Effects of Istat CDC (Centralised Data Collection) approach on the reduction of the Total Survey Error

Loredana De Gaetano – Istat - Directorate for Data Collection **(degaetan@istat.it)**

Pasquale Papa – Istat - Directorate for Data Collection **(papa@istat.it) (Speaker)**

BERGAMO- JUNE 10, 2019 – JUNE 12, 2019

TSE DEFINITION AND USE

Wide approach of conceptual framework aimed at enhancing all problems facing surveys beyond those of sampling error (Groves, Lyberg, 2010)

A value of TSE is in the survey design phase in which it takes on the role of criterion, in the presence of a set of alternative possibilities, of which it should be preferred

In this article the focus is not only on the planning phase of the survey but also to address several error-reduction strategies that can be used during the implementation phase (Biemer 2010)

The TSE can be considered as the difference between a population parameter and the estimate of the parameter based on a sample survey or on a census. Traditionally it has two components: sampling error and non-sampling error

TSE BREAKDOWN

The survey literature decomposes non-sampling errors into five general types or sources: (Biemer 2010).

Specification error

Frame error

Nonresponse error

Measurement error

Processing error

Modeling/estimation error

TSE and QUALITY

- TSE is not the only way to think about information quality
- The user orientation to quality is normally absent
- The TSE paradigm can be viewed as part of a much larger design strategy that seeks to optimize surveys by maximizing Total Survey Quality (Biemer 2010)
- There are different Survey Quality Frameworks, they include the following dimensions: Accuracy, Credibility, Comparability, Usability/Interpretability, Relevance, Accessibility, Timeliness/Punctuality, Completeness, Coherence

PERSPECTIVE OF PRESENTATION

In literature the concept of TSE normally refers to the breakdown of the error with reference to a single survey

A different perspective consists of focusing transversely on factors insisting on all surveys or on clusters of surveys of a NIS, considering technical, statistical and organizational solutions

In particular the scope of this work is to verify the effects of a centralized data collection model on the TSE

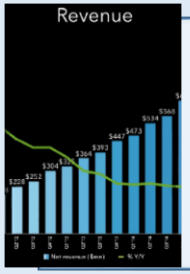
Centralised Data Collection (CDC) and Total Survey Error (TSE)

TSE identifies two major divisions based on sampling variance and bias on one hand and non sampling errors on the other

Errors of non- observation usually include coverage error, sampling error, and both unit and item nonresponse

Errors of observation involve differences between reported/recorded values of a survey variable and some ‘‘true’’ or underlying value

Introduction in Istat (Italian National Statistical Institute) of CDC allowed non- observation error reduction in surveys by increasing response rates



CDC involved the observation error reduction by fostering innovative data collection tools and the standardization and harmonization of procedures

Centralised DC in Istat: main characteristics

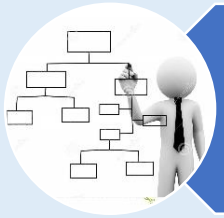
During 2016 Istat launched a Modernization Program

The program designed and implemented a new organizational set-up characterized by the centralization of all the support services, clearly separated from thematic statistical production

The new model restricts the role of production structures only to thematic aspects, while the “cross” expertises are all assigned to specialized sectors

The "transversalization" of many services fostered specialization of HR and harmonization of procedures notably in the field of Data Collection (DC)

Centralised DC in Istat: main characteristics



The introduction of a specialistic data collection, led to review of the organizational structure of data collection processes and to redesign of many of the management procedures

Before reorganization, statistical processes were organized according to the classical 'stovepipe' model, that involved independent, non-integrated, processes including all the necessary skills

The old approach involved low overall efficiency level due to overlapping redundancies and lack of integration among processes, increasing the costs of the surveys

Among the main Program there is also the valorization of administrative sources for statistical purposes and the construction of an integrated system of registers

Representative examples of process innovation introduced in different data collection domains

Examples concern mainly the implementation stage of the surveys

- Introduction of CDC in ISTAT CAWI business surveys that experienced a clearly increasing trend in response rates of the surveys associated to a significant reduction of the data collection period
- Reduction of survey coverage error deriving from an integrated approach to quality of survey lists, in terms of up-to-date information, completeness, normalisation, contact information available
- Application by ISTAT from the end of year 2018 of the mixed-mode technique in agricultural surveys

Effects on both observation and non observation errors

Average response rates for CAWI clusters before and after CDC introduction

Cluster: Short term economic surveys (CAWI)	• + 19.4 pp
Cluster: Structural economic surveys (CAWI)	• +11.0 pp
Cluster: «Sectorial» culture and agriculture surveys (CAWI)	• +6,3 pp

Main effects on non-observation error

Average response rates for CAWI surveys before and after CDC introduction

Representative examples of increase of the average response rates (rr) for single surveys carried out before and after CDC implementation	
Survey	**Increase of the average response rates (pp)**
Inward Fats survey	+24.2
Monthly survey on retail sales (MRS):	+28.0
Prodcom survey:	+ 11.6
Small and Medium business survey (SBS):	+11.0
Community innovation Survey (CIS):	+ 15.1
Fertilizing distribution	+23.0
Source: Elaboration on data extracted from Business Statistical Portal	

Effects on data collection periods

Average reduction for business structural surveys (CAWI)

- 37.2 solar days

Average reduction for "sectorial" surveys (CAWI)

- 53.3 solar days

Main effects on Total survey quality: timeliness

Innovations introduced in surveys by CDC

Innovations introduced in surveys are mainly based on:

1. The design and implementation of innovative **management tools and services**

2. Rationalization of the **management processes and procedures**: standardization and generalization

Example of innovative tool introduced in business surveys

The Business Statistical Portal (BSP)

Objectives :

- Establish bi-directional communication with units involved in surveys
- Streamline the operations required by respondents to fulfill their response obligations, with an overall reduction of the respondent burden
- Increase both ordinary and extraordinary (e.g. news) communications on the survey events and activities
- Standardize and harmonize data collection procedures in order to increase overall efficiency

Effects on both observation and non observation errors

Example of innovative management services

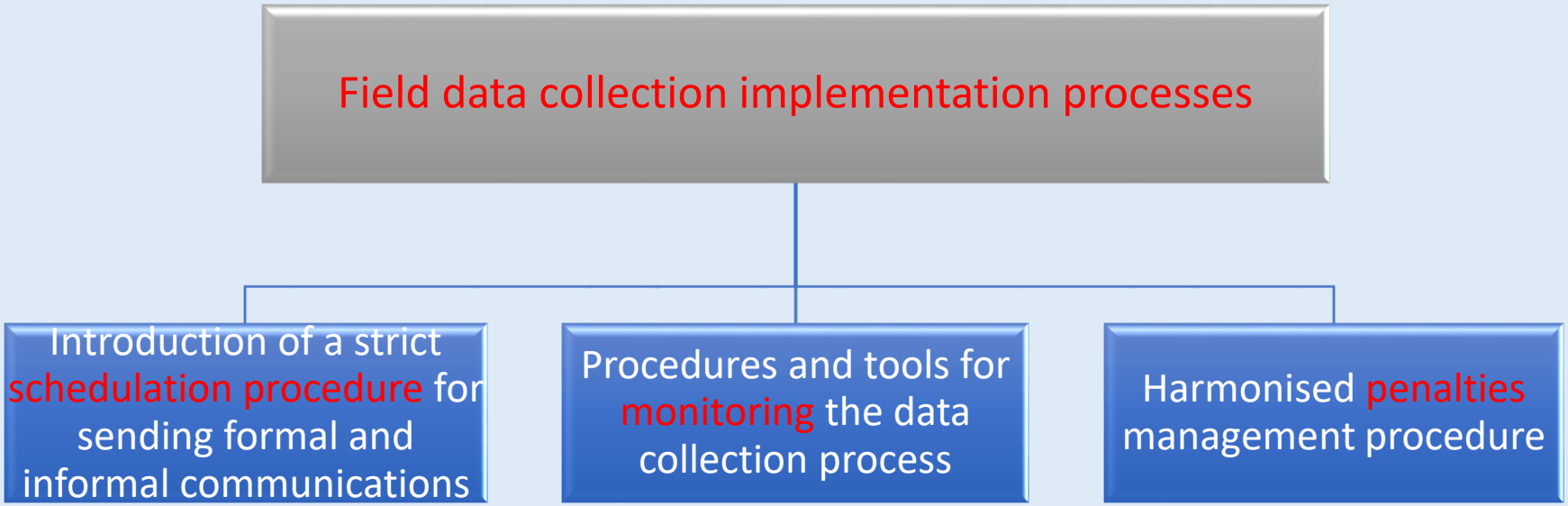
Centralised inbound and outbound Contact center services

- Progressive centralization of:
  - support and assistance services addressed to the units involved in the surveys (inbound)
  - telephone alert and reminders addressed to non-respondent units (outbound).

The coordinated management of the service ensures strong standardization

Effects on both observation and non-observation errors

Examples of process standardization



Effects on both observation and non observation errors

2-nd Example: generalized criteria for the definition of survey lists in agriculture surveys

The coverage of the list must be evaluated especially with reference to problems of under-coverage. Failure to update and inaccuracies in the information identifying the units in the list, can prevent contact, resulting in loss of accuracy of estimates and on TSE

- The lists management process includes the following steps:
  - acquisition from the Production Service that has in charge the survey
  - link to the Register, in order to verify the status of activity, any insolvency procedures or corporate events
  - check and correction of territorial references, updated periodically by ISTAT
  - formal control, standardization and correction of the certified mailing addresses (PEC), ordinary e-mails and telephone numbers available in the initial database
  - integration and updating of the most recent PEC addresses through information available in specific databases
  - punctual integration of the missing postal codes for farms without PEC
  - Normalization of "telephone" fields
- Example of TSE control in DC implementation stage

Main effects on non observation errors

Example 3. Application of a mixed-mode technique in agricultural surveys: design stage

- The introduction of a centralized data collection model has brought significant innovations in data collection processes for agricultural statistics
- Common feature of these processes is the adoption of a mixed technique data collection strategy which envisage the combined use of the CAWI and CAPI techniques or by CAWI and CATI techniques
- From the year 2018 a set of current surveys have started to use the mixed-mode technique according to the following modalities:
  - Sequentially and without overlap between the two techniques, use of the CATI / CAPI technique;
  - At the same time industrial companies (milk processing and processing of white meat and red meat) migrate into the Business Statistical Portal and adopt the single CAWI technique, as they are used to employ this tool;
  - Initially the survey is carried out using the CAWI technique, for a data collection period of two weeks;
- The possibility for the respondent to choose the preferred DC mode and the adaptation of the mode to the characteristics of the respondent is targeted to reduce observation errors
- TSE control mainly in the design stage

Main effects on observation error

Example 3. Application of a mixed-mode technique in agricultural surveys: implementation stage

- The introduction of a CDC approach and of a mixed-mode technique increased process efficiency, fostering standardization and harmonization of DC procedures. Application of CDC is expected to provide Total Survey Error reduction.
  - In the start-up phase of the on-line survey, the DC implementation concern Control and harmonization of the survey lists, Production of mailing lists
  - In the phase of conducting the on-line survey, the activity carried out concern assistance to farms by Centralized contact center services inbound and outbound, development of generalized procedures for monitoring data collection
  - In the closing phase of the on-line survey, the activity carried out concerns: implementation of generalised IT procedures for the definition of the lists of companies still not responding and possible integration with other information
  - During the CATI data collection phase by the external company, the activity carried out concerns the development of generalized procedures for monitoring, the specialised external company that carries out the activity, through indicators of provisional and definitive outcomes

Main effects on observation error

Conclusions

The CDC model has positive effects on the TSE of various surveys conducted by Istat, with particular regard to those conducted with the CAWI technique. The empirical results in fact attest a reduction of non-observation error due to significant increases in response rates. This with the same resources employed in the statistical survey processes

The same model also has positive effects on other dimensions of Total Quality of the surveys, such as timeliness, Usability / Interpretability and consistency, all aspects positively influenced by the introduction of a specialized approach to data collection

The issue that is proposed in this article tends to extend the concept of TSE from the reference to the single survey to a cluster of surveys conducted by an NSI. In this context, the effects of interventions aimed at reducing TSE that may be of an organizational or technological nature cannot be limited to the simple summation of the effects on single surveys. In fact many interrelation effects are activated

Conclusions

Following an approach already widely proposed in the literature (Biemer 2010) the commitment is not limited to reduce the TSE to the planning phase of the data collection, but to consider also the implementation phase in the field

The experiences reported in this document represent examples of solutions oriented to the standardization and harmonization of data collection management processes in diversified contexts, many other examples can be considered

The increased efficiency of RD processes and the subsequent gains free up resources to be used in process and product innovation activities, in generalized TSE reduction, in quality of the outputs and to respond to new needs for statistical information expressed by users. Furthermore, the results achieved in terms of increasing response rates can trigger a reduction in the statistical burden on respondents

Effects of Istat CDC (Centralised Data Collection) approach on the reduction of the Total Survey Error

Loredana De Gaetano – Istat (degaetan@istat.it)

Pasquale Papa - Istat (papa@istat.it) (Speaker)

BERGAMO- JUNE 10, 2019 – JUNE 12, 2019

Istat | Istituto Nazionale di Statistica

Questions

- Do you think that a centralized approach to data collection can positively influence the control of the TSE?
- Do you think that a cross approach in the control of the TSE, referring to clusters of surveys and not to individual surveys is applicable from an operational point of view and can give interesting results?
- Do you think that the analysis and assessment of the TSE should be concentrated on the planning phase of the investigation or that the implementation phase should prevail?