6th European Working Conditions Survey

External data quality assessment

Executive summary

Authors: Tamás Rudas, Dávid Simon, Barbara Zerinváry
Organisation: Tarki Social Research Institute, Hungary

I. Executive summary

The external survey quality assessment plays a vital role in ensuring the quality of the EWCS. In addition to a high number of internal quality checks prior to, during and after the fieldwork the external survey quality assessment provides an external perspective on the process of gathering data, the data itself as well as the internal quality system and the role of the actors, including Eurofound. The report gives an overall assessment of the survey quality in relation to international standards and comparable surveys and it provides suggestions for improvement of future waves of the EWCS.

The external quality assessment report was commissioned by Eurofound to Tarki Social Research Institute, which assessed the quality of the survey in relation to the five European Statistical System (ESS) quality components in close observation of their specification in the EWCS Quality Assurance Framework: relevance, accuracy, timeliness and punctuality, accessibility and clarity, and coherence and comparability.

In the external quality assessment of the 6th EWCS the following tasks were performed:

1. Prior evaluation of the Terms of Reference of the 6thEWCS.
2. Comparison of the Terms of Reference to the Tender Proposal of the winner participant.
3. Comparison of the survey process to the Proposal with the aim of the required Reports and paradata.
4. Calculation of output indicators assessing the reliability and validity of the survey data.
5. Comparing output indicators with previous waves and research in order to define extremity and assess validity.

The overall conclusion of the report is that the quality of the 6th EWCS is very high and complies with the quality standards defined in the ESS. Only minor issues were encountered and the assessment shows that the analysed output quality indicators of the 6th EWCS are similar to both other comparable European research projects and the previous wave of EWCS. This indicates that the fitness of the statistical output for the intended use remains high.

For future waves the consideration of the following proposals are recommended: in the sampling process to use sample proportions based on optimization taking into account the sampling error and cost in each country, and the overall sampling error, to prefer registries over enumeration, to check the quality of information about PSUs and to set the range and maximum range of the size of PSUs in a country. Regarding questionnaire processes additional summarized information about the process would be useful, concerning fieldwork and data processing further random back-checks are recommended.

I.1 Prior evaluation of the Terms of Reference

Eurofound is responsible for writing the Terms of Reference for the fieldwork of the EWCS. This is crucial stage, as the Terms of Reference dictate the work to be performed by an external contractor and cannot be changed after the contract is in place. The external survey quality assessment concludes that the Terms of References was comprehensive, well detailed and presented relevant methods, processes, an inventory of the components of the survey and as well as a list of requirements.

Some minor modifications may be considered for the next wave of the survey. Two modifications in connection with the sampling design are recommended: the direct sampling of individuals would be

preferable where high quality registers exist and are accessible, and the sample size's calculation should be based on a total sample size as well as a maximum confidence interval in each country and cost differences among countries in order to achieve a more efficient sample.

I.2 Comparing the Terms of Reference with the Tender Proposal

The high quality of the Terms of References is reflected in the Tender Proposal that largely uses and supplements the definitions mentioned in the Terms of Reference. The Tender Proposal met the requirements of the Terms of References, moreover it promised better quality in some aspects. Additionally, it adopts the recommended sample sizes, highlights the importance of the questionnaire's development, and the role of the new and modified questions. The questionnaire preparation processes (e.g. cognitive pre-test) fulfil the requirements written in the Terms of Reference, but also supplements it: the interviews are to be conducted by researchers, rather than interviewers.

I.3 Assessment of the sampling procedure

The 6th EWCS's sampling was a well-designed stratified multi-stage random sample (based on two attributes: regions and the degree of urbanization) of overall high quality. The sample size was 1000 in each country, except for larger countries (Germany 2000; Spain 1300; France 1500; Italy 1400, Poland 1200, UK 1600; Turkey 2000). The sample size was topped up in Belgium (2500), Slovenia (1600) and Spain (3300) according to the decision of the given countries. The number of PSU's in each country reached the minimum of 50, with net sample size within Primary Sampling Unit (PSU) was lower or equal to 20. There was a substantial change in sampling method compared to the previous wave as in the 6th EWCS net target PSU sample size was used instead of gross. This change in sampling made it impossible to substitute a missing respondent from another PSU. This change decreases the sampling error as the size of the PSU is not a random variable and makes the sampling more feasible, but it can also make it harder to reach the target in PSUs where the response rate is low. Analysis shows that the response rate has a moderate effect on the length of fieldwork and the cost of the change in the sampling method is worth the gain of more feasible and reliable samples.

An important aspect of the sampling in the case of a random sample is the sampling frame. The sampling was preferably drawn on the basis of registries (preferably population or household/address). When high quality registries were not available, the proposed method was enumeration. In the cases of non-EU countries there were two changes in the sampling approaches to the preferred method (address registry in Montenegro and Turkey). Among the EU Member States in seven countries less preferable approaches were used compared to those described in the Tender Proposal. Although more changes were made in the direction of the less preferred approaches, this did not necessarily lead to quality problems of the survey since the potential reason for the changes was the unavailability of high quality registries. Eurofound is recommended to ask for available registry information in the tender proposal in order to decrease the number of changes during the implementation.

In the sampling process some PSUs were replaced as a result of safety problems (Turkey, Italy, Lithuania, UK) and the low number of valid addresses (Spain), therefore the coverage was reduced and the replacement caused lower accuracy as well in these countries. The estimated reduction of coverage in proportion of PSUs was 1-2% in Italy, Lithuania and the UK, while 17.5% in Turkey.

I.4 Assessment of the questionnaire related processes

Overall, the questionnaire-related process was successful. The cognitive test fulfilled its goal and played an important role in enhancing intelligibility of the questions. The translatability assessment's process was well-documented. Eurofound should continue to pay special attention to the aspect of comparability over countries. The TRAPD procedure went well, however the reports could not reveal information about the extent and type of problems in translation. The translation's finalisation was one month late on average.

I.5 Assessment of data collection process

The data collection process was successful, there were no major problems and the required sample size was achieved in all countries. The data collection process includes the fieldwork pilot of the translation pre-test, the scripting and testing of CAPI tools, the selection and training of interviewers and the actual fieldwork itself. The fieldwork pilot was successful, because it provided insights that were converted to solutions, such as providing more clarification for interviewers on enumeration and the last birthday selection method, as well as training interviewers to handle distrustful people in order to increase response rates.

During the scripting one discrepancy occurred compared to the Tender Proposal: three computer programmes were used instead of the originally planned two, which could increase measurement error. In most of the countries the CAPI method was used for collecting metadata, except for Slovakia and the Czech Republic (and in further 9 countries partially), where only the paper and pencil method was available (due to data protection regulation). The latter caused delay and decreased clarity. It is recommended to use single CAPI software for recording metadata.

The selection and training of interviewers was appropriate. Some minor issues – missing documentation of requirements – occurred during selection of interviewers. While the interviewing phase was according to quality standards, high dropout among the interviewers along with delay in funding (in the accession countries) and other minor issues led to the delay of the fieldwork. The data collection process was not finished in the proposed and required time in June, but went on even in August. Higher dropout rates of interviewers caused the workload of the remaining interviewer to increase and Tarki analysed the effect of this increased workload, but could not find any negative quality effect. Tarki therefore recommends to relax the rule of maximum interviews per interviewer in order to prevent fieldwork delays. In one country (Spain) the fieldwork took a very long time. The elongated data collection in Spain was tested by Ipsos, and only small differences were found in the variables that may have been influenced by seasonality.

I.6 Assessment of data processing

The data checking, coding and weighting procedures were carried out according to high quality standards and in accordance to the Terms of Reference. There were several checks related to the quality control of the fieldwork process distribution (duplicate observation and back check). A further check was performed in connection with item nonresponse and the main goal of all other checks was to clean the data. During data checking one minor error was uncovered – in a variable's case 79 respondents should have answered the questions but they didn't (due to a script error, which was fixed). Other checks revealed very few and incidental issues that were solved.

In the coding procedure uniform coding software was used, and trained, experienced coders were employed. In this step a few important discrepancies were found: during the test phase local coder

agreement for ISCO was between 33% and 100% with an average of 74%, while for NACE the agreement was between 49% and 100% with an average of 78% (the target was 95% in each countries). It is recommended to document these discrepancies to further analyse the reasons for dissimilarities.

During the weighting procedure three weights were calculated: the design weight, the post stratification weight and the cross country weight. Concerning the weighting procedure the design weights were calculated on the basis of proxy information instead of eligible population in PSUs (as this information was not available). Tarki stresses the importance of data accuracy of PSU sizes and recommends to record inclusion probabilities. The post-stratification weights had noticeable variance that was trimmed.

I.7 Assessment of quality control

The design and implementation of the sixth EWCS managed to meet most of the broad quality criteria outlined in Eurofound quality assurance framework, as well as most specific criteria specified in the quality control plan. Arguably, the approach was successful in (1) making survey quality an integral element of the exchanges between Eurofound and Ipsos, (2) improving clarity on the definition of quality and of the quality targets that were to be achieved, and (3) increasing the level of detail with which each of the stages of the survey cycle were documented.

The process of quality control was parallel to all other processes; it can be therefore divided into three main parts: pre-field quality control, fieldwork quality control, and post-field quality control. During the quality control process three types of targets were defined: requirements (had to be met), real world targets (likely to achieve) and ideal world targets (would have been ideal to achieve). In total, 137 indicators were set for the 6th wave of the EWCS – 87 ‘requirements’ and a further 50 ‘real world targets’. Nearly all (90%) of the ‘required’ targets were achieved as were 30% of the ‘real world’ targets. For cognitive testing two targets were set and met and for the translation processes quality control 16 were set and met. For sampling, 9 requirements and 11 real world targets were set five real world targets were achieved and one requirement was not. Regarding enumeration out of 7 requirements and 1 real world target two requirements were not fulfilled, in this process a specific quality control tool was used: the PSUs’ at least 10% was back checked. The back check deviations were below 5% in 13 countries, and were over 10% in 4 countries. For fieldwork 18 requirements and 2 real world targets were set, from which 14 were achieved. Concerning fieldwork processes a special quality control was taken: the interviewers’ at least 10% was back checked. In connection with the latter, the documentation was incomplete - the error rate and the error type distribution found by back-check were not reported. Regarding data checking among 15 requirements all of them were achieved, for data coding 3 requirements and 1 real word target were set and all of them were achieved. In the weighting procedure’s case out of 15 requirements 1 was not achieved - due to technical reasons - and among 4 real world targets 3 were not achieved.

I.8 Fitness of the statistical output for intended use

Fitness of the statistical output for the intended use is suitable: output quality indicators of the 6th EWCS are similar to both other comparable European research projects and the previous wave of EWCS.

The design effect was calculated as a product of design effect due to different inclusion probabilities of respondents and other measurement errors as well as the design effect of clustering. The results

show that the variance of the estimations from the samples are 1.5-3 times higher than it would be in the case of a simple random sample, and the design effect of the clustered sampling is higher than the design effect due to the difference in inclusion probabilities. These conclusions highlight the role of the clustering part of the survey sample design. The comparable design effects are in the same range as in the previous waves and other known surveys, which shows 6th EWCS has similar sampling error as the previous wave and other surveys using stratified multi-stage random sampling.

Compared to the 5th EWCS the response rate decreased by 1.7 percentage point. Regarding the non-responding part: the cooperation rate increased by 8 percentage points, the contact rate decreased by 10.8 percentage points – which means the contacting step was harder than in the previous wave. The response rate was different across countries; the average of response rate in the comparable countries was lower compared to the 7th round of ESS (12.5 percentage points), however the target population were the workers in EWCS while the target population were citizens in ESS and it is harder to reach workers in a survey than citizens. Compared to the latter wave the response rate was increased in 23 out of 33 countries, while in 10 of them decreased (base: comparable countries). In Sweden further analysis was made due to the extremely low response-rate – the main reason was the method of contact.

The mean item non-response rate for each question was below 5% for each country. However, in Poland and Latvia, a high proportion of outlier questions (higher than average across countries item non-response) were found. Comparing the questions, the earning and health related ones had the highest rate of item non-response presumably due to these questions' sensitivity.

Further analysis was made to measure the coherence between last available LFS (2014) and EWCS. For working time and part-time workers the correlation between the 6th EWCS estimation and the LFS data across countries was high (0.9 and 0.94). In case of working hours in 8 out of 32 countries the estimations, in case of part-time worker rate in 12 out of 32 countries the estimations based on the EWCS sample were significantly different from the LFS data. These differences might be attributed to the fact that LFS data were available only for 2014.