



CONFERENCE OF EUROPEAN STATISTICIANS

WP.10

Workshop on Statistical Data Collection: Riding the Data Deluge

20 April 2015

29 April – 1 May, Washington D.C., United States of America

Efforts to Improve Quality of Statistics in Turkstat: Centralising Data Collection and Quality Check Operations via CATI

Bilal Kurban (Turkstat, Turkey)

bilal.kurban@tuik.gov.tr

Abstract and Paper

Centralising data collection and quality check operations are among the activities carried out in Turkstat to increase the level of quality of statistics.

Due to automated data acquiring systems and internet (Web and e-mail) survey mode, data collection process shifts rapidly to a mixed mode strategy. This shift requires National Statistics Offices (NSO) to reconsider of efficiency and flexibility of their statistical production architecture. Redesigning of statistical production process with an integrated and process-driven approach is now a necessity for NSO's. A non-integrated data collection architecture forces NSO's to put up with many complex processes parallel to each other that it is neither efficient nor flexible.

In Turkey, in the past, first method that comes to mind and at the same time most widely used traditional data collection method was face-to-face interviews which is expensive and time consuming and also causes a high response burden. Via face-to-face interviews the data is traditionally collected through door to door visits of field interviewers. Collected data consists of answers of respondents recorded on a computer during the visit. Transport costs and travelling time are unavoidable in this case. These costs are further increased if field interviewers have to revisit respondents.

It is clear that centralising and automation of data production procedures are necessary for data collection in order to reduce manpower needs, lower costs and accelerate the process of data collection. With the help of developing technology and increasing use of administrative sources, new data collection techniques to achieve further burden reduction is now a necessity.

On the other hand, quality check activities and techniques are applied to surveys in order to achieve a quality level associated with a specified target, goal or standard. These activities comprise the steps of control and operation process, inconsistency detection and analysis, producing statistics about inconsistencies at survey level, origin and cause of inconsistency analysis, notifying responsible units for systematic inconsistencies, reporting final quality measurement and forwarding the results to relevant management units.

Quality checks of field surveys help to produce better quality data, meeting the need for accurate and fast information, providing users with reliable information and monitoring, measuring and increasing level of quality on compiling data.

A variety of methods and techniques is used in checking survey quality in order to detect of inconsistencies and to reduce them in long-term process. One of these methods is checking quality of a field survey by the help of Computer Assisted Telephone Interviewing (CATI) mode.

The paper tries to explain recent developments and activities implemented on centralising data collection in business and social statistics in Turkish Statistical Institute (Turkstat). This paper also



presents objectives of quality check operations via CATI, reasons of choosing CATI as data collection mode for quality checks, operation process, measurement of inconsistencies, gaining of the system and overall assessment.



Centralising Data Collection and Quality Check Operations in Turkstat

Keywords: Survey Portal, EDI, Quality Check, CATI, inconsistency

1. INTRODUCTION

Due to automated data acquiring systems and internet (Web and e-mail) survey mode, data collection process shifts rapidly to a mixed mode strategy. This shift requires National Statistics Offices (NSO) to reconsider of efficiency and flexibility of their statistical production architecture. Redesigning of statistical production process with an integrated and process-driven approach is now a necessity for NSO's. A non-integrated data collection architecture forces NSO's to put up with many complex processes parallel to each other that it is neither efficient nor flexible.

In Turkey, in the past, first method that comes to mind and at the same time most widely used traditional data collection method was face-to-face interviews which is expensive and time consuming and also causes a high response burden. Via face-to-face interviews the data is traditionally collected through door to door visits of field interviewers. Collected data consists of answers of respondents recorded on a computer during the visit. Transport costs and travelling time are unavoidable in this case. These costs are further increased if field interviewers have to revisit respondents.

It is clear that centralising and automation of data production procedures are necessary for data collection in order to reduce manpower needs, lower costs and accelerate the process of data collection. With the help of developing technology and increasing use of administrative sources, new data collection techniques to achieve further burden reduction is now a necessity.

On the other hand, quality check activities and techniques are applied to surveys in order to achieve a quality level associated with a specified target, goal or standard. These activities comprise the steps of control and operation process, inconsistency detection and analysis, producing statistics about inconsistencies at survey level, origin and cause of inconsistency analysis, notifying responsible units for systematic inconsistencies, reporting final quality measurement and forwarding the results to relevant management units.

Quality checks of field surveys help to produce better quality data, meeting the need for accurate and fast information, providing users with reliable information and monitoring, measuring and increasing level of quality on compiling data.

A variety of methods and techniques is used in checking survey quality in order to detect of inconsistencies and to reduce them in long-term process. One of these methods is checking quality of a field survey by the help of Computer Assisted Telephone Interviewing (CATI) mode.

The paper tries to explain recent developments and activities implemented on centralising data collection in business and social statistics in Turkish Statistical Institute (Turkstat). This paper also presents objectives of quality check operations via CATI, reasons of choosing CATI as data collection mode for quality checks, operation process, measurement of inconsistencies, gaining of the system and overall assessment.

2. CENTRALISING DATA COLLECTION IN TURKSTAT

Internet survey and Electronic Data Interchange (EDI) technologies are among the cheapest survey modes for the reason that they do not require interviewers. “Web survey may be up to 25 times cheaper than face-to-face interviews per sample unit” (Barry Schouten and Gabriël van Dam, Statistics Netherlands, Mixed-Mode Data Collection–Recent Developments At Statistics Netherlands, Seminar on Statistical Data Collection, 2013). As well as being cheap, internet is a new communication tool we can reach some populations we missed with other survey modes. All the business surveys in Turkstat has been started to be carried out via internet and EDI technologies modes in 2014. The main motivation for this action is the reduction of costs without compromising the quality of the surveys as well as accurate and up to date data is desired to be achieved as soon as possible.

In order to keep pace with constantly evolving technology in the world and in our country, to reduce labour and time costs to a minimum and to make timely statistical data production, on the basis of the provision of article 8 of the Statistics Law of Turkey on "Statistical units, in the framework of the fundamental rights and tasks identified in the Constitution, are obliged to submit to the Presidency all the required information or data, which will be used for producing the official statistics on economy, social, demography, culture, environment, science, technology and other required fields, completely, accurately and free of charge in the form, period and standards specified by the Presidency.", Turkstat has been decided to conduct business surveys through only applications developed on the internet hereafter.

Receiving business and social information is essential for the generation of statistical data and economic indicators. Giving the requested information complete and accurate for households or individuals and the statistical units other than these is a legal requirement otherwise according to article 54 of the Statistics Law of Turkey; administrative fines are applied.

Statistics Law of Turkey guarantees that data collected from persons, institutions and organizations will be kept confidential and the confidential data compiled, processed and preserved for the production of official statistics cannot be delivered to any administrative, judiciary or military authority or person, cannot be used for purposes other than statistics or as an instrument of proof.

2.1. Business Statistics

In Business statistics, two different ways of business survey respondents answering facilities are provided: Survey Portal (SP) (self-completion of computer-assisted questionnaires via Web survey) and EDI Technologies (standard business reporting programs for reporting business data and electronic interchange)

2.1.1. Survey Portal

Broad changes in the social and technological context has required Turkstat to improve timeliness while reducing response burden by increasing the quality of data collection process of statistical production or at least maintaining the same level of quality. To meet this requirement, Turkstat Information Technologies Experts developed a SP which allows Web-based data collection.

SP is a single entry point platform for Web-based data collection from respondent units. Besides being a platform for statistical data provision from respondent units, SP also includes supportive information for respondent units related to how to use the platform. Also, reminders can be automatically sent to the e-mails of respondent units. SP provides respondent units to fill in electronic questionnaires by themselves in a comfortable and a safe manner. Thus, the number of surveys collected electronically via self-completion mode has increased. An effective technology covering all the statistical operations (business and social) and an integrated infrastructure which supports statistical production is targeted with this system. Following business surveys the system is designed also for social surveys which will come into operation in the next year.

Before taking part in a business survey, an “invitation letter” is being sent to enterprises to use SP, containing aim of survey, basic explanations related to survey, due dates and legal issues, explanations about access to survey and how survey questions are to be filled in. Respondent units are expected to do the data entry on the dates specified in the letter. Survey questionnaire is open for data entry on the Web for specific time intervals for each survey.

2.1.2. EDI Technologies

EDI application developed by Turkstat Annual Business Statistics Experts is an electronic data transfer system which allows needed business data to be the compiled from Accounting Applications by transferring them to Turkstat servers via Web services. EDI application is aimed to provide the followings;

EDI application is achieved in two ways; Web Service Integration and XML file upload. Through Web service integration, respondent units are able to fill in surveys electronically with the help of modules developed in their accounting applications. Related modules transmit the entered data to Turkstat systems by contacting EDI Web services. This process is achieved by integrating modules into accounting applications of respondent units by technical team of product manufacturer with the help of Turkstat guides. Instructions for accounting companies or producers have been developed. Web service functions are designed not only for a specific survey; it is also applicable for other surveys or projects when their XML structures are created in the future.

The respondent units which are not able to manage to integrate with Web services but are able to create needed XML file can complete their surveys via XML File Upload option. For the respondent units creating their own XML file, SP also provides an opportunity to send their survey responses to Turkstat system.

XML files are the same for both Web Service Integration and XML File Upload. Only the method of transmission of data to Turkstat is different. The participants using these two scenarios will be informed about questionnaire evaluation process via their e-mails. For example; participants will be notified by e-mail about the stages of transferring XML file to Turkstat system and result of checking XML structures for compliance.

2.2. Social Statistics

Turkstat is planning to redesign all its national social surveys in a way that data collection for them is achieved also via Web as well as effective use of CATI mode. Thanks to SP, Turkstat has a platform in order to implement these two modes too.

Currently, CAPI is used as data collection mode for all social surveys conducted in Turkstat. CATI mode is currently used for only quality check purposes of surveys in Turkstat.

2.2.1. Future Plans about CATI Mode Usage in Social Surveys of Turkstat

Labour Force Survey (LFS) and Consumer Tendency Survey (CTS) are planned to be initial surveys in which CATI mode usage will be tested. For both of these surveys Turkstat is planning to implement a comprehensive pilot study in 2015.

Turkstat is planning to carry out the LFS with CATI mode after first wave. In Turkey right now all waves of LFS are carried out by CAPI mode. And from the beginning of 2015 Turkstat intends to try transforming this process into mixed mode like that first wave will be still with CAPI mode but others with CATI mode. LFS is planned to be the first survey which will be carried out by mixed mode including telephone. LFS is planned to be the first survey which will be carried out by mixed mode including CATI.

On the other hand, CTS in Turkey is carried out in cooperation with Turkstat and Central Bank of Turkey since 2003 by using CAPI. According to the institutional target for CTS, it is planned to collect data with CATI as a single mode.

Main challenges for implementing CATI applications are how to handle the phone frame problem for those not panel surveys and how to change over the process from CAPI to CATI (steps and what needs to be done in transition to CATI application) for the panel ones. After implementing pilot study in 2015 solutions or answers to potential problems that may arise are expected to be clearer.

Accuracy Test Operation of Phone Records Of Private Companies

A test was conducted on the accuracy of the registered names and addresses of persons involved in the sample selected through an example of the records from two of four main GSM companies in Turkey. The result of the test that while the accuracy of names of persons is confirmed by a rate of 59%, the addresses of them are confirmed by the rate of 92% is promising.

3. QUALITY CHECK OPERATIONS VIA CATI

The process of quality check operations of field surveys via CATI is simply based on the control of whether the data collected from field by field interviewer are the same with the data collected via check operations on the phone by operator for the same reference respondent and time.

3.1. Objectives of CATI Quality Check Operations

The aims of CATI data collection for the purpose of quality checks of field surveys are to monitor, to measure and to improve the quality level of the collected data. These operations contribute to both supporting to produce higher quality data and presenting more reliable information to users.

3.2. Reasons of Choosing CATI as Data Collection Mode for Quality Checks

CATI is a system which allows collecting data by conducting surveys centrally via telephones with the help of computer-assisted questionnaires. In data collection phase of statistical process, among the reasons why the CATI mode can be preferred for the purpose of measuring the quality of the collected data via field interviewers and performances of them is that the results of quality checks are obtained very fast and easily with a low cost, relatively less workers are needed, whole process can be controlled from one center and it provides more standardised data.

3.3. Quality Check Process Stages

A preliminary analysis of the survey if it is appropriate for the selection, quality check structure design and process planning, detection of the questions to be checked, determining types and weights (if exist) of inconsistencies, preparation of operation flow, development of application program, test, training and improvement activities, CATI operation, internal control of the operation, generating inconsistency statistics and reporting and evaluating results create all stages of the process, respectively.

3.4. Type of the Characteristics to Be Questioned

The characteristics to be questioned in the course of quality check operations can be grouped under two main headings. First; permanent questions which are asked in each project in the same way. Second; survey questions which are convenient for CATI mode and picked out of related survey in order to be checked for consistency. Among permanent questions, as well as general impression of the field interviewer and the survey itself also identity and contact information of respondent are questioned. As for survey questions, they are the questions from content of survey having key importance.

3.5. Selection of Respondents

Selection of respondents for quality checks of a survey may be full enumeration or sampling depending on the size of the survey.

The survey frame of quality check operations consists of response units having at least one meaningful phone number from that survey.

When selecting samples for a survey for CATI quality check operations, number of local units (e.g. regional offices) and field interviewers in each of them involving that survey should be taken into account. Sample sizes for each local unit may be determined proportional to the one in the master survey which will be checked for quality.

3.6. Measurement of Inconsistencies

After completion of the data collection process via CATI, resulting data set is compared to the one collected by field interviewer for selected questions. It is very critical to have a minimum time-lag between master survey and CATI quality check one. If there is a difference between the responds for the same characteristics in these two data sets then there occurs an inconsistency. Uncertain or doubtful cases cannot be evaluated as an inconsistency. As for permanent questions asked during CATI operations, negative responses from respondents to these may be treated as the same as in the case of survey questions or can be evaluated for only information purposes for further analyzes.

After getting results of quality check interviews analyzes are performed. In these analyzes, resulting inconsistencies, as well as information on the cause of inconsistencies, can be reported at the level of the survey itself, the units collecting field data (e.g. regional offices) and field interviewers.

The calculation method of inconsistency rate is simply the division of the frequency of detected inconsistencies after comparing two data sets to the total number of questions subject to this consistency measurement asked to respondent on the phone. This rate can be multiplied by hundred so it can be expressed in percentage when reporting.

3.7. Minimizing Mode Effect

During CATI quality checks of the surveys which are conducted via face-to-face interviews it is inevitable to have mode effects. It is impossible to avoid these mode effects completely but it should be tried to minimise them at the quality check questionnaire-design stage of the survey cycle.

3.8. Gaining of the System

Inconsistency rates of changes by year resulting from the aggregation of ongoing quality checks of field surveys in Turkstat since 2005 are presented in the figure 1.

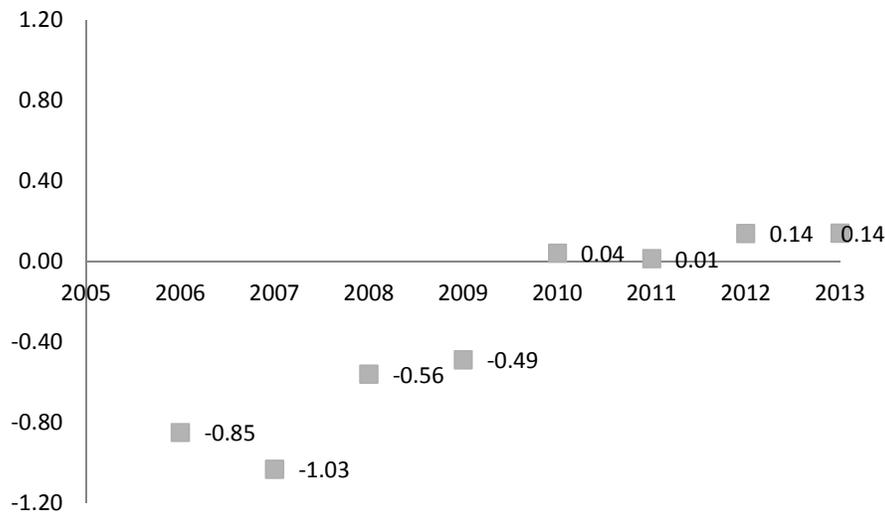


Figure 1. The change of inconsistency rates by year

As can be seen from the graph that from 2005 in which quality check operations were introduced in Turkstat each year inconsistency rates have declined regularly and dramatically until 2010. After 2010, the inconsistency change rates have become more balanced and have begun to focus on a certain inconsistency rate. Also after 2010 to the present a slight increase in the change of inconsistency rates can be observed and this may be because of steadily increasing number of surveys involved into the check operations and excluding the surveys which are no longer needed to be checked. According to the figure 1, it can be also said that, due to the quality of check operations, variation of inconsistency rates of field surveys of Turkstat has been reduced.

With quality checks via CATI, beyond detecting inconsistencies it is intended to find the sources of these and then to alert related units about them. Once it is announced that a

quality check operation for a specific survey will be performed, field interviewers of that survey work more carefully and with fewer (or even zero) errors than before.

4. CONCLUSIONS

Turkstat aims to gather and process information more quickly, more accurately, and more effectively than before. Because of these reasons, Turkstat always needs for an accurate and up-to-date data in very short time. Accordingly, to support the right decision-making process, Turkstat maintains trying and utilizing these new data collection modes efficiently.

In business statistics, new methods of data collection such as Web surveys and EDI applications will be continuously promoted in Turkstat. As respondent units become familiar with these new modes it is expected to reduce application problems such as password reminders in the coming years.

In social statistics, CAPI is still predominant practice. But in the close future, it is planned to utilize other central data collection modes.

Regarding CATI quality check activities, these contribute to improve quality of field surveys. In the projects where these checks are performed generally positive developments have been provided. With quality check operations via CATI, answers of some selected key questions from a field survey are checked once again and according to resulting frequency of inconsistencies survey items (such as survey methodology, adequacy of the trainings provided, clarity of survey definitions and concepts) can be revised again. Moreover through these checks impression about a survey and/or a field interviewer can be obtained from respondents. Field interviewer perception that there will be a quality check operation just after field application contributes to reduce the number of field interviewer-induced errors and thus improve general quality of surveys. To sum up, CATI quality check operations help to provide a reasonable level of error arising from field.

One of the objectives of field survey check activities by phone is monitoring, measuring and reporting of quality of field application part of statistical production. Thus; it helps to reduce problems of data collection process from field. Other main objective is that it contributes to find and improve weaknesses (if any) of statistical production process with active participation of all units. In this context, the quality of check operations can be considered as a part of the statistical system.

REFERENCES

- [1] Paulo Saraiva dos Santos and Almiro Moreira, Statistics Portugal, Creating A Data Collection Department: Statistics Portugal's Experience, Seminar on Statistical Data Collection (2013)
- [2] Turkstat Web pages
- [3] Tiaray Razafindranovona (Insee, France), Gaël de Peretti (Insee, France) and Alexis Barrau (ENS Cachan, France), The Philosophy Of French Experiments on Internet And Mixed-Mode Data Collection, Seminar on Statistical Data Collection (2013)
- [4] Natale Renato Fazio, Manuela Murgia, Andrea Nunnari, ISTAT – Italian National Statistical Institute, The Business Statistical Portal: A New Way Of Organizing And

Managing Data Collection Processes For Business Surveys In ISTAT, Seminar on Statistical Data Collection (2013)

- [5] İstatistiklerle Kalite Kontrol, Sorularla Resmi İstatistik Dizisi 11, 2011, Turkstat
- [6] B.Kurban and A.Özdemir, TÜİK'te Kalite Denetim Çalışmaları, 1. Genç İstatistikçiler Sempozyumu – Hacettepe Üniversitesi – ANKARA, 2013
- [7] E. Mawby, H. Smith and C. Hewat, Statistics New Zealand Navigates Towards Online Data Collection: Planning And Implementing An Internet-Based Survey, Seminar on Statistical Data Collection (2013), 1 – 6