Subject: Survey and questionnaire design

Title: Polish LFS person questionnaire design after IESS FR

Małgorzata Długołęcka, Hanna Strzelecka – Statistics Poland

Introduction

The IESS FR implementation has caused the necessity of designing quite a new LFS questionnaire according to the EU recommendations. The presentation will be devoted to the process of elaboration of the revised version of the Polish LFS questionnaires.

Description of the general approach to outline the revised Polish LFS questionnaires

In the Polish LFS there are used two questionnaires:
- ZG Household File – questionnaire concerning the household and basic demographic data of its members
- ZD questionnaire – individual questionnaire addressed to particular persons aged 15-89 years covered by the survey.

At the beginning, there was carried out a detailed analysis on the provisions of the Regulations (IESS FR 2019/1700 and its implementing regulations) and Eurostat guidelines concerning methodology, topics and variables, as well as the applied filters regarding the domain "labour force". There were also analysed methodological guidelines regarding, i.e. definitions of a household, resident population, etc., which allowed verification of the hitherto procedure for qualification of persons for the survey. To achieve this goal, a path of questions was prepared (ZG 2021 qualification) qualifying particular persons for the survey and particular sections of the ZG File.

Due to the fact that the Polish LFS also uses a paper version of the questionnaires and elaboration of the paper version is more difficult than the electronic one (especially when using complex filters), there was initiated developing a preliminary "paper" version of our forms.

First, the variables required by the new regulations were assigned to the appropriate forms used in the Polish LFS.

All variables from the topic "Person and household characteristics" were placed in the ZG Household File. The variables from the thematic blocks: "Labour market participation", "Educational attainment and background", "Job tenure and previous work experience", "Participation in education and training", "Working conditions including working hours", "Health status and disability" and "Income, consumption and health" were placed in the ZD questionnaire.

Then, the main part of the work began, consisting in appropriately arranging the variables in a given questionnaire so as to meet both the specified conditions regarding the order of variable blocks (flow-charts) and the filters used (so that the appropriate populations were covered by the required variables). Variables obtained for national needs were also added - they had to be placed in a way that did not violate EU variables, but was consistent within given thematic areas, addressed to appropriate populations (i.e. using appropriate filters that did not conflict with EU requirements).

After developing the initial layout of a given questionnaire, the variables were translated into appropriate questions. Explanatory notes for interviewers were also developed. The questionnaire questions were elaborated in accordance with the methodology and recommendations obligatory in the LFS from 2021 onwards. In order to precisely formulate questions for the questionnaires and provisions in the explanatory notes for the interviewers, there were used developed by Eurostat "Explanatory notes" for particular thematic blocks and available model questions for variables.

It should be emphasized that the variables used in the EU LFS are often so complex that more than one question is necessary to obtain information for a given variable or, in the case of a paper form, due to the use of complex filters for some variables, repeating the same question in several places so that it reaches different groups. Some of the final variables are therefore derived secondarily on the basis of several questions.

A preliminary paper version of the new tools was dedicated to testing by statistical interviewers. The testing purpose was assessment of the correctness of the design of the questionnaires (transition paths and specification of adequate populations included in particular questions), transparency and correct understanding of the questions, as well as detecting any appearing difficulties, problems and inaccuracies (noticed by both interviewers and respondents). Reported comments and mentioned errors in completion of questionnaires were related to a large number of complex filters (in the case of the survey implementation through the use of the application CAPI/CATI such problems will not occur). The comments regarding correction of the wording of questions, lack of answer "I don't know" were analysed by the Polish LFS team and possible corrections were introduced to the questionnaires. These tests also helped to improve the elaborated explanatory notes for interviewers.

The IT application was designed on the basis of the elaborated paper version of the tools.

It should be added that in Poland it was decided to implement wave approach only to the module variables (what is obligatory since 2021) – all annual and biennial variables of the core LFS survey are still collected with a quarterly frequency. There are also not introduced simplified data collection principles for surveying selected populations. When making such a decision, the lack of reported problems regarding the burden of these groups of respondents with questions in the survey carried out in 2021 was taken into account, therefore, organisational changes implemented within this scope would not result in potential

advantages, whereas they would pose additional hindrance connected with preparation of new additional solutions in the survey implementation. In addition, they would affect some limitations on obtaining information - which is not desired, particularly in the case of the data on disabled persons for whom the LFS in Poland comprises a valuable source of information on economic activity of this population.

In the Polish LFS all variables are collected by interview. The questionnaire ZD is structured according to the recommendations included in the Commission Implementing Regulation (EU) 2019/2240 – the labour status module is placed at the beginning of the personal questionnaire (the questions on the demographic characteristics are generally included in the ZG Household File that is fulfilled as the first one) and questions on labour status have been prepared according to the flowcharts and model questionnaire elaborated by Eurostat. It should be noted that in the Polish LFS we also included optional submodules – "small jobs" and questions connected with agricultural employment (production mostly for own consumption). The information on main activity status (self-perceived) is collected after the labour status module – it is placed in the last section of the core LFS questionnaire.

Due to the fact that in Poland agricultural work is still significant and in order to meet the provisions of the "Resolution on statistics of work, employment and labour underutilization" regarding employed population, we decided to include questions allowing to obtain information on persons employed in private agriculture (farms owned by natural persons) who allocate the effects of their work exclusively or mainly for own consumption and simultaneously do not have any other job. We didn't use the "agricultural employment recovery" submodule prepared by Eurostat, but we elaborated our own solution how to "detect" these persons in order to exclude them from the employed population. It was caused by the fact that this submodule wasn't included in the flowcharts and wasn't covered even by the model questionnaire (while "small jobs" submodule was included in the model questionnaire). But after obtaining it from Eurostat we noticed that proposed approach is rather opposite than this one applied by us – we ask questions to detect if a respondent can be considered as employed and finally we <u>exclude</u> persons who have only one job and work in private agriculture (NACE) and perform work exclusively or mainly for own consumption (we have additional question on the purpose of producing agricultural goods).

This can be described as follows:

- We check submodule "At work" and, if necessary - "Absences from work" (for persons who declare not performing work in the reference week).
- If a person can be considered as employed >> submodule "Second or multiple jobs" and then we ask about very general NACE activity categories:

    1) other activity than agricultural one,

    2) agricultural activity on a private farm and

    3) agricultural activity outside private agriculture.

- If a person declares that has only one job and works in agriculture on a private farm, we ask additional question what is the purpose of produced agricultural goods.
- If the person's answer is "exclusively or mainly for own consumption", **she/he is excluded from the employed and is being redirected to the path for not employed people**, i.e. to submodules connected with searching for employment.

The approach we use to "correct" the final employed population – i.e. **excluding** the identified people producing agricultural goods exclusively or mainly for own consumption is, in our opinion, a better solution than the opposite one – i.e. <u>including</u> people producing agricultural goods intended for sale or barter.

Some examples of variables which made the process of questionnaire designing more problematic

When designing questionnaires used to collect data, an important issue is not only the wording of the questions, but also the filters for variables that influence the question arrangement. The more complex the filters, the more difficult it is to build the appropriate questionnaire, especially if it is in a paper form, but this complicates the preparation of the correct electronic tools as well. Writing down the conditions for filters is definitely easier than translating them into an appropriate question path in a questionnaire and then in the reverse process - transcoding the data gathered during the survey into required variables.

Some examples of LFS variables with filters which made the process of questionnaire designing more problematic are presented below.

1. NEEDCARE

In the Polish ZD questionnaire we have one question for this variable, but in order to fulfil the filter conditions, it is placed in different places (with separate numbers) for each population included in the filter – questions S6A, S6B, S6C, P13A and P13B. Since 2021 this variable has become very complicated by extending the filter, because now the same people will go through the question twice (we block it in the application, but in the paper version it needs additional notes like before question P13B, because some persons may have answered the P13A question).

The table below shows the transcoding of this variable – from PL LFS into EU LFS codes:

Variable identifier	Variable name	Code	Labels	Filter	Polish LFS – question numbers
NEEDCARE	Main reason why care for children or incapacitated relatives limits labour market participation	1	Relevant care services not available	SEEKREAS = 4 OR WANTREAS = 3 OR FTPTREAS = 3 OR AVAIREAS = 3 Persons whose reason for not searching for employment or for not wanting to work or for working part- time or for not being available to start work is having care responsibilities	S6A=1 or S6B=1 or S6C=1 or P13A=1 or P13B=1
		2	Relevant care services not affordable		S6A=2 or S6B=2 or S6C=2 or P13A=2 or P13B=2
		3	Want to provide care themselves		S6A=3 or S6B=3 or S6C=3 or P13A=3 or P13B=3
		4	Other factors were decisive		S6A=4 or S6B=4 or S6C=4 or P13A=4 or P13B=4
		Blank	Not stated		lack
		9	Not applicable		SEEKREAS ≠ 4 and WANTREAS ≠ 3 and FTPTREAS ≠ 3 and AVAIREAS ≠ 3

2. WAYJFOUN and ad hoc module 2022 (on job skills) variables

There are examples of very detailed conditions included in these filters related to the last 12 months period.

Filter for the WAYJFOUN variable:

(EMPSTAT = 1) AND ((YSTARTWK = REFYEAR) OR (YSTARTWK = REFYEAR — 1 AND 01 <= MSTARTWK <= 12 AND MSTARTWK > REFMONTH))

Filter for the AHM 2022 variables:

AGE = 15–74 AND (EMPSTAT = 1 or (EXISTPR = 2, 3 and (REFYEAR-YEARPR = 0, 1 or (REFYEAR-YEARPR = 2 and 1<=MONTHPR<=12 and MONTHPR ≥ REFMONTH))))

It is extremely difficult (or even impossible at all) to introduce such filtering of population in a paper version of a questionnaire. Such a detailed condition is possible to implement in the electronic application but this requires more complicated programme algorithms – and we must bear in mind that the application consists of many conditions connected with used filters (for appropriate target population and questionnaire paths) as well as other additional elements like e.g. validation rules. And all of them have to work well when interviewing a respondent.

Therefore, in the Polish LFS in such a case we implement a slightly wider filter relating to the whole years instead of checking precisely required number of months – i.e. the questions related to the variables are asked, respectively, to all respondents who started working in their current main job in the current or previous year (WAYJFOUN) or terminated their last work in the year of the survey and the two previous calendar years (AHM 2022). In this way we obtain a slightly larger population than the one assumed for these exemplary variables, but this is made consistent with the filter at a later stage, after data collection.

The use of such detailed filtering may have another disadvantage – we can lose important information for a given variable for respondents for whom there is a lack of information on the exact month, because we can't calculate the exact period required in the filter.

3. ABSHOLID, ABSILLINJ and ABSOTHER

Variables regarding absenteeism from main work have been added in EU LFS from 2021 for the purpose of better capturing information about hours worked during the reference week. It was a result of work of the *Task force on the measurement of absences and working time in the LFS* that found out that under-reported absences had a marked impact on the reported hours actually worked. The model questionnaire developed by the TF team for the LFS variables on working time (agreed by LAMAS WG in December 2015) was intended, mainly, to remind respondents about their absences in the reference week in order to improve the reporting of hours actually worked.

According to the model questionnaire, questions within these variables are directed to persons who worked during the reference week. It is reasonable, because we are interested in getting correct information on actual hours worked during the reference week (so the hours >0; for persons who didn't work the hours=0). But taking into account currently valid filter for these variables, they are ultimately have to be coded for all employed people (EMPSTAT=1). This means that we have to code these variables also for persons who declared not performing any work during the reference week which is unnecessary - they were absent from work a whole reference week. It seems that this filter is too wide – a reference to persons who worked

during the reference week (what was applied in the model questionnaire: WSTATOR[1]=1) would be more rational (it also limits potential errors arising when transcoding the set of domestic questions into the EU variables).

It is worth to be noticed that these variables relate to the main job only. Among persons who declare performing work during the reference week, there are also persons who worked in their additional job only, not in the main one. This situation causes a problem, most of all during developing a questionnaire. The model questionnaire does not provide solutions for this case.

Transformation table from national variables to the EU ones

The list of EU LFS variables with their description and technical format, definitions used in the survey as well as other important guidelines and recommendations included in applicable regulations provide the basis for designing national LFS questionnaires. Other useful materials like model questions/questionnaires and explanatory notes are also very helpful and they are used during preparing data collection tools. Data collected in a given country during each round of the survey forms a domestic database (its structure usually corresponds to the construction of the questionnaires used). In order to transmit the domestic survey results to Eurostat, it is necessary to prepare an additional dataset according to the specified structure (which is different from a domestic one). This requires the development of an appropriate transition key allowing the data to be recoded into the required EU file.

In Poland we prepare one file with transition key for quarterly datasets and the second one where there are included also variables with yearly and biennial frequencies as well as module variables for a given year.

Depending on the complexity of a given variable, transcoding can be very simple or, on the contrary, very complicated. Simple variables with uncomplicated filters that usually need only one question implemented in a questionnaire are the easiest to transcode (it a situation when we have coding like 1:1).

The coding of the NEEDCARE variable (presented above) can be an example of a transition key for a variable with a few groups of target respondents included in its filter – it is coded using one question that was placed in different places in the questionnaire.

The table below shows the transcoding (from PL LFS into EU LFS codes) for one of a complex variable – AGEPENSO from 2023 module on pension and labour market participation:

Variable identifier	Variable name	Code	Labels	Filter	Polish LFS – question numbers
AGEPENSO	Age at which the person started receiving an old age pension	0-120	Age in completed years (3 digits)	PENSTYP1=1-7 Persons receiving an old age pension	1) If MP1=1 and MP2≠99 then AGEPENSO=MP2 2) If MP1≠1 and MP4=1 and MP6=1 and MP5≠99 and MP7≠99 then AGEPENSO=minimum (MP5, MP7) 3) If MP1≠1 and MP4=1 and MP6=1 and MP5≠99 and MP7=99 then AGEPENSO=MP5 4) If MP1≠1 and MP4=1 and MP6=1 and MP5=99 and MP7≠99 then AGEPENSO=MP7 5) If MP1≠1 and MP4=1 and MP6≠1 and MP5≠99 then AGEPENSO=MP5 6) If MP1≠1 and MP4≠1 and MP6=1 and MP7≠99 then AGEPENSO=MP7
		Blank	Not stated		(PENSTYP1=1-4 and MP2=99) or (PENSTYP1=5 and MP5=99 and MP7=99) or (PENSTYP1=6 and MP5=99) or (PENSTYP1=7 and MP7=99)
		999	Not applicable		PENSTYP1=8, blank, 9

Preparing the transition key is not only needed to transcode data for sending to Eurostat, but also plays an important role in the stage of developing a questionnaire - thanks to such an "exercise", we can verify the correctness of a designed tools for our data collection.

---

[1] This variable is the "old" one – since 2021 it has been replaced (with modifications) by WKSTAT

Some concluding remarks

Preparing a set of variables with all technical details for the EU LFS and then developing a national questionnaire allowing to collect the data in accordance with the requirements imposed by regulations is certainly not an easy task.

However, a positive aspect of the implementation of uniform requirements by all Member States is certainly a greater harmonisation of the survey and the obtained results at the European level. It should be emphasized that the additional materials being developed at a European level by Eurostat (like explanatory notes and model questionnaires) are also important for this process, because they help in developing data collection tools by particular countries. Possibility to consult with Eurostat as well as other MS countries and getting help with any problems or concerns encountered during creating a domestic survey tools is also very important.

It is definitely good practice to carry out pilot tests of the introduced changes as well as to prepare a transition key from developed questions for the national survey into variables required by a given regulation that should be transmitted to Eurostat. As mentioned above, this is a good solution allowing at a stage of developing survey tools for verification whether the developed questionnaire ensures to obtain all required EU LFS variables in accordance with their specifications. It is worth noting that it is not any additional work – the previously prepared transition key is ready to use during preparation a dataset for Eurostat stage.

Our work on implementing changes in the PL LFS since 2021, also provides some conclusions that can be used for improvement of introducing changes in the future.

It is important that the provisions relating to a given issue would be consistent between different documents (e.g. flow charts, model questions and explanations included in the explanatory notes). From the point of view of construction of the survey questionnaire, proper cooperation of different submodules (blocks of questions) is extremely important. These thematic blocks are often elaborated separately (even by different task forces) and there is a lack of trying to put them together and see if everything is working well together. But a questionnaire is a set of different thematic blocks that must create a coherent, well-functioning whole and such problems have to be solved while developing national survey tools. It is worth mentioning that an attempt to combine separate blocks of model questions (a kind of a model questionnaire for a greater part of EU LFS) would allow to identify potential problems with developed variables (e.g. with filters) and to solve them before entering the provisions to a given regulation.

From the above-mentioned problems resulting from the use of complex filters for variables, we can draw the conclusion that limiting the population for which responses are to be collected at the stage of creating the list of variables (what these precise but complicated filters are intended to serve) may seem beneficial (due to limiting the burden on respondents). In fact it causes a significant increase in the burden on NSIs at the stage of implementing these requirements when constructing the survey and at the same time may cause implementation errors (incorrect coverage of the variable). Sometimes it is better to use a broader filter to make it easier and more complete to collect data, which can ultimately also be processed for a smaller group - by selecting the appropriate population from the collected data.

Annex:

Polish LFS questionnaires used in 2023:

- ZG Household File – questionnaire concerning the household and basic demographic data of its members
- ZD individual questionnaire - addressed to particular persons aged 15-89 years covered by the survey.