A paradata-driven statistical approach to improve fieldwork monitoring: the case of the Non-Profit Institutions census

Gabriella Fazzi, Manuela Murgia, Alessandra Nuccitelli, Francesca Rossetti, Valentino Parisi, Roberta Piergiovanni, Luigi Arlotta, Maura Giacummo

1. Introduction

A complex process requires relevant information on the crucial nodes of the process itself to make more effective decisions. This is the case for large complex surveys where, among the several causes of wrong or inappropriate interviewers' behaviours, only the crucial ones have to be identified and corrected to avoid a knock-on effect. An example of such a survey is the Non-Profit Institutions (NPIs) census, for which fieldwork monitoring is improved by using a paradata-driven approach based on quality control tools (Jans *et al.*, 2013).

The complexity of the NPIs census is due to the variety of unit-typologies: from large and structured institutions to very small associations. The complexity depends also on the different data collection modes and on the several communication channels. Besides, two questionnaires with different research aims – to assess the quality of statistical registers (short form) and to collect information (long form) – contribute to boosting the complexity.

The use of computer-assisted survey instruments offers the opportunity to automatically record paradata, making it possible to apply statistical procedures that allow for near real-time monitoring. To this end, a set of performance indicators is defined to assess the adequacy and observance of survey protocols and to uncover any problematic situations that need to be addressed quickly. Once indicators are defined, control charts can be used to display them (Reed and Reed, 1997).

This work focuses on the system of indicators and control charts developed for the 2022 NPIs census carried out in Italy by the National Statistical Institute (Istat). The paper is organized as follows. Section 2 provides a brief introduction to the survey. Section 3 describes the data collection system. Then, the procedure specifically developed to monitor the interviewers' work is presented, focusing on indicators (section 4), control charts (sub-section 5.1), and possible interventions for the main types of out-of-control events (sub-section 5.2). Finally, some conclusions are drawn (section 6).

2. The Non-Profit Institutions census

The NPIs census aims to expand the extent of information available on the non-profit sector by investigating specific issues and by verifying and supplementing the data from the Statistical Register of NPIs, which is based on various administrative sources.

The survey runs from March 10 to November 23, 2022, and involves a sample of approximately 110,000 NPIs. A letter, signed by the president of Istat, is sent to all the sample units to inform them about the purpose of the census, the modes of participation, the deadline, the obligation to participate, and the penalty in case of no participation in the survey.

The survey sample is drawn from the Statistical Register of NPIs and is divided in two sub-samples that differ in terms of units' characteristics, data collection mode, and questionnaire. Besides, each sample is associated to different aims.

The first sub-sample includes about 11,000 NPIs, selected among those units with "weak"

Gabriella Fazzi, ISTAT, Italian National Institute of Statistics, Italy, fazzi@istat.it, 0000-0001-5661-3963
Manuela Murgia, ISTAT, Italian National Institute of Statistics, Italy, murgia@istat.it, 0000-0003-4154-3784
Alessandra Nuccitelli, ISTAT, Italian National Institute of Statistics, Italy, nuccitel@istat.it
Francesca Rossetti, ISTAT, Italian National Institute of Statistics, Italy, frrosset@istat.it, 0000-0002-5974-0826
Valentino Parisi, ISTAT, Italian National Institute of Statistics, Italy, parisi@istat.it, 0000-0001-8429-4250
Roberta Piergiovanni, ISTAT, Italian National Institute of Statistics, Italy, piergiov@istat.it, 0000-0003-2489-8786
Luigi Arlotta, ISTAT, Italian National Institute of Statistics, Italy, arlotta@istat.it
Maura Giacummo, ISTAT, Italian National Institute of Statistics, Italy, magiacum@istat.it

administrative signals in the Register: these are mainly small units that are assigned to the CAPI[1] mode for the administration of a short questionnaire (for this reason, such a sub-sample is called "short"). The main aim is to assess the quality of the Statistical Register of NPIs.

The second sub-sample includes about 99,000 NPIs selected among those units with "strong" administrative signals in the Register. The aim is to collect new information or consolidate the existing one: all these NPIs are initially assigned to the CAWI[2] mode to complete the full version of the questionnaire (for this reason, such a sub-sample is referred to as "long").

To boost cooperation, CAWI non-respondents are sent a maximum number of four reminders. The reminder letter restates the purpose of the census, the modes of participation, the deadline, and the regulatory framework. In case the NPIs prefer to change the survey technique, they can request the support of a CAPI interviewer by calling the contact center or by accessing a dedicated survey page. To reach all the sample units, the CAPI mode is used also for those NPIs that do not receive the information letter.

The CAPI mode is implemented by an external company on behalf of Istat. Each interviewer from the external company is instructed to find the NPIs and to conduct a targeted survey in a given geographical area. Specifically, the interviewer has to find the NPIs by following online traces (such as website, pages on social media, *etc*.) and by visiting them at their postal addresses. If no signs of "activity" emerge, then the interviewer is required to make an in-person visit at least three times, trying to obtain useful contact information and to administer the interview itself. The units without digital and physical signs of activity are registered as "untraceable"; the units with signs of activity, but untraceable after three visits, are coded as "impossible to be interviewed".

3. The data collection system

The NPIs census is a complex survey. From a technical point of view, the complexity is related to the presence of various actors (respondents, interviewers, fieldwork supervisors, survey managers) with different views on data, the management of several communication channels, and the use of mutually exclusive techniques (CAPI or CAWI). Besides, each unit is assigned a data collection mode, but the unit can ask to change it during the survey.

An integrated web-based information system supports all the different stages of the survey process. The system consists of two web applications that can be customized for any type of survey (they were already used for the Agriculture and Population censuses):

i) the data acquisition application, *i.e.*, the online questionnaire used by both respondents (CAWI) and interviewers (CAPI);

ii) the management and monitoring application (SGI), accessible to all census operators.

The two applications interact in such a way that they look like a single one to the end user.

SGI is designed to support the various activities of the data collection process. Each actor has a specific profile associated to an appropriate view of data, functions, and outcomes. In this way, each actor can only process data or enter information for which he/she is responsible or authorized. In particular, each authorized actor can enter and manage his/her own data collection network and assign units to the interviewers. This makes it possible to intervene at any time to avoid, for instance, work overloads that might compromise the data quality. In addition to the profile, a key element of SGI is the user-entered outcome, which allows actions to be activated or deactivated via a previously configured workflow. This also enables a unit to be assigned to a different technique.

The information system automatically collects a variety of paradata. As regards the accesses to the data acquisition application (*i*), the number of work sessions and the timestamps of the first and last visit to the online questionnaire are stored for each user.

As for SGI (*ii*), the application records and historicizes each transaction, collecting paradata at the unit level. They are stored in tables that are updated weekly and include the survey technique,

---

[1] Computer-Assisted Personal Interviewing
[2] Computer-Assisted Web Interviewing

the delivery status of the information letter, the address changes, the date and the author of each contact attempt, the latest – temporary or final – outcome of the various contact attempts (*e.g.*, completed interview, refusal, break-off, eligibility status). This information can be used both during the survey to monitor the fieldwork, intervening promptly if necessary, and at the end of the data collection process to understand what needs to be improved.

4. The monitoring indicators

A set of indicators can be adopted to monitor the work of each interviewer involved in the NPIs census. This set is defined taking into account the constraints dictated by both the available information and the interview protocol (which was agreed with the external company).

The monitoring indicators are defined as outcome rates based on the main survey disposition codes, namely the set of codes that SGI uses to record the outcomes of the various contact attempts (section 3). Of the several indicators that can be derived from the available paradata[3], the following are considered the most effective in highlighting any anomalies in fieldwork:

  a) *response rate*, the no. of completed interviews divided by the no. of eligible units in the sample;
  b) *activity rate*, the no. of units with a final outcome divided by the no. of contacted units;
  c) *eligibility rate*, the no. of eligible units divided by the no. of total (eligible plus ineligible) units in the sample;
  d) *non-interview rate*, the no. of units for which the interview could not be carried out divided by the no. of units with a final outcome.

Rates (*a*) and (*b*) are sufficient to monitor the scheduling and carrying out of the interviews, while indicator (*c*) makes it possible to verify that when those rates are high, it is because the interviewer is working well, in the sense that he/she is not making up ineligible units (these are given a short form that is paid as a completed interview). Moreover, some problems in contacting the NPIs may be detected by an excessive proportion of non-interviews (*d*).

All the above rates are produced at regular time intervals (weekly) during the fieldwork period, only for those interviewers who have been working in the last four weeks. In fact, it may happen that, also due to the difficulties experienced in the data collection, some interviewers stop carrying out the field activity. Besides, the indicators are calculated by province to understand whether problems arise in specific areas of the country – and are therefore common to all the operators working in those areas – or whether the problems concern certain interviewers only.

Finally, given the relevant impact that both the type of administrative signal and the questionnaire length have on the fieldwork (section 2), the set of rates is produced separately for:

  - the NPIs in the short sub-sample;
  - the NPIs in the long sub-sample that did not receive the information letter[4].

In this way, any anomalies more directly attributable to the interviewer's behaviour are better highlighted.

5. The monitoring procedure

5.1 Control charts

The monitoring procedure for the NPIs census is mainly aimed at understanding whether the CAPI operators are working in compliance with the interviewing protocol or, if not, what actions must be taken to improve their work. Besides, it tries to simplify the monitoring activities so that

---

[3] It is worth noting that the time interval between the first and last access to the online questionnaire is a too rough estimate of the interview duration and, therefore, is of little help in monitoring the interviewers' work.
[4] The indicators are not calculated for the NPIs (long sub-sample) that ask for a change of technique (from CAWI to CAPI), as for these units both the contact phase and the interview are less troublesome (response rate and eligibility rate very close to 1).

costs and efforts of this phase of the data collection process are reduced: thanks to this procedure, survey and fieldwork managers can immediately detect any potential problem interviewers might encounter and take the proper actions to solve it in due time.

The procedure is designed as an alternative monitoring tool to the traditional contingency tables that report the values of performance indicators by interviewer, week, 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 be monitored increases. Displaying the values of each indicator on a control chart, instead, makes it much easier to find critical situations, as out-of-control cases are highlighted by statistical evidence. Moreover, in this way, contingency tables can only be produced for a restricted number of variables and cases.

Each indicator introduced in section 4 is displayed using a Shewhart *p*-chart, where the central line represents the mean, and the upper and lower limits – respectively, UCL and LCL – bound the range of variation of the mean when the process is in statistical control (Montgomery, 2009).

The control charts are implemented with SAS/QC software (SAS Institute Inc., 2018) and are produced weekly in two steps:

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

What differs in the two types of charts are the sub-groups of elements for which the mean is calculated: in the screening charts, the sub-groups are the provinces or the interviewers, while for the in-depth charts they are the fieldwork weeks.

The in-depth charts are fundamental to understand whether an out-of-control event that has occurred in the last four weeks is occasional or systematic. In the latter case, the survey manager can decide whether and how to intervene on each interviewer.

Some examples of charts are reported below to better explain how they work.

Figure 1 provides the screening chart of the eligibility rate for the interviewers who have been working in the four weeks preceding June 13, 2022. Three interviewers have out-of-control rates: for CRR the value falls below the LCL, while for both interviewers MDS and RSS the value is 1.

Figure 1. Screening control chart of the *Eligibility rate* for all interviewers (up to June 13)
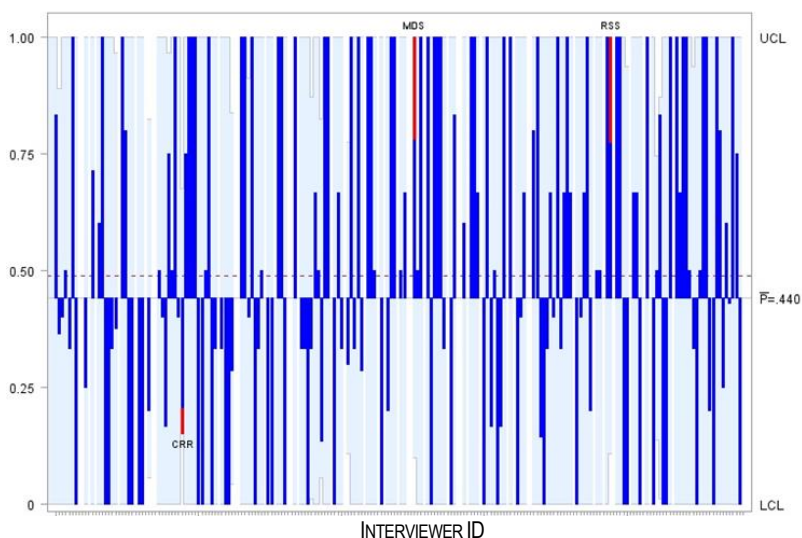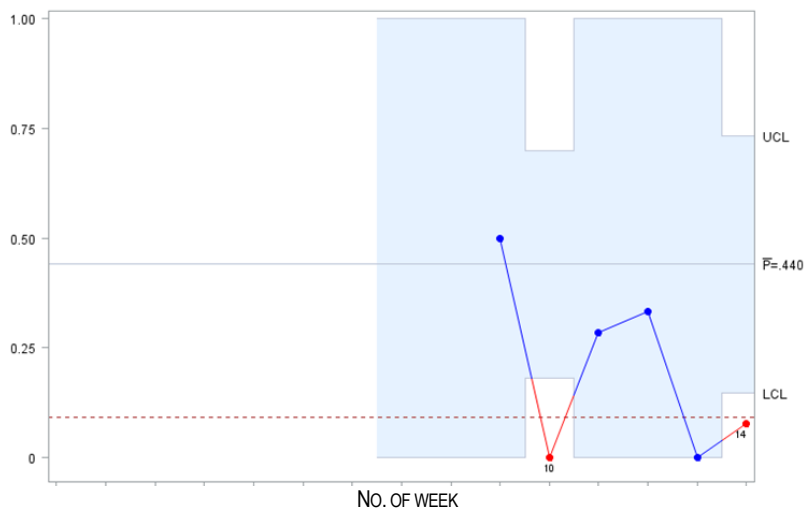

Source: NPIs census data, Short sub-sample, 2022

The control limits[5] are calculated with respect to the mean value $\bar{P}=0.44$, which is referred to all the interviewers who have been active for at least one fieldwork week (from March 10 to June 13). The average rate in the last four weeks is plotted as a dashed red line.

To understand whether the out-of-limits values are occasional or systematic, an in-depth chart is produced for each of the three interviewers. For the sake of brevity, only the in-depth chart for CRR – who started working from the 9th week of fieldwork – is shown (Figure 2). In this chart, the average eligibility rate for the interviewer (dashed red line) is below the mean value ($\bar{P}=0.44$), suggesting that the NPIs surveyed by him/her are mostly ineligible (especially in the last two weeks). It is important to analyse the charts for the other indicators before taking a proper decision.

In the case of interviewer CRR, if the activity rate is very high and, at the same time, the non-interview rate falls below the LCL, further investigation is required to exclude that he/she is making up interviews. Instead, for MDS and RSS, if the response rate turns out to be excessively low, it is quite likely that they need to be trained again on the contact strategy with the respondents.

Figure 2. In-depth control chart of the *Eligibility rate* for interviewer CRR (up to June 13)

Source: NPIs census data, Short sub-sample, 2022

5.2 Out-of-control events and types of intervention

In addition to the above-mentioned indicators (section 4) and charts (sub-section 5.1), the monitoring procedure automatically produces two tabular reports listing, respectively, the provinces and the interviewers with at least an out-of-control event, along with the limit values at which each out-of-control event occurs. The absolute values of the variables used to build the indicators are also reported to take in due account those "signals" based on many units.

The information from the two reports helps to understand whether the out-of-control events subtend a structural issue affecting the entire province (when no interviewer is flagged within a flagged province) or an interviewer-specific problem (when the interviewer is flagged regardless of whether the province in which he/she operates is flagged or not). In the latter case, targeted actions, such as de-briefing or additional training sessions, might be undertaken. Some of the interventions suggested by the output of the procedure are summarized in Table 1.

Any doubts about the actions to be undertaken are removed by analysing all other available information – traditional reports and questionnaires – and/or by randomly contacting some NPIs for feedback on whether the interview was actually conducted and/or whether some of the data reported in the questionnaire are accurate.

---

[5] The control limits are 3 times the standard error, above and below the central line, and depend on the sub-group size.

Table 1. Possible interventions by the main results of the monitoring procedure

| TYPES OF OUT-OF-CONTROL EVENT | | | | POSSIBLE ACTIONS |
RESPONSE RATE	ACTIVITY RATE	ELIGIBILITY RATE	NON-INTERVIEW RATE	
Below the LCL or in control	Above the UCL	Below the LCL (too many ineligible units)	Below the LCL or in control	*Interviewer to be checked:* he/she might cheat in surveying ineligible units (they are given a short questionnaire that is paid as a completed interview)
Below the LCL	Above the UCL	--	Above the UCL or in control	*Interviewer to be re-trained or generalized problem:* he/she shows a high activity rate (*i.e.*, he/she assigns a lot of final outcomes), but this is due to a high non-interview rate. It is important to understand if he/she faces any difficulties in the contact phase or if the contact information is not updated. If the entire province is flagged, then the problem of outdated contact information is generalized
Below the LCL	--	Above the UCL	--	*Interviewer to be re-trained:* units surveyed by him/her are mostly eligible, but he/she has problems in completing the interview
Below the LCL	Below the LCL	--	--	*Interviewer to be checked:* he/she is slow and possibly lazy and should be invited to put more effort in his/her work

6. Conclusions

The monitoring procedure for the NPIs census is developed to understand whether the operators 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 to support the survey-specific monitoring goals and then assist in finding inefficiencies in the 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.

The procedure is used next to the traditional reports and under a close cooperation among methodologists, fieldwork supervisors, and survey managers. This allows 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 is required.

Finally, this experience will be extremely important to understand whether this approach is suitable for other censuses or any other survey that needs to monitor the fieldwork.

References

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

Montgomery, D. C. (2009). *Introduction to statistical quality control*, *Sixth edition*, John Wiley & Sons Inc., Hoboken, (NJ).

Reed, S. J, Reed, J. H. (1997). The use of statistical quality control charts in monitoring interviewers, in *Proceedings of the Joint Statistical Meetings*, *Survey Research Methods Section*, American Statistical Association, Alexandria, (VA).

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