the type of survey design -sequential or concurrent or both - and need to be tailored with it.

The advance letter is always sent, despite cost constraints. It is in fact used to inform sample units not only about the survey, but also about the data collection modes, and eventually on whether these can be chosen. It has therefore an extremely important role in raising participation and promoting the success of a survey. Paper letters look like the most suitable means, even though also flyers are used sometimes. As to timing, advance letters need to arrive close to the start of data collection and, if possible, on specific days of the week, those that appear to be the most effective in terms of survey participation. The choice of which days of

the week depends on the culture and habits of each country (for example: beginning of the week for Slovenia, Friday for the Netherlands). To capture respondents' interest, the advance letter may contain an incentive (or the notice that this will be received or raffled after questionnaire completion) or may be sent along with an information brochure, where there is detailed information on the survey and some interesting results from previous surveys.

3.3. Reminders and follow-up phase in multiple modes

An efficient tool to increase response rates is reminders, and, therefore, they are commonly used. Sometimes reminders employ the same mode of contact, for instance a postcard in a mail survey, an e-mail in an Internet survey, or a telephone reminder in a telephone survey. Sometimes a different mode of contact is used for the follow-up contact. For instance, costs and time constraints may prohibit in-person follow-ups in a face-to-face survey. A second, different mode for the follow-up may also lead to additional information about the sampling frame or improved contact information, (e.g., a telephone reminder to an Internet or mail survey) and changing modes may improve the attention or novelty value of the reminder. Persuasion letters are another form of a mixed-mode follow-up contact. Reluctant respondents in face-to-face and telephone surveys are sent a special persuasion letter, emphasizing the importance of complete measurement for the survey. Persuasion letters should be tailored to subgroups of refusers if possible, and in general communicate the legitimacy and importance of the study to the reluctant respondent. If the mode change only involves the follow-up reminder, potential measurement errors due to mode effects are avoided, while the potential benefits of a second mode of communication can be exploited: a win–win situation, just as when using prenotifications. However, when the follow-up is also used to collect additional data, a potential mode effect may occur. If the full questionnaire is administered in another mode to reduce nonresponse, it is a sequential mixed-mode approach. A special case is when a shortened questionnaire is used to reduce response burden when collecting data on nonrespondents. These data may shed some light on selectiveness of nonresponse and can be used for weighting and adjustment. In this case the researcher should again decide which source of error is the most important, and whether it is mode effects or nonresponse that is the worst of two evils.

Reminders to non-respondents, that include also break-offs, are generally sent by post (paper letters), whatever data collection mode is used. E-mails are mostly used when CAWI is involved. In general, no more than one or two reminders are used; usually the second one depends on the duration of data collection. More than two reminders are rarely used. Generally, reminders are sent after seven days since the beginning of the survey or, for CAWI break-offs, since the last access to the questionnaire has occurred. Reminders scheduling shows some differences according to the data collection designs: in general, for sequential design, the "7 days rule" is often used; for concurrent designs when modes can be chosen by respondents, reminders are sent after a longer period of time (10-14 days): this better allows to give respondents enough time to express their preference and to the INS to organize the data collection accordingly. Besides, for sequential or 'partly sequential-partly concurrent' designs, reminders are used to inform non-respondents about the availability of new data collection modes. CAWI break-offs receive a higher number of reminders than other modes break-offs because of availability of e-mail addresses. When CAWI is used, the last reminder is sometimes sent just few days before the close of the survey.

3.4. Incentives

Some European INSs offers incentives to increase response rates and to thank respondents for their time. Some countries, usually apply this strategy for almost all the ESS surveys, while the remaining ones only to specific situations (one or two surveys).

Incentives typically comes in two main forms: monetary, in general voucher, and, more frequently, non-monetary. Examples of non-monetary incentives are: calculators, pens, shopping bag, books of stamps, etc. Sometimes respondents are proposed to participate to lotteries of iPads, Tablets or gift cards. Whatever the form, the common practice is to give incentives only to those households/individuals that have completed the questionnaire. This is, in fact, a way of containing costs and, at same time, to stimulate the participation in the survey. In few cases incentives are offered unconditionally. In general, incentives are offered to all respondents with no diversification. In two cases different strategies are used for specific sub-groups of sample units. Specifically, Germany and Austria for EHIS differentiate the incentives according to the different propensity to participate in the survey, in order to try to reduce differences among strata. Specifically: · In Germany, "younger participants were offered a 10€-voucher after completing the questionnaire. This was done because participation rates in younger age groups are remarkably lower than in older age groups. Older age groups were offered to participate in a lottery (50€-voucher), once the questionnaire was filled out". · In Austria, "Up to 30€-vouchers are planned after finish. There is currently an internal discussion about paying different amounts depending on hard-to-reach areas. This was necessary during last data collection period in order to get equal regional distribution."

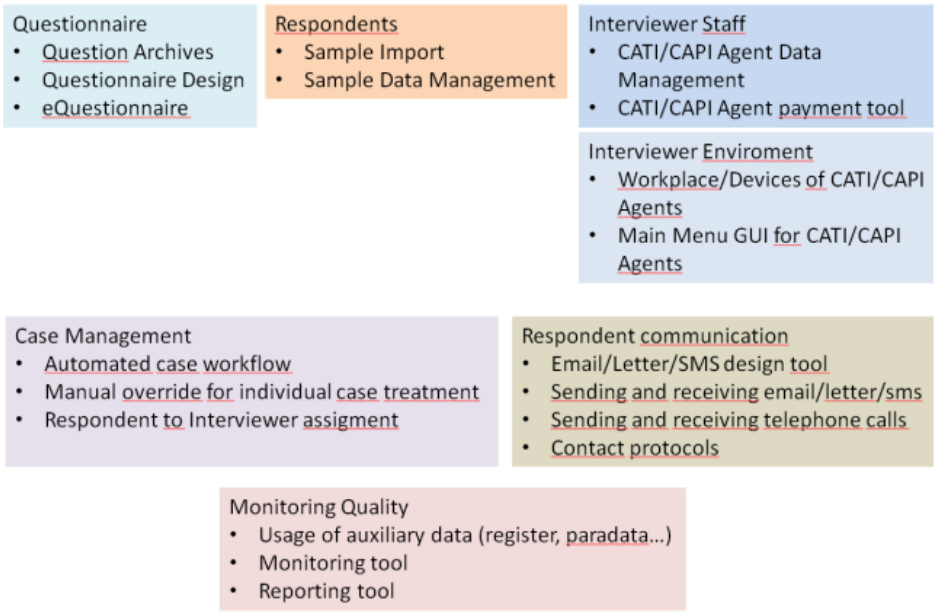
The use of incentives seems to be strongly linked with the data collection technique. In fact, among surveys giving an incentive, there is a prevalence of mixes with self-administered modes, among which CAWI is more frequently used. Moreover, the habit of offering incentives is more common in concurrent or 'partly concurrent- partly sequential' designs respect to sequential ones.

Some conclusions can be drawn from the EU INSs experiences in the use of incentives: the use of incentives divides the INSs in two almost equal parts: those offering and those not offering incentives. Moreover, the majority of INSs that offer incentives prefer to do it for some specific surveys especially those with a high response burden. This might suggest that i) their efficacy in enhancing response rates is not given for granted by all INSs and ii) incentives might have a high impact on costs and, therefore, in case of budget constraints, they are limited to those surveys where respondents cooperation is more difficult to obtain (i.e. high burden); · Incentives can be differentiated per sub-groups of sample units, in order to reduce differences in participation rates or proposed only to specific groups of respondents to limit costs; · Whatever mixed-mode strategy is used, incentives are more frequently offered when self-administered modes are adopted. This is probably because of the absence of interviewers who generally boost respondents' cooperation.

3.5. Not only data collection: IT components and general recommendations.

As a final suggestion for the chapter, we provide a chart containing possible IT components that can be involved in data collection.

Figure 3 - Chart of IT components for data collection



There is great variation in terms of available contact modes and contact information, and the constantly changing (albeit at different paces) technological, social, economic, legal, and other conditions make it difficult to offer very concrete recommendations regarding survey communication strategies. Clearly, access to quality register information on addresses, mobile phone numbers and e-mail addresses is of great benefit for the contact and follow-up phases.

INSs should integrate information on each contact attempt in each contact mode in the case administration systems to be better able to evaluate and make evidence-based decisions during the data collection process.

Five tentative recommendations are:

1. Consider using all available contact and response modes
2. Use proper and coherent design principles in information materials and questionnaires
3. Be prepared to experiment and continuously develop contact strategies for initial contact and follow-up phases
4. Design questions and questionnaires for mixed-mode data collection; do not constrain them to one mode or type of communication. For mixed-mode involving web, consider shortening questionnaires and avoid formats that are not mobile or web friendly
5. Integrate communication with respondents for all contact modes in case administration systems.

4. Multimode effect

Well-designed multi-mode surveys may reduce costs and non-sampling errors (coverage, nonresponse, and measurement errors). However, possible mode selection effects (resulting from errors of non-observation), 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 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 always remains unknown. 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. Körner (2014) produced a report on the definition, identification and analysis of mode effects.

Identification of mode effects refers to describing and explaining why different modes may exhibit relative selection and measurement effects. 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 straight lining answering behavior in different modes and the measurement effects this can induce in survey estimates. Many reasons can be conjectured or shown to cause mode effects. Here, focus is on the assessment of mode effects, and on adjustment methods. We want to stress 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 conclude 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).

4.1. Assessing the mode effect

Mode effects (both selection and measurement effects) need to be properly assessed and adjusted in order to ensure accurate estimates. Both the ESS country experiences and the literature review on methods for mode effect assessment and adjustment show that activities and published literature on mode effect assessment are more widespread than on mode effect adjustment techniques. Mode assessment analyses are sometimes limited to quantifying the total mode effect. 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. Sometimes assessment of mode effects may be sufficient, but when detected, some effects may need to be corrected for measurement effects. Methods to disentangle measurement and selection mode effects are needed.

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 sociodemographic 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.

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, characterized using another or several other modes. Assessing mode effects presents several issues, e.g., confounding of selection and measurement is a key problem (Jäckle et al. 2010). 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. Some assessment methods extend naturally to adjustment techniques; hence, adjustment methods usually incorporate – implicitly or explicitly – an assessment of mode effects.

To disentangle selection and measurement bias in face-to-face and telephone data collection modes was done applying an interview re-interview approach and analyzed with a latent class model (Biemer 2001). 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.

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 (Van den Brakel, 2008) 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. In experiments aimed to assess differences between data collection modes, 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.

Alternatively, a propensity score matching approach can be used where respondents from one mode are matched to respondents from a different mode, as in Lugtig et al. 2011. The difference in survey estimates on the matched samples is taken to be the measurement effect. This assumes that the covariates in the propensity score models explain the selection fully. Similarly, one can consider splitting samples and their follow-ups in a disjoint fashion and compare selection effects of the resulting mixed-mode designs, as in Klausch et al. (2015). In this way, total bias, and bias resulting from selection and measurement can be evaluated. The key to success in these analyses is the experimental design allowing for fully explaining selection effects.

It is also possible to investigate whether mode preference could be a helpful covariate to explain selection effects in mixed-mode designs (Vandenplas et al. 2016). 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. 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. Important measurement differences seem to exist between interviewer and non-interviewer modes.

The common theme and challenge is still 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

Mode assessment studies are more insightful when they separate the total mode effect into selection and measurement components. Experimental designs specifically aimed at disentangling the two effects, are preferable, but costly, and hence less common. Such designs include parallel, independent surveys, embedded experiments, and re-interview studies.

4.2. Adjusting for the mode effect

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 re-weighting and calibration approaches, imputation, and prediction approaches.

Appropriate adjustment methods require the separation of selection and measurement effects in order to correct each, potentially by different types of approaches. Adjustment methods in the context of mixed-mode designs are aimed at correcting survey estimates for undesired mode effects, typically bias resulting from 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 constructs 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 need adjustments to remove the bias with respect to the benchmark.

Adjustments can be applied by using different approaches:

- Weighting approaches seek to correct through applying adjustments to the usual survey weights.
- In some situations, one could use an imputation approach where counterfactuals are imputed: it consists in the application of prediction methods that attempt to predict at the item level 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 additional correction.

Measurement and selection mode effects are sometimes confounded in mixed-mode designs. The two effects can be separated in experimental studies. Experimental studies are rather rare because of the associated costs. However, 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 (which are assumed to be mode insensitive) 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.

Re-interview designs are a potential tool to estimate and adjust for mode-specific measurement bias. In the MIMOD ESSNet (see https://www.istat.it/en/research-activity/international-research-activity/essnet-and-grants) you can find the results of a cost-benefit analysis for two surveys, the Dutch Health survey and the Dutch Labour Force survey, discussing the utility and validity of re-interviews. The conclusions of the study are that for the Labour Force survey

a re-interview may not be useful due to relatively small measurement differences, while for the Health survey it may be useful. The paper considers mixed-mode re-interview designs, in which a sample of respondents to the regular survey is invited to participate in one of the other modes that is employed. More specifically, it considers sequential mixed-mode designs, where some of the modes are offered only to non-respondents in the other modes.

However, re-interview designs may be very costly, especially when face-to-face is included as a survey mode. The crucial question is whether benefits outweigh costs, i.e., whether the potential increase in accuracy of survey statistics is worth the investment. The answer to this question depends heavily on the purpose of there-interview, i.e.. assessment versus adjustment, the size of the mode-specific measurement biases, and the relative costs of the modes. Re-interview designs also make several assumptions that will not hold for every setting. The two case studies mentioned show that a re-interview can be profitable under both the design perspective and the adjustment perspective

Mode adjustment methods are methods that adjust survey estimates obtained from mixed-mode designs to correct for mode effects induced using 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 here 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, 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 minimizes 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) proposes 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 16 differences between modes. Which covariates can be used for this purpose remains an ad-hoc choice.

4.3. Strategy to control mode effect

The key elements that characterize a methodological strategy to deal with mode effect estimation and adjustment are: i) the design (experimental design or observational studies); ii) auxiliary data (from administrative data/frame data/paradata) or covariates; and iii) a set of assumptions.

This section provides some general guidance related to the design of strategies to control for potential mode bias/mode effect in mixed-mode surveys. In general, in deciding if and how to estimate mode effects and/or to adjust for their biasing effects on survey results, there are three key decisions to be made:

- how to assess whether mode effect adjustment is beneficial?
- the multi-dimensionality of a survey: what key estimates and population parameters of interest need to be evaluated?
- the time perspective: is the survey repeated and can effects be assumed constant?

Without a consensus on how quality and costs are quantified, it is, generally, hard or impossible to make an objective choice between unadjusted and adjusted estimates. Since true values are often unknown, one inevitable sub-question is what mode is chosen as benchmark for measurement. In other words, to what benchmark is the adjustment made. The answer to this question may ideally be different for different survey variables. However, in practice, a single choice must be made. Surveys obviously contain many questions, so that it is imperative that stakeholders select the most crucial variables in order to limit complexity of decisions. It makes a big difference when surveys are repeated and decisions to adjust may stretch over a longer time period. When defining a methodological strategy to deal with mode effect estimation and adjustment, there are essentially three main requirements to be defined:

- a design
- auxiliary data (from administrative data/frame data/paradata) referred to as covariates
- a set of assumptions.

Concerning the design, Table 3 below summarizes the type of design, experimental and non-experimental, within which it is possible to carry out analyses aimed at either assessing or adjusting selection and/or measurement effects.

Experimental designs allow controlling for selection effects, and hence the unbiased assessment of measurement differences between modes. An experimental design is of course optional and not standard practice.

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 that are observed without error are available, potentially obtainable from external data sources.

There are two types of **auxiliary data**: data informative about selection into modes and data informative about measurement within a mode. In causal inference literature, the two have been referred to as backdoor and front door variables and may be employed to improve external and internal validity. Data informative of selection typically consist of linked frame data and administrative data and paradata from the contact and participation processes. Data informative of measurement consist of record check or validation data, repeated measurements and paradata from the answering process.

Table 1 - General scheme of survey settings and objective of the analyses

Design type	Objective
Experimental	
Parallel independent surveys (single mode and mixed mode)	Mode assessment Mode adjustment
Re-interview study - repeated measurement designs	
Other (Embedded experiments, Split sample designs)	
Non-experimental	
Observational studies (Mixed-mode design only)	Control for selection effects through weighting or regression-based inference methods Adjusting for measurement effect

Assumptions may be divided into three types: assumptions about the explanation of the missing data mechanism due to mode selection, assumptions about the explanation of measurement differences due to modes, and assumptions about the absence of experimental influence on (non)respondents in experimental designs. It is straightforward to mention that when the available covariates do not fully explain the selection mechanism, the decomposition of the total mode effect into selection and measurement effects is incorrect. Concerning the assumptions, they depend on the type of auxiliary data and type of design. It is straightforward to mention that, with the same auxiliary data and design, different estimation

strategies/methods should/must not be too influential. In mode effect estimation, the following steps may be followed:

1. Identify the main quality and cost criteria
    a. What benchmark is chosen for measurement?
    b. Is it sufficient to consider accuracy (i.e., MSE) or also comparability in time and/or between subgroups?
    c. What is the time horizon for which the mode design and budget are fixed, and mode effects are estimated?
    d. What are the key variables/population parameters of interest?
2. Decide whether mode effect estimation serves explanation only, design choice or adjustment
3. Identify available auxiliary data that is informative about
    a. Mode selection
    b. Mode measurement
4. Evaluate anticipated validity of assumptions for mode selection, mode measurement and absence of experimental influences
5. Decide whether an experimental design (such as re-interview or parallel run) is required and feasible to serve the purposes of the mode effect estimation
6. Conduct experimental designs if deemed feasible and necessary

Based on these assumptions, in the following of this section general guidance when selecting methods to deal with the mode effects are provided. We make two side remarks. First, we note that mode effects do not refer to biases only but may also affect precision. Modes may affect, for example, motivation and concentration. Less motivated or concentrated respondents may give less reliable, i.e., noisier, answers, leading to a loss of precision. Furthermore, interviewer effects have been a widely studied source of potential variation in survey statistics. In this report, we focus on bias adjustment. Second, we note that mode-specific measurement differences must be prevented above all through careful questionnaire design and testing. We, thus, assume that mode effect estimation is conducted to estimate remaining differences that are hard to detect and prevent in questionnaire design.

In practice, biases and mode effects can be estimated according to two main approaches:

1. record check approach when the true scores are available from an external source;
2. measurement benchmark mode approach that requires the choice of a reference mode to produce best answers for a question.

The first approach is rarely feasible in practice, but it allows estimating all biases and effects. The second approach assumes the benchmark measurement equals the true scores. The methods referred here mainly follow this last approach.

The following schemes outline the methods which can be applied for different objectives of the study in different survey/experimental contexts, given the following types of analyses to be carried out:

- Analysis of total mode effect (Table 2), and
- Analysis to disentangle measurement and selection effects (Table 3).

The objectives of the study considered are:

1. assessing differences between estimates obtained based on data collected through different survey designs (single-mode and mixed-mode), in order to evaluate the total mode effect and the measurement equivalence

2. analyzing the response processes and evaluation of the bias caused by the total nonresponse (selection errors)
3. assessing mode effect - disentangling measurement and selection effects.

Table 2 - Analyses of total mode effect

Objective of study: Assessing differences between estimates obtained based on data collected through different survey designs (single-mode and mixed-mode), in order to evaluate the total mode effect and the measurement equivalence		
Method	**Analysis**	***Context* / Conditions**
Regression modelling approach to test whether design has a significant effect on the mean or distribution of the item (Martin and Lynn, 2011)	Univariate analysis of items to evaluate the impact on marginal distributions of mixed-mode design	– *Parallel independent surveys* Appropriate statistical models and tests
Tests on differences in the estimates (Martin and Lynn, 2011)	Univariate analysis to highlight significant differences in the estimates calculated on the two sample designs	– *Parallel independent surveys* Appropriate statistic tests for independent samples
Tests on indicators of *completeness* (item nonresponse) Tests on indicators of *accuracy* (comparisons with external data) (Jackle *et al*., 2010)	Analysis on differences in the quality indicators	– *Parallel independent surveys* Appropriate statistic tests
Multi-group confirmatory factor analysis (Martin and Lynn, 2011; Hox *et al*., 2015)	Analysis of the measurement equivalence when concepts are measured through more than one variable	– *Parallel independent surveys* – *Mixed mode survey design* Identification of the latent structure of the phenomenon, Control of selection effect
The proportional odds modelling technique (or parallel regression model, grouped continuous model) (Jackle *et al*., 2010)	Analysis to assess measurement equivalence of ordinal data on comparable samples	– *Parallel independent surveys* – *Mixed mode survey designs* Control of selection effect Validity of model assumption about covariates (covariates "shift" the distribution of responses proportionately across all categories)
Regression modelling approach whit one or more predictor variables and a binary indicator of single-mode and mixed-mode respondents (Martin and Lynn, 2011)	Multivariate analysis on estimates of the association between variables	– *Parallel independent surveys* Appropriate statistical models and tests on significant interaction effects

Objective of study: Analysing the response processes and evaluation of the bias caused by the total nonresponse (selection errors)		
Method	**Analysis**	**Context / Conditions**
Tests on the response rates respect to some characteristics of sample units (Jackle *et al*., 2010)	Analysis on the response rates	– *Parallel independent surveys* – *Single and mixed mode designs* Appropriate statistic tests for independent samples
Summary statistic tests	Analysis of deviations from mode independence (absolute and relative selection error per benchmark variable)	– *Parallel independent surveys* – *Comparison between single mode and mixed mode designs* Appropriate statistic tests
R-indicator, Conditional and Unconditional partial *R*-indicator (Klausch *et al*., 2015; Schouten *et al*., 2011; Shlomo and Schouten, 2013; Schouten, *et al*., 2017)	Analysis of the representative response (absolute selection error for sets of benchmark variables)	– *Parallel independent surveys* – *Single and mixed mode designs* MAR assumption for Response model
Tests on the differences between benchmark variables (true value) and estimates (Roberts and Vandenplas 2017)	Analysis on benchmark variables known for selected sample units	– *Parallel independent surveys* – *Single and mixed mode designs* Appropriate statistic tests

Table 3 - Analyses to disentangle measurement and selection effects

Objective of study: Assessing mode effect - disentangling measurement and selection effects			
Method	**Analysis**	**Conditions**	***Context***
Weighting • Propensity score (PS) • Calibration • Post-stratification (Vandenplas *et al.*, 2016; Rosenbaum and Rubin, 1983; Vannieuwenhuyze, *et al.*, 2014)	Analysis based on response model to control for respondent characteristics (comparable samples in MM)	MAR assumption Mode-insensitive auxiliary variables Balancing assumption in PS	– *Mixed mode survey designs (observational studies)*
Regression model (Kolenikov and Kennedy, 2014)	Model analysis to estimate measurement and selection errors	Mode-insensitive auxiliary variables in the model to control selection effect	– *Mixed mode survey designs (observational studies)*
Other methods *-double robust estimation* that combines an outcome regression with a propensity score model - *matching*	Model to estimate causal effect	Appropriate statistical models	– *Mixed mode survey designs (observational studies)*
Instrumental variable approach (Vannieuwenhuyze *et al.*, 2010)	Analysis based on benchmark single-mode design	Validity of comparability and representativity assumptions	– *Parallel independent surveys*
Re-interview (Biemer, 2001)	Analysis based on re-interview data, administrative data and paradata. The response of each mode is calibrated to the combined response of the re-interview and follow-up. Measurement effect (ME) is estimated as remaining difference between modes. Selection effect (SE) is estimated using mix of re-interview data, administrative data and paradata.	Re-interview does not affect measurement behavior of respondent. Nonresponse to re-interview is unrelated to survey variables of interest given administrative data and paradata.	– *Re-interview of subset of mixed-mode respondents (experimental design with sequential mixed mode survey)*

The following Table 4 outlines the methods which are applicable for adjusting for mode effect in experimental contexts (re-interview, parallel single mode), or when auxiliary data from either administrative data or paradata are available, or in the case of longitudinal or repeated over time surveys. Table 5 presents, for the standard covariate-based adjustment approach, a set of methods that can be used to correct selection and/or measurement effects.

Table 4 - Approaches to adjust for mode effects

Objective of study: Adjustment methods			
Method	**Data requirements**	**Assumptions**	**Advantages/Disadvantages**
Standard Covariate-based adjustment	• Sampling frame data • Paradata • Survey responses	Missing at random potential outcomes (MAR) Exogeneity of auxiliary data	Too strong assumptions in many settings (-) Adjustment on individual level possible (+)
Time-series stabilization/ mode calibration (Buelens and van den Brakel, 2015, 2017)	Repeated cross-sectional / longitudinal survey	Independence of measurement and selection error Time-stability of measurement error (ME)	Does not decompose (-) Avoids ME estimation problem (+) Strong assumption on mode contributions (not fluctuate) (-)
Instrumental variable method (Vannieuwenhuyze *et al.*, 2010)	Single-mode reference survey parallel to mixed-mode	Single-mode and mixed-mode survey have same selection bias (SB)	Avoids MAR and exogeneity assumption (+) Representativeness assumption usually implausible (-) Not available for >2 modes
Re-interview method (Klausch *et al.*, 2017)	Re-interview of subset of mixed-mode respondents	Measurement equivalence	More plausible MAR assumption (+) MNAR estimators available (+) Measurement equivalence traded off against true score time-stability (-)

Table 5 - Methods to adjust for mode effect

Objective of study: Adjusting selection/measurement effects in MM (observational studies)		
Method	**Aim**	**Conditions**
Weighting - Propensity score - Calibration - Post-stratification (Vandenplas *et al.*, 2016; Rosenbaum and Rubin, 1983; Austin, 2011; Vannieuwenhuyze, *et al.*, 2014)	To equate samples To correct selection effect	Ignorability of selection mechanism (MAR) Mode-insensitive auxiliary variables Measurement error negligible
Regression (Kolenikov and Kennedy, 2014)	To estimate measurement and selection effects To correct measurement error	Appropriate statistical models
Other methods *-double robust estimation* that combines an outcome regression with a propensity score model - *matching*	To estimate causal effect To correct measurement error	Appropriate statistical models
Multiple imputation		
1.Multiple (standard) imputation	To predict counterfactual data (potential outcomes) To correct measurement error	Choice of benchmark mode MAR assumption
2.Multiple imputation with response and selection models proposed by Suzer-Gurtekin *et al.* (2012)		Choice of benchmark mode Sequential design and two modes (Possibility – non-ignorability of selection mechanism)
3.Fractional multiple imputation proposed by Park *et al.* (2016)		Sequential design and more than two modes Possibility – non-ignorability of selection mechanism

It must be reminded 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 results from mode-specific selection 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 conclude 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).

4.4. A case study

Here we present a set of analyses for assessing and adjusting mode effects in a specific survey context. The methods are framed in the review of methodologies to cope with mode effects. Mode effects are analyzed for the ISTAT "Multipurpose Survey on Households-Aspects of daily life", (Multipurpose survey hereafter), for which an experimental design was considered. Here you find some comments on the results of the applied methods and discusses the advantages and limitations of the proposed approaches. In 2017, the mixed-mode was introduced for the first time in the Multipurpose survey; the web technique was added to the traditional PAPI technique in a sequential design. A parallel single mode PAPI design was planned to allow for an assessment of mode effect on two independent samples collected with different techniques.

This experimental design allows the researchers to disentangle selection and measurement effects by using mode insensitive auxiliary information. The results highlight that the mixed mode design catches better the overall population, being more "representative" than the single mode design. When the assessment of mode effect is carried out for specific variables, the results generally provide an explanation for breaks in the series due to both selection and measurement effect. The detection of measurement effects provides a useful advice for the planning of future survey editions, i.e., the increased coverage due to mixed-mode design.

The set of analyses presented in the MIMOD ESSnet deliverable (see details at https://www.istat.it/it/files//2011/07/WP2-deliverables.zip in Deliverable 2), can be viewed as a sequence of steps usable by researchers of other INSs to carry out an assessment of mode effects in similar application contexts. They cover the different approaches applicable in this specific survey context, though without claiming to be exhaustive. This experience was very useful to ISTAT because it was the occasion for experimenting several methods for assessing and adjusting mode effects in an experimental context, usually not very frequent. A similar research path can be followed to evaluate the impact of the switching from single to mixed-mode. Indeed, the underlying effort is hardly compatible with the usual resources and the timing of a statistical process. For repeated surveys an accurate planning of the data collection phase is advisable, in order to prevent the measurement effect, which is the main drawback of the mixed-mode.

References

Bethlehem, J. (2009), Applied Survey Methods: A Statistical Perspective, Wiley Series in Survey Methodology, 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. 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

Calinescu, M., Schouten, B. (2015), Adaptive survey designs to minimize mode effects. A case study on the Dutch Labour Force Survey, Survey Methodology, 41 (2), 403 – 425

Calinescu, M., Schouten, B. (2016), Adaptive survey designs for nonresponse and measurement error in multi-purpose surveys, Survey Research Methods, 10 (1), 35 – 47

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

Corness, C., Bosnjak, M. (2018), Is there an association between survey characteristics and representativeness? A meta analysis, Survey Research Methods, 12 (1), 1 - 13

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

De Leeuw, E.D., Hox, J., Dillman, J. (2008), International Handbook of Survey Methodology, European Association of Methodology Series, CRC Press Books

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). Internet, phone, mail, and mixed-mode surveys: The tailored design method. Hoboken, NJ: John Wiley & Sons

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

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R. (2009) Survey Methodology, Wiley & Sons

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

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

Kreuter, F. (2013), Improving Surveys with Paradata. Analytic Uses of Process Information. WIley Series in Survey Methodology, Wiley & Sons

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

Lundquist, P. Särndal, C.E. (2013), Aspects of responsive design with applications to the Swedish Living Conditions Survey, Journal of Official Statistics, 29 (4), 557 – 582

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

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

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, J.G., Cobben, F., Bethlehem, J. (2009), Indicators for the representativeness of survey response, Survey Methodology, 35 (1), 101 – 113

Schouten, B., Peytchev, A., Wagner, J. (2017), Adaptive Survey Design, Series on Statistics Handbooks, Chapman and Hall/CRC

Schouten, B., Shlomo, N. (2017), Selecting adaptive survey design strata with partial R-indicators, International Statistical Review, 85 (1), 143 – 163

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

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

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

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

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

Competence makes a difference!

Project selected under the Administrative Capacity Operational Program, co-financed by European Union from the European Social Fund