Working Group Position Paper

on mixed-mode data collection in household surveys
Minutes of the 8th meeting
September 2nd, 2021

Participants: Fiona O'Callaghan (Ireland), Clelia Romano (Italy), Claudia de Vitiis (Italy), Nadja Lamei (Austria), Thomas Burg (Austria), Patrick Sillard (France), Zoltán Vereczkei (Hungary), Ferenc Mújdricza (Hungary), Petra Fekete-Nagy (Hungary), Gwennaëlle Brilhault (France), Andreja Smukavec (Slovenia), Martina Stare (Slovenia)

Excused: Fiona O'Riordan (Ireland)

Minutes of the 7th meeting are validated (with a correction by **G. Brilhault**).

P. Sillard reported on the contributions concerning the analysis of the questionnaire that had been received before the meeting:

- **T. Burg** on the SILC wave 1

- **A. Smukavec** on LFS wave 1

- **C. de Vitiis** on HBS

- **G. Brilhault** on ICT

and concerning the first part of the questionnaire, the analyses of LFS wave2 and SILC wave 2 are missing, waiting for a validation of the waves 1 templates. Concerning the end of the questionnaire:

- **A. Smukavec** produced an analysis of questions 8.1 to 8.3 .

- an analysis was made by **F. Mújdricza**, **Z. Vereczkei** and **P. Fekete-Nagy** for questions 8.4 to 8.9 which mixes raw facts about the questionnaire and recommendation that could be included in the position paper so we will probably have to separate the two aspects.

So, at this stage, apart from LFSW2 which should come from Slovenia and SILCW2 which should be done by Austria, after agreeing on how to analyze the corresponding wave 1s, the only missing point is section 7 (subsections 7.1 and 7.2) which will be taken over by **T. Burg** after the meeting (confirmed).

We also have the motivation part of the position paper written by **Z. Vereczkei** (sent before the 7th meeting of the group). A more detailed analysis of section 1.4 of the questionnaire was also prepared by **Z. Vereczkei**, **F. Mújdricza**, **P. Fekete-Nagy** before the 7th meeting of the group and a summary of this analysis was sent before the 8th meeting.

A. Smukavec presents briefly the conclusions she draw from the analysis it the LFSW1 of the questionnaire. The switch to CATI is clear on Q2 of 2020. In Q3, some country tried to come back to the use of CAPI. Some countries had some problems with responses: the response rate was low, there were some inconsistencies (misunderstandings) when we switch from CAPI to CATI. Concerning the intention of keeping the changes, 10 countries out of 30 told that they are not going to keep the changes, 6 are going to keep the changes. The reasons for keeping are safety reasons, and cost-effectiveness. The main reasons for not keeping are worse data quality and less efficient field monitoring. There is no particular change in the the possibility of choosing the mode of interview. The largest decrease in response rates is observed between Q2 2019 and Q2 2020. For wave 2, the change is lower.

In Wave 1, 18 countries introduced new modes: 13 introduced CATI and 6 introduced CAWI. 13 countries introduced these new modes exclusively due to the pandemic; 4 were already planned. CAWI, when introduced in the first wave, was already used for example in wave 2 or its use was planned before the pandemic.

A. Smukavec also asked to the group two questions about the way of interpreting some questions in the questionnaire and the way of reporting the results.

- Concerning questions 3.2 and 3.4, the reference period is the one mentioned in the previous questions to which 3.2 and 3.4 are connected: that is, respectively, questions 3.1 (before pandemic) and 3.3 (after pandemic).

- In the report, for the year 2019, the figures are almost the same for each quarter, so she proposes to show only Q1 2019. The changes occur from Q1 2020 onwards, so when showing 2020, instead, we show all quarters. And when calculating a change between 2020 and 2019, we should use the correct quarter for 2019, not the first one for 2019. The group agrees with these presentation/calculation rules.

A. Smukavec also suggests adding totals in the tables to show, where relevant, that the percentages shown should, for example, sum to 100.

P. Sillard says that, surprisingly, the changes in response rate levels, in terms of overall mean or median, are not very high between 2019 and 2020 (Table 4.1). We have more or less the same picture for all the surveys. Another interesting point is that the impact of the crisis is already visible in the first quarter of 2020, a little more important in the second quarter, while in the third quarter the impact has decreased a lot (the increases and decreases of the responses are balanced). The LFS is also interesting because it allows the impact of the crisis to be followed as it develops throughout 2020. And of course, in parallel, we have a very clear change in the combination of collecting modes, mainly due to the crisis. This change was mainly not anticipated by the countries. The message is more or less the same for all surveys. The fact is that the LFS gives us a more detailed picture over the period because it is organized quarterly.

C. de Vitiis, concerning HBS, suggests for the first table not to consider the percentage. The same comment applies to Table 2.3. In these two cases, we would therefore put the number of respondents rather than the percentage in each modality (because the numbers are too small).

Another consideration regarding HBS is the confusion made by some countries between "mixed-mode" and "multimode" survey. HBS contains a self-administered diary that implies a way of collecting the information that differs from the one used for other questions in this survey (like the one on the structure of the household for example) collected for example by face-to-face interview. Some countries saw this inner combination of collection mode as a mixed-mode survey which is not the case for us. We know that because some countries reported this point in the note. In the counts, the survey was therefore considered by us to be single mode. Normally these doubtful cases were corrected. But these countries where misunderstandings were detected were removed in the tables where combinations of modes are calculated. Then the number of contributing countries varies from table to table.

Concerning the template that in table 4.2 for example, reporting the changes in response rates before/after through a split into two sets: one for decreases and the other one for increases, **C. de Vitiis** thinks that we should add two columns with the overall change in order to highlight the general tendency.

In general, mixed mode was introduced for HBS because of the crisis, with a decrease in the response rate during the crisis. And countries do not intend to keep the changes in data collection, as these changes were due to the crisis and not desired.

Concerning Table 2.3, **A. Smukavec** wished to clarify what is presented in this table in terms of counting the different combinations of modes. The idea of **T. Burg** and **P. Sillard** was to put the two major modes, considering that the presence of a mode in the set of collected modes should be weighted by the number of associated responses. Here mechanically, a third mode is less significant in the analysis than the first two. One option is therefore to carry over the two main modes. Another option, undoubtedly cleaner, consists in only mentioning the modes whose weight, in terms of respondents, is higher than a lower limit. For example: 5%. In conclusion, for this table, we consider the main modes, up to a maximum of three, provided that all modes considered account for more than 5% of actual responses. We will write this in the reading note of the table. And if there are specificity (like the one pointed out by **A. Smukavec** concerning a country that puts "other" when the mode used is CATI), then they should be mentioned (if deemed necessary by the author responsible for this part of the text) as a comment to the table in the text.

Concerning ICT, **G. Brilhault** explains that the general messages are the same: development of mixed mode (CATI) due to the pandemic. The change was smaller than for other surveys because ICT is a survey for which using "technological" modes seem to be more natural. Then the proportions of the modes already in use changed, rather than a real introduction of new modes due to the pandemic. **P. Sillard** adds that it is interesting to compare ICT with other surveys and it clear that ICT is specific. ICT was already a multimode survey before the crisis (see table 2.2 for example): 19 countries were already using CATI before the crisis and 17 were using CAWI. And the

consequence was probably that the change in response rate was smaller than for other surveys. There is another specificity: it is the possibility of choice let to the respondent that is much higher in ICT than in other surveys.

Concerning the proposal of slight changes in the tables proposed by **C. de Vitiis**, the position of the group is to keep these changes. Generally speaking, as a contribution to the annex, it would be great if the comments would be a bit enriched. We are going to aggregate all the contributions in the annex to the position paper and we will see then what harmonization is still needed at that time. We can also accept that there are slight differences in the reporting tables for each survey.

For the position paper, the main messages coming from the questionnaire should be put in part 3 of the position paper and the annex should present the raw results in the survey.

The analysis produced by **Z. Vereczkei** concerning the end of part 8 (questions 8.4 and following) is almost a contribution to the position paper. It would be great if some tables (reporting somehow the raw results) could be provided for the annex. Maybe tables are not quite relevant because these questions are open-ended. So we may have to work a bit on how the "raw results" should be presented, but that's the idea. The main recommendations made to ESS based on the responses we get from the questionnaire are: (1) the organization of a workshop on good practices, (2) the change of focus of some of the training given at European level, (3) the possibility of setting up an expert group dedicated to these issues.

Concerning the beginning of section 8 (questions 8.1 to 8.3), **A. Smukavec** presents the tables she has prepared. The corresponding questions refer to the access the NSIs have to telephone numbers and e-mail addressees. 4 countries improved the quality of the data base on e-mail addresses. The most interesting part is the open question on coverage. 17 countries reported some difficulties to get access to phone numbers, for young households, etc. 9 countries have access to e-mails addresses. In terms of presentation of the answers for open questions, the choice made by **A. Smukavec** is to report the full text. **P. Sillard** says that care should be taken about the sensitivity of these responses for countries, but the way the responses are reported seems to be good.

The main message in this part is that all countries have some coverage issues even if they get access to e-mail addresses or phone numbers. So **A. Smukavec** suggests that maybe at the EU level, we could propose, in the position paper, something like a basis for statistical institute to get this access.

P. Sillard thinks that it would be great to have tables (or any relevant way) that present the results of questions 8.3 and following.

P. Sillard proposes to prepare a draft for the next meeting (22nd of September). Taking into account that we need discussion on this draft, **P. Sillard** suggests to plan another meeting by the end of September or beginning of October. This meeting at the end of September or beginning of October would then be the last meeting of the group. The group agrees with this scheme. **P. Sillard** asks for the contributions concerning the annex (raw results of the questionnaire): in detail, sections about

waves 2 of LFS and SILC, section 7, tables (or any way of presenting raw results) for questions 8.3 and after; and comments that are seen useful should also be added along the text.