

## **Towards the adoption of adaptive contact strategies of units involved in business surveys**

Giampaola Bellini<sup>1</sup>; M.C. Casciano<sup>2</sup>; S. Filiberti<sup>3</sup>; M. Piaggese<sup>4</sup>; M. Rinaldi<sup>1</sup> (Istat)

[bellini@istat.it](mailto:bellini@istat.it); [casciano@istat.it](mailto:casciano@istat.it); [filibert@istat.it](mailto:filibert@istat.it); [piaggese@istat.it](mailto:piaggese@istat.it); [mrmarina@istat.it](mailto:mrmarina@istat.it)

### *Abstract*

Since data collection activities have been centralized in Istat in 2016 in a specific Directorate, several actions have been taken in order to standardize procedures adopted at different level in different surveys, grouped by sector. First aim was to increase response rate, while adopting similar strategies to treat units of the same sector, from the beginning of the data collection process till the end, using a centralized and dedicated Contact center providing assistance.

Recently data collection activities have been focusing also on data quality while running data collection. As propensity of enterprises to participate to surveys can't be longer increased, more attention has to be paid to quality of data collected. The first attempt to adopt adaptive and responsive contact strategies has been run in the extraordinary survey called Business situation and prospects during and after the Covid-19 health emergency, run in year 2021. A reason to make the experiment on this survey was determined also by shortness of the data collection phase and the necessity to increase data collection effectiveness.

Monitoring response rate of targeted groups of involved enterprises was of fundamental relevance in order to assess trends and related estimated data quality. In fact in order to guarantee a prefixed level of accuracy for the final estimates, alternative approaches for the re-contacting strategy, based on the interactive analysis of expected sampling errors and observed response rates, have been adopted. Main results achieved adopting such strategies will be depicted.

---

<sup>1</sup> Directorate for data collection – Istat – Italian National Statistical Institute.

<sup>2</sup> Directorate for methodology and statistical process design – Istat.

<sup>3</sup> Directorate for economic statistics – Istat.

<sup>4</sup> Authors: Bellini G. § 1, 2, 3.2, 3.2.2 (partially), 4; M.C. Casciano § 3.2.1 (partially); S. Filiberti § 3.2.1 (partially); M. Piaggese §§ 3.1, 3.2.2 (partially); M. Rinaldi § 3.2.1 (partially).

## **1. Introduction**

Field implementation of data collection for direct structural survey on businesses has been managed since 2016, when centralization of data collection (DC) activities has been set up, in a standardized manner. Mainly centralized DC was focused on adoption of new procedures, centralized IT tools, a unique inbound and outbound service run by an external dedicated society, standardized communication strategies and calendar, implementation of monitoring tools and strategies to obtain non respondent units questionnaire, and, not lastly, clear penalties management procedures.

Response rates of the surveys involved increased, if compared with previous ones, as centralized data collection took place. Nevertheless, if characteristics of the survey remain unchanged (as to say, length of data collection phase, numbers of reminders sent, kind of units involved in the survey's lists) response rate remained on average stable during the most recent years.

So far, the standard applied in soliciting non-respondent survey units has been the same for all of them, assuming that they would have the same behavior in propensity to survey participation, nevertheless experience showed that there are several factors affecting such propensity. Dealing with sample survey a bias on final estimates is introduced as the non-respondent units have peculiar characteristics. That is the reason for considering total unit non-response as a component of the non-survey error. Thus, a more targeted kind of communication has been adopted in some surveys.

This kind of approach, that is adopted when specific situations occurs, is considered one of the possible application of adaptive and specifically responsive survey. According to this approach, the reminder activity can be run on specific target units whenever specific parameters reach pre-defined thresholds. As reminders are sent to specific non-respondent units to compile the questionnaire, at the end this approach can lead to the reduction of the bias on the estimates due to the unit non-response.

An interesting case study in this field is represented by the new survey called "Business situation and prospects during and after the Covid-19 health emergency" (Covid survey in the following). It was set up to collect information in order to define new policies and strategies to support private enterprises to face constraints imposed by restrictions and social distancing during the pandemic.

This new survey has been implemented and run in three separated waves: two of them were run in year 2020, in May and October, whereas the third one was run in October 2021.

## **2. Overall survey characteristics and strategies adopted**

The questionnaire content of this new survey was oriented to collect a reasonable number of information on how companies were reacting to Covid-19 pandemic effects. Kind of measures they were taking to secure workers from health risks or initiatives – such as new businesses, more digital processes, new business models, etc. - adopted in short and medium-long terms, became object of investigation and analysis as they were not all already covered by official statistics. Thus, the latest run included more issues on recovering actions taken in order to overcome that particular period still characterized by general economic uncertainty, considering also relevance of the public support introduced with this purpose. The questionnaire content was defined in agreement with many stakeholders even if the planning process was quite quick, especially for the first run.

The random sample included 90.468 companies with enterprises (with 3 and more persons employed) active in industry, in trade and services, representative of a population of about 970 thousand units.

As the survey participation was not mandatory, in order to reduce non-survey error especially in relation with unit non-response, those enterprises were selected among the ones already answering to Businesses census run in year 2019.

The survey achieved 45.5 percent response rate on the first wave, 44.3 in the second one and 46.0 percent in the third one.

Despite the differences with other surveys of the same category, the Covid survey can be classified as a structural survey. Nevertheless, it is possible to note some main differences with other structural surveys:

- 2.1 Length of data collection phase. Number of working days (wd) devoted to this phase were between 17 and 24 thus a very short period of time compared to standard survey which length is around 60 wd. The shortness of the time interval had an impact on strategies adopted for re-contacting units.
- 2.2 Number of units involved. As the survey wanted to be representative also for small and medium enterprises, the number of units involved was quite large, reaching 90.468 units. Only the Businesses census had a larger number of units involved.
- 2.3 Burden constraint. Due to the particular and difficult time period, the collection of new information was considered a main task and objective of a National statistical office but this new needs had to cope with burden increment on involved enterprises. Thus the questionnaire was very short and included only main issues. Analyses done on time spent to fill in the questionnaire gave on average a result of 14 minutes.
- 2.4 Communication strategies. In general, compared to standards, the number of communication sent to non-respondent units done was lower. Referring to kind of units contacted, communication strategies were quite standard in the first run but changed smoothly moving to second and third one. If at beginning main strategy was to reach all the survey units with generalized reminder, later the main aim moved to the identification of targeted groups of non-respondent units, especially to reach the ones with less response rate or a higher value of the estimated sample error.
- 2.5 Response rate. Response rate was lower than a standard survey but similar to a survey including enterprises having similar size such as the Survey on Enterprise Accounting System. In fact, EAS survey, especially the module devoted to Small and medium enterprises (EAS – SME), includes a large amount of small to medium size businesses, having generally a lower propensity to survey participation. The overall response rate and the one for units with less of 250 persons employed are similar in both surveys (Table 1).

**Table 1 – Survey response rate (RR), length (in working days- wd) and presence of legal obligation (LO) and penalty (P)**

SURVEY	SAMPLE UNITS						DC length wd	Legal obligation (LO) and penalty (P)
	< 250		≥ 250		Total			
	N	% RR	N	% RR	N	% RR		
EAS - SME	77,611	43.6	-	-	81,608	45.7	69	only LO
EAS	-	-	3,997	85.9	-	-	77	LO and P
COVID-19 - w3	86,962	45.3	3,506	61.6	90,468	46.0	24	none

2.6 Legal obligation to respond and penalties. As the organization of the survey has been done in few weeks, it was impossible to include in it the legal act process that define legal obligation and penalties in the statistical survey, as it normally takes longer time. Thus, questionnaire compilation was on voluntary basis. This affected enormously response rate especially for businesses with at least 250 persons employed. In fact, analyzing differences between the response rate of those kind of units in EAS - where they are subject to penalties - the response rate reached 85.9 percent, whereas for the same kind of units in Covid survey the percentage was only 61.6 percent (Table 1).

### **3. Monitoring survey outcomes and impacts of reminder communication on businesses**

During data collection, main activity is devoted on one hand to give assistance to the respondent units and on the other hand to solicit non-respondent units. Reminders consist of written communications normally sent massively through two different channels: certified and ordinary email. On the last days of the data collection phase an outbound activity is also run to solicit the most relevant but also reluctant units.

Monitoring survey outcomes, taking into account respondent characteristics, is of fundamental relevance in order to define and adopt strategies to obtain non-respondent units questionnaires, in general and specifically.

In the following paragraphs strategies adopted and results obtained, in terms of response rate, are described.

#### **3.1 Massive reminders and response rate trends**

In the Covid survey, after the Informative letter was sent to the units sampled for the survey, then the reminders consisted of one certified and two ordinary massive emails (Scheme 1). The ordinary emails are sent to the email addresses registered for each specific enterprises that already had access to the Businesses statistical Portal.

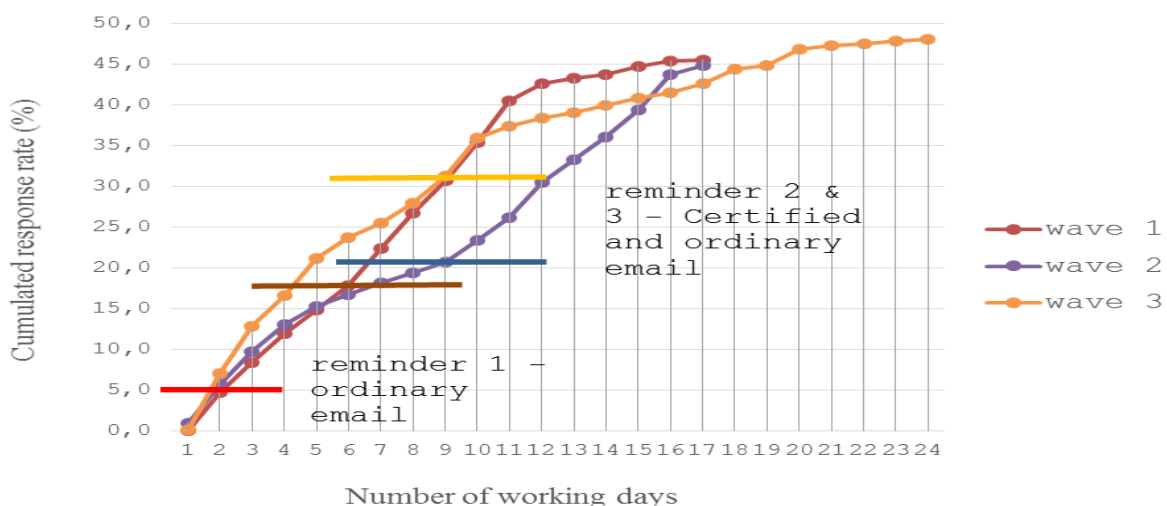
After those massive reminders, other specific reminders were sent to targeted non-respondent units: a) group of units with incomplete questionnaire; b) group of non-respondent units with high priority; c) group of units belonging to strata with a low response rate and/or with a high estimate error. If a) and b) are strategies normally adopted, c) is the the one adopted in the Covid survey and it has to be underlined that it requires a strict connection between the data collection sector and the methodological one, especially for calculating estimates of sampling error during data collection phase and for selecting the units to be solicited.

### Scheme 1 – Massive and targeted communication per wave, date and type

KIND OF COMMUNICATION	WAVE 1		WAVE 2		WAVE 3	
	date	Survey units in list (N)	date	Survey units in list (N)	date	Survey units in list (N)
Informative letter	08-12/05/2020	90.468	23-24/10/2020	90.468	16-17/11/2021	88.624
Reminder - ordinary mail 1	11-13/05/2020	83.722	23-28/10/2020	90.524	17-18/11/2021	87.954
Reminder - certified mail 1 (a)	16-17/05/2020	75.719	04-05/11/2020	74.427	25-26/11/2021	67.574
Reminder - ordinary mail 2 (b)	15-18/05/2020	76.293	05-10/11/2020	72.666	25-26/11/2021	67.049
Reminder - ordinary mail 3 (c)	21/05/2020	7.061	12/11/2020	5.893	03/12/2021	20.000
Reminder - ordinary mail 4 (d)	28/05/2020	5.800	13/11/2020	17.559	10/12/2021	20.000
Reminder - ordinary mail 5 for uncompleted questionnaire	-	-	16/11/2020	3.667	13/12/2021	5.896
Outbound service (number of contacted units) (e)	-	-	02-12/11/2020	4.525	06/12-10/12	3.297
a) In wave 3 communication of new data collection deadline						
b) In wave 3 communication of new data collection deadline						
c) In wave 1 and 2 communication to uncomplete questionnaire, in wave 3 to target group						
d) In wave 1 communication to uncomplete questionnaire, in wave 2 and 3 to target group						
e) In wave 2 recall of priority not responding units in wave 3 recall of units in the target group						

The massive reminders have an effect on the cumulated response rate trend, as showed in graph 1.

**Graph 1 – Cumulated response rate (%) per wave and working day**



### 3.2 Targeted reminders and effects on response rate

Strategy adopted to plan targeted reminders was strictly connected to the length of data collection phase, very short if compared with other surveys, to the timeliness in disseminating results and to technological constraints, as the time necessary to send the electronic reminder.

Thus the maximum results, quantitative and qualitative one, had to be achieved in short time. Referring to criteria adopted to identify critical strata and units, more details are reported in the following. Referring to technical constraint, the maximum number of submissions per day was around 20 thousand emails, thus this was chosen as maximum number of units to be re-contacted.

As already mentioned, in terms of calendar, targeted reminders were sent after the massive ones, and close to the end of the data collection phase. In wave 2, the reminders were sent just ones and very close to the end of the

data collection phase, whereas in wave 3 there were still several days of data collection to go and the reminders were sent twice and moreover a subgroup of them, the units with higher priority, was chosen to make a specific recall.

### **3.2.1 Detecting estimation domains and strata for targeted reminders**

Given the constraints of the data collection phase characterizing this survey, the main aim was to get the highest possible response rate and, at the same time, to get good quality responses. So, in order to organize the reminders, the following approach has been adopted to detect non-respondent units to be re-contacted. It was decided to concentrate resources on acquiring useful information about relatively underrepresented estimation domains and strata. So, the attention was focused on the quality of the responses: the goal was to obtain a sufficient number of respondents per stratum in order to achieve an acceptable sampling error, both at the level of total and at the level of estimation domains.

Two different criteria were implemented in order to define the methods of re-contacting the non-responding units, the first one based on empirical information and on the interviewer's experience and second one centered on statistical information.

It has to be underlined that in wave 2 a first attempt of targeted solicitation was done, thus only the first approach was adopted, whereas in wave 3 the first and the second criteria were adopted jointly, in order to identify non-respondent units to be solicited.

The first criterion was based on the response rates obtained up to that day of data collection at estimation domain and stratum level. It was important to ensure a minimum number of respondents per stratum in order to obtain reliable estimates. A further element to consider was to re-contact units on the basis of a priori knowledge of the behavioral characteristics of enterprises. In fact, it is known that the response behavior is connected to some structural characteristics of the economic units.

In wave 2, the population of interest was stratified according to some structural characteristics of the economic units: size class (4 sizes) and territorial areas (21 regions). The priority units to be solicited were the ones included in strata in which the response rate - up to that point - was lower than prefixed values.

In wave 3, the structural characteristics considered were: Nace Rev.2 division, size class (4 sizes), territorial areas (21 regions).

In this case, two degrees of priority were assigned to each economic unit to be solicited: high priority and low priority.

High priority of re-contact was assigned to non-respondent units with at least 99.5 persons employed and to units belonging to strata with no respondents units.

Low priority of re-contact was assigned to non-respondent units belonging to:

- small size strata in which number of respondent units is up to 5 sampled units in the stratum;

- medium size strata in which number of respondent units is less than 1/3 of sampled units in the stratum;
- large size strata in which number of respondent units is less than 1/8 of sampled units in the stratum.

The second criterion – adopted only in wave 3 – was based on more purely statistical information. Analysis of the distribution of response rates by stratum provided indications on which strata had to be paid more attention in order to plan reminders. At the same time, the study of the coefficients of variation helped to identify the strata for which it was necessary to concentrate data collection.

So, also in this case, it was assigned a flag to each unit with respect to the priority of being contacted again. High priority to be re-contacted was assigned to:

- units belonging to strata with quite low response rate and sample rate percentiles over predefined thresholds;
- units belonging to strata or estimation domains with relatively high coefficient of variation.

Within the same selection group with the same level of priority, the companies to be re-contacted were selected at random.

### **3.2.2 Effects on response rate**

Hereafter, more detailed information on the quantitative effect of specific communications on the target group and on the overall response rate, in wave 2 and 3, are given.

First of all, it has to be underlined that in wave 2 the solicitation message was sent to 32 percent of the non-respondent units at that time of sending, whereas in wave 3 the target group included 43 percent of non-respondents.

Going to the results, in the case of wave 2, in the group of solicited units (a basin of about 17.5 thousand units), 10.5 percent of them compiled the questionnaire, while for the unsolicited units (about 38 thousand units) the value reached 8.1 percent. So, analyzing the two groups separately, the alert seems to have had a positive effect on the RR, but looking to the effect at the overall RR, it emerges that the respondents of the targeted group contributed with an increase of 2.0 percent while the other respondents with 3.4 percent. The overall RR increase is therefore equal to 5.4 percent, which has allowed in the last two working days of data collection to rise the RR from 38.9 to the final 44.3 percent. In this case, a deepening is necessary to assess whether the increase in the targeted group is relevant in terms of increased data quality.

Looking at wave 3, the effect of the specific reminder on targeted group - at quantitative level - is as follows: in this phase there was a return of 20.2 percent of the solicited units (a basin of about 21.5 thousand units), while for the unsolicited units (about 28 thousand units) the return rate was only 11.8 percent. In terms of overall response rate, respondents in the targeted group gave a contribution of 4.8 percent while the other respondents of 4.1 percent. However, generally the unit non-responses are conditioned by business characteristics and propensity of the units to provide the requested information. In fact, main factors affecting propensity to respond are dimension, economic activity, location, degree of reachability (correct information on addresses, e-mail addresses and telephone numbers used to reach the survey units). Thus as the group reached by targeted communication includes units with a larger number of persons employed who tend to participate more actively in the surveys, so in part they would have responded even without solicitation in a larger percentage than other units.

Overall, therefore, in the 11 working days remaining at the end of the data collection period there is an overall increase of 8.9 percentage points, which allows to reach a final RR of 46.0 percent.

Thus, it is evident that the solicitation activity was certainly more effective in wave 3. The reasons for that can be found in:

- Interval time between sending reminders and end of data collection, that was longer in wave 3 than in wave 2;
- Number of solicitation activity, that was repeated twice in wave 3, with the adoption of the same priority criteria also for the recall activity;
- More robust criteria, that made possible a more appropriate identification of strata to be solicited.

#### 4. Lessons learned and future activity

The experience done with units selected in the re-contacting strategies showed that the reminder has positive effects on response rate, depending also on the days remaining to the end of data collection phase. Thus, due also to shortness of data collection interval of the considered survey, the effectiveness of solicitation has to be maximum, thus the correct selection of the units to be solicited is of capital relevance.

Nevertheless, the adaptive responsive approach needs to be further investigated, in order to assess if data quality and error determined by unit non-response have been positively affected as well.

Not last, a similar experience should be run in a standard business structure survey, in order to set up the details of the entire procedure, as to say criteria to be adopted, appropriate calendar for sending solicitation, total number of units to re-contact, in order to have a reasonable result.

Generally speaking, the adaptive responsive survey is a challenging approach but it requires a standardized model to analyze interactively the non-respondent units and their parameters, as an efficient system of communication and data exchange between three different sectors: the data collection, the economic and methodological one.

#### 5. References

Bellini G., Monetti F., Papa P. *The impact of a centralized data collection approach on response rates of economic surveys and data quality: the Istat experience*. Statistics and Economy Journal Vol. 100 (1) 2020 ISSN 1804-8765 (Online) ISSN 0322-788X (Print).

Bosso P., S. Curatolo, G. Fazzi, M. Murgia and A. Nuccitelli. *The survey on the impact of the pandemic on Italian companies: the data collection's point of view*. Data Collection in times of COVID-19 - UNECE2020. 05 – 08 Oct 2020. Virtual Geneva Switzerland meeting.

[https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.58/2020/mtg4/DC2020\\_D1-3\\_Istat\\_-\\_Bosso\\_et\\_al\\_def\\_P.pdf](https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.58/2020/mtg4/DC2020_D1-3_Istat_-_Bosso_et_al_def_P.pdf)

Cicchitelli G., A. Herzel, G. E. Montanari. *Il campionamento statistico*. Bologna, 1997.

Eurostat. *Handbook for Monitoring and Evaluating Business Survey Response Burdens*. Luxembourg, 2003. <https://ec.europa.eu/eurostat/documents/64157/4374310/12-handbook-for-monitoring-and-evaluating-business-survey-response-burden.pdf/600e3c6d-8e8d-44f7-a8f5-0931c71d9920>



Istat. *Censimento permanente delle imprese 2019: i primi risultati*. Roma, 2020.  
<https://www.istat.it/it/files//2020/02/Report-primi-risultati-censimento-imprese.pdf>

Istat. *Situazione e prospettive delle imprese dopo l'emergenza sanitaria Covid-19 Statistica report*. Roma, 4 febbraio 2022. [https://www.istat.it/it/files/2022/02/REPORT-COVID-IMPRESE\\_2022.pdf](https://www.istat.it/it/files/2022/02/REPORT-COVID-IMPRESE_2022.pdf)

Schouten B., M. Calinescu and A. Luiten. What are adaptive survey designs? In *Survey Methodology*. Volume 39, Number 1. Statistics Canada. Code 12-001-X. June 2013.