ROMANIA

National Institute of Statistics

Quality improvement of the statistical processes on data collection and data processing

Standard guide Romania

Bucharest, 2008

Contents

1. Introduction

During the last six years our statistical institute has made a lot of progresses in the field of quality. Quality awareness should be a solid statement in all statistical institutes. The TQM (Total Quality Management) is the main quality concept that will be implemented by NSI – Romania.

The quality concepts can be a bit difficult to understand and implement, but with a solid study of all the materials available, a specialized training of all staff, from the top management to bottom down, all difficulties can be overcome.

TQM is a complex concept that incorporates a series of core values that can be summarized as four concepts: Leadership, Staff involvement, Organization oriented to quality and Review and continuous improvement. The current study (guide) is not intended to analyze all four concepts, not even one of the concepts. It is only intended to analyze a small part of the whole TQM concept.

When we analyze the quality of a survey we have to take into consideration many aspects, but the two main aspects that are usually taken into consideration are the quality of the product and quality of the process. A theory states that a product is a result of different processes. A statistical product is the result of a series of statistical processes. The quality of the statistical processes has a direct impact on the final product and its quality.

The current guide is intended to analyze two very important key processes used in statistical production: data collection and data processing. Both processes can influence the quality of the final product and thus improving the process quality can improve the quality of the final product.

In order to determine the quality of a statistical process we have to identify a series of key process variables and then determine the best measurement tools for the key process variables. After the measurement of the process variables, some adjustments, if needed, will be performed on the processes. Finally the product quality will be reviewed and an analysis should be made to see how much of the quality of the final product was affected.

The current study was divided into four main sections. The first two sections contain informations on the key process variables chosen for the study and measurement tools used for the analysis of the key process variables. The last section contains the study. It is divided into subsections that refer to the two chosen surveys, an analysis of the key process variables used in the data collection, an analysis of the key process variables used in data processing, evaluation of the results, conclusions and recommendations.

2. Identifying key process variables

Statistical quality control refers to the use of statistical methods in the monitoring and maintaining of the quality of products and services. One method, referred to as statistical process control (SPC) or statistical process quality control uses graphical displays known as control charts to determine whether a process should be continued or should be adjusted to achieve the desired quality.

Statistical key process variables are used to determine whether a process is according to the specifications and should be continued or it should be adjusted to achieve the desired quality. Identifying statistical key process variables can be a very difficult process, but they are essential for the statistical process quality control.

The key process variables analyzed in the study were taken from the "**Handbook on improving quality by analysis of process variables**", Eurostat. Since our study refers to data collection and data processing, the variables taken into consideration were those belonging to these processes. The variables are presented in the following table.

Statistical process	**Key process variable**	**Ability to measure the variable at NIS ROMANIA**
Data collection	Ability of respondents to answer a problem question	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Percentage of item non-response rate	Calculated
	Percentage unnecessary response	We will be able to calculate this key process variable after an analysis of other data sources
	Percentage with a mark entered in both yes and no boxes	Calculated
	Percentage without a mark in either yes or no boxes	Calculated
	Percentage with a numeric value in a mark box	Can be calculated
	Percentage with a mark in a numeric field	Can be calculated
	Percentage with more than one mark box completed where only one is expected	Can be calculated
	Percentage with a non-relevant mark in or across mark boxes	Can be calculated
	Distribution of number of complaints received by size of responding business	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Percentage of ineligible sampling units found in the sample	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Refusal rate	Calculated
	Temporarily away rate	Calculated

Statistical process	Key process variable	Ability to measure the variable at NIS ROMANIA
	Number of interviewers performing interview 2 and 5 weeks after reference period	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Interviewing time by survey	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Travel time of interviewers by survey	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Other time (eg planing) by survey	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Working hours by survey	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Total interview time by survey	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Number of planned hours by survey and survey manager	Can be calculated
	Number of successful refusal conversion attempts divided by total number of attempts	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Number of contact attempts by time period	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Number of final code units by time period and domain	Some modifications to the survey are necessary
	Interviewing time by respondent and survey	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Proportion of monitored interviews by survey	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Proportion of re-interviews by survey	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Proportion of field observations by survey	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Number of editing errors by item and survey	We are unable to calculate the variable
	Proportion of responses obtained from modes other than the main one by survey	We are unable to calculate the variable
	Proportion of proxy interviews by survey	We are unable to calculate the variable
Data processing	Time spent in manual examination of questionnaires	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Number of analysts manually examining questionnaires	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Years of experience of survey analyst in the specific survey (field editing)	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Runtime of automatic editing adjusted	We are unable to calculate the variable

Statistical process	Key process variable	Ability to measure the variable at NIS ROMANIA
	by the sample size	
	Percentage of errors detected	Calculated
	Percentage of errors corrected	Calculated
	Percentage of new errors	Can be calculated
	Reference material available to the analyst	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Timeliness of external information	We are unable to calculate the variable
	Years of experience of survey analyst in the specific survey (manual examination of errors)	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Number of failures of data at each validation gate	Some modifications to the survey and database application are necessary
	Staff performance at examining errors identified by data editing	We will be able to calculate the variable, but only after the implementation of the new questions into the surveys
	Overall coding accuracy rate	Calculated
	Accept rate	Some modifications to the survey and database application are necessary
	Number and percentage coded by mode	Some modifications to the survey and database application are necessary
	Rate of incorrectly assigned 'uncodables'	Some modifications to the survey and database application are necessary
	Frequency of types of error in coding	Some modifications to the survey and database application are necessary

As it can see from the table an analysis was done to see if we can calculate the key process variables or some modifications should be done. The key process variables were chosen based on the possibility to calculate the variables. Since the study was done on two different types of surveys, different key process variables were chosen for each type of survey. The key process variables were also differentiated based on the process.

I. **<u>For the data collection phase</u>** we have chosen for each of the surveys three key process variables.

a) For the self-administered questionnaire based survey we determined the following three data collection variables:

1. percentage of items non-response
2. percentage of items with a mark entered in both yes and no questions
3. percentage of items without a mark in either yes or no boxes

The first key process variable measures the rate of non-response for each question, thus giving us a picture of the way the respondent understands the question. A high non-response rate results in a poor quality. In case of a high rate of non-response some adjustments to the questions are necessary in order to increase quality rate.

The other two key process variables are a completion of the first key process variable. They help us measure the way a respondent understands the questions or is able to provide us with the information's that we require and can help us improve the quality of the questionnaire.

b) For the second survey which is a survey administered by an interview operator we have identified three key process variables:

1. percentage of items non-response rate
2. refusal rate
3. temporary away rate

The first key process variable is a quality indicator of the way a respondent understands the questionnaire or is able to provide required data. The other two key process variables are very different and can be calculated only for the surveys administered by an interview operator.

The refusal rate and temporary away rate are observed by the interview operator and consented in the questionnaire. Usually, the measurement of these key process variables requires information's on the status of the sampled units. The key process variables are useful for monitoring the quality of the response to surveys and can help us to apply improvement actions. One possible action is when the sample is extracted a buck unit would be extracted, where possible.

II. **<u>For the data processing phase</u>** we have identified three key process variables:

1. percentage of errors detected,
2. percentage of errors corrected
3. overall coding accuracy rate.

Although there are many more key process variables we have chosen these three because they were much easier to calculate and to asses.

The first two key process variables used in the measurement of data processing are usually done computer assisted and are an indicator of the quality of various sub-processes of data editing.

The third process variable is usually calculated where a sample of codes is available. The use of this indicator is an effective method to identify systematic errors, assess reported accuracy and improve accuracy over time.

Although, we were able to calculate these key process variables a proper assessment can be done only over time. In order to see the impact of the adjustments another assessment should be done over the next two years, using the same key process variables.

3. Measurement tools for the key process variables

Statistical process control uses sampling and statistical methods to monitor the quality of an ongoing process such as a production of statistical data. A graphical display referred to as a control chart provides a basis for deciding whether the variation in the output of a process is due to common causes (randomly occurring variations) or to out-of-the-ordinary assignable causes. Whenever assignable causes are identified, a decision can be made to adjust the process in order to bring the output back to acceptable quality levels.

Control charts can be classified by the type of data they contain. For instance:

- ***$\bar{x}$-chart*** is employed in situations where a sample mean is used to measure the quality of the output; it can also be used to monitor quantitative data such as :
  - length;
  - weight;
  - temperature.
- ***range or R-chart*** can be used to monitor process variability.
- in cases in which the quality of output is measured in terms of the number of defectives or the proportion of defectives in the sample, an **np-chart or a p-chart** can be used.

All control charts are constructed in a similar fashion. For example:

- the centre line of an $\bar{x}$-chart corresponds to the mean of the process when the process is in control and producing output of acceptable quality;
- the vertical axis of the control chart identifies the scale of measurement for the variable of interest;
- the upper horizontal line of the control chart, referred to as the upper control limit, and the lower horizontal line, referred to as the lower control limit, are chosen so that when the process is in control there will be a high probability that the value of a sample mean will fall between the two control limits.

Standard practice is to set the control limits at three standard deviations above and below the process mean. The process can be sampled periodically. As each sample is selected, the value of the sample mean is plotted on the control chart:

- if the value of a sample mean is within the control limits, the process can be continued under the assumption that the quality standards are being maintained;
- if the value of the sample mean is outside the control limits, an out-of-control conclusion points to the need for corrective action in order to return the process to acceptable quality levels.

The Pareto chart (p-chart) is one of the seven basic tools of quality control, which include the histogram, Pareto chart, check sheet, control chart, cause-and-effect diagram, flowchart, and scatter diagram.

Typically on the left vertical axis is frequency of occurrence, but it can alternatively represent cost or other important measurement unit. The right vertical axis is the cumulative percentage of the total number of occurrences, total cost, or total of the particular unit of measure. The purpose is to highlight the most important among a (typically large) set of factors. In quality control, the Pareto chart often represents the most common sources of defects, the highest occurring type of defect, or the most frequent reasons for customer complaints, etc.

The **Ishikawa diagram** (or *cause and effect diagram*) are diagrams, that shows the causes of a certain event. A common use of the Ishikawa diagram is in product design, to identify potential factors causing an overall effect.

Causes in the diagram are often based on a certain set of causes. Cause and effect diagrams can reveal key relationships among various variables, and the possible causes provide additional insight into process behaviour.
Causes should be derived from brainstorming sessions. Then causes should be sorted through affinity-grouping to collect similar ideas together. These groups should then be labeled as categories of the fishbone. Causes should be specific, measurable, and controllable.

Most Ishikawa diagrams have a box at the right hand side, where the effect to be examined is written. The main body of the diagram is a horizontal line from which stem the general causes, represented as "bones". These are drawn towards the left-hand side of the paper and are each labeled with the causes to be investigated, often brainstormed beforehand and based on the major causes listed above.
Off each of the large bones there may be smaller bones highlighting more specific aspects of a certain cause, and sometimes there may be a

third level of bones or more. These can be found using the '5 Whys' technique. When the most probable causes have been identified, they are written in the box along with the original effect. The more populated bones generally outline more influential factors, with the opposite applying to bones with fewer "branches". Further analysis of the diagram can be achieved with a Pareto chart.

The key process variables represent the main component of the graphic representations used as measurement tools.

- ***Data collection phase***

The key process variables used for the *data collection*, calculated for the *self-administered survey* were:

1. percentage of items non-response rate
2. percentage of items with a mark entered in both yes and no boxes
3. percentage of items without a mark in either yes or no boxes

The measurement of these key process variables is done by selecting a sample of returned forms identified as containing errors. Since our analysis was done mainly on the database we were able to take into consideration the whole number of units. After the analysis on the whole database and the counting of the items non-responses, we were able to assess the number of items without a mark in either yes or no boxes. After the count of the item non-response all the data was converted into percentages for analysis.

For the second survey, administered by an interview operator, the key process variables were:

1) percentage of item non-response rate
2) refusal rate
3) temporary away rate

The first key process variables was calculated taking into consideration the whole number of records from the database.
The last two key process variables (refusal rate and temporary away rate) can be determined if the collection of data is done based on the status of the sampled units. Based on the status of the sampled units we can extract from the database the number of refusals and the number of temporary away units, and then calculate the rates.

- ***Data processing phase***

The key process variables used for the data processing were the same for both surveys:

1. percentage of errors detected
2. percentage of errors corrected
3. overall coding accuracy rate

The first two process variables, *percentage of errors detected and percentage of errors corrected*, were determined from the existing data in the database, by the assistance of the computer. These two variables provide useful informations on the quality of various sub-processes of data editing. The measurement of these variables is done by counting the number of errors and then we can calculate the percentage. The number of errors corrected is done with the help of the computer assisted error correction software, and then the number of errors is recounted. The number of errors corrected is calculated from the difference between the old number of errors detected and the new number of errors detected after the computer assisted corrections. Then the percentage is calculated.

The third key process variable was the *overall coding accuracy rate*, which represents the number of correct codes divided by total number verified. In order to obtain an accurate figure we didn't take a sample of codes, we took into consideration all the codes, and with the help of the computer we have calculated the variable.

4. Case study

4.1 General presentation of the statistical surveys used in the study

As presented in the theoretical part (chapters 1-3), of the guide on "**Improving the quality of the statistical processes and production**", in order to obtain results and to be able to make a correct analyses, we have chosen two questionnaire based surveys onto which we have applied the theoretical part. The two surveys are very different, so the results obtained were very different.

The two surveys chosen for the study were:

1. "**The effective number of pig population on 1st august 2008**"
2. "**The use of IT&C products in enterprises**"

For the current study, we used two types of surveys:

- a self-administered questionnaire based survey;
- a survey administered by an interview operator based on a questionnaire.

The differences between the two types of surveys are obvious and the results obtained from the study, reflect the fact that the two surveys are very different. The two surveys are harmonized and realized according to the European regulations.

- **The effective number of pig population on 1st august**

The general objective of the survey is to obtain data on the number of pig population on 1st august. The survey is done on a yearly basis and completes another survey which obtains data on 1st April 2008. The survey is done according to the European Directive 93/23/EEC and it is sample based.

The statistical units are the individual agricultural exploitations and agricultural exploitations organized as enterprises. The size of the sample is 14.217 exploitations out of which 13.419 are individual agricultural exploitations and 798 agricultural exploitations, organized as enterprises.

For the individual agricultural exploitations, the survey is administered by an interview operator and for the agricultural exploitations organized as enterprises, the survey is self-administered.
The interview operators are mainly agricultural specialists recruited by the territorial statistical offices. The training of the interview operators is done by the statistical experts from the territorial offices.

The data obtained by the survey refers to the number of pig population and the questionnaire is a balance. The questionnaire has two sections: section A: which refers to the effective number of pig population, by types and weights and section B: which refers to the evolution of the effective number of pig population from 1st January to 1st August.

The survey is composed from the questionnaire and the methodological guide. The methodological guide contains important informations on the data completion process in the field. It also contains all the informations on the correlations between the rows.

- **The use of IT&C products in enterprises**

The general objective of the survey is to obtain data on the use of IT&C products and services in enterprises. The survey is done on a yearly basis and it is according to the European Directive 808/2004 and is sample based. The sample is extracted from the total number of active enterprises.

The statistical units are the active enterprises and the sample size is 14082 units. The survey is self-administered and the collected data refers to the use of IT&C products.

The questionnaire is divided into nine modules:

- First module (**A**) collects general informations on the use of IT&C;
- Second module (**B**) collects data on the use of internet;
- Module (**C**) collects data on the automated data exchange;
- Module (**D**) collects information referring to the electronic information partition referring to the management of the suppliers' network;
- Module (**E**) collects data on the automated information transfer inside the enterprise;
- Module (**F**) collects data on e-commerce;
- Module (**G**) collects data on the benefits from the use of IT&C;
- Module (**H**) collects data on the investments and expenditures on IT&C products;
- Module (**X**) collects general informations.

At the end of the questionnaire there are some methodological indications that explain the specific IT&C terms. The correlations, between specific rows and questions, are presented at the beginning of each module.

4.2 Analyses of the key process variables

4.2.1 Analyses of the key process variables used in the data collection

As presented in the first part of the guide we have identified some key variables used in the process of data collection. These key process variables were applied on the two chosen surveys. The results obtained were very different and we believe that the differences are both a

consequence of the type of survey used and of the length and difficulty of the questionnaires.

The chosen key process variables were:

- **percentage item non-response**;
- **percentage with a mark entered in both yes and no boxes;**
- **percentage without a mark in either yes or no boxes**.

Although there are many more key process variables that can be calculated, in order to be able to compile the data necessary to obtain those variables, we would've had to create a new questionnaire.

The three key variables chosen were applied on both of the surveys. For the survey "**The effective number of pig population on 1st august**" we were able to calculate two more key process variables:

- **Refusal rate**;
- **Temporary away rate**.

We were able to calculate these two key process variables thanks to the small chapter in the questionnaire, at the end, that gathers data on outcome of the interview.

First we will analyze the *self-administered questionnaire*: "**The use of IT&C products in enterprises**".

In order to be able to compile the necessary data to calculate the key process variables, we have analyzed both the questionnaire and the database.

First key process variable analyzed was *the item non-response*, then *the percentage of item with a mark entered in both yes and no boxes*, and last *the percentage of item without a mark in either yes or no boxes.*

The questionnaire collects data for 108 indicators. After initial analyses we have decided that we will take into consideration only the item *non-response rates* that are higher than 20%. This decision was taken because 78 indicators have a 0% item non-response rate, 4 indicators have a 0.01% item non-response rate and 19 indicators have a maximum of 0.71% item non-response rate. Out of the 108 indicators only 7 indicators have an item non-response rate higher than 20%. The full list of item non-response rate is presented in annex 1.

First two indicators that had an item non-response rate higher than 20% were items **E21** and **E22** which 1) **collect data on whether the enterprise have made buying orders (electronic or not) and 2) for which of the departments (stock level management or accounting), were the informations, on the electronic or automated transfer, relevant**.
For these two indicators the item non-response rate was about 55%. Next we have begun an analysis of the possible causes that triggered an item non-response rate so high. The conclusion was that the question has to be rewritten because it is not very clear, thus making it very difficult for the responders to answer.

Next indicator was item **E31** which **collects data on whether the enterprise uses ERP software applications for resource planning**.
The item non-response rate was 55%. The causes that determined a high item non-response rate, after careful analysis, was the fact that not every one who answers this questionnaire knows what ERP software is. Although at the end of the questionnaire there are methodological indications, these indications are very unclear and hard to understand. There should also be given examples of the possible software applications used. Also the fact that the indications are presented at the end of the questionnaire is an important fact in the high item non-response rate.

High item non-response rate had also the indicators **E41** and **E42**, which **gather informations on the use of CRM software applications for management (E41) and sharing, among departments, of clients' informations (E42)**. The item non-response rate was 55% and the main cause was the fact that there are insufficient informations and indications on what are the CRM software applications. A possible fix could be the insertion of some explanations about item requirement in front of the questions.

The item **F41** has also a high item non-response rate of about 57.66%. The item **gathers data on orders made through computer networks** (excluding e-mails). The high item non-response rate is caused by the little understanding of the terms. In the final section that contains methodological indications, there are no indications on orders made through computer networks. Some methodological indications and the change of the question could improve the response rate.

The last item with a high item non-response rate is item **F51**. This item **requires that the respondents make an estimate of the percentage of the expenses on electronic orders out of total expenses**, and has a non-response rate of 94%. The causes that led to a so high item non-response rate are multiple, like the fact that usually the accounting system doesn't register these type of expenses, and it would be very hard to estimate. Another cause is the fact that there are no methodological indications on how a respondent could estimate this figure. Also on a second look at the question, a rewrite is required in order to increase the response rate.

The second key process variable verified on the survey was the percentage of items with a mark entered in both yes and no boxes.
After a close analysis we concluded that there were no items with a mark entered in both yes and no boxes.

The third and last key process variable verified for the survey on the use of IT&C products in enterprises was the percentage without a mark in either yes or no boxes. By analyzing the questionnaires and the database we were able to calculate the percentage. The percentage obtained was 48.19%. The percentage obtained is very high, and reflects the fact that the respondents do not understand the questions and the methodological indications, which should help the respondents. A general rule in creating a good survey is that the questions be clear and easy to understand.

The second survey analyzed was "**The effective number of pig population on 1st august"**.

For this survey we were able to analyze only percentage item non-response rate. The key process variables: "**percentage of items with a mark entered in both yes and no boxes**" and "**percentage of items without a mark in either yes or no boxes**" were impossible to calculate since there are no questions that require an answer of either yes or no. But we were able to calculate other two key process variables: "**Refusal rate**" and "**Temporary away rate**" thanks to the fact that the survey gathers data on the status of sampled units.

Since the whole questionnaire is a balance, after a close analysis of the database we were unable to find any item non-responses, so the item non-response rate is 0%.

The second key process variable was "**refusal rate**". The refusal rate has been calculated separately on the whole number of units and on the each type of unit. The refusal rate is very low (0.042%), thanks to the fact that the survey is administered by an interview operator. Basically there were no refuses for the agricultural exploitation organized as enterprises. The solid training of the interview operators helps maintain a low refusal rate.

The last key process variable was "**temporary away rate**". The key process variable was calculated on each type of units and on the whole number of units. For the whole number of units the temporary away rate is 5.07%, for the agricultural exploitations organized as enterprises the temporary away rate is 18.54% and for the individual agricultural exploitations the temporary away rate was 4.31%.

Although the temporary away rates are low, these numbers can be improved with a careful selection of the units that are part of the sample.

4.2.2 Analyses of the key process variables used in the data processing

Analysis of the key process variables used in data processing is much more difficult than analyzing key process variables used in data collection, because there aren't many measurement tools available and the measurement tools available are time consuming.

We have chosen for our study three key process variables:

1. percentage of errors detected;
2. percentage of errors corrected;
3. overall coding accuracy rate.

The first two key process variables aren't usually measurable in practice. The third key variable is measurable only if the database has a key code for verification purposes. Both surveys have key codes and thus we were able to calculate the overall accuracy rate.

The first key process variable "*percentage of errors detected*" has been calculated for both surveys.
For the survey on *the use of IT&C products in enterprises* the percentage of errors detected was 0.085%, as for the *survey on the effective number of pig population on 1st august* the percentage is 0.091%. Although the

percentage of errors detected is low on both surveys, still some improvement can be made.

The second key process variable was *the percentage of errors corrected* (computer assisted error correction).
For the survey on *the use of IT&C products in enterprises* the percentage of errors corrected was 94.9%. We were unable to reach a 100% percentage because the rest of 5.1% of uncorrected errors, were errors that required a manual correction.
For the second survey *on the effective number of pig population on 1st august,* the percentage of errors corrected (computer assisted) was 100%, since all of the errors were coding errors.

The third key process variable was *overall coding accuracy rate*.
The process variable represents the number of corrected codes divided by total number verified.
For the both surveys the *overall coding accuracy rate* was over 99%, more exactly 99.92% for the survey on the use of IT&C products in enterprises and 99.91% for the survey on the effective number of pig population on 1st august.

All of the coding errors can be corrected automatically through computer assisted error correction and manual examination of errors. If a systematic error is identified in time, then a computer solution can be implemented to improve the overall coding accuracy rate. In order to identify systematic errors a quality check must be performed for at least 2 years.

4.3 Evaluation of the results

The evaluation of the results can be a very difficult process without the feedback from the respondents. But another good method of detecting and making the necessary adjustments is the analysis of the data obtained through brainstorming. The feedback is an important part of the quality improvement of the processes. As organizations seek to improve its performance, feedback helps it to make required adjustments.

We have presented a lot of results in the chapters on the analysis of the key process variables, both on data collection and data processing. But only the results obtained from the analysis on the key process variables on data collection can benefit from the brainstorming sessions with the specialists from the territorial offices.

During the analysis of the key process variables on data collection and the brainstorming sessions helped us identify some of the causes that led to the high item non-response rate.

The main causes that led to a high item non-response rate can be gathered into five:

- unclear questions;
- lack of specialized personnel in data provision;
- insufficient methodological indications;
- lack of data sources;
- response burden.

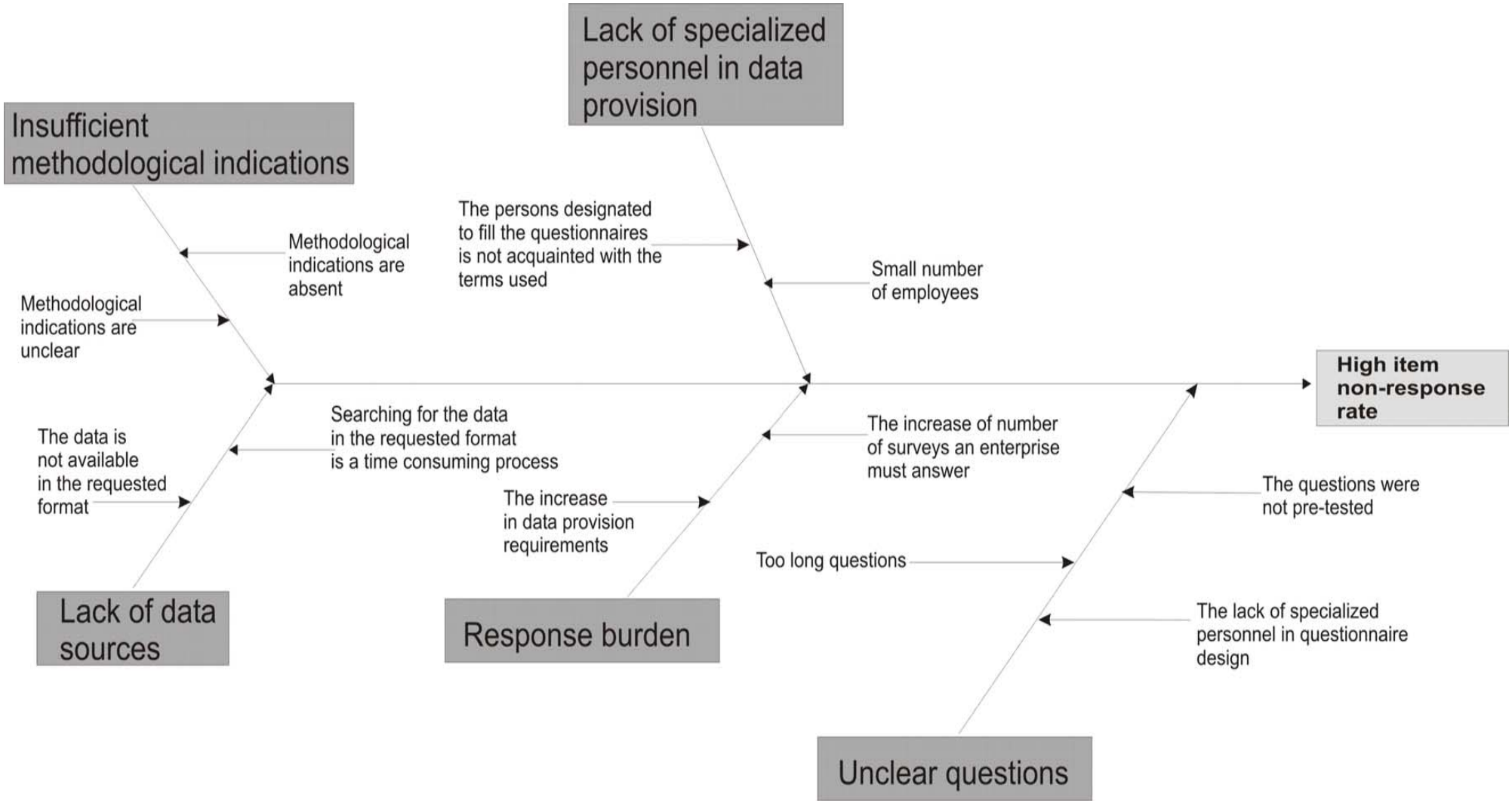
They have been presented in the following diagram, also called "*The cause and effect diagram or Ishikawa diagram".*

A further analysis has revealed that each of the main causes that led to a high item non-response rate can be detailed:

- Unclear questions:
    - The questions were not pre-tested;
    - The questions were too long;

    - The lack of specialized personnel in questionnaire design.
- Lack of specialized personnel in data provision:
    - The persons designated to fill the questionnaires in not acquainted with the terms used;
    - Small number of employees.
- Response burden:
    - The increase in the number of surveys an enterprise must answer;
    - The increase in data provision requirements.
- Insufficient methodological indications:
    - Methodological indications are absent;
    - Methodological indications are unclear.
- Lack of data sources:
    - Searching for the data in the requested format is a time consuming process;
    - The data is not available in the requested format or it is not available in any format.

Cause and effect diagram for “**High item non-response rate**”



4.4 Conclusions and recommendations

In order to improve the quality of the statistical products, some improvements are necessary to all statistical processes. Data collection and data processing are some of the most important processes. Most of all, data collection must be optimized in order to reduce the response burden.
Although, some of the key process variables presented can be calculated, some require additional data.

As a recommendation we propose to add some **questions to the existing surveys**. Before adding the questions to a survey an analysis must be performed. But, basically the following questions should be applied to all surveys in order to be able to calculate **the key process variables**.

The questions have been divided into two categories:

- questions on data collection;
- questions on data processing.

The questions on ***data collection*** have been divided into two categories:

- respondents ease of understanding the questionnaire;
- questionnaire for survey analyst and survey manager.

The questions on ***data processing*** have also been divided into two categories:

- questionnaire for survey analyst;
- questionnaire for survey manager.

- **Questions on data collection**

➢ Respondents ease of understanding the questionnaire:

1) To which of the following questions did you encountered difficulties providing an answer?
2) Which were the main causes that led to difficulties or impossibility in answering a question?
   a) Unclear questions;
   b) Unclear / lack of methodological indications;
   c) Lack of data sources;
   d) Lack of know-how.
   e) Others…….
3) Please provide us with some comments on the problems encountered in providing the requested data? ……

➢ Questionnaire for survey analyst and survey manager:

1) Please estimate an average time spent with the interview: ……
2) Please provide us with an estimate of the travel time to the interviewee: …….
3) Please specify, how many times have you tried to contact the household/unit?....

- **Questions on data processing**

➢ Questionnaire for survey analyst:

1) What is the average time spent on the examination of a questionnaire? ……..
2) Please, provide us with an estimation of the number of errors encountered in a questionnaire? …….

3) In case of errors, please estimate the average time spent on correction of the data:
   a) Time spent on correction of data……..
   b) Time spent trying to reach the respondents…….

➢ Questionnaire for survey manager:

1) Please provide us with the following informations:
   a) Number of years and months of experience of the survey operator…….
   b) Number of years and months of experience of the survey analyst…….
   c) Number of planned hours for the survey (per whole survey)……..
   d) Number of effective hours spent for the survey by the survey operator……
   e) Number of effective hours spent for the survey by the survey analyst……..

2) Please provide us with some informations on the availability of reference materials:
   a) Are there enough reference materials for survey operators? YES/NO
   b) Are there enough reference materials for survey analysts? YES/NO
   c) In your opinion are there enough methodological indications? YES/NO
   d) If no, please provide us with some feedback on what improvements do you think there should be made to the reference materials…….

In order to increase the item response rate the following **recommendations** should be taken into consideration:

- reanalyze the questionnaires and pre-test all the questions;
- analyze the possibility to obtain the data from administrative sources in order to reduce response burden;

- organize training sessions for data suppliers in order to cultivate the statistical culture;
- if data is not available or difficult to obtain and the item non-response rate is very high, some of those questions should be discarded;
- for the items with a high item non-response rate the methodological indications should be reviewed;
- the methodological indications should be moved from the end of the survey in front of each module and item with a high item non-response;
- the study should be done on a survey at least two years in a row in order to obtain relevant results.

One important characteristic of a standard should be the applicability to all statistical surveys not only to those that were studied in the case study. The chosen process variables and process methods used in the current study could be used for any statistical survey as long as the required data for the calculations is available in the required format.
As presented, in the above paragraphs, some small adjustments to the statistical questionnaires would be necessary in order to calculate more variables.
In order to have a very clear picture of the efficiency of the calculated key process variables and of the methods used, the study should be done at least for one more year, on the two surveys used in the current study. In the following years the standard will be used on other statistical surveys and the results will be analyzed and used to bring improvements to the standard.
The current Romanian standard guide could be also used by any statistical agency or statistical data producer as long as the required data is available, in order to be able to calculate the key process variables.

5. Annex

Annex 1: Item non-response rate for Survey on use of IT&C products in enterprises

Question code	Item responses	Maximum number of responses	Item non-response rate
A11	14082	14082	0.000
A41	11446	11446	0.000
A51	11446	11446	0.000
A52	11446	11446	0.000
A53	11446	11446	0.000
A54	11446	11446	0.000
A81	11446	11446	0.000
B11	11446	11446	0.000
B31	10494	10494	0.000
B32	10494	10494	0.000
B33	10494	10494	0.000
B34	10494	10494	0.000
B35	10494	10494	0.000
B41	10494	10494	0.000
B51	10494	10494	0.000
B52	10494	10494	0.000
B61	10494	10494	0.000
B611	6887	6887	0.000
B612	6887	6887	0.000
B613	6887	6887	0.000
B614	6887	6887	0.000
B615	6887	6887	0.000
B71	10494	10494	0.000
B81	4902	4902	0.000
B82	4902	4902	0.000
B83	4902	4902	0.000
B84	4902	4902	0.000
B85	4902	4902	0.000
B86	4902	4902	0.000
B91	10494	10494	0.000
B92	10494	10494	0.000
B93	10494	10494	0.000
B94	10494	10494	0.000
B95	10494	10494	0.000
C11	11446	11446	0.000
C21	3573	3573	0.000
C22	3573	3573	0.000
C23	3573	3573	0.000
C24	3573	3573	0.000
C25	3573	3573	0.000
C26	3573	3573	0.000
C27	3573	3573	0.000
C28	3573	3573	0.000
C31	3573	3573	0.000
C32	3573	3573	0.000

Question code	Item responses	Maximum number of responses	Item non-response rate
C33	3573	3573	0.000
D11	11446	11446	0.000
D21	1227	1227	0.000
D22	1227	1227	0.000
D31	1227	1227	0.000
D32	1227	1227	0.000
D41	1227	1227	0.000
D42	1227	1227	0.000
E11	11446	11446	0.000
E12	11446	11446	0.000
E13	11446	11446	0.000
E14	11446	11446	0.000
F21	454	454	0.000
F31	454	454	0.000
G11	11446	11446	0.000
G12	11446	11446	0.000
G13	11446	11446	0.000
G14	11446	11446	0.000
A31	11445	11446	0.009
A32	11445	11446	0.009
A61	11445	11446	0.009
B211	10493	10494	0.010
F11	11444	11446	0.017
A33	11443	11446	0.026
X01	14077	14082	0.036
X02	14077	14082	0.036
X03	14077	14082	0.036
X04	14077	14082	0.036
X06	14077	14082	0.036
X05	14076	14082	0.043
A21	11406	11446	0.349
A71	11393	11446	0.463
A72	11393	11446	0.463
A73	11393	11446	0.463
A74	11393	11446	0.463
A75	11393	11446	0.463
A76	11393	11446	0.463
A77	11393	11446	0.463
A78	11393	11446	0.463
A79	11393	11446	0.463
B21	10420	10494	0.705
H112	11233	14082	20.232
H19	11227	14082	20.274
H111	11226	14082	20.281
H12	11219	14082	20.331
H110	11219	14082	20.331
H11	11218	14082	20.338
H13	11218	14082	20.338
H14	11216	14082	20.352

Question code	Item responses	Maximum number of responses	Item non-response rate
H15	11215	14082	20.359
H16	11214	14082	20.366
H17	11214	14082	20.366
H18	11213	14082	20.374
H115	11212	14082	20.381
H113	11211	14082	20.388
H114	11210	14082	20.395
E21	5137	11446	55.120
E22	5137	11446	55.120
E41	5137	11446	55.120
E42	5137	11446	55.120
E31	5122	11446	55.251
F41	4846	11446	57.662
F51	649	11446	94.330