UNECE

CONFERENCE OF EUROPEAN STATISTICIANS

Expert Meeting on Statistical Data Collection Session 1

27-30 September 2021, Online 27 July 2021

Streamlining data collection monitoring activities through the use of control charts: an application to the EU-SILC survey

Manuela Murgia, Alessandra Nuccitelli, Francesca Rossetti (Istat, Italy)

murgia@istat.it, nuccitel@istat.it, frrosset@istat.it

Abstract

Data collection monitoring plays a key role in containing the total survey error and ensuring the data quality.

Monitoring procedures based on a non-statistical approach include direct observation of the data collection process (especially in telephone surveys) and re-contacting respondents. However, the increasing use of computer-assisted survey instruments offers the opportunity to automatically record paradata and to apply monitoring procedures based on a statistical approach that allows for near real-time controls. To this end, a set of performance indicators can be defined and implemented to assess the adequacy and observance of the survey protocols and to uncover any problematic situations that need to be addressed quickly.

A basic set of performance indicators generally includes outcome rates – calculated on the basis of survey dispositions (e.g., completed interviews, refusals, breakoffs, non-contacted units, ineligible units) – at regular time intervals during the field period. In addition to the basic set, ad-hoc indicators can be defined using other recorded paradata or response data to support the survey-specific monitoring goals and then assist in finding inefficiencies in data collection.

Once indicators are defined, control charts can be used to display them. Control charts help balance cost and thoroughness of monitoring activities by using statistical principles to differentiate potentially problematic cases from those that vary naturally around a process average. In this way, survey managers are guided in making targeted interventions, without spending time exploring false alarms. The decision rules for when to intervene are based on empirical control limits that are derived from data on the process itself, rather than arbitrary specifications. The operational definition of each performance indicator and its level of measurement (e.g., categorical, continuous, or count) define the specific type of chart to be used.

The strategy outlined above has several strengths, as it can be combined with non-statistical procedures to better optimize the selection of cases for direct observation or re-contact, making the monitoring more cost-efficient. Besides, this approach can uncover various forms of departure from the survey protocols (e.g., a glitch in the interviewing software, an intentional error in coding the answer to a question to avoid follow-up questions or other hard-to-identify falsifications), and thus compensate for the weaknesses of non-statistical monitoring procedures.

This work focuses on the system of performance indicators and control charts developed for the EU-SILC (European Union Statistics on Income and Living Conditions) survey carried out in Italy.

Some illustrative examples are given using data from the 2019 survey edition. The data were collected by an external company on behalf of the Italian National Statistical Institute in CAPI or CATI mode. The system of

control charts for the EU-SILC survey is mainly aimed at understanding whether interviewers are working in compliance with the interviewing protocol or, if not, which actions have to be taken to improve their work.

In defining the set of performance indicators, two kinds of aspects are considered: on one hand, the constraints dictated by both the collected information and the interview protocol, which were agreed upon with the external company; on the other hand, the effects of undesired sources of variability in the process data.

The monitoring procedure is designed to be carried out at regular time intervals during the field period. Two sequential types of control charts allow the consistency of each performance indicator to be assessed:

1. among all interviewers;

2. at the interviewer level, for any interviewer with at least one point outside the control band in the previous step.

Besides, particular emphasis is placed on monitoring the coding of textual variables, such as occupation, by developing an ad-hoc procedure to assess for each interviewer:

- the presence of possible concentrations of classification codes;

- the tendency to confirm previously assigned codes, with reference to the respondents involved in past editions of the survey.

Finally, in addition to the charts, the monitoring procedure produces a tabular report listing the interviewers with at least an out-of-control event, along with the threshold values at which the outof-control event occurs. For the flagged interviewers, other information and statistics are also reported to help decide on the type of intervention to be implemented before the end of the field period (e.g., supplemental training for interviewers who systematically assign inappropriate codes).

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE CONFERENCE OF EUROPEAN STATISTICIANS

Expert Meeting on Statistical Data Collection,

27-30 September 2021

Session 1: **Managing and modernizing data collection**

Streamlining data collection monitoring activities through the use of control charts: an application to the EU-SILC survey

Manuela Murgia, Alessandra Nuccitelli, Francesca Rossetti
Istat (Italian National Statistical Institute)

1. Introduction

Data collection monitoring plays a key role in containing the total survey error and ensuring the data quality.

Monitoring procedures based on a non-statistical approach include direct observation of the data collection process (especially in telephone surveys) and re-contacting respondents. However, the increasing use of computer-assisted survey instruments offers the opportunity to automatically record paradata and to apply monitoring procedures based on a statistical approach that allows for near real-time controls. To this end, a set of performance indicators can be defined and implemented to assess the adequacy and observance of the survey protocols and to uncover any problematic situations that need to be addressed quickly.

A basic set of performance indicators generally includes outcome rates – calculated on the basis of survey disposition codes (*e.g.*, completed interviews, refusals, breakoffs, non-contacted units, ineligible units) – at regular time intervals during the fieldwork period. In addition to this set, ad-hoc indicators can be defined using other recorded paradata or response data to support the survey-specific monitoring goals and then assist in finding inefficiencies in data collection (Jans *et al*., 2013).

Once indicators are defined, control charts can be used to display them. Control charts help balance cost and thoroughness of monitoring activities by using statistical principles to differentiate potentially problematic cases from those that vary naturally around a process average. In this way, survey managers are guided in making targeted interventions, without spending time exploring false alarms. The decision rules for when to intervene are based on empirical control limits that are derived from data on the process itself, rather than arbitrary specifications. The operational definition of each performance indicator and its level of measurement (*e.g.*, categorical, continuous, or count) define the specific type of chart to be used.

The strategy outlined above has several strengths, as it can be combined with non-statistical procedures to better optimize the selection of cases for direct observation or re-contact, making the monitoring more cost-efficient. Besides, this approach can uncover various forms of departure from the survey protocols – *e.g.*, a glitch in the interviewing software, an intentional error in coding

the answer to a question to avoid follow-up questions, or other hard-to-identify falsifications – and thus compensate for the weaknesses of non-statistical monitoring procedures.

This work focuses on the system of performance indicators and control charts developed for the EU-SILC (European Union Statistics on Income and Living Conditions) survey carried out in Italy.

The paper is organized as follows. Section 2 provides a brief introduction to the survey. Section 3 describes the monitoring strategy designed to fulfill the survey-specific goals (sub-section 3.1) and how it has been adapted to consider the constraints dictated by the available data (sub-section 3.2). Two procedures specifically developed to monitor the interviewers' work and the coding of the occupation are presented in sections 4 and 5 respectively, focusing on the following topics:

- indicators (sub-sections 4.1 and 5.1);
- control charts (sub-sections 4.2 and 5.2);
- possible interventions for the main types of out-of-control events (sub-sections 4.3 and 5.3).

The use of control charts is illustrated by means of examples based on the data from the 2019 survey edition. Finally, some conclusions are drawn (section 6).

2. EU-SILC survey

EU-SILC is the EU reference source for comparative statistics on income distribution and social exclusion at European level, particularly in the context of the "Programme of Community action to encourage cooperation between Member States to combat social exclusion" and to produce structural indicators for the annual Spring Report to the European Council (Eurostat, 2007).

The survey, carried out in Italy by Istat in accordance with Regulation (EC) no. 1177/2003 of the European Parliament and of the Council, is also included in the National Statistical Program.

The main aim of the survey is to collect data on individual and households' incomes, together with a set of information on their living conditions (housing conditions, household expenses, economic difficulties, etc.).

The reference population is made up of all the households and their members residing in Italy at the time of the interview. Individuals living in institutions are excluded[1].

The survey is carried out yearly and provides both cross-sectional and longitudinal information; household and individual characteristics are collected at the time of the interview, expenses refer to the last 12 months, while the income reference period is the previous calendar year.

The information is collected through an electronic questionnaire structured in three parts:

1. a general form, which collects demographic information of all the household components;
2. a household form;
3. an individual module, for each member aged 16 or over.

The sample design is based on a two-stage scheme (municipalities-households) with stratification of municipalities (primary sampling units) by population size. A rotating design is used for households (secondary sampling units): each year about one quarter of the households are drawn from the population registers of the municipalities selected at the 1st stage; the remaining three quarters of the sample include all households and individuals interviewed in the previous year, who

[1] A household is defined as a group of people who usually live in the same dwelling and are linked by ties of kinship, affinity, adoption, protection, affection, or friendship.

are then re-contacted for the second, third or fourth time. The annual sample therefore comprises households of wave 1 (first interview) and households of wave 2, 3 or 4.

Data are collected by an external company on behalf of Istat, in CAPI[2] or CATI[3] mode. CAPI and CATI interviews are carried out in parallel throughout the survey period and are based on the same set of questions.

The data collection technique is assigned to each household according to some general rules: households with no telephone number or whose members are all foreign citizens are assigned to CAPI, while the remaining households are assigned to CATI.

During the interview of wave 1, households are asked for their telephone number so that it can be used in the following years. Therefore, the share of households assigned to CAPI is highest in the first wave and decreases in the subsequent ones.

Interviews of wave 1 start from scratch, while from the second wave on, some questions are posed according to the dependent interviewing approach, which asks for confirmation of the information previously collected. As a consequence, the duration of the interview is longer in the first wave than in the others.

In 2019, the overall sample was made up of 20,831 households for a total of 43,400 individuals living in 636 municipalities (Istat, 2020). As displayed in Table 1, a similar percentage of individuals were interviewed using one of the two techniques; on the other hand, with reference to the first wave, CAPI individual interviews are about twice as many as those from CATI. Besides, a total of 192 interviewers were involved in data collection: 158 and 34 for CAPI and CATI respectively.

Table 1. Percentage distribution of individuals in the sample, by data collection technique and wave, and number of interviewers by data collection technique

DATA COLLECTION TECHNIQUE	WAVE		TOTAL	INTERVIEWERS
	I	II, III, IV		
CAPI	20.35	27.25	47.60	158
CATI	10.08	42.31	52.40	34
TOTAL	30.44	69.56	100.00	192

Source: EU-SILC data, 2019

3. The strategy for monitoring fieldwork activities

3.1 Overview

The monitoring system based on control charts for the EU-SILC survey is mainly aimed at understanding whether interviewers are working in compliance with the interviewing protocol or, if not, which actions must be taken to improve their work. Besides, it tries to simplify the monitoring activities to reduce costs and efforts of this phase of the data collection process: thanks to this system, survey and fieldwork managers can immediately detect any potential problem interviewers might encounter and take the proper actions to solve it in due time.

[2] Computer-Assisted Personal Interviewing

[3] Computer-Assisted Telephone Interviewing

The system is designed to be an alternative monitoring instrument to the classical contingency tables reporting the values of performance indicators per interviewer, weekday, geographical area, etc.. Contingency tables are extremely useful in monitoring data collection, but they might become hard to read when the number of variables and cases to monitor increases and only expert "eyes" are able to quickly find potential problems among a long list of values. Displaying the values of each indicator on a control chart, instead, makes it much easier to find problematic situations, as out-of-control cases are highlighted on the base of statistical evidence. Moreover, in this way, contingency tables can only be produced for a restricted number of variables and cases.

The monitoring system for the EU-SILC survey is made up of two procedures, one to oversee the interviewers' work and the other one to supervise those interviewing behaviors that might badly affect the coding of the occupation variable.

The monitoring strategy underneath the first procedure is aimed at keeping under control the response, breakoff, and refusal rates, together with the number of attempts to complete an interview and the interview duration. These are the classical indicators used in monitoring the collection of data with interviewer-administered modes; besides, the duration of the interview is a target variable required by Eurostat.

The control charts used to display the above indicators are produced weekly in two steps:

1. a first set of control charts (called *screening charts*) is produced for all the interviewers working in the reference week;
2. for each interviewer with at least an out-of-control event from the first step, a second set of control charts (called *in-depth charts*) is produced to monitor each indicator over the whole period the interviewer has been working.

In-depth charts are fundamental to understand whether an out-of-control event that occurred during the last week (reference week) is occasional or systematic (Murgia and Simeoni, 2005). In the latter case, the field supervisor and/or the survey manager can decide whether and how to intervene on each interviewer.

For example, in case an interviewer registers out-of-limits values due to a too low response rate in the reference week, an in-depth chart is produced. If this shows similar results over the whole period the interviewer has been working, it is important to analyze the control charts for the other indicators before taking a proper decision. If the control charts for the refusal or breakoff rates show values that fall above the upper limits and, at the same time, the interview duration is in control (each point between the limits), then it is quite likely that the interviewer needs to be trained again on the contact strategy with respondents.

The same holds if the number of contact attempts per completed interview is too high: the interviewer might have some difficulties in completing the interviews (perhaps he/she has an improper approach with respondents) or in using the list of sample units (for example, the list may contain a high proportion of wrong telephone numbers).

Similar analyses can be performed using the control chart for the duration of the interview: too long interviews associated with a too low response rate and too frequent breakoffs might mean that the interviewer needs a de-briefing to understand which problems make the interview longer than the average and induce respondents to interrupt it before its very end.

The second procedure is aimed at monitoring the coding of the occupation. The purpose is to detect any potential improper interviewing behavior that might lead to assign a wrong code to the occupation declared by the respondent and collected by an open-ended question. To code the

occupation, interviewers use an online coding system that looks for a match of the textual description inside a database containing the codes of the official classification[4].

Signals of potentially wrong coding behaviors are represented by:

- a too short time spent in completing the coding operations;
- a too frequent use of incomplete codes (less than five digits);
- a too frequent use of the same code for different occupations;
- a significant attitude in confirming previously assigned codes, for waves after the first one.

It is worth noting that interviewers are trained in completing the coding in the presence of the respondent and, only if they fail to assign a code in a reasonable time, they can complete their task once the interview is over. In this case, the interviewers have to report in a special field, called *Notes field,* all the information useful to code the occupation at a later time. The presence of notes and their length are used in the monitoring procedure as additional signals for potential problems in the coding task.

The indicators of interest can be monitored using *screening charts* produced at regular time intervals during the fieldwork period. For example, if an interviewer shows a systematic attitude to spend a too short time in coding, it is useful to analyze the type of codes he/she assigns (complete or incomplete codes), to understand how to intervene:

- if the proportion of complete codes is above the upper limit, it might mean that the interviewer tends to assign the same code to different kinds of occupation to speed up the coding process. If it is the case, then he/she must be re-trained, otherwise he/she is just more efficient than the others;
- if the proportion of incomplete codes is above the upper limit, then a tight control on the interviewer is required because he/she might assign meaningless codes on purpose to quickly end the interview.

Other useful information can be obtained by integrating the results of the two monitoring procedures. For example, if an interviewer is out-of-limits for both the time spent in coding (too long) and the level of completeness of the assigned codes (only completed codes are assigned), it is important to ascertain whether he/she also registers a too high breakoff rate, and on which questions this usually happens. If breakoffs occur on questions administered after the coding of the occupation, then it is crucial to re-train the interviewer on the entire coding activity to limit the negative impact on the response rate.

3.2 Data and survey constraints on the monitoring procedures

The two monitoring procedures have been tested on the final data of EU-SILC 2019 and will be used in the 2021 survey edition. They were meant to run for both data collection techniques but, as data from CAPI on contact attempts were incomplete, the first procedure has been tested on CATI data only. Besides, data from CATI contained a very small amount of refusals to make it impossible to use control charts for the relative indicator. The same happened for the breakoff rate: CATI data contained no breakoffs at all, since interviewers were paid per completed interview and not by the hour. Therefore, in implementing the first procedure the response rate indicator has been defined

[4] This classification is a national implementation of ISCO-08 (International Standard Classification of Occupations 2008).

as the proportion of completed interviews in relation to the amount of eligible units, as explained in section 4.

As far as the coding of occupation is concerned, the monitoring procedure has been tested for both data collection techniques. However, the time spent in the coding activity was not recorded due to a technical problem of the external company that conducted the interviews on behalf of Istat. Furthermore, the interview protocol made it mandatory to code each collected occupation (under penalty of no interview payment) and to assign it a complete code (five digits). For this reason, the second procedure does not use the time spent in coding and the proportion of complete codes as signals of potential problems, but only relies on other available variables, as described in section 5.

4. Procedure to monitor the interviewers' work

4.1 Indicators

Specific indicators can be defined to monitor the work of each interviewer. They are as follows:

- *response rate*, calculated as ratio between the number of completed interviews (at household level) and the amount of eligible households in the sample;
- *productivity rate*, calculated as ratio between the number of completed interviews and the amount of contact attempts (both at household level);
- *duration of interview*, namely the time spent to complete a household interview[5].

In defining the set of indicators, two kinds of aspects are considered: on one hand, the constraints dictated by both the collected information and the interview protocol, which were agreed upon with the external company (section 3); on the other hand, the effects of undesired sources of variability in the process data. As regards the latter aspect, special measures are adopted to control for the effect of other variables, such as data collection mode, wave, household size. In this way, more accurate control limits are obtained, better highlighting any anomalies more directly attributable to the interviewer's behavior.

Therefore, as the data collection mode has a relevant impact on fieldwork, the monitoring procedure is designed to produce all the above indicators – and associated control charts – by technique. Besides, considering the significant effect that the wave has on the number of completed interviews and on the time spent to complete an interview (section 2), the response rate and the interview duration are calculated separately for interviews from the first wave and for interviews from the second wave on.

It is worth pointing out that the interview duration also depends on the number of eligible household members. This is because the number of questionnaire modules to be administered increases as the household size increases and the interviews from the second wave on are conducted according to the dependent interviewing approach, which asks for confirmation of the information previously collected.

However, by further stratifying the interview duration by household size, the amount of control charts to be monitored would be too high; moreover, the control limits would be inaccurate due to the very small number of interviews in each stratum.

---

[5] For those sample units that needed more than one day to complete the interview, the time spent to complete an interview is calculated as the sum of the partial daily durations.

For this reason, the time spent to complete an interview is adjusted for the household size. More in detail, the interview duration (*D*) is made equivalent to that of a one-person household according to the following formula:

$$Adj\_D = D/N^{\beta}$$

where:

- *Adj_D* is the adjusted duration of the interview;
- *N* is the number of eligible household members;
- β is an adjusting parameter[6] ranging from 0 (*Adj_D* = *D*/*N*) to 1 (*Adj_D* = *D*).

The monitoring procedure automatically calculates the *adjusted duration of interview* according to the above method and uses this indicator in place of the original *duration of interview*.

4.2 Displaying indicators

For each of the indicators defined in sub-section 4.1 two types of control charts can be plotted: screening charts and in-depth charts. They are both Shewhart charts, where the central line represents the average of the summary statistic (that is the statistic used to summarize the indicator values) and the upper and lower limits bound the range of variation of the summary statistic when the process is in statistical control. What differs in the two types of charts are the sub-groups of elements for which the summary statistic is computed and monitored: in the screening charts the sub-groups are the interviewers, while for the in-depth charts they are the fieldwork days.

The control limits are calculated as 3 times the standard error above and below the central line and are referred to as 3σ limits[7]. It is worth noting that they depend on the sub-group size because the standard error of the summary statistics is a function of the sample size.

The Shewhart charts have been implemented with SAS/QC software (more details can be found in SAS Institute Int., 2018). In particular, *p*-charts are used for the indicators defined as proportions (*i.e.*, the response rate and the productivity rate), while $\bar{X}$ and *s* charts are used for the only indicator measured on a continuous scale, that is the adjusted duration of interview.

Some examples of control charts produced by the monitoring procedure are reported below to better explain how they work.

Figure 1 shows the screening chart to monitor the adjusted duration of interview (for wave>1) in the 5th week of the fieldwork period. In this week, four interviewers register out-of-control events: for operator 3002 the average duration value falls above the upper limit, while the other operators – 2006, 3032 and 3042 – show a too short average duration (below the lower limit). The control limits are computed with respect to the mean value $\bar{\bar{X}}$=28.5, which is referred to the entire fieldwork period (from the first day up to the end of the 5th week). The average duration of the interviews in the reference week is plotted as a red dashed line, showing that during this period the interviewers take a shorter time in completing an interview, probably because they are getting more and more

---

[6] β is estimated from all the available data using the following regression model:

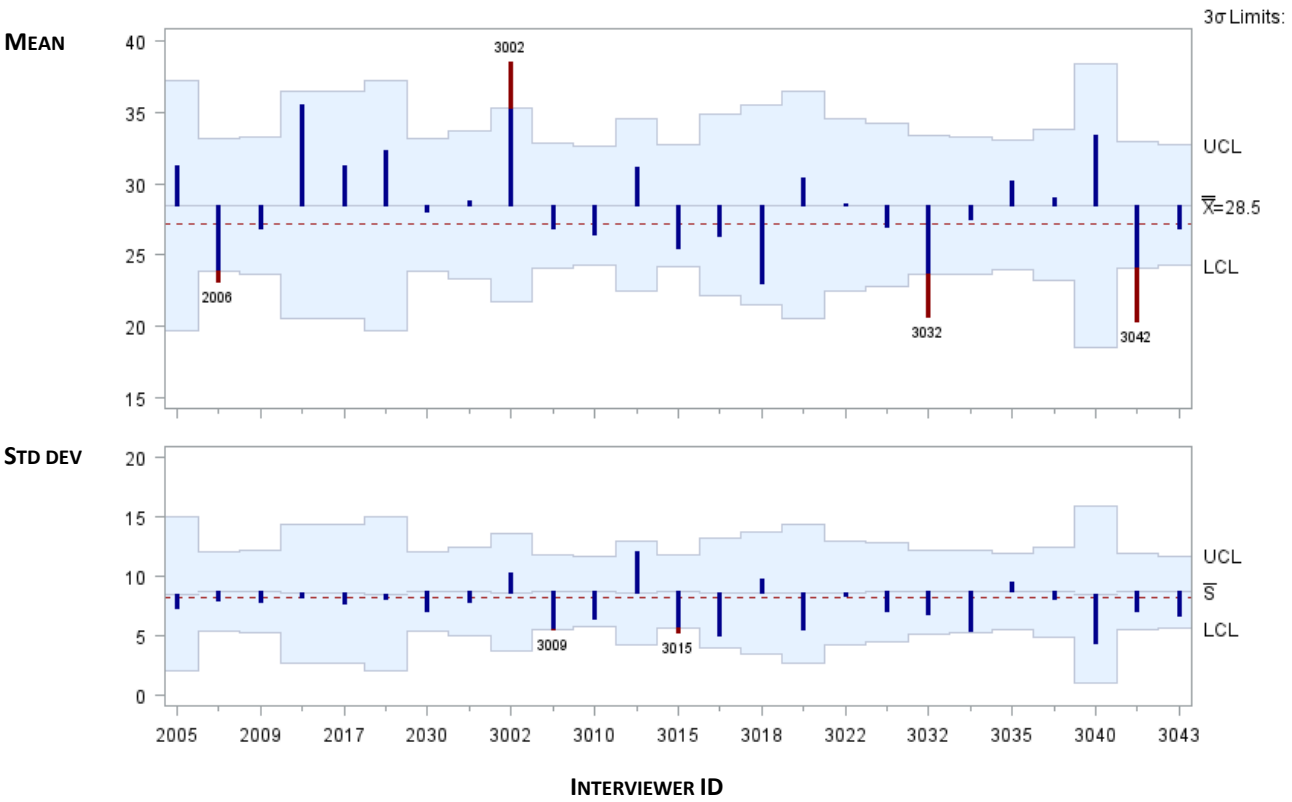
$$ln(D) = \alpha + \beta \cdot ln(N) + \gamma \cdot Wave + \varepsilon$$

where *ln* is the natural logarithm and *Wave* is a dummy variable that takes value 1 if wave=1, and 0 otherwise (in this way, the estimate of β is net of the wave effect).

[7] Constant values different from 3 can be used. Besides, control limits can also be determined in terms of α, a specified probability that the summary statistics exceeds the limits. Probability limits are used in the procedure to monitor the coding of occupation (section 5).

experienced. In the bottom part of the figure, the control chart for the standard deviation is displayed, indicating that, except for few cases, the interviewers' behavior is almost homogeneous.

Figure 1. Screening control charts of the *Adjusted duration of interview*, in the 5th week of the fieldwork period (wave>1)



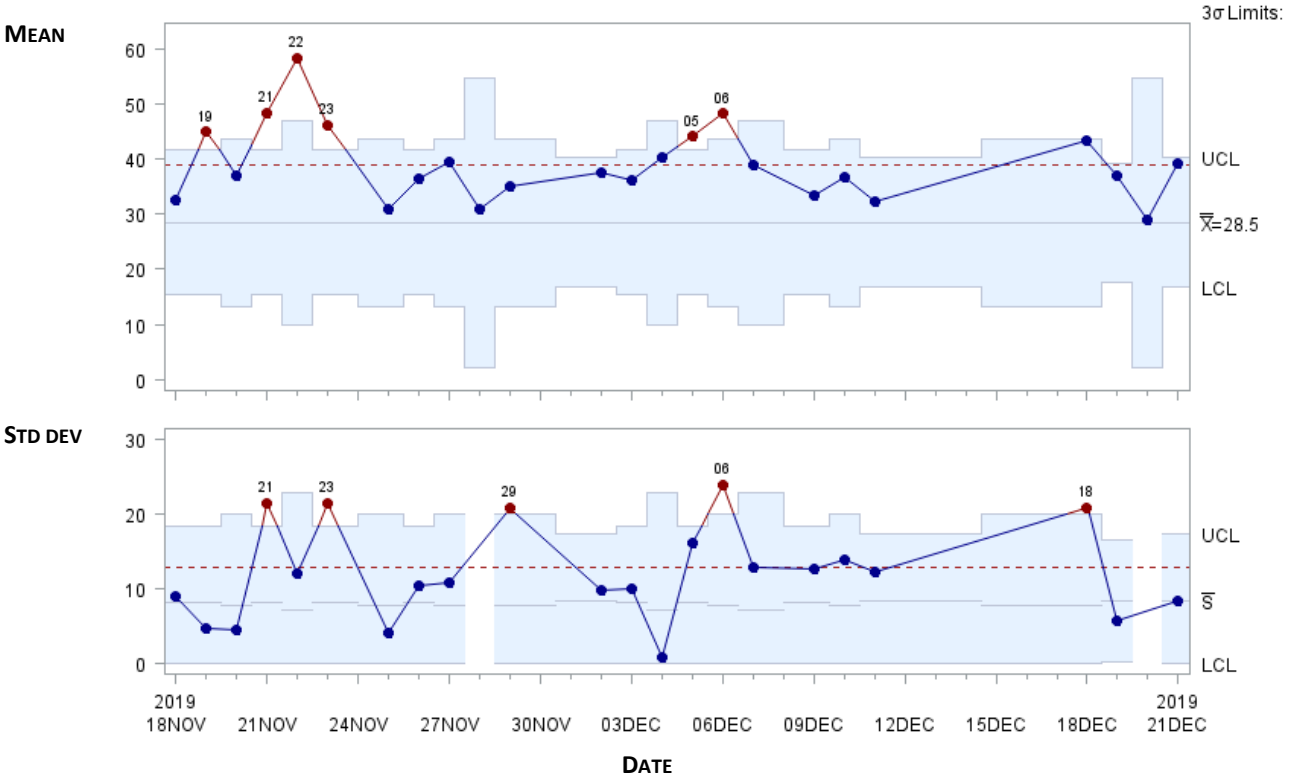
Source: EU-SILC data from CATI, 2019

To understand whether the out-of-limits values are occasional or systematic, in-depth charts are produced for each of the four interviewers. Figures 2 and 3 show in-depth charts for interviewers 3002 and 3042 respectively. In these charts, the limits are computed with respect to the average duration of the interviews conducted by all the interviewers who are active along all the fieldwork period. In each chart, the average duration for the flagged interviewer is also displayed (dashed red line) to better understand the situation.

Operator 3002 seems to need, in general, more time to complete the interview, as his/her own average duration (dashed red line) is above the general mean. However, he/she shows some out-of-limits values only at the beginning of the fieldwork period and not towards the end, meaning that, as time goes by, he/she has learned how to manage the interviews, although the time spent per interview is still quite high. The opposite happens for operator 3042 who is, in general, quicker than the average, with points below the lower limit also in the last week of fieldwork.

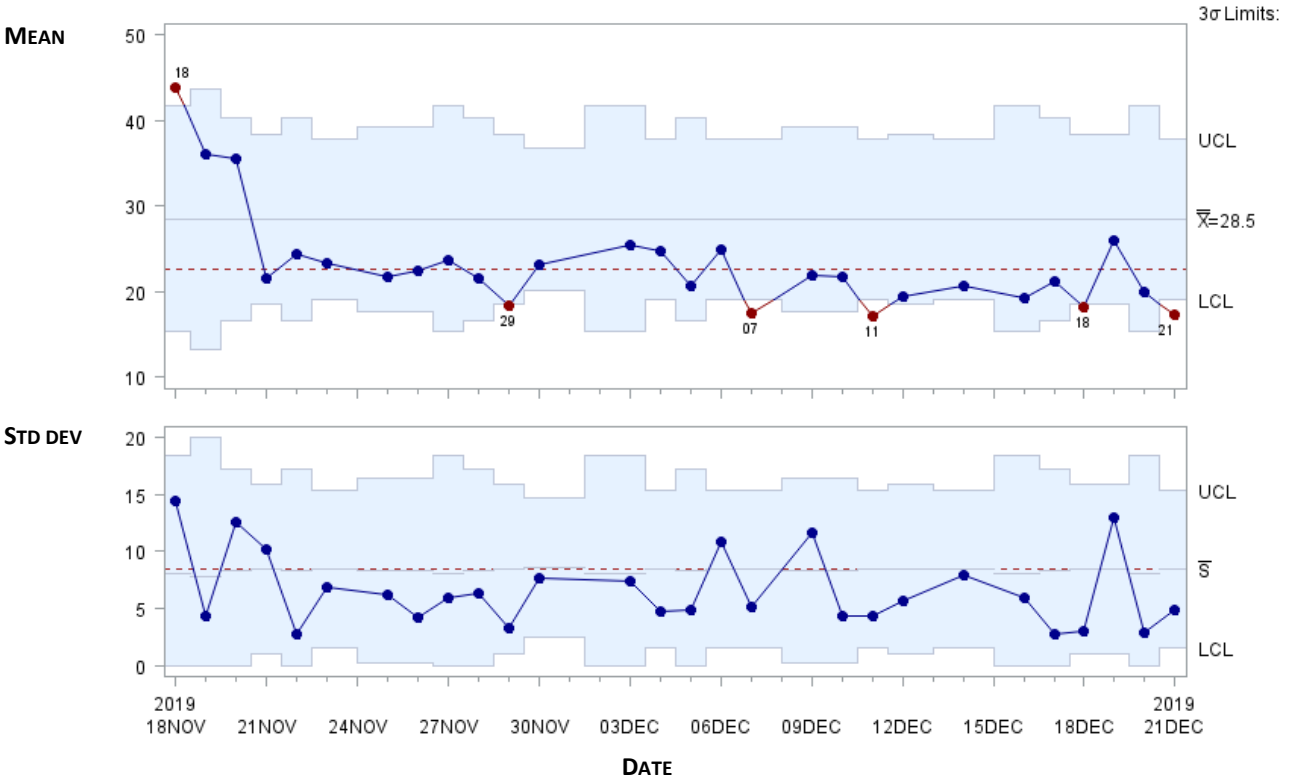
Besides, the high variability in the standard deviation for operator 3002, as opposed to a more homogenous behavior of operator 3042, suggests that it should be advisable to invite interviewer 3002 for a de-briefing – to understand whether there are some problems to be solved – and to strictly monitor operator 3042.

Figure 2. In-depth control charts of the *Adjusted duration of interview* for interviewer 3002, at the end of the 5th week of the fieldwork period (wave>1)



Source: EU-SILC data from CATI, 2019

Figure 3. In-depth control charts of the *Adjusted duration of interview* for interviewer 3042, at the end of the 5th week of the fieldwork period (wave>1)



Source: EU-SILC data from CATI, 2019

4.3 Flagged cases and type of intervention

For each indicator, the monitoring procedure automatically produces a tabular report listing the interviewers with at least an out-of-control event, along with the limit values at which each out-of-control event occurs. This eases the job of the fieldwork supervisor who does not have to examine long reports listing all the operators.

Other information and statistics are also reported by interviewer to help decide on the type of intervention to be implemented. They include the *day of the interview*, the *number of interviews per day*, and the *average value of the indicator*.

The report about the adjusted duration of interview also contains the main information on the out-of-control events related to the other indicators. This helps in analyzing the results and in prioritizing the interventions, as the operators with out-of-limits situations for most of the indicators are the first ones to be checked or re-trained.

Some of the interventions suggested by the output of the procedure are summarized in the following table.

CASES	TYPE OF OUT-OF-CONTROL EVENT			POSSIBLE ACTIONS
	ADJUSTED DURATION OF INTERVIEW	RESPONSE RATE	PRODUCTIVITY RATE	
1	Below the lower limit	Below the lower limit	Below the lower limit (too many contact attempts per interview)	*Further analysis requested before re-training the interviewer*: he/she completes a small number of interviews, but not because of their short duration; the interviewer might need a re-training on the contact strategy with respondents or the list of units assigned to him/her might have some problems (*e.g.,* not updated phone numbers)
2	Below the lower limit	Below the lower limit	Above the upper limit or in control	*Interviewer to be checked:* he/she is slower and possibly lazy and should be invited to put more effort in his/her work
3	Below the lower limit	Above the upper limit	--	*Interviewer to be checked*: he/she might cheat in completing interviews
4	Above the upper limit	Above the upper limit or in control	--	*No action needed:* the interviewer works well, and a longer interview might mean he/she is just a more dedicated interviewer
5	Above the upper limit	--	--	*Interviewer to be re-trained:* a debriefing is needed to understand which questions are difficult for him/her to administer

5. Procedure to monitor the coding of occupation

5.1 Indicators

Taking into account the constraints dictated by both the available information and the interview protocol (section 3), the coding of the occupation is monitored by assessing two aspects:

- the presence of possible concentrations of classification codes;
- the tendency to confirm previously assigned codes, for interviews from the second wave on.

For this purpose, specific frequency-based indicators by interviewer are defined. As for the first monitoring procedure (sub-section 4.1), they are also produced per technique since the data collection mode has a relevant impact on fieldwork. The indicators are as follows:

1. *proportion of classification codes* in nine major occupational groups according to the national implementation of ISCO-08;
2. *proportion of confirmed classification codes*, for interviews from the second wave on.

The confirmed codes are excluded from the calculation of the first indicator, as they may have been previously assigned by a different interviewer.

5.2 Displaying indicators

The indicators introduced in sub-section 5.1 are displayed using *analysis of means* (ANOM) charts, implemented with SAS/QC software. The ANOM method identifies the means that are significantly different from the overall mean and has a graphical representation that is similar to a Shewhart chart[8].

In this context, each ANOM chart refers to a specific indicator and is based on the data collected up to a time point in the fieldwork period. In particular, the indicator values for the various interviewers are plotted as deviations from the overall mean and compared with upper and lower decision limits to identify and flag the interviewers with at least one point outside the limits (out-of-control event). The decision limits are determined in terms of α, a specified probability that the summary statistics exceeds the limits under the null hypothesis of no differences among the subgroup means. Furthermore, the limits are functions of the sub-group size (more details are in SAS Institute Inc., 2018).

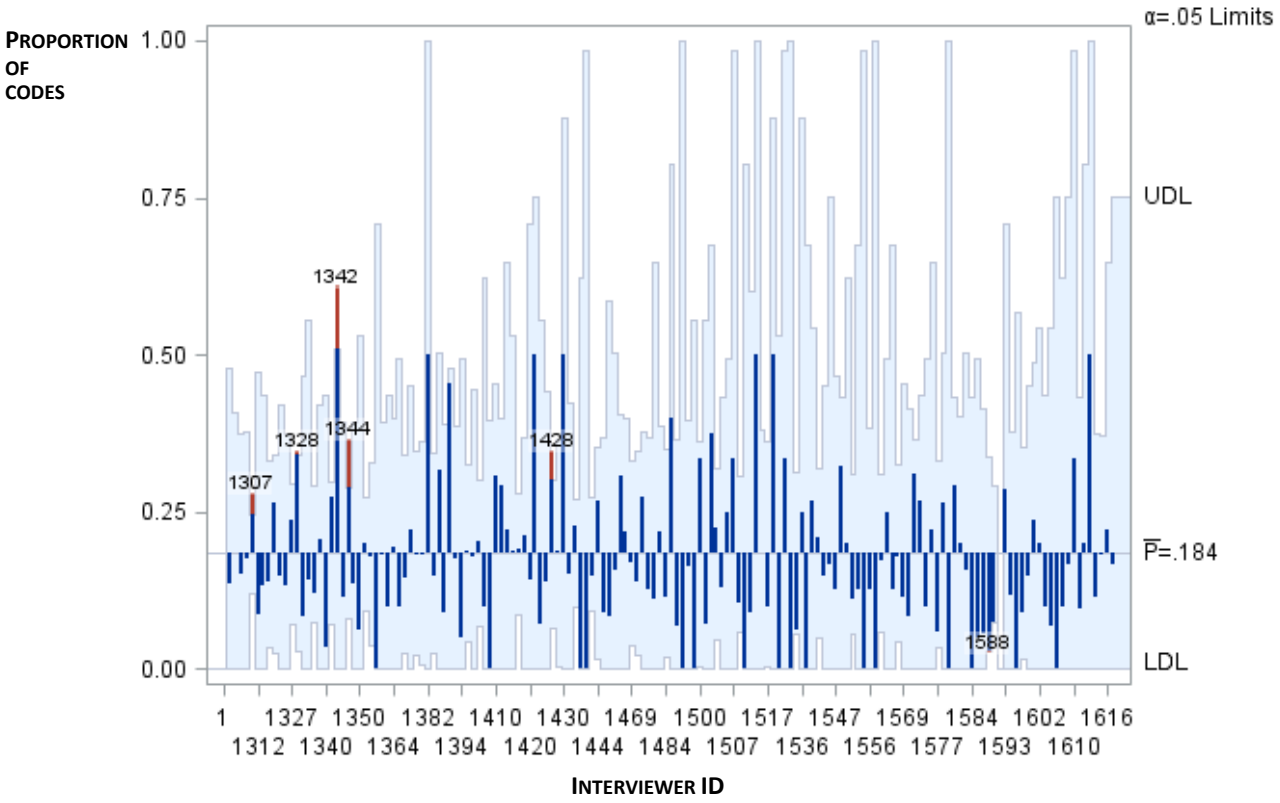
As regards the first indicator introduced in sub-section 5.1, a specific chart is produced for each occupational group. In this case, the values above (below) the upper (lower) limits highlight significantly high (low) concentrations of classification codes within the occupational group, indicating for the flagged interviewers a possible inadequate effort or a poor understanding of how the coding activity should be carried out. However, especially in the case of CAPI interviews, an out-of-control event might also be linked to specific characteristics of the territorial entity assigned to the interviewer (for example, the existence of an industrial district might show a higher-than-average frequency of certain occupational profiles).

With reference to the CAPI interviews, an example of chart for the proportion of classification codes within the 6th occupational group is shown in Figure 4. The central line represents the overall

[8] Although the term “analysis of means” suggests that the method is intended for means of continuous measurements, the technique is also applicable to means of attributes data, including proportions.

average. The decision limits correspond to a significance level[9] of 0.05 (note that the decision limits vary with the number of codes assigned by each interviewer, and the widest range corresponds to the interviewer with the fewest codes). Based on this chart, 6 out of 158 interviewers are flagged.

Figure 4. Chart of the *Proportion of classification codes* in the 6th occupational group, at the end of the 17th week of the fieldwork period



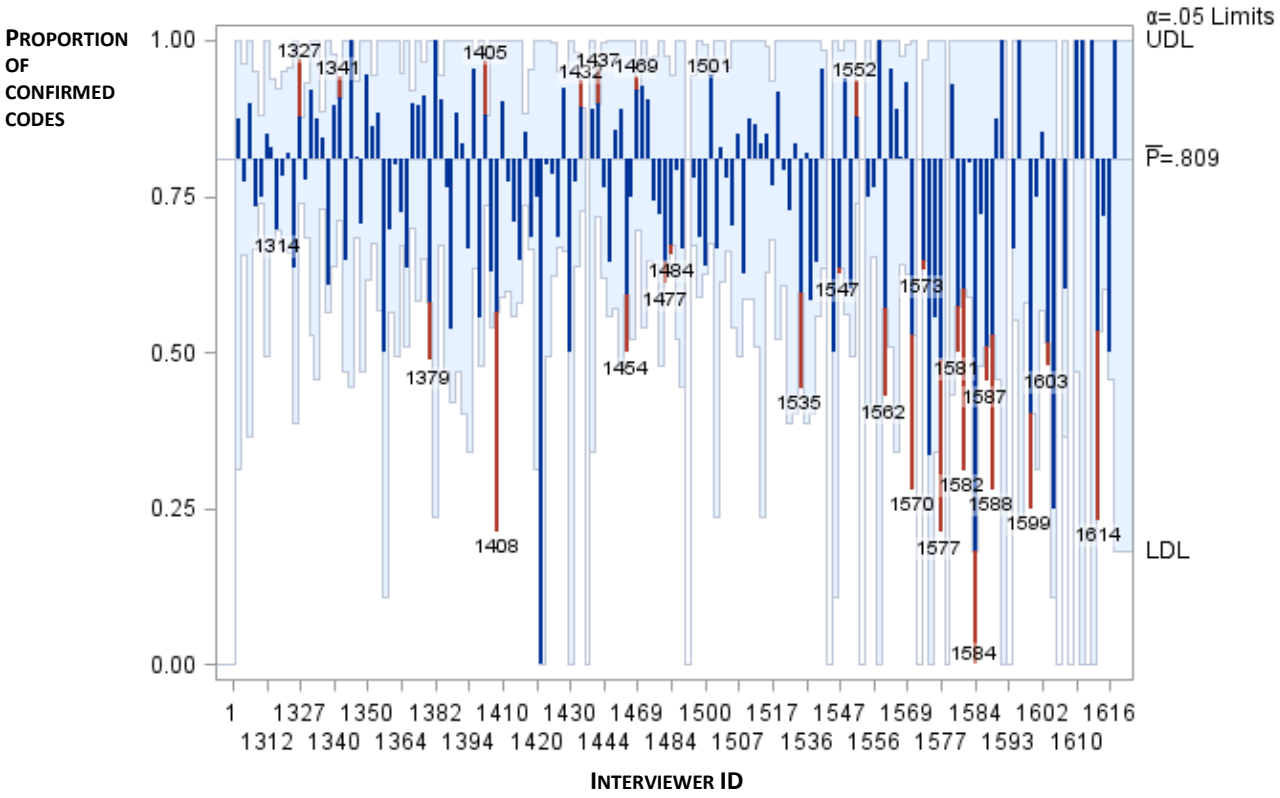
Source: EU-SILC data from CAPI, 2019

As regards the proportion of confirmed classification codes (second indicator in sub-section 5.1), the values above the upper limits point out the interviewers with a significant tendency to confirm previously assigned codes, suggesting a probable insufficient commitment in the coding task. On the other hand, the values below the lower limits reveal above-average efforts, likely associated with a lack of understanding of how the specific activity should be performed or with difficulties in using the assisted coding software.

With reference to the CAPI interviews, an example of chart for the proportion of confirmed codes is shown in Figure 5. In this case, 8 out of 158 interviewers exhibit a significant tendency to confirm previously assigned classification codes. The integration of these results with those deriving from the other charts and other available information makes it possible to understand whether that tendency is linked to a fraudulent attitude, directing towards the type of intervention to be implemented before the end of the fieldwork period.

[9] This means that, assuming all interviewers have the same proportion in the 6th occupational group, there is a 0.05 probability that one or more of the decision limits would be exceeded purely by chance.

Figure 5. Chart of the *Proportion of confirmed classification codes*, at the end of the 17th week of the fieldwork period



Source: EU-SILC data from CAPI, 2019

5.3 Flagged cases and type of intervention

In addition to the above indicators and charts, the monitoring procedure automatically produces a tabular report listing the interviewers with at least an out-of-control event, along with the limit values at which each out-of-control event occurs.

Other information and statistics are also reported by data collection mode and interviewer to help decide on the type of intervention to be implemented. They include:

- *the Chi-square test statistics* and corresponding p-value, which provide an indication about the goodness of fit of the distribution of the classification codes assigned by each interviewer to the overall distribution, related to all the interviewers (the interviewers whose distribution is different from the overall distribution at the significance level of 0.05 are highlighted in the report);
- *a relative index of dissimilarity*, as an additional measure of how much the distribution of the classification codes assigned by each interviewer is "different" from the overall distribution, related to all the interviewers;
- *the median and mean length of the text in the Notes field.* This text can only be present if the interviewer assigns a code after the interview is over. Too short texts might reveal a poor accuracy of codes assigned after the interview, while long texts, and therefore, in theory, suitable for coding after some time, might indicate that the interviewer is committed to his/her task, but he/she probably has some difficulty in classifying the occupation (at least an out-of-control event is recorded for the interviewer, being listed in the report).

The main interventions suggested by the output of the procedure are summarized in the following table.

CASES	TYPE OF OUT-OF-CONTROL EVENT		TEXT IN THE NOTES FIELD	POSSIBLE ACTIONS
	PROPORTION OF CODES BY OCCUPATIONAL GROUP	PROPORTION OF CONFIRMED CODES		
1	Above (or below) the upper (or lower) limit	Above the upper limit or in control	Absent or short	*Interviewer to be checked*: he/she does not put enough effort into the coding activity
2	Above (or below) the upper (or lower) limit	Below the lower limit	Adequate	*Interviewer to be re-trained*: he/she engages in coding but tends to assign codes from the same occupational group
3	--	Above the upper limit	Absent or short	*Interviewer to be checked*: he/she over-confirms previously assigned codes; he/she might assign inappropriate codes after the interview is over
4	--	Above the upper limit	Adequate	*Interviewer to be re-trained*: he/she engages in coding but over-confirms previously assigned codes

6. Concluding remarks

The monitoring system for the EU-SILC survey has been developed to understand whether interviewers are working in compliance with the interviewing protocol or, if not, what actions must be taken to improve their work.

Specific indicators are defined using recorded paradata and response data to support the survey-specific monitoring goals and then assist in finding inefficiencies in data collection.

The system of control charts, which is used to display the proposed indicators, helps balance cost and thoroughness of monitoring activities by using statistical principles to differentiate potentially problematic cases from those that vary naturally around a process average. In this way, fieldwork supervisors and survey managers are guided in making targeted interventions, without spending time exploring false alarms. Two monitoring procedures have been implemented, one to control the interviewers' work and the other one to supervise those interviewing behaviors that might badly affect the coding of the occupation variable.

The two procedures will be used for the 2021 edition of EU-SILC running since August 2021. They are going to be used next to the classical reports and under a close cooperation among methodologists, fieldwork supervisors, and survey managers. This will allow the latter – fieldwork supervisors and survey managers – to get acquainted with the new instrument and the former to understand whether any improvement in terms of usability or efficacy of the system is required.

Finally, this experience will be extremely important to understand whether this approach is suitable for any other survey that needs to monitor the interviewers' performance and/or specific aspects like the coding of textual variables.

References

Eurostat, 2007. *Comparative EU statistics on income and living conditions: issues and challenges*. Proceedings of the EU-SILC conference, Helsinki, 6-8 November 2006. Luxembourg: Office for Official Publications of the European Communities.

Istat, 2020*. Income, living conditions and fiscal burden of households – Year 2019*. Istat Press Release*,* December 23, 2020*.*

Jans M., Sirkis R., Morgan D., 2013. *Managing data quality indicators with paradata based statistical quality control tools: the keys to survey performance*. In: Kreuter F. (ed.), 2013. *Improving surveys with paradata: analytic uses of process information*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Murgia M., Simeoni G., 2005. *Improving the quality of the assisted coding of occupation in CATI surveys through control charts.* SIS conference - Classification and Data Analysis - Cladag 2005, Parma, June 6-8, 2005.

SAS Institute Inc., 2018. *SAS/QC® 15.1 User's guide*. Cary, NC: SAS Institute Inc.