



EUROPEAN COMMISSION  
EUROSTAT



# **Handbook on improving quality by analysis of process variables**

---

Alexis Aitken, Jan Hörngren, Nia Jones,  
Daniel Lewis, Maria João Zilhão

*General Editors: Nia Jones, Daniel Lewis*

# Acknowledgements

## ***Instituto Nacional de Estatística Portugal***

Carlos Marcelo  
Albano Miranda

## ***National Statistical Service of Greece***

Aikaterini Livaditaki

## ***Office for National Statistics***

Susan Full  
Marta Haworth  
Jacqui Jones  
Ann Lewis  
Frank Nolan  
Bernice Reeks  
Andrew Stephens  
Nigel Stuttard  
Andrew Walton  
Helen Williams  
Jayne Wilson

## ***Statistiska Centralbyrån (Statistics Sweden)***

Leopold Granquist

## ***Reading Group for the Handbook***

Svein Gåsemyr  
Johanna Laiho  
Donna Livesey  
David Marker  
Hans Viggo Saebo

# Contents

<b>1</b>	<b>INTRODUCTION TO THE HANDBOOK.....</b>	<b>4</b>
<b>2</b>	<b>GUIDANCE ON IMPROVING PROCESS QUALITY .....</b>	<b>6</b>
2.1	INTRODUCTION .....	6
2.2	MOTIVATION FOR MONITORING STATISTICAL PROCESSES .....	6
2.3	AN APPROACH TO IDENTIFYING, MEASURING AND ANALYSING KEY VARIABLES OF STATISTICAL PROCESSES .....	8
2.3.1	<i>Identify Critical Product Characteristics</i> .....	10
2.3.2	<i>Develop a Process Flow Chart</i> .....	11
2.3.3	<i>Determine Key Process Variables</i> .....	12
2.3.4	<i>Evaluate Measurement Capability</i> .....	13
2.3.5	<i>Determine Stability of Critical Processes</i> .....	14
2.3.6	<i>Determine System Capability</i> .....	16
2.3.7	<i>Establish a System for Continuous Monitoring of Processes</i> .....	17
2.4	REFERENCES .....	19
<b>3</b>	<b>EXAMPLES OF IMPROVING STATISTICAL PROCESS QUALITY .....</b>	<b>21</b>
3.1	INTRODUCTION .....	21
3.2	THE STATISTICAL VALUE CHAIN .....	23
3.3	APPLYING QUALITY IMPROVEMENT METHODS TO STATISTICAL PROCESSES .....	26
3.3.1	<i>Data Collection</i> .....	30
3.3.1.1	Data Collection by Paper Self-Completion Questionnaires (ONS).....	31
3.3.1.2	Face-to-face Interviewing (INE-Pt) .....	41
3.3.1.3	Interviewing activities (SCB).....	55
3.3.1.4	Literature review .....	66
3.3.2	<i>Accessing Administrative Data</i> .....	69
3.3.2.1	Accessing Administrative Data (INE-Pt).....	69
3.3.2.2	Literature Review .....	73
3.3.3	<i>Data Processing</i> .....	75
3.3.3.1	Editing and Validation (NSSG) .....	76
3.3.3.2	Validation (ONS).....	80
3.3.3.3	Continuous Quality Improvement of surveys from an editing perspective (SCB) .....	84
3.3.3.4	Coding (ONS).....	95
3.3.3.5	Literature review .....	112
3.3.4	<i>Weighting and Estimation</i> .....	117
3.3.4.1	Measuring nonresponse bias (SCB) .....	117
3.3.4.2	Literature review .....	125
3.3.5	<i>Analysis of Primary Outputs</i> .....	126
3.3.5.1	Tabulation (NSSG) .....	126
3.3.5.2	Literature review .....	130
3.3.6	<i>Time Series Analysis</i> .....	131
3.3.6.1	Reviewing Seasonal Adjustment (ONS) .....	131
3.3.7	<i>Confidentiality and Disclosure</i> .....	135
3.3.7.1	Disclosure (NSSG) .....	135
3.3.7.2	Literature Review .....	139
3.4	CONCLUSION.....	140
	<b>ANNEX 1 – GLOSSARY OF KEY CONCEPTS IN PROCESS QUALITY .....</b>	<b>143</b>
	<b>ANNEX 2 – FLOW CHARTS .....</b>	<b>146</b>
	<b>ANNEX 3 – CAUSE AND EFFECT DIAGRAMS .....</b>	<b>147</b>
	<b>ANNEX 4 - PARETO CHARTS .....</b>	<b>148</b>
	<b>ANNEX 5 - CONTROL CHARTS .....</b>	<b>149</b>
	<b>TECHNICAL ANNEX: CONTROL CHART METHODOLOGY.....</b>	<b>150</b>
	<i>A. CONSTRUCTING CONTROL CHARTS</i> .....	150
	<i>B. INTERPRETING CONTROL CHARTS</i> .....	155

# 1 Introduction to the Handbook

## Background to the Handbook

During the last few decades, the importance of quality has become increasingly evident, as organisations realise that continuous improvement is necessary to stay in business. Statistical organisations are no exception, and steps have been taken in Europe to focus on improving and developing a systematic approach to quality in National Statistical Institutes (NSIs).

One important step in 1999 was the formation of a Leadership Expert Group (LEG) on Quality. The LEG aims to attain improved quality in the European Statistical System (ESS), which comprises Eurostat and the statistical offices, ministries, agencies and central banks that collect official statistics in EU and EEA EFTA Member States.

Following Group discussions the LEG defined its task in more detail, and provided a list of recommendations proposing future actions for the ESS, all related to quality. To follow up on this, an Implementation Group for the LEG on Quality was formed to support projects working on various recommendations.

The LEG's final report (Eurostat (2002)) highlights the need to distinguish between different types of quality. Product quality is the quality of the output. In the case of a statistical organisation this is the quality of the data and services provided. These products are generated by an underlying process or sequence of processes, and so the product quality is likely to be affected by the process quality. The report states that 'in theory, good product quality can be achieved through evaluations and rework. However, this is not a feasible approach since it is costly and time-consuming. Instead, it is believed that product quality will follow from improvements in process quality.'

So improving process quality is a key aim. The report goes on to explain how 'the process quality is improved by identifying key process variables (i.e. those variables with the greatest effect on product quality), measuring these variables, adjusting the process based on these measurements, and checking what happens to product quality. If improvements do not materialise, alternative adjustments are made or new key variables are identified and measured. This is an example of the so-called PDCA (Plan, Do, Check, Act) cycle advocated by the late W. Edwards Deming in the spirit of continuous improvement.'

This theory led to the third recommendation of the LEG, relating to process quality: 'Process measurements are vital for all improvement work. A handbook on the identification of key process variables, their measurement, and measurement analysis should be developed.'

And so a project was set up in June 2002 to produce this handbook for NSIs to use. The project team consisted of members from the NSIs of Greece, Portugal, Sweden and the United Kingdom.

The handbook describes a general approach and useful tools for the task of identifying, measuring and analysing key process variables. This includes practical examples of the application of the approach to various statistical processes. The handbook does not aim to provide a list of recommended key process variables across all statistical processes.

## **Structure of the Handbook**

After this general introduction, the main body of the handbook is split into two main sections. The first contains the relevant theory and the second some examples of using the methods in practice.

Section 2: 'Guidance on Improving Process Quality' provides guidance on how to identify, measure and analyse process variables. The methods described are relevant to any process, statistical or non-statistical. Annexes to the handbook provide more detail on quality management concepts, and the tools used in the proposed approach.

Section 3: 'Examples of Improving Process Quality' looks in more detail at some specific statistical processes. Each NSI working on the project applied some process quality improvement methods to a few statistical processes, and the results are reported here. The information presented should be useful for statistical output managers (those responsible for a set of statistics - based on a survey, or administrative sources etc) and others working on the quality of their processes, but also helps with the understanding of methods in described in the previous guidance section.

References containing further information on the theory of process quality are found in sub-section 2.4, and references with examples of applying process quality ideas to different statistical processes are included at the ends of sub-section 3.3.1 to 3.3.7.

## **2 Guidance on Improving Process Quality**

### **2.1 Introduction**

This section of the handbook provides a general description and guidance on methods in Continuous Quality Improvement (CQI), focusing on identifying, measuring and analysing process variables. Annex 1 provides a definition of CQI, and other quality management concepts.

Literature on applying these methods in statistical institutes is limited. However, Biemer and Lyberg (2003) is a key and recent publication in the area of survey process quality. The handbook summarises the main ideas arising in the various documents reviewed. The aim is to learn from the literature a way forward for identifying, measuring and analysing process variables.

Sub-section 2.2 describes the motivation for identifying, measuring and analysing process variables, and introduces some general ideas in CQI. Sub-section 2.3 details the common approach to monitoring statistical processes found in our literature review. The methods involved are described in general terms, hence will be applicable in a variety of situations.

### **2.2 Motivation for monitoring statistical processes**

Monitoring and improving process quality in statistical operations is a key part of achieving CQI. Biemer and Caspar (1994) outline three important aspects of the CQI approach:

- the use of teams to identify problems, determine solutions and implement the corrective measures;
- quantitatively evaluating components of a statistical operation using process variables;
- identifying and addressing root causes of instances of unacceptable quality.

This handbook will focus on and provide guidance for the second point above.

Dippo (1997) mentions that the application of methods in quality control and CQI to a statistical service requires a wider approach than in manufacturing. This is because the processes to be addressed are typically not physical products, but human or machine action, decisions and the paths these decisions take. Literature on applying CQI methods to statistical operations is scarce, however, the methods are still applicable and the benefits are evident from the examples below, and further examples described in section 3.

Although it has not been common for National Statistical Institutes (NSIs) to collect and analyse process data in a systematic way, Sundgren (2001) explains how the need for such efforts is becoming increasingly evident. This is partly due to a growing interest in systematic quality work such as Total Quality Management (TQM, see Annex 1.)

Despite a lack of focus on process quality in NSIs in the past, some are developing work in this area. Statistics Norway has recently adopted a systematic approach to process quality. Sæbø et al. (2003) explain that so far, they have not applied statistical methods in the analysis of process quality to a great extent, partly because of a lack of process variable measurement. They suggest that the development of methodology and tools such as flow charts is important to aid the process quality approach. This handbook describes the most common methodology and tools in sub-section 2.3 and in the annexes.

Some reasons for systematically monitoring processes are listed below:

**To improve product quality, cost-efficiently.**

Sæbø et al. (2003) assert that 'improving process quality is a precondition for better product quality at an acceptable cost'. Mudryk et al.(2001) explain how the quantitative process variables measured during the data capture process allow them to ensure high quality outputs for the entire capture process.

**To allow managers to be responsive to problems.**

Haselden and White (2001) suggest that a process quality approach allows early identification of problems that occur during a statistical operation. This enables the manager to take measures to counter those problems and still produce high quality data. These fixes can be applied to other statistical outputs with common processes before they experience the problems. Similarly, Mudryk et al.(2001) explain how their approach to the data capture process allows them to take action at appropriate times during production by isolating problem documents. They repair or re-process documents as required, and prevent the reoccurrence of problems.

**To allow objective measuring and monitoring of quality over time.**

Mudryk et al.(1996) describe how they monitor the Computer Assisted Telephone Interviewing (CATI) process in Statistics Canada. Their approach to process quality ensures an ability to track and analyse performance objectively over time.

**To aid future improvement projects.**

Sæbø et al. (2003) suggest that defining and measuring process variables is vital for future improvement projects. These variables allow statistical output managers to evaluate changes to the statistical process. Cost efficiency is improved by enabling improvement resources to be focused where most required.

**To provide effective feedback and training to staff.**

Mudryk et al.(1996) state that a quality control approach ensures more effective feedback, and enables training resources to be focused where most required.

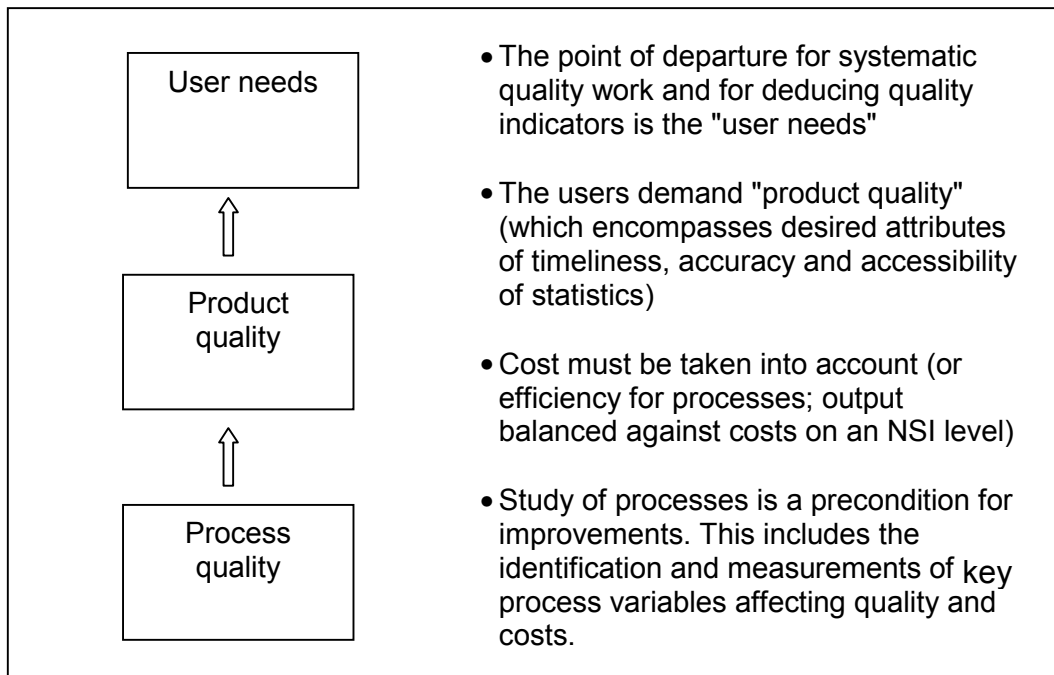
**To provide customers with Quality Assurance (QA, see Annex 1).**

Defining and monitoring process variables ensures credible quality assurance for customers. Mudryk et al.(2001) explain how measuring process variables allow them to provide estimates of incoming and outgoing quality.

## 2.3 An approach to identifying, measuring and analysing key variables of statistical processes

From the relevant literature reviewed it was apparent that a common approach has been used in most examples. Figure 1 below, taken from Sæbø et al. (2003), describes how process quality, product quality and user needs are linked. The identification and measurement of key process variables is included as an important element of quality improvement.

**Figure 1:** User needs related to quality



An approach to process quality is fully described in Morganstein and Marker (1997). It contains a flow chart that acts as a guide to identifying and monitoring process variables, and which fits in well with approaches found in other documents reviewed. For example there are three stages at which checks occur and necessary changes are made, which fits in with the PDCA (Plan, Do, Check, Act) cycle advocated by Deming.

The approach employs ideas from Statistical Process Control (SPC, see Annex 1). SPC techniques are used widely in industry, and it is as yet unclear how useful they are for statistical processes. This question is revisited in sub-section 3.4.

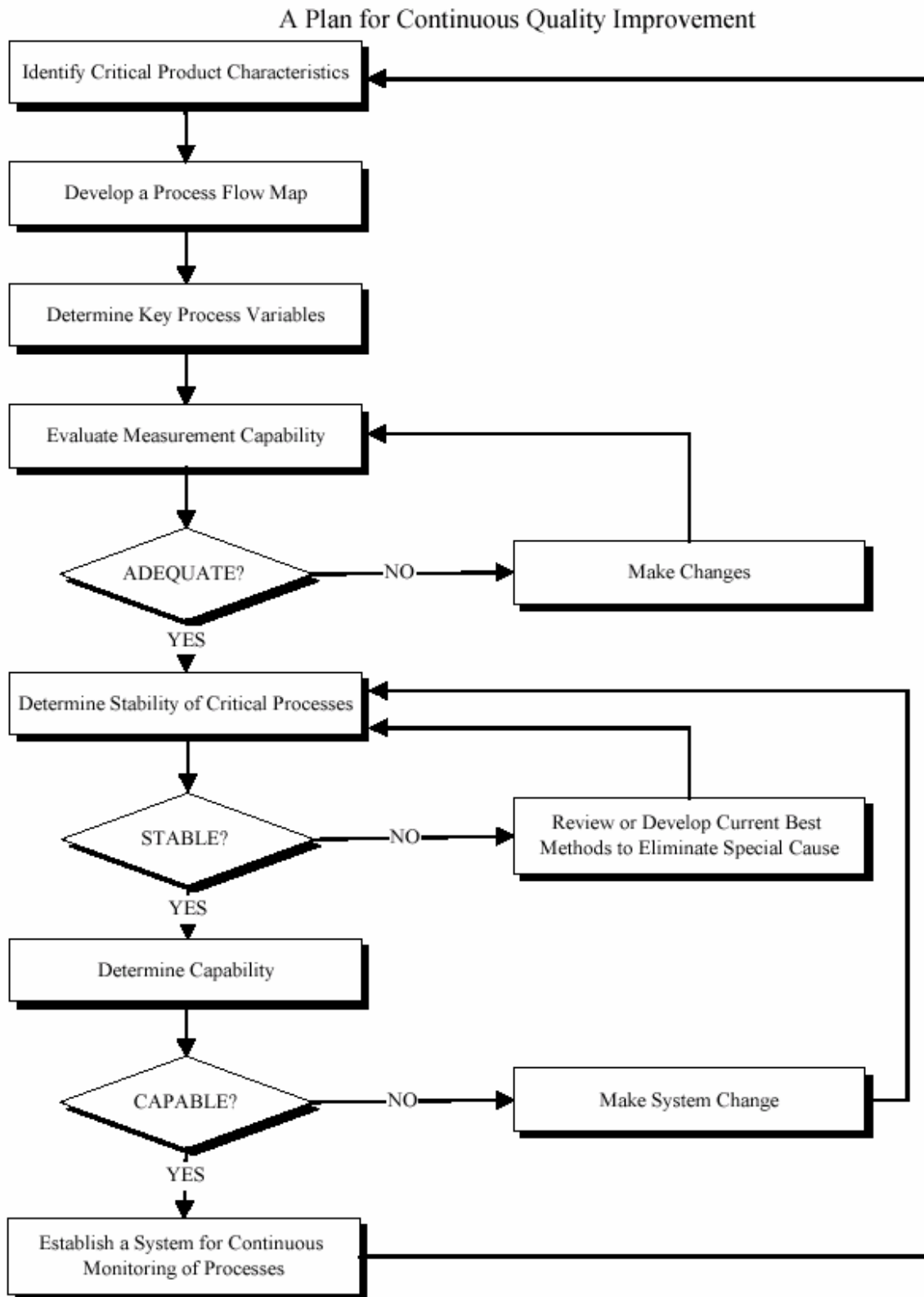
This flow chart is given below in figure 2. The flow chart has seven steps:

1. Identify critical product characteristics;
2. Develop a process flow chart;
3. Determine key process variables;
4. Evaluate measurement capability;
5. Determine stability of critical processes;
6. Determine system capability;
7. Establish a system for continuous monitoring of processes.

The first three cover identifying, the fourth measuring, the fifth and sixth analysing key process variables. The seventh point covers further, wider issues in CQI. A detailed description of each step follows.



Figure 2: A Plan for Continuous Quality Improvement, taken from Morganstein and Marker (1997),



### 2.3.1 Identify Critical Product Characteristics

**Tools:** Customer Satisfaction Survey.

An understanding of the product characteristics that are important to customers is essential to effectively achieving CQI. Early on in any project, Morganstein and Marker (1997) advise having 'meetings with the customers to establish priorities among conflicting survey goals that affect such critical elements as the design, time schedules, and budgets'. A useful tool here is a Customer Satisfaction Survey, which can be used to determine the customers' definition of quality and their perception of specific products and services. On distributing a customer satisfaction survey Pikounis et al. (2001) found that their customers were pleased that an effort to improve quality was being made, and even requested periodical re-surveying. Statistics Finland participated in a customer satisfaction survey to evaluate major public and private organisations in Finland. This enabled them to assess their strengths and weaknesses relative to other organisations.

CQI can be applied to a specific part of a statistical process or the complete process. For the complete process the next step, not explicitly mentioned by Morganstein and Marker (1997), is to identify the key parts of the overall production process which affect product quality (as defined by the customer), and are thus important to monitor. For this Handbook, the ONS Statistical Value Chain (SVC) has been used to define the statistical processes. The SVC is described in full in sub-section 3.2.

An important consideration for this step is to ensure adequate coverage of the six ESS Data Quality dimensions (see Eurostat (2003)) through the processes to be monitored. The dimensions are:

- Relevance
- Accuracy
- Timeliness
- Accessibility and Clarity
- Comparability
- Coherence

It is difficult to measure most of these output quality dimensions, but it is possible to relate some process variables to the dimensions. Sub-section 3.4 discusses process variables in the context of these dimensions.

Colledge and March (1993) warn that 'it is difficult to establish target levels of product quality in customer terms. Desirable features - for example, range of products or timeliness - that lead to customer satisfaction are difficult to identify due to the wide variety of unknown and potential customers. The impact of defects (i.e. errors) leading to customer dissatisfaction are hard to assess because the magnitudes of errors are often unknown (due to their multiplicity), as are their impact on customers (due to the wide range of uses). It follows that the appropriate allocations of resources across products and across process steps for a given product are difficult to determine because the effects of changes to these allocations cannot be easily measured in terms of the output quality.'

Note that there is a separate Eurostat Leadership Expert Group report dealing with the design, implementation and analysis of customer satisfaction surveys. See Cassel et al. (2003).

### 2.3.2 Develop a Process Flow Chart

**Tools:** Process flow chart (see Annex 2); Quality Improvement Team.

Once a process has been identified for monitoring, the next step is to map the process by developing a comprehensive process flow chart, which can be used to identify sources of variation in the process. At this stage, a 'Quality Improvement Team' (QIT) could be set up, including all levels of staff involved in the process, key process suppliers and customers, and an independent quality advisor (see section 2.3.7 for more details on QITs and their function). Morganstein and Marker (1997) suggest including three components in a process flow chart:

- the sequence of processes is delineated, indicating decision points, the flow of the process, and the customer(s) for each step;
- the owners of each process are identified;
- the key process variables, decisions, or actions that can be taken by those involved in the process are listed. Section 2.3.3 explains how key process variables can be identified.

Filippucci and Calia (2001) employed the methods of Morganstein and Marker (1997). They stress the importance of describing and understanding the process, and of distinguishing between - and identifying those responsible for - actions and decisions. Haselden and White (2001) similarly describe elements of a process map as:

- what each of the processes are, their inputs and outputs and how they fit together;
- who the process owners are as well as their customers;
- what constitutes quality for each of the processes.

The involvement of representative relevant staff in a QIT ensures that these three components in the process flow chart are accurate and comprehensive. Also, QIT members will gain an understanding of each part of the processes (eg who is responsible for this process?) and how these relate to each other (eg where does this component begin and end?). Sæbø et al. (2003) describe how Statistics Norway have trained 'quality pilots' in 'techniques for mapping the processes involved and in identifying and measuring critical process parameters'. They assist QITs in mapping their processes.

Haselden and White (2001) explain how different statistical processes use slightly different methodologies, making it difficult to describe a process in generic terms. Potential sources of information during process mapping are: existing documentation eg survey procedures manuals and specifications; statistical output managers; those who carry out the tasks - the process owners.

Haselden and White (2001) note that a process can be broken down into a number of sub-processes which each have their own inputs and outputs. Each sub-process should produce a quality report for its customers further down the chain. Morganstein and Marker (1997) suggest that using different levels of charts can simplify the process flow chart, and help to achieve a balance between adequate and excessive detail. A more detailed micro-level chart is useful for new staff to understand the overall process, while macro-level charts show how the individual components interact.

More detail on process flow charts is provided in Annex 2.

### 2.3.3 Determine Key Process Variables

**Tools:** Pareto diagram (see Annex 4); Cause and effect diagram (see Annex 3), Quality Improvement Team.

*Definition: **Key process variables** are those factors that can vary with each repetition of the process and have the largest effect on critical product characteristics, i.e. those characteristics that best indicate the quality of the product.*

However process variables are in general different to quality indicators, which are more closely related to output quality. A definition of quality indicators is provided in Annex 1.

This step aims to identify these critical process variables. The earlier step of identifying critical product characteristics (3.1) will input to this step. Examples of process variables are: reinterview results, time or resource used, edit failures, coder error rates, number and type of customer complaints, and number of cases where disclosure control techniques fail to protect the data. Before describing methods for identifying key process variables, some more explanation is provided below.

#### **Further explanation of key process variables**

A variable is defined by the Cambridge dictionary of statistics (Everitt (2001)) as 'some characteristic that differs from subject to subject or from time to time'. In the context of a process, we can think of variables as things that can change with each repetition of the process. All processes have numerous variables, and so it is important to identify the most useful (or 'key') variables for our purpose.

But what differentiates a key process variable from any other process variable? The key variables are those judged to have the largest effect on pre-defined critical product characteristics. This judgement of which variables are 'key' may be based on evidence, or else may be purely subjective. Different people may have different opinions on the importance of variables. We may have to measure and analyse several variables before deciding which are key.

Often there will be several potential variables to choose between, of varying importance and varying ease of measurement. For example, if customers of a certain statistic were concerned about bias in the data, an estimate of non-response bias would be key to them. However, in reality they may be provided with the response rate: a more easily obtained process variable, in many situations correlated with the non-response bias. So although an estimated bias is our ideal variable to measure (the most key), we measure a response rate which is less key but much easier to measure.

In summary, in the words of Biemer and Lyberg (2003), 'the best indicators of quality are process variables that can be observed conveniently and continuously during the survey process and that are highly correlated with the components of error that need to be controlled.'

#### **Identifying key process variables**

A simple and effective method for identifying key process variables is the Pareto diagram (see Annex 4). This tool tries to find the relatively few error types that account for the majority of all errors, hence enabling staff to be more effective in allocating resources. The cause-and-effect diagram (see Annex 3) was designed for the purpose of identifying key process variables when numeric information on the variables is not available. This diagram

is also called the fishbone or Ishikawa diagram (see Annex 3.) Morganstein and Marker (1997) explain that, 'from all of the factors on the fishbone, the QIT selects the five or six they believe to be most important. These are the factors to measure over time and whose variability should be reduced.'

Haselden and White (2001) describe how they decided, initially, to consider any possible variables whether their primary use would be as a management tool or as a quality indicator to be passed on to clients, or both. They asked process owners 'what they needed in order to be able to do their job well, using the process maps as a guide.' This enabled them to 'determine what all the aspects of quality are for each process, even if they are not apparent to the direct customer for that process'. They describe how 'it became clear that the quality of the output from one process often did not become apparent until several processes down the line.' For example, consider how edit failures may reflect bad questionnaire design. Thus the manager of a specific process may need to analyse variables derived from processes further down the production line.

Sæbø et al. (2003) describe how Statistics Norway have mapped processes for the Consumer Price Index (CPI). For data collection, the rates of missing price observations, the number of out of range items and inconsistent price observations are variables useful for analysing error sources such as questionnaire design, and data entry. During process mapping, it was found that the rate of non-response is a critical process variable. The authors identified a need to analyse the distribution of non-response (and its relation to variance) in a more systematic way, so that efforts to reduce non-response can be focused on outlets having the largest impact on CPI.

#### **2.3.4 Evaluate Measurement Capability**

After identifying the key process variables, it is important to know how accurately they can be measured. If there are significant errors in the system used to measure the process variables, the measurements will be unreliable and may invalidate any analysis of process stability and capability. Therefore evaluating measurement capability is essential to ensure a good basis for process improvement.

The measurement process used is capable if the mean squared error is small relative to the overall error requirements. Morganstein and Marker (1997) warn that 'a common mistake is to collect data and reach conclusions about process stability without any knowledge of measurement error. Alternatively, researchers often select a process because it is easy to measure, rather than choosing a more important but harder-to-measure process.'

Once measurement capability is established, ideally the measurement of process variables should require as little effort as possible on the part of the producers. Sundgren (2001) would prefer processes to be designed such that when implemented they automatically generate relevant, basic process data.

Two commonly used measurement systems are discussed below: customer satisfaction surveys, and sampling and verification.

#### **Customer Satisfaction Surveys**

Customer satisfaction surveys commonly use poor measurement systems such as limited scales, or scales that are anchored at the extremes. In such cases the measurement error

limits the use of the survey for the purpose of continuously improving quality. A systematic process for acquiring (i.e. measuring) the data is needed to achieve stability of the critical processes (see 3.5 below).

Pikounis et al. (2001) describe how they measured the quality of statistical reports in a pharmaceutical company, using a customer satisfaction survey. Results of a pilot survey of clients led to changes in the questionnaire, to improve the measurement of client opinion. Responses to the revised questionnaire were then collected from a sample of clients, before and after making changes to the process of producing reports. Where these two measurements showed a change in client opinion, statistical tests (using the sample sizes) were used to assess the significance of the apparent change.

### **Sampling and Verification**

Mudryk et al.(2001) describe their measurement of quality of the document preparation process. Employing the method of sampling and verification, they describe how 'A supervisor checks a sample of work at regular intervals, completing a 'spot check control form' to identify results... The forms are aggregated on a weekly basis to monitor and document progress, which, if sufficient, results in a decrease in the frequency of sampling.'

This technique of inspecting and assessing a sample of work from the process is a common method of assessing quality. When the work is judged to be either acceptable or not, this is called 'acceptance sampling.' In acceptance sampling, if the batch of work is below the acceptance quality level, it is reworked. This is a different approach to process control. In process control, quality indicators input to an ongoing monitoring system. Unacceptable performance is dealt with by investigating and eliminating its cause, without necessarily reworking incorrect work.

Cevis and Peregoy (1984) briefly describe acceptance sampling, and explain dependent and independent inspection methodologies. In dependent inspection, as opposed to independent inspection, the inspector is aware of the original result (eg a code allocated to a survey response). Independent inspection tends to be the more reliable of the two. Although the independent method generally demands more resources, it can eliminate the need for costly and time-consuming dependent adjudication.

Colledge and March (1993) stress that 'quality control of repetitive clerical procedures should be promoted, based on acceptance sampling with feedback of results to help improve the procedures. Decreasing sampling rates and skip-lot sampling can be used as quality improves, with the ultimate goal of stabilising the procedures so that process-control methods can be applied', leading us to the sub-section on stability of processes.

### **2.3.5 Determine Stability of Critical Processes**

**Tools:** Pareto diagram (see Annex 4); Control chart (see Annex 5); Quality Improvement Team.

Once key process variables have been identified and measured, we move on to their analysis. First, they are tested for stability: process stability is a state where the process variation consists entirely of random components. The state of stability is often referred to as being a state of 'statistical control'. Bowerman and O'Connell (1989) explain that 'a statistically controlled process is a process that displays a consistent amount of variability about a constant mean'. When statistical processes are reasonably stable (or 'in statistical

control'), the variables can provide a basis for comparison after changes are made to the system.

It is important to note that stability does not necessarily imply that the process is operating well enough to meet customer requirements. Stability (or control) only tells us that no unusual process variations are being observed. It also is a requirement for prediction about the ability of existing systems to consistently meet targets in the future.

### **Tools for determining stability**

Various statistical tools and methodologies are described in the literature, and provide a systematic way to analyse data. The main tools are control charts (see Annex 5) and Pareto analyses (see Annex 4). Through this analysis step, we can identify the sources of process and product variations, and examine the effects changes to the process have on variation.

Control charts have control limits that are typically calculated at three standard deviations from a centre line (possibly a group average). Observations that fall out of the control limits are deemed incapable or 'out-of-control'. A special or common cause for out-of-control observations should be identified and addressed. Tools such as control charts can be used to measure the effectiveness of improvements and accurately predict likely outcomes in the future.

However, as Tapiero (1995) explains, control charts alone do not lead to quality nor improve process performance. Rather, they induce actions that improve quality, and help monitor and maintain a process in control.

The use of control charts in evaluating a statistical operation, such as data keying or coding, is straightforward. However, Dipppo (1997) points out that 'we have not yet developed a way to determine when the processes of developing a survey question or deciding on the sample allocation among strata are in statistical control.' For these processes where it is difficult to establish stability, it may still be worthwhile to analyse data collected, to gain an understanding of the system capability and improvements in quality.

Filippucci and Calia (2001) use simple histograms and scatterplots as evidence to identify the best and worst performing municipalities with respect to their process variables.

### **Causes of variation**

Process variability is commonly classified into two types, special cause and common cause. In some cases, a specific, identifiable problem may be the cause of an unacceptable product, service or process condition. These sources of variation are known as special causes. For example, a single batch of questionnaires may have been improperly coded by one individual. In contrast, common cause variation affects all processes. For example, poor recruiting and training practices apply to all staff working on all statistical processes.

One function of a Quality Improvement Team (QIT) meeting is to identify the source of an unacceptable variation. Following this identification, it is important to distinguish a special cause from a common cause, as responsibility for correcting these two causes often rests on different staff. A specific one-time, local action by operating staff may be sufficient to remedy a special cause of variation. However, to address common cause variation, typically management should take action to change the system. Making these distinctions between special and common cause variation is a primary reason for using control charts (see annex 5 and the technical annex).

### 2.3.6 Determine System Capability

**Tools:** Pareto chart (see Annex 4); Quality Improvement Team.

For processes that are reasonably stable, staff can determine the limits of the expected process variation, or evaluate the capability of the process to predictably meet specifications. Examples of process specifications are minimum response rates, production deadlines, and maximum coefficients of variation. A stable process is capable if its random variation is such that the system will consistently meet the customer's requirements or limits for that process.

Mudryk et al.(2001) describe how, throughout their data capture process, QC reports are generated providing an indication of the incoming and outgoing quality levels. These include:

- control charts by scanner, operator and field type on a daily and weekly basis;
- Pareto charts by operator, scanner and overall, showing distribution of errors by page and field type on a weekly basis;
- run charts of key quality indicators aggregated on a daily or weekly basis;
- special reports showing estimates of error rates by operator and scanner and overall error rates, initially on a daily and then weekly basis.

However Biemer and Caspar (1994) point out that there may be disadvantages to creating control charts for individuals, in that focusing on an individual as an assignable cause of error may affect morale.

In a similar vein, Haselden and White (2001) describe how, in their work on quality indicators, there was some concern that recording and analysing some information might lead to perverse incentives. For example if recording the number of times a questionnaire is redistributed to interviewers, researchers might be reluctant to do this even when there is a problem with the questionnaire, for fear that it may reflect badly on them. To address these concerns, accompanying documentation should be written, giving advice as to how the indicators should be interpreted, and outlining their limitations.

System changes to reduce variation are needed for processes that are not capable. For example, the variation can be reduced through adherence to standard procedures (see section 2.3.7 below). After making changes to the system, the stability of the variables should be re-evaluated before determining the new process capability.



### **2.3.7 Establish a System for Continuous Monitoring of Processes**

Morganstein and Marker (1997) suggest that 'achieving reliable and capable processes is only the beginning of the improvement process.' A continuous monitoring system is needed to keep staff informed, to provide feedback, to assist in controlling process variation, to help achieve continuous reduction in process variation through improved methods, and to evaluate process changes.

Marker and Morganstein (2004) note that 'monitoring, by itself, will not result in continuous improvements. Procedures must be established for responding to this information in a manner consistent with the numerical results.' Effective procedures for managing the implementation of improvements identified are important. Specific guidance on how to plan and monitor the implementation of improvements is not included in this handbook.

The following paragraphs provide an overview of process improvement work at Eurostat, followed by three key elements for managing process improvement, and finally mentioning three situations where it may be unclear how to proceed with CQI techniques.

#### **Experience at Eurostat**

Eurostat has developed a system of continuous improvement - the 'Process Improvement Methodology' - facilitated by a network of Process Improvement Co-ordinators drawn from operational areas. Co-ordinators work closely with Process Improvement Teams, which are set up specifically for the analyses and are representative of stakeholders. The methodology is based on the principles of detailed process description, stakeholder analyses, risk analyses, and improvement planning. Facilitated workshops, flow charts and process indicators are tools employed and a cycle of continuous improvement with regular formal reviews is encouraged. An intranet site, e-learning package, manual, standard electronic templates, formal training courses and regular forums of the Co-ordinators have all been established to promote and share best practice. The methodology has been piloted and used on a voluntary basis to date but an exercise to identify and analyse all critical processes is underway and is planned for completion by end 2004. So far statistical methods have not been employed and Eurostat will aim to strengthen the methodology based on the work of this project.

#### **Standard procedures**

Standard procedures such as current best methods (CBMs) - described in detail in Morganstein and Marker (1997) - help minimise variation, increase the likelihood that best practice is followed, and make monitoring processes easier.

#### **Quality Improvement Teams**

Regular meetings of Quality Improvement Teams (QITs) allow staff to benefit from feedback at a group level, as well as encourage the discussion and solution of other problematic issues. In this way, the external environment is improved as well as the skill-level of staff. Biemer and Caspar (1994) suggest that QITs should include the operators, inspectors, the supervisor of the operation, and a quality advisor. The advisor should be trained in CQI, and could also provide survey methodology advice, help prepare reports, and liaise with higher management or staff in related operations. Involvement of key process suppliers and customers will improve communication with the process staff.

### **Individual feedback**

If individual operators are involved in the process, individual feedback in the form of a report (containing tailored Pareto or control charts etc) or a discussion with an inspector, supervisor, or quality adviser can help improve future performance. For the CATI process, Mudryk et al.(2001) describe feedback provided weekly and intermediately on an individual basis, as well as in group settings.

### **New processes**

Techniques in CQI are useful for modifying well-established processes, where measurements of process variables are possible. However many of the improvements done in NSIs involve developing completely new processes, for example development of electronic data collection and use of new data sources. A common question is how do the CQI methods – such as process stability – apply to this situation where our process is brand new? Tools such as flow charts and cause-and-effect diagrams will still be useful, and it is important to identify process variables and set up a measurement system to inform future improvements.

### **Changes in the process**

To evaluate process changes, Filippucci and Calia (2001) suggest measuring the effect on the key variables and deciding whether to keep the changes on the basis of the findings.

### **Production backlogs**

When production backlogs occur, Cevis and Peregoy (1984) stress that it is important 'to resist the temptations to modify or eliminate QC.' In that kind of situation, errors are more likely to occur and the process may become out of control. QC will help to identify problems earlier, and prevent them from become unmanageable.

## 2.4 References

- Biemer P and Caspar R (1994) Continuous Quality Improvement for Survey Operations: Some General Principles and Applications, *Journal of Official Statistics*, vol 10, no 3, pp 307-326.
- Biemer P P and Lyberg L E (2003) *Introduction to Survey Quality*, Wiley.
- Bowerman B L and O'Connell R T (1989) *What to Teach about Control Charts and Process Capability in Basic Statistics*, proceedings of the section on Survey Research Methods, American Statistical Association.
- Burr I W (1976) *Statistical Quality Control Methods*, Marcel Dekker, Inc.
- Cassel C, Järnbert M, Holzer W, Barbieri G, Tiddi A (2003) *State-of-the-art regarding planning and carrying out Customer/User Satisfaction Surveys in NSIs (LEG on Quality Recommendation No. 7) - Final Report*, Eurostat.
- Cevis T O and Peregoy R A (1984) *An Overview of Quality Control Procedures in the Economic Area of the Bureau of the Census*, proceedings of the section on Survey Research Methods, American Statistical Association.
- Colledge M and March M (1993) Quality Management: Development of a Framework for a Statistical Agency, *Journal of Business and Economic Statistics*, vol 11, no 2, pp 157-165.
- Deming W E (1991) *Out of the Crisis*, Cambridge University Press.
- Dippo C S (1997) Survey Measurement and Process Improvement: Concepts and Integration, in Lyberg L et al. (eds), *Survey Measurement and Process Quality*, Wiley, pp 457-474.
- Eurostat (2002) *Quality in the European Statistical System – The Way Forward*, Office for Official Publications of the European Communities: Luxembourg.
- Eurostat (2003) *Definition of Quality in Statistics*, working paper for the 6th meeting of the Working Group of Assessment of the Quality in Statistics, 2-3 October 2003.
- Everitt B S (2001) *The Cambridge Dictionary of Statistics*, 2<sup>nd</sup> ed. Cambridge University Press.
- Filippucci C and Calia P (2001) *Towards Process Quality Control in the Italian Consumer Expenditure Survey*, presented at the International Conference on Quality in Official Statistics, Stockholm, May 14-15, 2001.
- Hapuarachchi K P and Wronski A (1994) *Control Charts for Non-response Rates in the Canadian Labour Force Survey*, proceedings of the section on Survey Research Methods, American Statistical Association.
- Hapuarachchi P, March M and Wronski A (1997) Using Statistical Methods Applicable to Autocorrelated Processes to Analyze Survey process Quality Data, in Lyberg L et al. (eds), *Survey Measurement and Process Quality*, Wiley, pp 589-600.
- Haselden L and White A (2001) *Developing New Quality Indicators in Social Surveys*, presented at the International Conference on Quality in Official Statistics, Stockholm, May 14-15, 2001.
- Ishikawa K (1976) *Guide to Quality Control*, Asian Productivity Organization: Tokyo.
- ISO (2000) *ISO Standards Handbook: Statistical methods for quality control* (2 volumes).

- Jeskanen-Sundström H (2003) *Identifying the needs for change and adjusting them in the management systems of NSIs*, presented at the GUS/ISI Satellite Conference, Szczecin, Poland, August 8-9, 2003.  
[http://www.stat.gov.pl/isi/Jeskanan-Sundstrom\\_Paper-Identyfing\\_the\\_needs\\_for\\_changes.pdf](http://www.stat.gov.pl/isi/Jeskanan-Sundstrom_Paper-Identyfing_the_needs_for_changes.pdf)
- Marker D A and Morganstein D R (2004) Keys to Successful Implementation of Continuous Quality Improvement in a Statistical Agency, *Journal of Official Statistics*, forthcoming paper.
- Mendenhall R and Beaver (1989) *Statistics for Management and Economics*, 6th ed. PWS-KENT: Boston, MA.
- Morganstein D and Marker D A (1997) Continuous Quality Improvement in Statistical Agencies, in Lyberg L, Biemer P, Collins M, De Leeuw E, Dippo C, Schwarz N, and Trewin D (eds.), *Survey Measurement and Process Quality*, New York: Wiley, pp. 475-500.
- Mudryk W, Bougie B and Xie H (2001) *Quality Control of ICR Data Capture: 2001 Canadian Census of Agriculture*, presented at the International Conference on Quality in Official Statistics, Stockholm, May 14-15, 2001.
- Mudryk W, Burgess M J and Xiao P (1996) *Quality Control of CATI Operations in Statistics Canada*, proceedings of the section on Survey Research Methods, American Statistical Association.
- Pikounis V B, Antonello J M, Moore D, Senderak E T and Soper K A (2001) Practising What We Preach: The Application of Continuous Improvement in a Preclinical Statistics Department at a Pharmaceutical Company, *Journal of Official Statistics*, vol 17, no 1, pp 187-206.
- Reed S J and Reed J H (1997) *The Use of Statistical Control Charts in Monitoring Interviewers*, proceedings of the section on Survey Research Methods, American Statistical Association.
- Sæbø H V, Byfuglien J and Johannessen R (2003) Quality Issues at Statistics Norway, *Journal of Official Statistics*, vol 19, no 3, 2003, pp 287-303.
- Sundgren B (2001) *Documentation and Quality in Official Statistics*. presented at the International Conference on Quality in Official Statistics, Stockholm, May 14-15, 2001.
- Sytsma S and Manley K (1999) *The Quality Tools Cookbook* [online], Ferris State University. Available from <http://www.sytsma.com/tqmtools/tqmtoolmenu.html> . [Accessed 1 December 2003].
- Tapiero C S (1996) *The Management of Quality and its Control*, 1st ed. Chapman & Hall: London.

## **3 Examples of Improving Statistical Process Quality**

### **3.1 Introduction**

This section of the handbook contains several reports on the application of process quality improvement methods to statistical processes.

#### **Aim and Scope of this Section**

It is important to note that the handbook does not aim to list recommended process variables for all statistical processes. There are many gaps in the overall statistical production process, either because no suitable material could be found, or due to the resource constraints of producing this handbook.

However, there are many other examples of process quality improvement in existence or under development. Some of these may cover gaps in this handbook. In future, it may be possible to produce further editions of the handbook, with added examples and guidance.

Where process variables are defined, these should be interpreted as suggestions only. Although many process variables are defined, the main purpose of this section is to illustrate the use of a Continuous Quality Improvement approach with statistical processes.

#### **Use of this Section**

This section should be useful for statistical output managers (those responsible for a set of statistics - based on a survey, or administrative sources etc) and other experts working on the quality of statistical processes, but also helps with the understanding of process quality improvement methods. Reports on a particular process may be useful for those directly involved in the process in their organisation, although some issues will not be transferable between different organisations. For example some process maps will not be generic enough to apply to all NSIs.

#### **Content**

The next sub-section describes the ONS Statistical Value Chain (SVC). This was used as a basis for splitting the whole statistical production process into sub-processes, for application of process quality ideas.

Sub-section 3.3 contains the reports for each sub-process examined. This is split into chapters such as 'data collection' and 'data processing', which contain one or more examples of applying the approach.

The conclusion in sub-section 3.4 summarises the key findings of the chapters in 3.3, drawing attention to areas where the approach was considered a success, and to those where the work proved difficult. Some ideas for further work, which could not be undertaken for this handbook, are also presented.

## **Reading this Section**

This section is not intended to be read from cover to cover. Some guidelines on how best to read it are provided here. It is useful but not essential to read the next sub-section (on the SVC) and the introduction to sub-section 3.3 before reading the individual reports. The reports are complete in themselves, and can be read in any order. The concluding sub-section is a useful summary of the findings, and may direct you to particular reports of interest.

## **Note on Acronyms**

Throughout section 3, the National Statistical Institutes (NSIs) involved in the project will be referred to in acronym form as follows:

- Greece - NSSG;
- Portugal – INE-Pt;
- Sweden - SCB;
- UK – ONS.

### 3.2 The Statistical Value Chain

The ONS Statistical Value Chain (SVC) was developed during 2001 as part of the organisation’s Statistical Infrastructure Development Programme. The SVC describes the key processes used within the ONS to produce statistics. There exist several other similar representations of the statistical production cycle, for example Eurostat uses a tool called the Cycle de Vie de Données (CVD) or Data Life Cycle.

This ‘splitting up’ of the overall statistical production process made it easier to identify potential areas with useful input to the handbook. Each NSI working on the project applied the methods to a few statistical processes (or ‘groups’) in the SVC. There is some overlap between examples, which gives us a wider perspective and helps to see the difference in possible approaches between NSIs.

Figure 3 shows the 15 links in the SVC. Below that is a more detailed table describing the more specific component activities associated with each link.

Figure 3: The Office for National Statistics’ Statistical Value Chain

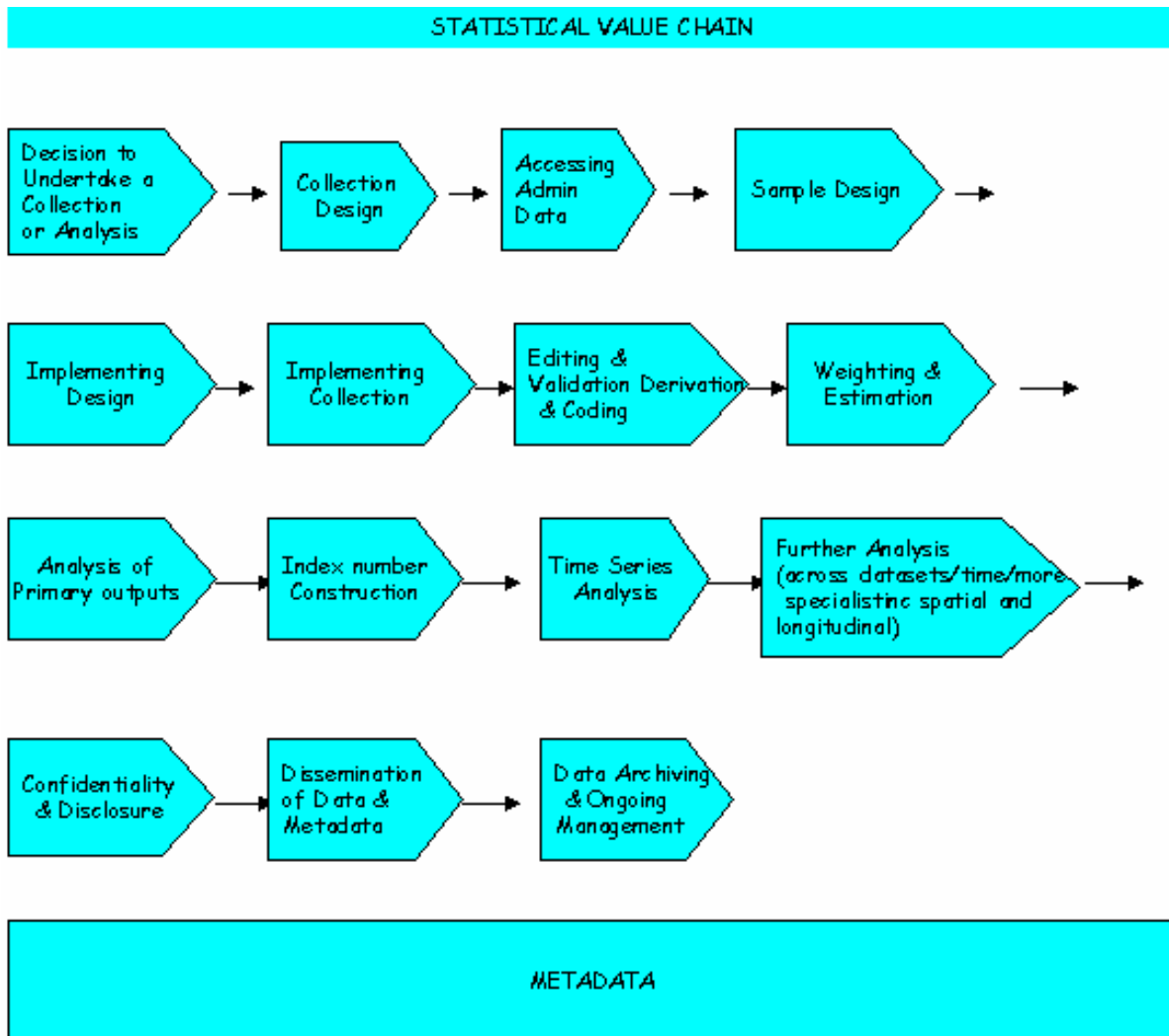


Table 1: Components of the Statistical Value Chain

	SVC Group	SVC Component Activities
1.	decision to undertake a collection or analysis	<ul style="list-style-type: none"> <li>analysis of user needs</li> <li>formulate research hypotheses</li> <li>research on related studies (within NSI and international)</li> </ul>
2.	collection design	<ul style="list-style-type: none"> <li>clarify objectives</li> <li>research past and related work</li> <li>develop measurement instruments, including testing</li> <li>develop field procedures, including testing</li> <li>develop edit/ imputation strategies</li> <li>develop data management strategies</li> <li>develop dissemination strategies</li> <li>document</li> </ul>
3.	accessing administrative data	<ul style="list-style-type: none"> <li>clarify objectives</li> <li>arrange access (including any legal issues)</li> <li>document data that is accessed</li> </ul>
4.	sample design	<ul style="list-style-type: none"> <li>clarify objectives</li> <li>research past and related work</li> <li>determine target population, frame, selection and estimation methods</li> <li>design and allocate sample</li> <li>document</li> </ul>
5.	implementing design	<ul style="list-style-type: none"> <li>frame creation/ cleaning</li> <li>sample selection</li> <li>sample cleaning</li> <li>allocation of sample to interviewers for interview based surveys and despatch of workloads</li> <li>allocation of sample to validators (self completion forms) and identification of workloads on corporate database</li> </ul>
6.	implementing collection	<ul style="list-style-type: none"> <li>despatch of mail based or electronic questionnaires to respondents</li> <li>interviewing for interview based surveys</li> <li>management of respondent relations and feedback to frame information</li> <li>resolution of queries relating to selected units</li> <li>management of collection, including quality assurance of processes and monitoring of progress</li> <li>follow up procedures, including re-issue of sample to interviewers and reminders to mail based respondents</li> <li>document procedures and outcome of processes</li> </ul>
7.	editing and validation, derivation and coding	<ul style="list-style-type: none"> <li>unit level editing and validation</li> <li>imputation and construction</li> <li>derivation of variables</li> <li>quality assurance of processes</li> <li>document procedures and outcome of processes</li> </ul>
8.	weighting and estimation	<ul style="list-style-type: none"> <li>estimation (weighting and grossing)</li> <li>outliers</li> <li>sampling errors</li> <li>special adjustments</li> <li>quality assurance</li> <li>document procedures and outcome of processes</li> </ul>



	<b>SVC Group</b>	<b>SVC Component Activities</b>
9.	analysis of primary outputs	<ul style="list-style-type: none"> <li>• macro editing and drill down to unit data</li> <li>• tabulation</li> <li>• exception reporting</li> <li>• assessment of results against related information</li> <li>• document including quality report</li> </ul>
10.	index number construction	<ul style="list-style-type: none"> <li>• index construction (including deflation, chain linking)</li> </ul>
11.	time series analysis	<ul style="list-style-type: none"> <li>• interpolation</li> <li>• seasonal adjustment</li> <li>• trend analysis and extrapolation</li> <li>• document including quality report</li> </ul>
12.	further analysis (across data sets/ over time/ more specialist analyses, includes spatial and longitudinal analysis)	<ul style="list-style-type: none"> <li>• identify and access relevant series</li> <li>• identify available methods</li> <li>• develop and evaluate new methods and extensions of existing methods</li> <li>• statistical analysis including tabulation, exploratory data analysis, spatial analysis and longitudinal analysis</li> <li>• adjusting data series for further analysis</li> <li>• validate results</li> <li>• document including quality report</li> </ul>
13.	confidentiality and disclosure	<ul style="list-style-type: none"> <li>• identify user requirements for outputs and priorities</li> <li>• identify potentially disclosive information</li> <li>• apply solutions to avoid disclosure</li> <li>• evaluate results</li> <li>• document</li> </ul>
14.	dissemination (data and metadata)	<ul style="list-style-type: none"> <li>• dissemination of standard aggregated outputs including text, diagrams, numbers etc</li> <li>• dissemination of non standard aggregated outputs</li> <li>• dissemination of micro data externally or within ONS under controlled conditions</li> <li>• dissemination of metadata</li> <li>• customer inquiries and complaints</li> <li>• content management</li> <li>• document processes and report on quality</li> </ul>
15.	data archiving and ongoing management	<ul style="list-style-type: none"> <li>• identify and maintain contact information, particularly data custodian</li> <li>• implement archiving policy</li> <li>• document policy and practice</li> </ul>

### 3.3 Applying quality improvement methods to statistical processes

This sub-section has several chapters based on the SVC groups, each containing one or more reports on an application of the process quality approach to a statistical process. Most chapters also contain a literature review as a concluding sub-section. These consist of a list of useful references, often with short descriptions of the content.

The examples described in the reports contain a mixture of new work on the process (by the project team), and existing or past work related to process quality. The past work may not have been undertaken with the proposed approach in mind, but nevertheless fits into the framework and helps to illustrate and promote the methods.

Reports vary in length and detail. For example some will focus on identifying process variables through process maps and fishbone diagrams, whilst others will concentrate on analysing measured variables.

Further details on how the chapters relate to the SVC groups are given in table 2 below.

**Table 2:** Chapters related to Statistical Value Chain Groups.

<b>SVC Group</b>	<b>Chapter</b>	<b>Title of Chapter</b>
1. Decision to undertake a collection or analysis	None	n/a
2. Collection design	1	Data collection
3. Accessing administrative data	2	Accessing administrative data
4. Sample design	None	n/a
5. Implementing design	None	n/a
6. Implementing collection	1	Data collection
7. Editing and validation, derivation and coding	3	Data processing
8. Weighting and estimation	4	Weighting and estimation
9. Analysis of primary outputs	5	Analysis of primary outputs
10. Index number construction	None	n/a
11. Time series analysis	6	Time series analysis
12. Further analysis	None	n/a
13. Confidentiality and disclosure	7	Confidentiality and disclosure
14. Dissemination (data and metadata)	None	n/a
15. Data archiving and ongoing management	None	n/a

Note that SVC groups 'collection design' and 'implementing collection' have been merged into a single chapter on 'data collection'. This is because it is difficult to separate the process variables for the two aspects of data collection. That is, many process variables relating to the collection design will not be measurable until implementation takes place.

SCV group 7 has been renamed to the simpler 'data processing' for the purpose of this handbook.

There are seven SVC groups with no corresponding chapter (and therefore no reports). Difficulties were found in considering how to apply the process quality approach to groups 1, 10 and 12. In theory the approach may be appropriate for the remaining groups 4, 5, 14 and 15. For example it should be possible to measure process variables for the timeliness of dissemination (group 14). However at present, no known examples of suitable process quality work in these areas exist in the project teams' NSIs. As experience of using process quality methods on statistical processes grows, it may be possible to expand this body of examples, filling in some further gaps in the SVC.

The reports below give an in depth description of the examples considered for the handbook. As a quick reference, we present the main details of the process variables identified in table 3.

**Table 3: Process variables identified in the handbook.**

Process	Process variable	Measurement	Experience	Conclusion
Data Collection	Ability of respondents to answer a problem question.	Analysis of responses and comments relating to the question.	ONS Purchases Inquiry.	Easy to implement. Gives information on reasons for item non-response which can be used to improve questionnaire.
Data Collection	Percentage item non-response. Percentage unnecessary response. Percentage with a mark entered in both yes and no boxes. Percentage without a mark in either yes or no boxes. Percentage with a numeric value in a mark box. Percentage with a mark in a numeric field. Percentage with more than one mark box completed where only one is expected. Percentage with a non-relevant mark in or across mark boxes.	A sample of returned forms identified as containing errors.	ONS E-Commerce Survey.	Easy to implement – need some resource to do the analysis. These indicators can identify issues that can be tackled to improve the quality of the questionnaire.
Data Collection	Distribution of number of complaints received by size of responding business.	Analysis of database containing complaints to business survey questionnaires.	ONS 'Response to Public Inquiries' database.	Implementation requires store of information on complaints received from respondents. Indicator allows analysis of response burden. This can assist in improving response.
Data Collection	Percentage of ineligible sampling units found in the sample. Refusal rate. "Temporarily away" rate. Number of interviewers performing interview 2 and 5 weeks after reference period.	Analyse indicators using bar charts and Pareto charts.	INE-PT study on improving the quality of surveys using face to face interviews.	Implementation requires collection of data on status of sampled units. Indicators are useful for monitoring the quality of response to surveys. Can lead to the implementation of improvement actions.

Process	Process variable	Measurement	Experience	Conclusion			
Data Collection (interviewing activities)	Interviewing time by survey.	A process database with information from the CATI system and the interviewing system which feeds into a set of standard reports.	Statistics Sweden project.	Implementation requires CATI and interview reporting systems that measure the relevant information. A system would need to be set up to produce the standard reports. Indicators allow the identification and analysis of key process variables relating to nonresponse errors, measurement errors and productivity in interviewing activities.			
	Travel time of interviewers by survey.						
	Other time (eg planning) by survey.						
	Working hours by survey.						
	Total interview time by survey.						
	Number of planned hours by survey and survey manager.						
	Number of successful refusal conversion attempts divided by total number of attempts.						
	Number of contact attempts by time period.						
	Number of final code units by time period and domain.						
	Interviewing time by respondent and survey.						
	Proportion of monitored interviews by survey.						
	Proportion of re-interviews by survey.						
	Proportion of field observations by survey.						
	Number of editing errors by item and survey.						
Proportion of responses obtained from modes other than the main one by survey.							
Proportion of proxy interviews by survey.							
Data Processing (field editing)	Time spent in manual examination of questionnaires.	Not measured in practice.	Suggested by flow chart of data editing process derived by NSSG. Not carried out in practice.	These indicators would provide a useful assessment of the quality of various sub-processes of data editing if implemented in practice.			
	Number of analysts manually examining questionnaires.						
	Years (or months) of experience of survey analyst in the specific survey.						
Data Processing (automatic editing)	Runtime of automatic editing adjusted by the sample size.	Log of data failing each test. Gates with the highest number of failures are examined individually to check they are identifying errors correctly.	Data Validation Branch for ONS business surveys.	Easy to implement. This provides an effective method for regularly checking the efficiency of validation gates.			
	Percentage of errors detected.						
Data Processing (computer assisted error correction)	Percentage of errors corrected.						
	Percentage of new errors.						
Data Processing (manual examination of errors)	Reference material available to the analyst.						
	Timeliness of external information.						
	Years (or months) experience of survey analyst in the specific survey.						
Data Processing (validation)	Number of failures of data at each validation gate.				Quality checks built into results system. 10% checked by manager.	ONS business surveys.	Implementation requires building quality checks into results system. Checks allow monitoring of staff performance. These can be used to highlight weaknesses, which can be improved by training.
Data Processing (data editing)	Staff performance at examining errors identified by data editing.						

Process	Process variable	Measurement	Experience	Conclusion
Data Processing (coding)	Overall coding accuracy rate (number of correct codes divided by total number verified).	Take a sample of codes for verification and use to estimate process variables.	2001 UK Census.	Implementation requires a sample of codes. Using these indicators gives an effective method to identify systematic error, assess reported accuracy and improve accuracy over time.
	Accept rate.			
	Number and percentage coded by mode.			
	Rate of incorrectly assigned 'uncodeables'.			
	Frequency of types of error in coding.			
Weighting and estimation	Estimated standard errors.	Formulae can be derived for estimating standard errors in many cases.	Widely applied across NSIs.	These indicators of accuracy can be used to assist in designing surveys with improved quality.
	Relative standard errors.	Calculated as ratio of standard error estimate to point estimate.	Widely applied across NSIs.	
	Confidence intervals.	Easily derived from standard error estimate.	Widely applied across NSIs.	
	Mean square errors.	Calculated as variance + bias <sup>2</sup> . Difficult to measure in practice.	Often estimated as special study.	
	Nonresponse bias.	Difficult to measure in practice.	Estimated as special study. eg Statistics Sweden Activity After Graduation survey.	
Analysis of Primary Outputs (tabulation – checking and correction of weights)	Time spent in manual examination of weights.	Calculated from analysis of data.	NSSG Retail Sales Value Index.	Implementation would require a record of time spent, numbers of errors etc. These process variables can be used to monitor the quality of two sub-processes of tabulation – checking and correction of weights, and correction and modification of data.
	Percentage of errors due to wrong weights with respect to the total number of records.			
Analysis of Primary Outputs (tabulation – correction and modification of data)	Percentage of errors detected with respect to the total number of records.			
	Percentage of enterprises not modified.			
	Percentage of modifications at time t.			
	Percentage of modifications at time t-1.			
Time Series Analysis	Closeness to optimal seasonal adjustment.	Review of Seasonal Adjustment methods across organisation. Grades given in each case to indicate quality.	Seasonal adjustment review by ONS Time Series Analysis Branch.	Implementation requires a large scale program to assess the quality of seasonal adjustment across an organisation. If seasonal adjustment is found to be of low quality, intervention (eg centralising seasonal adjustment) may lead to improved quality.
Confidentiality and Disclosure	Number of cases where statistical disclosure control methods failed to protect the data.	No experience of measuring these indicators.	Suggestion not tried out in practice.	Further work needed to assess the suggested process variables.
	Indicators derived from feedback on suitability of disclosure-limited data products.			
	Number of impingements by researchers using restricted data divided by total number of inspections of research data centres.			

### 3.3.1 Data Collection

This chapter includes examples relating to data collection. There are reports from three NSIs, each dealing with a slightly different aspect of the process. In the early stages of the collection design it is necessary to choose between administrative data, survey data or a combination of sources. This chapter concentrates on survey data and does not deal with data collection from administrative data or combined sources.

Report 3.3.1.1 (from the ONS) deals exclusively with paper self-completion questionnaires, as used in ONS business surveys. The questionnaire design process is considered in detail, as well as feedback from respondents.

In contrast reports 3.3.1.2 (from INE-Pt) and 3.3.1.3 (from SCB) examine a different data collection mode: interviewing. There is less focus on questionnaire design in these two reports, which mainly concentrate on the work of the interviewer. However there is little overlap between the two reports, which look at slightly different aspects of interviewing as described below.

The INE-Pt report describes critical aspects of face-to-face interviewing. Non-response is broken down and measured in different categories, including ineligibility. This highlights an important link between the quality of the sampling frame and data collection. The results of a re-interviewing study are also summarised, assessing the 'accuracy' of interviews.

On the other hand, SCB analyse interviewing activities in general, including central activities and productivity. Process variables are measured and presented in standard reports, for the use of managers in tracking the progress of their collection. The emphasis here is on productivity and therefore timeliness and cost.

Despite their differences, a key message from all three reports is that non-response is an important issue – and a useful process variable - in all data collection.

### **3.3.1.1 Data Collection by Paper Self-Completion Questionnaires (ONS)**

#### **Introduction**

This example focuses on applying our process quality approach to data collection by paper self-completion questionnaires. The remainder of the introduction provides a short description of the process, followed by an overview of the content of the report.

#### **Description of the process**

A common way of retrieving information from respondents is to send out a paper self-completion questionnaire, usually by mail. The questionnaire may be accompanied by a cover letter providing background information, and instructions on how to fill in the form. The respondents either do not respond, or respond to all or some questions, returning the questionnaire to the survey organisation by mail or by fax.

Other means of data collection include telephone data entry, electronic data capture, and face-to-face interviewing. The latter mode is examined in reports 3.3.1.2 and 3.3.1.3. Data collection by paper self-completion questionnaire has different sources of error in comparison with face-to-face interview. For example, although the former mode has no interviewer effects, it is more susceptible to misreading and misinterpretation of questions and instructions by the respondents. To combat these effects, it is important to monitor and improve the quality of questionnaire design and formatting.

#### **Summary of the report**

The two main parts of this report are:

- process maps for data collection by paper self-completion mail questionnaires;
- some relevant examples from within ONS where teams have identified, measured and analysed process variables to improve process quality.

## Develop a process flow map

### *Process map for data collection*

Figure 4 shows a high-level process map or flowchart for the data collection (via paper self-completion questionnaire) process. The map helps us to understand what is involved in the process, what dependencies there are between sub-processes, and to identify potential areas where it may be important to monitor quality. This map in particular follows the general procedure of data collection in the ONS, and may not be applicable to all specific data collections.

There are six action points in the map, which may be viewed as the sub-processes. One of the action points is 'data entry and validation', which is generally agreed to be out of scope of 'data collection', but is included here for completeness of the map, and to lead to our ultimate output of a 'clean' data set.

Many simple process variables could be identified for the printing and despatching sub-processes, although these would typically be based on machine (such as printer) performance. In this report we are more interested in the questionnaire design as a statistical process, and its links to the analysis of feedback from respondents (in the forms of response and non-response). A more detailed map of the questionnaire design and testing sub-process is described below.

### *Questionnaire design and testing*

Questionnaire design and testing has been identified as a key sub-process of data collection via paper self-completion questionnaires. Figure 5 presents a detailed process map or flowchart for the sub-process, which aims to follow the typical procedure for questionnaire design and testing used within the ONS. Further details on this procedure are provided in Jones (2003).

To summarise the map, a draft questionnaire may go through several stages of checking. Customers, questionnaire design experts, and cognitive interviewees are given the opportunity to provide comments, and appropriate changes are incorporated at each stage. The next possible stage is to conduct a field pilot to test the operational aspects of the data collection. At each stage of checking, we can consider the comments or results of that check to be process variables, which we analyse to determine if further improvements are necessary. Eventually a final questionnaire will be created, signifying the end of the sub-process.

However, there is further input to this sub-process later on during the actual data collection. This is shown in Figure 4 by the feedback loop from analysis of response and non-response. The examples provided in this report highlight how feedback from survey responders can be used as informative process variables, and in particular can contribute to improvement in the questionnaire. Feedback obtained during data collection may also reflect on other parts of the overall survey operation, for example the sample design and response burden, or confidentiality procedures. This is illustrated to some extent by the last ONS example on analysis of complaints from respondents.

At the end of the data collection process, it is important to evaluate the experiences gained. 'Lessons learned' could be collected, and the information could be shared with others, for example through lunchtime discussions.



**Figure 4:** Flow Chart of part of the Data Collection process

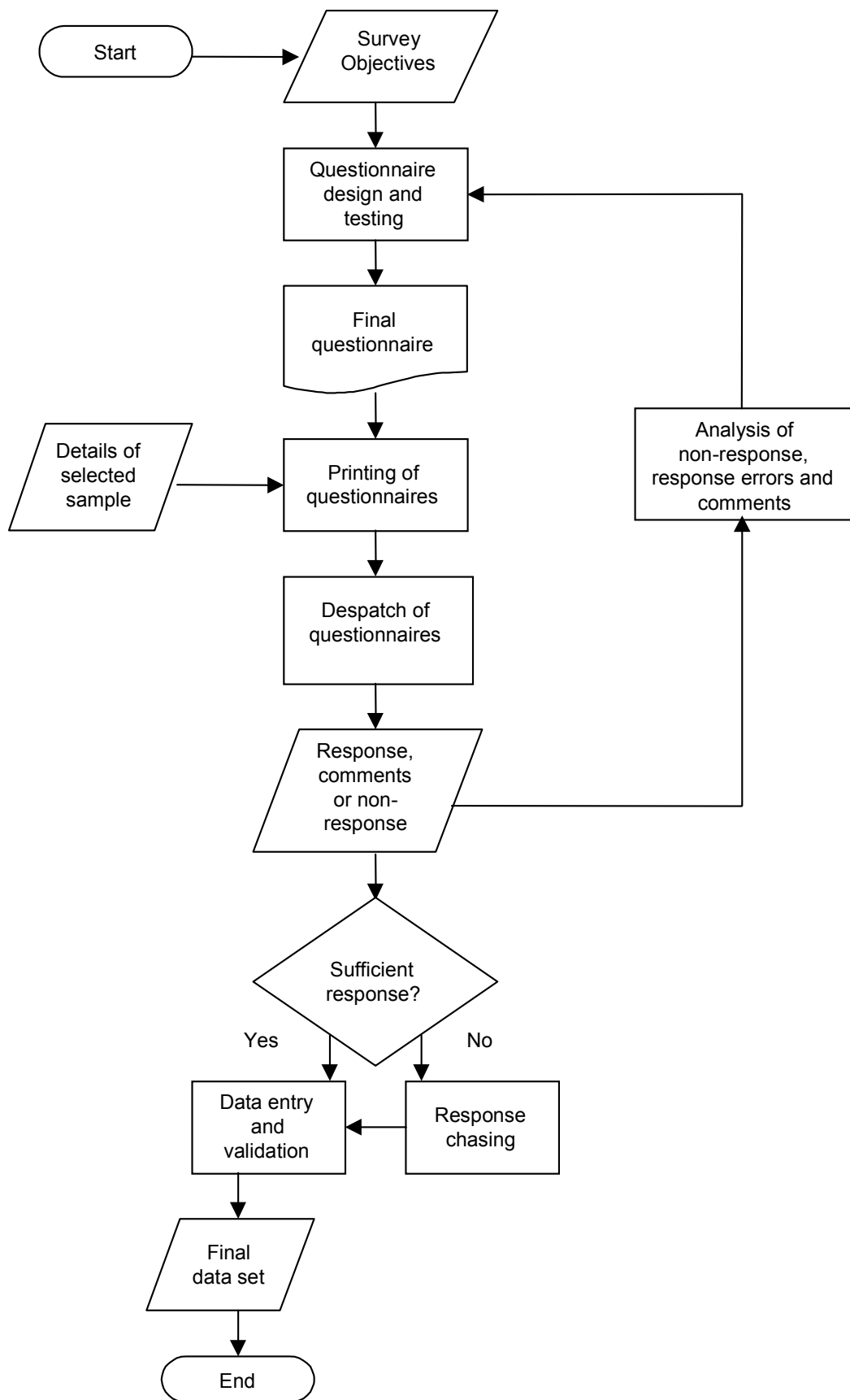
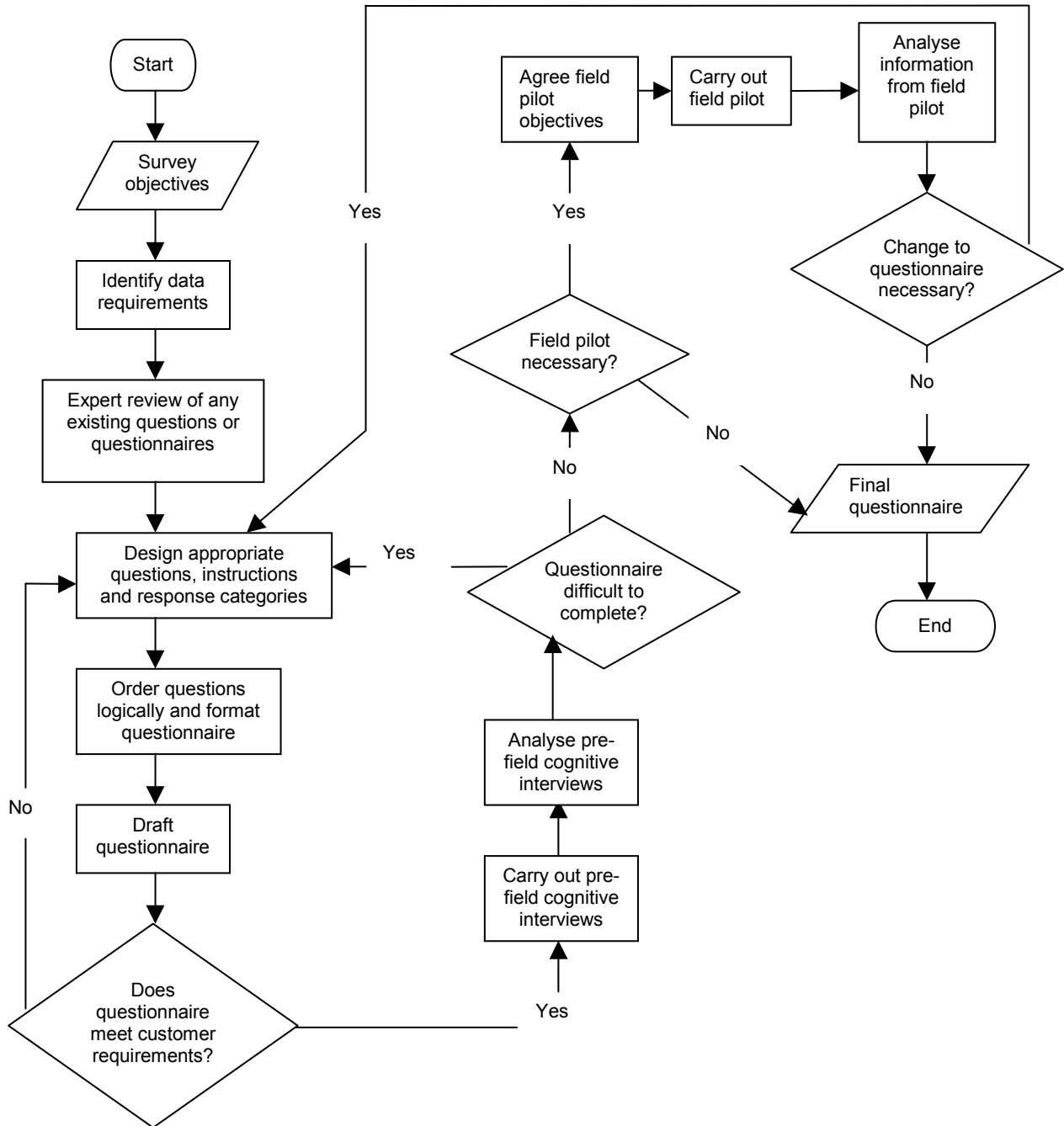


Figure 5: Flow Chart for the Questionnaire Design and Testing process in ONS



## **Some examples from within ONS of receiving feedback from the survey**

As explained in the discussion surrounding the process maps presented above, it is likely that the quality of survey operations and hence data can be improved by acting upon feedback from respondents during the data collection stage. Often the feedback can be put into the form of quantitative process variables for analysis.

This report describes three examples where this has happened in practice on ONS surveys. In the first example, analysing non-response led to improvements in the questionnaire for that survey. The second example examines non-response and other response errors for the purpose of questionnaire development. The third example describes a system for logging and analysing complaints from respondents, received during the data collection period.

### Using item non-response analysis to improve a questionnaire

#### ***Identifying key process variables***

In 2001 the ONS Purchases Inquiry was extended to include the construction industry. This met with problems, as many of the contributors were finding it difficult to provide the level of information required. This was a particular problem for a question relating to purchases of services. Therefore the key process variable for this example is the ability of businesses to respond to the question, represented by item non-response.

#### ***Measuring key process variables***

ONS put a lot of resources into chasing up responses for the 'purchases of services' question. The problem was particularly evident for the construction industry, as one team deals with all of the questionnaires for the industry, and it is a relatively small industry. There was a worry that even after chasing there was insufficient data to make this question meaningful. The Data Validation Branch (DVB) investigated this question further, to see whether it was worthwhile to include it in the survey. They analysed the responses to the question on purchases of services to see whether there was a problem with item non-response.

#### ***Analysing key process variables***

There were 522 businesses from the construction industry selected for the Purchases Inquiry. Only 108 (21%) of the businesses were able to give the required breakdown for the question on purchases of services, and many of these were only obtained after extensive phone calls. 70 (13%) of the businesses were unable to provide any breakdown at all. Many of these businesses sent back comments stating that it is either impossible or extremely time consuming to provide information at this level, as their accounting systems were not set up to provide it. This confirmed that there was a problem with the question. The results of this analysis were used as evidence to drop the question and thus improve the questionnaire.

#### ***Evaluation***

In this example, a survey team noticed that respondents were experiencing difficulties answering a certain question. This was followed by an analysis of item non-response for use as evidence to support a decision to drop that question. This was an isolated analysis, but responses to the Purchases Inquiry are regularly monitored and any questions with very low response over a two-year period are dropped.

## Post-Implementation Evaluation of the E-commerce survey questionnaire

E-commerce is likely to have a big impact on the way we do business. In recognition of its significance, the UK Government set itself the target of becoming 'the best environment in the world to do e-commerce.' In response to this policy need, the ONS has developed measures that will help monitor the UK's progress towards this aim. One strand of the strategy is a survey of UK business that asks about their use of the internet, and other communication technologies, and publishes estimates of the value of electronic sales and purchases using ICTs (Information and Communication Technologies.) The focus of this example is recent work on the E-commerce survey questionnaire.

### ***Identifying key process variables***

#### *Identify critical product characteristics*

Data collection for the E-commerce survey is carried out via paper questionnaires. Like all survey questionnaires, it is important that the content and phrasing of questions are clear to respondents and remain relevant to data requirements. The questionnaire design has an impact on the quality - in terms of accuracy, relevance and comparability - of the survey responses, and so is a critical product characteristic.

#### *Determine key process variables*

As part of work to improve the quality of the E-commerce questionnaire, the Data Collection Branch of Methodology Group fully reviewed the questionnaire for 2002. This included carrying out a post-implementation evaluation, aiming to:

- assess how successful revisions to the questionnaire had been in terms of:
  - meeting data requirements;
  - respondent response to the various questions;
- highlight areas for further improvement.

To achieve these aims the evaluation team gathered feedback from the Forms Processing Centre (FPC) who scan and capture data, and the Data Validation Branch (DVB) who validate the data and answer respondent queries. As well as common key process variables related to response rates, the evaluation team used feedback from the FPC (following analysis of batch errors in scanning) to identify several key process variables to measure and analyse.

The full list includes percentages of:

- a) item non-response;
- b) unnecessary response;
- c) entered a mark in both yes and no boxes;
- d) left both yes and no boxes blank;
- e) entered a numeric value in a mark box;
- f) entered a mark in a numeric field;
- g) completed more than one mark box where only one is expected;
- h) entered a non-relevant mark in or across mark boxes.

## ***Measuring key process variables***

To measure process variables, the evaluation team drew an 11% systematic sample of the returned forms that were identified as containing errors. The evaluation team inspected images of these forms, and counted the number of non-responders (a above) or unnecessary responses (b above), as well as the various types of errors affecting scanning (variables c to h above). These counts were converted into percentages for analysis.

In theory, it is possible to estimate sampling errors for estimates derived from the 11% systematic sample, if we can assume that the ordering of the forms is random. However insufficient information was available to calculate sampling errors for this report.

Whilst measuring the process variables, the evaluation team identified further useful breakdowns of process variable h, relating to the entering of non relevant marks in or across mark boxes. This problem was split into more detail as follows:

- all ticks (where crosses are required, not ticks);
- partial ticks;
- crossing out of questions;
- unclear tick/cross marks;
- unclear numerical response.

## ***Analysing key process variables***

### *Response issues, variables a and b*

Using data collected on response for the 11% sample, item non-response rates were calculated for each question. The Pareto chart in figure 6 below displays the contribution (in percentages) to the total item non-response from each of the thirty questions, in descending order. This quickly shows us which of the thirty questions are causing most of the item non-response in the sample, highlighting areas for improvement. For example, question 11 seems the most problematic, causing 12% of the total item non-response. Questions 11, 16, 25 and 27 together account for one third of the total item non-response.

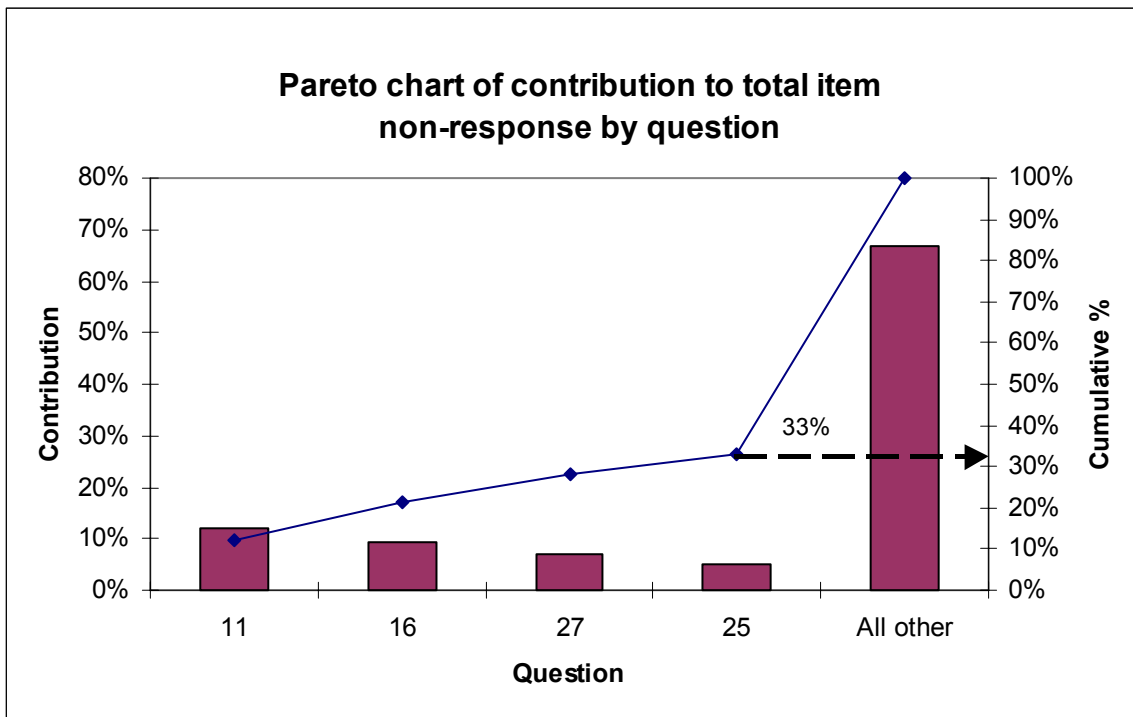
Independently of this data, feedback was received from the DVB. The comments from the DVB indicated specific questions where respondents were having problems. As would be expected, in general the problems reported occurred on questions with a relatively high item non-response or unnecessary response. For example, the DVB provided feedback for questions 16 and 25, which have the second and fourth highest contribution to overall item non-response, respectively. Both questions use the term 'ICT' (Information and Communication Technologies), that the DVB suggest would benefit from further explanation. Other problems identified included unclear definitions, a need for everyday examples so that respondents can relate the issues to their businesses, incorrect inclusions or exclusions, and on the respondents' part, a number of different people completing the questionnaire.

### *Scanning issues, variables c to h*

Analysis of the remaining variables enabled a better understanding of the mistakes made by respondents, which led to errors in the scanning process. Errors often occurred when a respondent attempts to correct mistakes made by: marking the wrong box; ticking before realising a cross is required; or entering the wrong digits in a numerical field. In the first case of marking a wrong box then correcting the mistake by marking the correct box in bolder ink, the evaluation team suggests that using densities in the scanning process might overcome the problem.

Feedback from the FPC revealed concern that the questionnaire contained inadequate instructions on how to give responses that could be read by a scanner. The instruction to 'please put a cross in the box that applies' was given under each appropriate question.

Figure 6:



### Evaluation

As mentioned in the paragraphs on response issues above, the evaluation suggested which questions have most need for improvement in the next review. Turning to scanning issues, development work on the instruction to 'please put a cross in the box that applies' led to an improvement. The instruction was changed so that it specified whether to mark one box only or to mark all applicable boxes.

In summary, analysis of the collected data helped to identify several issues that could be tackled to improve the quality of the questionnaire. These included the:

- layout of the questionnaire;
- formatting of the questionnaire;
- instructions on the questionnaire;
- guidance notes;
- question wording;
- validation;
- respondent burden.

Continued measurement and analysis of the process variables will help identify new and recurring problems, as well as highlight improvements or deterioration over time.

### Improvements to the complaints database

The ability of a statistical organisation to provide accurate and timely statistics is dependent on the organisation having a good, trusted relationship with the suppliers of the data. For example, producers of statistics need to balance user needs with the response burden on the providers. To understand the issues affecting responders, it is important to accurately monitor the volume and type of complaints raised.

#### ***Identifying key process variables***

The Correspondence Unit in ONS uses a Response to Public Inquiries database for recording the complaints that have been made by respondents to business surveys. The system enables monitoring of the volume and type of complaints received, which are the key process variables in this example.

#### ***Measuring key process variables***

There are three types of information recorded on the database for each complaint received:

- information on the complaint: this includes the survey involved, the date the complaint was made and the issue of the complaint;
- information on the respondent making the complaint: this includes the name and Standard Industrial Classification (SIC) code;
- information on the person responsible for dealing with the complaint and the progress they have made.

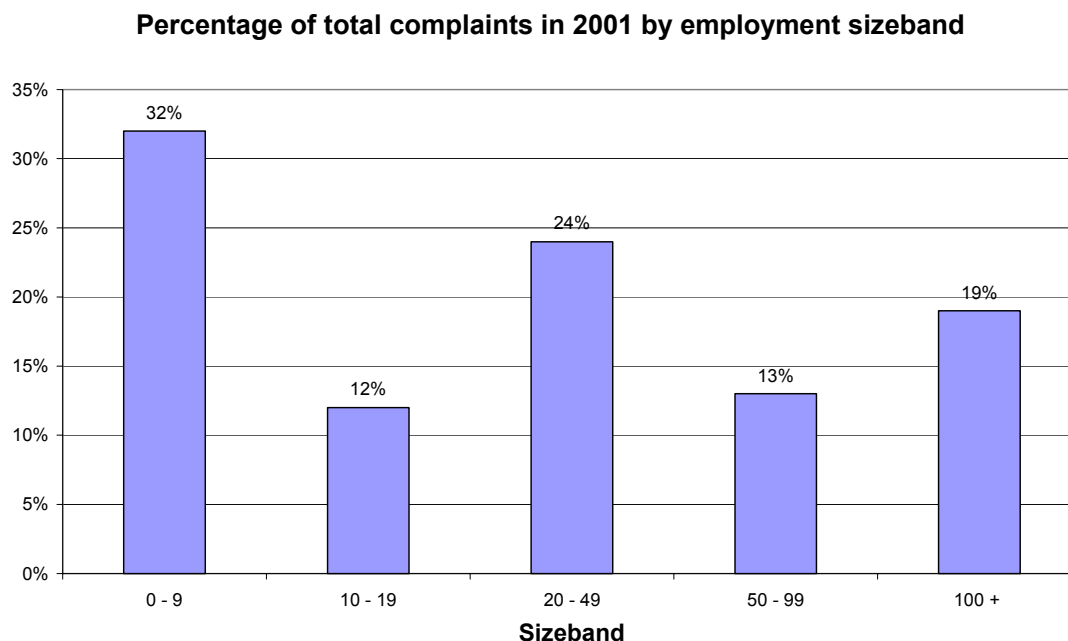
The accuracy of this information is dependent on staff recording the complaints consistently and entering the details of businesses accurately.

#### ***Analysing key process variables***

The Correspondence Unit analyse the data collected by the complaints database, including the number of complaints received. The chart in figure 7 below shows the distribution of complaints in 2001 by size of responding business. Note that this chart assumes that each size band has approximately the same number of businesses. We have not been able to ascertain the accuracy of this assumption. If the assumption does not hold, this will have an impact on the analysis below.

This distribution is roughly as was expected. The smallest businesses made most complaints, as they may not have the resources to fill in questionnaires. However, there was an unexpectedly high number of complaints received from businesses with 20-49 employees. Around 61% of the complaints from these businesses were about questionnaire filling burden. There are measures in place to control the number of questionnaires sent out to businesses with less than 20 employees. No such measures exist for larger businesses, but it appears from the volume of complaints that many businesses with 20-49 employees have insufficient resources for completing questionnaires. The Correspondence Unit decided to monitor this in case it became a serious issue.

**Figure 7:**



### ***Evaluation***

The Correspondence Unit has recently launched a new database to replace the Response to Public Inquiries database. Improvements in this database include options to note the source and type of the complaint and how the complaint was received. There are four new analysis options available:

- A table showing how the complaints were received and by what method;
- A table showing the number of complaints received in each size band;
- A section showing who the complaints were received from;
- A section showing how quickly the complaints were dealt with.

The new database will enable the Communication Division to easily identify the most common causes for complaint, and to relate these to the size of the responder, as was done in the analysis above.



### **3.3.1.2 Face-to-face Interviewing (INE-Pt)**

#### **Introduction**

This report considers applying our process quality approach to data collection by face-to-face interviewing. The remainder of the introduction provides a short description of the process, followed by a summary of the example from INE-Portugal that will form the focus of the report.

#### **Description of the process**

A comprehensive description of face-to-face interviewing is given below, taken from Office of Management and Budget, U.S.A. (2001).

*“ Face-to-face interviewing is the mode in which an interviewer administers a structured questionnaire to respondents. Using a paper questionnaire, the interviewer completes the questionnaire by asking questions of the respondent. This method, the paper and pencil personal interview (PAPI) method, has a long history of use. Although this method is generally expensive it does allow a more complex interview to be conducted. This mode allows the use of a wide variety of visual aids to help the respondent answer the questions. A skilful interviewer can build rapport and probe for more complete and accurate responses.*

*The advent of lightweight laptop personal computers has resulted in face-to-face interviewing being conducted via computer assisted personal interviewing (CAPI). Interviewers visit the respondents’ homes and conduct interviews using laptop computers rather than paper questionnaires. The use of CAPI permits editing data for accuracy and completeness at the time of the interview and provides for the correct following of skip patterns”.*

#### **Summary of the report**

This report presents the results of a study conducted by INE-Portugal with the aim of improving the quality of surveys using face-to-face interviewing as a mode of data collection. This study started in 1999, and improvement actions are still occurring, based on these results.

The method used to study the process and to analyse its critical aspects closely follows the approach recommended in this handbook.

## ***Identifying key process variables***

### *Identify critical product characteristics*

Interviewers are considered to be customers of the data collection process, in the sense that they use the information and tools that methodologists and design teams have chosen for data collection. In this study, their opinion on tools such as CAPI, training, difficulties when approaching households, etc were found to be important to consider when improving the quality of the process.

To investigate interviewer characteristics and opinions, a survey of interviewers was conducted. The questionnaire included questions that helped us to characterise the fieldwork teams, and to ask their opinion on the different surveys on which they have worked.

The structure of the questionnaire was:

1. *Function*: Interviewer, re-interviewer; region.
2. *Individual information*: age; sex; education level.
3. *General working environment aspects*: professional experience within (and outside) interviewing.
4. *Interviewing*: time used in interviewing, time used working at home, time used with supervisors; usual times for interviewing on each day of the week with reasons; opinion on the adequacy of the deadlines imposed.
5. *Motivation*: Motivation level; factors that influence motivation; expectations.
6. *General opinion on surveys (Global)*: problem solving and clarification of doubts; procedures across surveys (e.g. coding); CAPI – easy to use?
7. *Survey-specific opinions*: training and procedures; tools – CAPI, Handbooks.
8. *Comments and suggestions*.

The questionnaire was sent to 497 interviewers, and the response rate was 70.2%.

The study also involved a team of two people observing interviews in every region for every survey that was in the field. About 100 interviews were observed involving 42 interviewers.

These two phases (survey of interviewers and observation of interviews) helped us identify critical aspects of the process that could influence the quality of the product. In this case the product was data collected in the field. These aspects were organised into six categories or 'issues' that were analysed in detail to identify improvement opportunities and key process variables.

Critical aspects of the process:

- Issue 1 – Response rates
- Issue 2 – CAPI – software development
- Issue 3 – Methodological standardisation among surveys
- Issue 4 – Training Interviewers
- Issue 5 – Quality control in the field work
- Issue 6 – Institutional relation between interviewers and the Organisation

### *Develop a process flow map*

The Process Flow Map in figure 8 on the following page helped us analyse the way the process works. In the upper part of the map the process is centralised, with a survey coordinator who mostly works in the central part of the organisation in Lisbon. In the lower part of the map, the work is de-centralised by Region (there are five Regions in the mainland), where the fieldwork takes place.

Several aspects of this Process Flow Map are interesting to analyse: one example is the importance of “handbooks and CAPI” during the process, as it appears in the flow map three times; also “training” appears as an important task to ensure the performance of interviewers.

### *Determine key process variables*

The six critical product characteristics identified above were analysed with the help of Cause-and-Effect diagrams. Since most of the issues identified are not easily represented by numeric information, this tool was chosen as it displays descriptive information clearly for potential readers. The Cause-and-Effect diagrams for the six critical aspects of the process follow the process map.

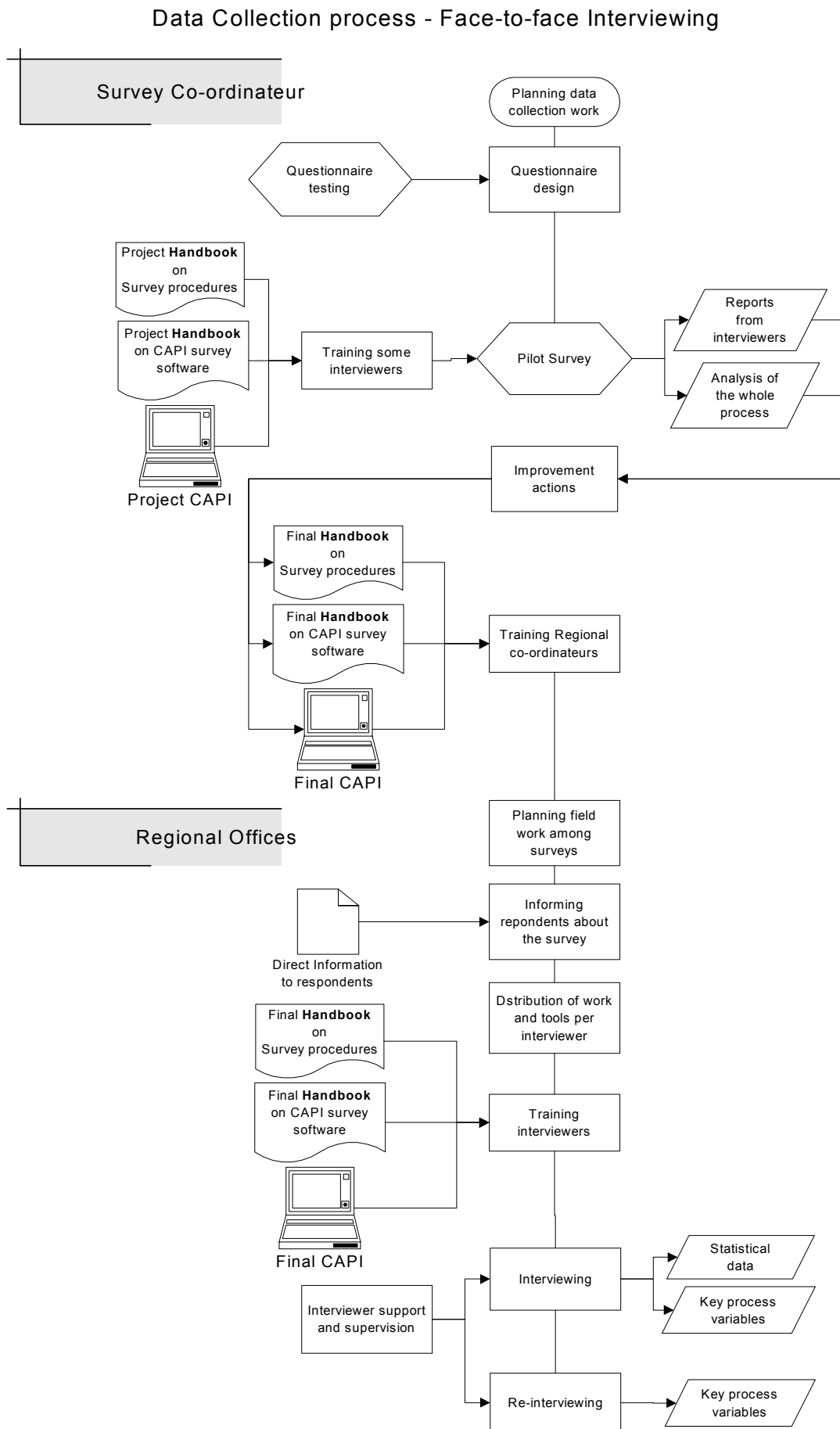
The analysis of the diagrams helped us to identify the factors that we thought would have most influence on the process. These factors were divided in two major groups, Institutional and Operational Factors. The institutional factors concern the organisation as a whole, while the operational factors relate to the survey itself.

Not all factors identified resulted in the identification of key process variables to measure over time. However, for some factors with no process variables, improvement actions were proposed and implemented. These improvement opportunities, once implemented, will have an effect on the quality of the product.

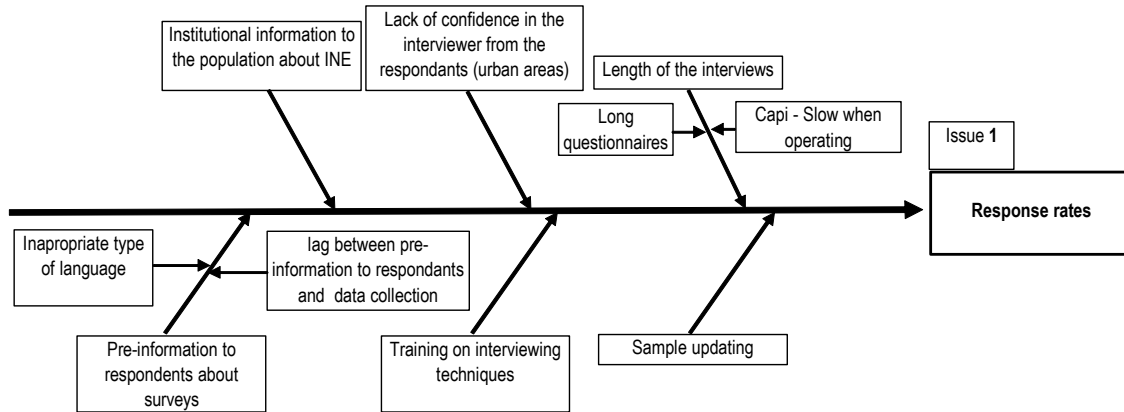
Measurement of key process variables for other factors may indirectly measure the effect of these improvement actions. For example, response rates can be improved if improvement actions for advertising the Organisation are implemented.

The improvement actions identified for institutional and operational factors are presented in tables 4 and 5 following the diagrams.

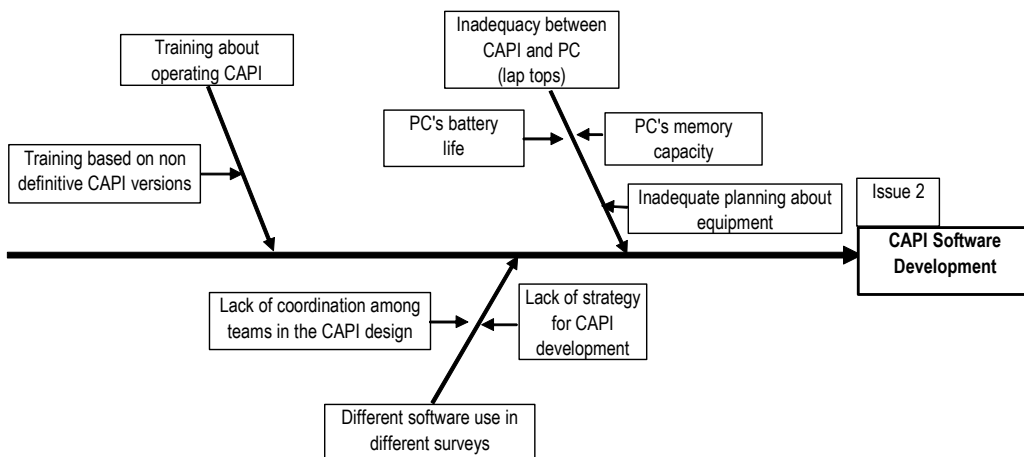
Figure 8: Flow Chart for the face-to-face interviewing process in INE-Portugal.



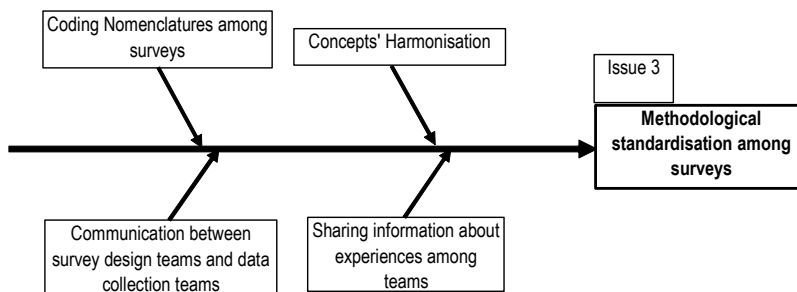
**Cause-effect Diagram 1 - Response rates**



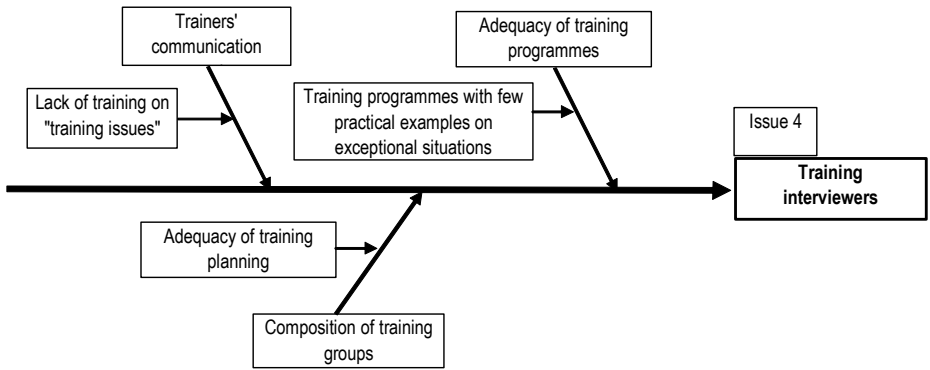
**Cause-effect Diagram 2 - CAPI Software Development**



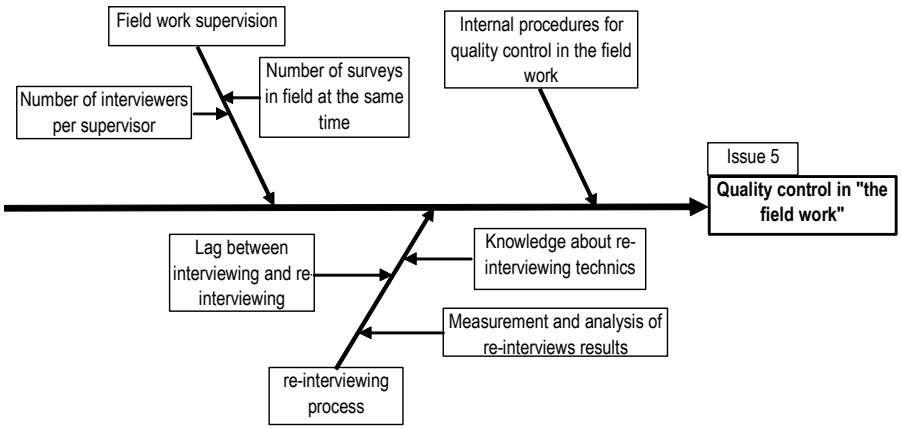
**Cause-effect Diagram 3 - Methodological Standardization among surveys**



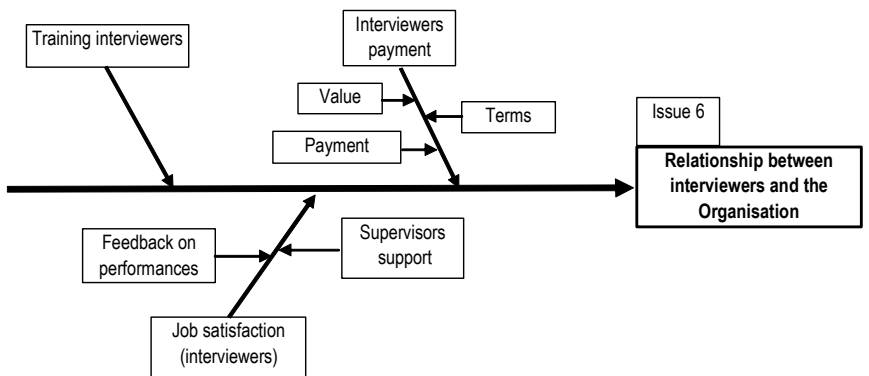
**Cause-effect Diagram 4 - Training interviewers**



**Cause-effect Diagram 5 - Quality control "in the field work"**



**Cause-effect Diagram 6 - Relationship between interviewers and the Organisation**



**Table 4:** Table of improvement actions identified for the institutional factors

Institutional factors		Improvement actions
I	Institutional information to the population about INE	Communication about INE and the most important surveys – to increase the dissemination of information about what INE is and the importance of respondents.
		Distribute promotional gifts from INE as a way to thank respondents (eg pens, pencils, watches, note pads, etc).
II	Relationship between interviewers and INE	Develop a handbook on interviewing (Interviewing techniques and general information about the organisation).
		Give interviewers the statistical publications of the surveys on which they have worked.
		Evaluate systematically (on an annual basis) the performance of interviewers.

**Table 5:** Table of improvement actions identified for the operational factors

Operational factors		Improvement actions
III	Questionnaire design	Methodological harmonisation in “coding” across surveys.
		Shorten questionnaire length.
IV	CAPI – Software development	Harmonisation and generalisation of CAPI Software – Blaise.
		Improve handbooks on CAPI survey software for interviewers (contents and timings in delivering the handbooks).
V	PCs – hardware	Adequate equipment (“Laptops”) for software applications.
VI	Training	Improve training planning and other organisational aspects related to interviewers.
		Train internal trainers on “pedagogical issues”.
		Improve content and timings of handbooks on survey procedures.
		Train interviewers and supervisors on methodological issues (eg sampling, quality issues, support documentation).
VII	Field work supervision	Implement an automatic data transmission system (a system that allows direct electronic data transmission between interviewers and INE). This system will allow supervisors to have more time to support interviewers.

### Measuring key process variables

Analysing the Cause-and-Effect diagrams above led to the identification of key process variables for several of the operational factors. The variables measured are outlined in table 6 below, along with details on their formulae.

**Table 6:** Key process variables identified for the operational factors

Operational factors		Key process variables	
VIII	Sampling frame updating	Ineligibility index due to non-effective sampling frame updating – percentage of sampling units found not eligible in the total sample. $\{(4) / [(1)+(2)+(3)+(4)]\} * 100$	
IX	Interviewing	Response rates	Net response rate – percentage of responses in eligible units. $\{(1) / [(1)+(2)+(3)]\} * 100$
			Gross response rate - percentage of responses in total units. $\{(1) / [(1)+(2)+(3)+(4)]\} * 100$
			Refusal rate – percentage of refusals in total eligible units. $\{(3) / [(1)+(2)+(3)]\} * 100$
			“Temporary away” rate – percentage of temporary away units in eligible units. $\{(2) / [(1)+(2)+(3)]\} * 100$
	Performance of interviewers	Deviation 2 weeks after reference period – measures the delays in performing the interview with respect to the reference period of the interview.	
X	Re-interviewing	Observation data entry errors	Consistency rate per question in the questionnaire – measures the number of equal answers between interview and re-interview for the same statistical unit (Number of equal answers related to total interviewed units).
		Response rates	Gross Response rate - percentage of responses in total units in re-interviewing sample (*).
			Net Response rate - percentage of responses in eligible units in re-interviewing sample (*).
			Refusal rate - percentage of refusals in total eligible units in re-interviewing sample (*).
			“Temporary away” rate - percentage of temporary away units in eligible units in re-interviewing sample (*).
		Performance of re-interviewers	Deviation 2 weeks after interviewing - measures the delays in performing the re-interview with respect to the performance date of the interview.
Deviation 5 weeks after reference period of the data - measures the delays in performing the re-interview with respect to the reference period of the data.			

(\*) identical to interview rates

**Key on next page.**



## Key to table 6

Eligible units: (1)+(2)+(3)

(1) Responses

(2) Temporary away units

(3) Refusals

(4) Ineligible units : (5)+(6)+(7)+(8)+(9)

(5) Secondary residences

(6) Empty household units

(7) Non-localisable household units

(8) Demolished household units

(9) Others

## Analysing key process variables

For factors VIII, IX and X of table 6, the process variables measured are analysed every quarter for the Labour Force Survey (LFS) relating to the 'Lisboa e Vale do Tejo' region.

Reports containing the analysis of these variables are disseminated in the organisation, and in particular to the areas involved in the survey process.

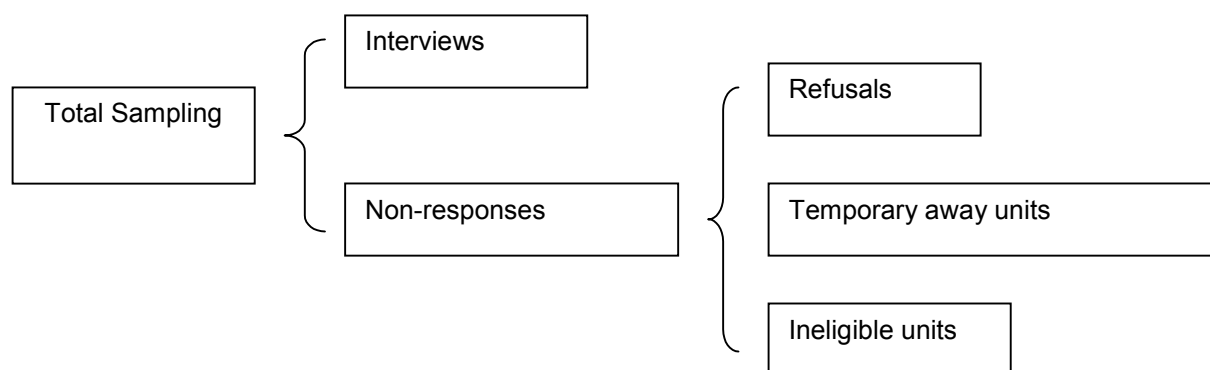
We also give interviewers, re-interviewers and supervisors feedback on their performance and results, making some recommendations for improvement work. The remainder of the report describes some analysis of response rates and of re-interviewing studies.

## Analysis of response rates using Pareto Charts

An example of analysing response rates using bar charts and Pareto Charts for the LFS is shown below.

The reasons for analysing response rates are not restricted to measuring the success of getting interviews. It is also important to analyse the non-response rate and how it is composed, as shown in figure 9.

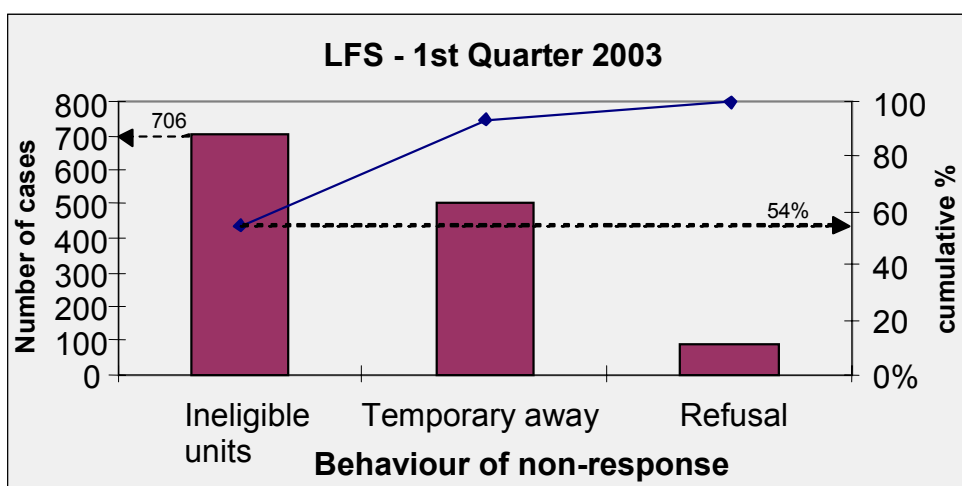
**Figure 9:** Composition of the non-response rate



Analysing non-response should lead to an understanding of its behaviour, which is portrayed in the diagram above. For household surveys, non-response is due to refusals, temporary away units and ineligible units. Ineligible units include secondary residences, empty household units, non-localisable household units, demolished household units. The presence of these units in our sample indicates ineffective sampling frame updating procedures. To improve the response rate we have to understand the composition of non-response, and the reasons for the different elements of non-response.

The rates referred to in table 6 are calculated every quarter in the case of LFS. Let us take the example of the LFS for the 4<sup>th</sup> quarter of 2002 and the 1<sup>st</sup> quarter of 2003. The Pareto Chart in figure 10 below shows the composition of non-response for the 1<sup>st</sup> quarter of 2003. Of the total non-response rate of 25.3% (corresponding to 1296 units), the chart shows that 54% is due to lost units. It is important to investigate why this rate is so high and take action to improve it.

**Figure 10:** Non-response behaviour for the Labour Force Survey, Lisboa e Vale do Tejo Region, first quarter of 2003.

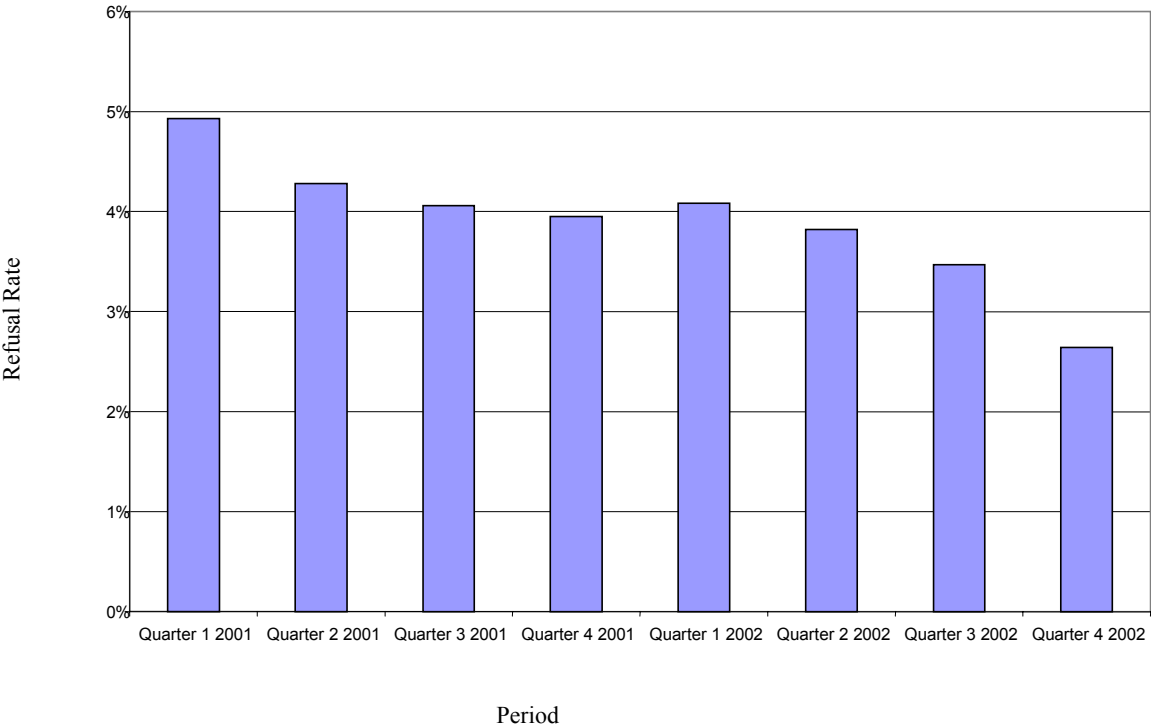


The two bar charts below – figures 11 and 15 - display the behaviour over time of two key process variables: the refusal rate and the ineligible index.

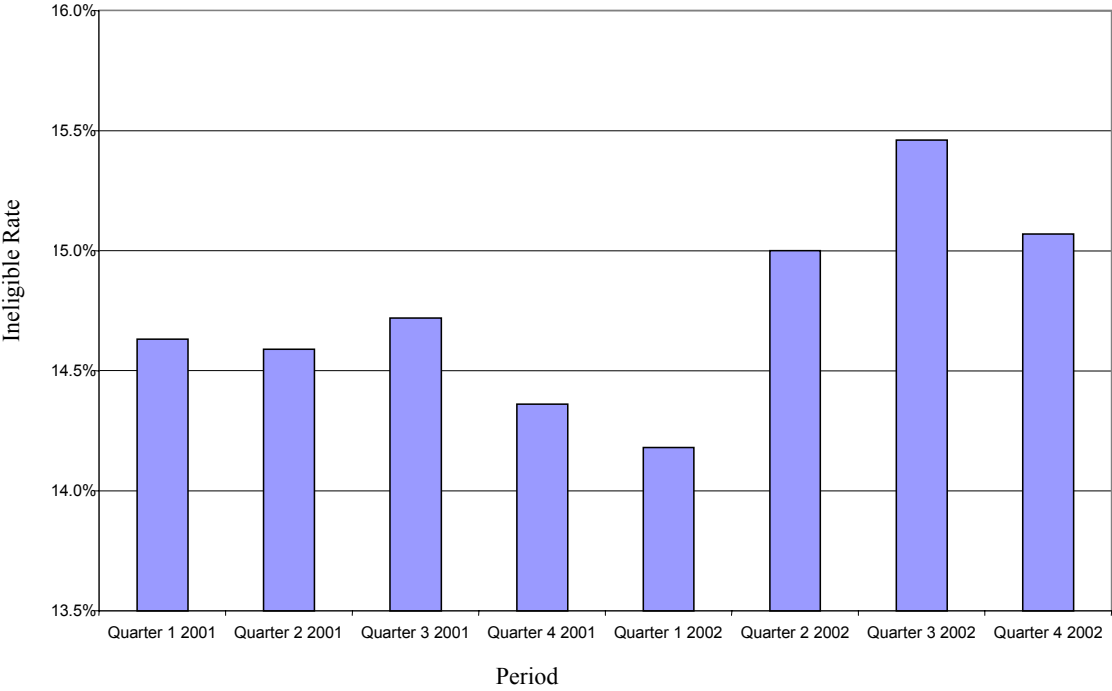
The refusal rate appears to be steadily decreasing throughout 2001 and 2002. This may indicate the success of improvement actions from tables 4 and 5, especially actions relating to the institution, such as improving promotion of the organisation.

The ineligibility index does not exhibit an obvious trend, and appears to have increased towards the end of 2002. This may indicate that more effort is required to improve the process of updating the sampling frame.

**Figure 11:** Chart of the refusal rate for the Labour Force Survey, Lisboa e Vale do Tejo Region



**Figure 12:** Chart of the ineligible index for the Labour Force Survey, Lisboa e Vale do Tejo Region



### Analysis of consistency rates on re-interviewing

A description of the re-interviewing process is given below, taken from “Measuring and Reporting Sources of error in Surveys”, Statistical Policy, Working Paper 31, Statistical Policy Office, Office of information and Regulatory Affairs, Office of Management and Budget, U.S.A., July 2001.

*“Reinterview – a replicated measurement on the same unit in interview surveys – is a new interview which re-asks the question of the original interview (or a subset of them) for a small subsample (usually around 5 percent) of a survey’s sample units. It is conducted for one or more of the following four purposes:*

- *To identify interviewers who are falsifying data;*
- *To identify interviewers who misunderstand procedures and require remedial training;*
- *To estimate simple response variance; and*
- *To estimate response bias.*

*The first two purposes provide information on measurement errors resulting from interviewer effects. The last two purposes provide information on measurement errors resulting from the joint effect of all four sources (i.e. interviewer, questionnaire, respondent and data collection mode). Reinterviewers do not usually provide an estimate of response variance/bias attributed to each source.”*

In our case, all four purposes for re-interviewing are relevant: to have a general quality measure, and to give feedback to the fieldwork team, at supervisor and interviewer levels. The detailed information resulting from re-interviewing is crucial to understand the real difficulties in the fieldwork and help us to define the principal issues in training programs we prepare.

In this analysis we compare the answers collected in the first and second interviews for each question. Assume that there are n possible answers to a question. In table 7, the answers are represented by ‘modality’ 1 to n, and the x-values represent the number of interview-re-interview pairs corresponding to each combination of answers.

**Table 7:** Template table of reinterview versus interview results

Variable		Reinterview				
		answer 1	answer 2	...	answer n	Total
Interview	answer 1	<b>x<sub>11</sub></b>	x <sub>12</sub>	...	x <sub>1n</sub>	N <sub>1.</sub>
	answer 2	x <sub>21</sub>	<b>x<sub>22</sub></b>	...	x <sub>2n</sub>	N <sub>2.</sub>
	...	...	...	...	...	...
	answer n	x <sub>n1</sub>	x <sub>n2</sub>	...	<b>x<sub>nn</sub></b>	N <sub>n.</sub>
	Total	N <sub>.1</sub>	N <sub>.2</sub>	...	N <sub>.n</sub>	<b>N<sub>nn</sub></b>

The correct answers are represented in the diagonal, and we can measure the quality of the variable using the formula below, where N is the total number of interview-re-interview pairs.

$$\text{Consistency Index} = \frac{100}{N} \sum_{i=1}^n x_{ii}$$

The teams of interviewers of the first (interview) and the second (re-interview) collection processes are independent. We assume that the re-interviewers provide the correct answers, and so we need to assure that the re-interviewers are as skilled as possible. This includes promoting special training sessions for them.

As an example, table 8 below presents the results for the variable “Educational Level” from the 4<sup>th</sup> Quarter of the Portuguese Labour Force Survey.

**Table 8:** Table of reinterview versus interview results for the variable “Educational Level” from the 4<sup>th</sup> Quarter of the Portuguese Labour Force Survey

Educational level		Reinterview										Total
		1	2	3	4	5	6	7	8	9	10	
Interview	1	160	15		1							176
	2	16	277	8	4	1	1					307
	3		14	69	9	1						93
	4		1	11	100	8	10	1				131
	5		1	3	9	73	1	1	4			92
	6				3	6	6					15
	7		1			3		8	4			16
	8					1		4	51	5		61
	9								2	1		3
	10										1	1
	<b>Total</b>	176	309	91	126	93	18	14	61	6	1	<b>895</b>

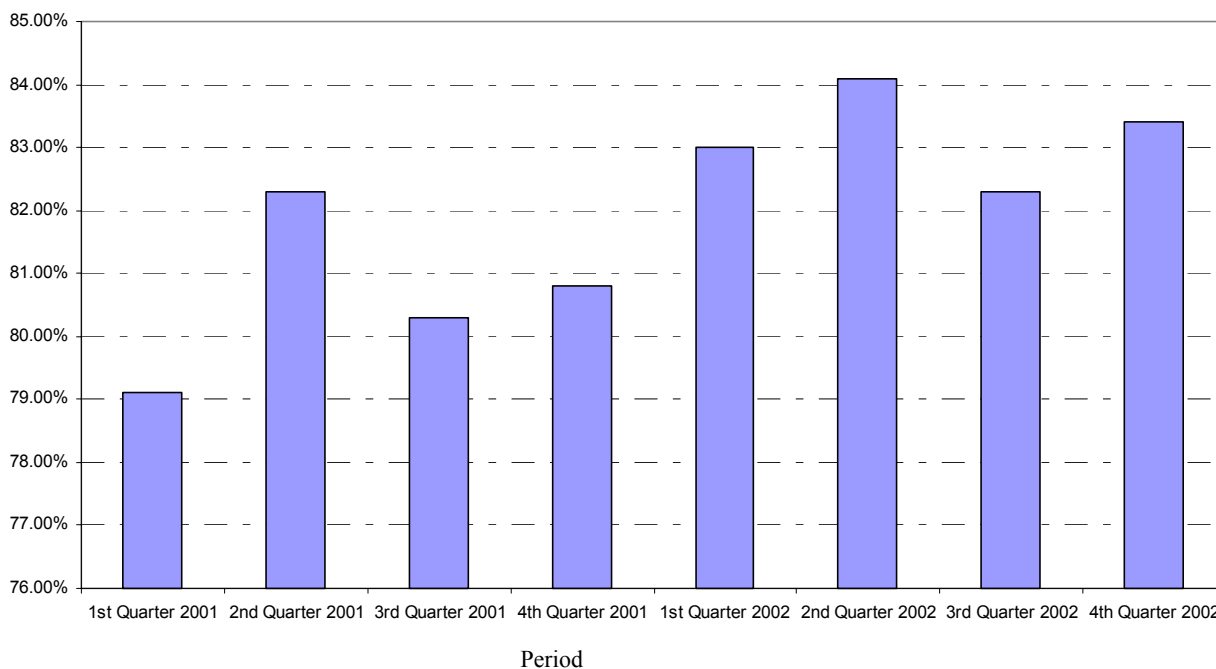
1-None; 2-Basic education (1<sup>st</sup> cycle); 3-Basic education (2<sup>nd</sup> cycle); 4-Basic education (3<sup>rd</sup> cycle); 5-Secondary education (general education); 6-Secondary education (vocational or technical education); 7-Higher education - 1<sup>st</sup> degree (“Bacharelato” - 3 years); 8-University education – 1st degree (“Licenciatura” – average 4 years); 9-University education – 2<sup>nd</sup> degree (“Mestrado” - 2 years); 10-University education – Doctorate

Source: Portuguese Labour Force Survey, 4<sup>th</sup> Quarter 2002

In this case, the Consistency Index is  $746/895 = 83.4\%$ . Moreover, we observe that the wrong answers are very near the diagonal. Exceptions to this closeness to the ‘true’ value may be consequences of data entry errors, because data collection uses the CAPI system involving laptop computers. It is not possible to separate interviewer error from data entry error in the re-interview analysis. We also give the team responsible for the survey access to the information in the table above. They will have a good understanding of the difficulties involved in the fieldwork, so may be able to infer more from the table than those less familiar with the data collection process.

The following bar chart in figure 13 displays the progress of the consistency index through the quarters of 2001 and 2002, for the same variable “Educational Level”.

**Figure 13:** Chart of the consistency index for the variable 'Educational level' for the Portuguese Labour Force Survey, from 1<sup>st</sup> Quarter 2001 to 4<sup>th</sup> Quarter 2002



We observe an increase in the values of the index, suggesting that efforts made to explain the real meaning of each of the answers to interviewers were successful. This was done using training sessions or through communication between supervisors and interviewers.

### ***Evaluation***

The identification of important sources of quality for the data collection (by face-to-face interviewing) process led to the implementation of improvement actions and to routine measurement of key process variables. Analysis of these measurements for the LFS highlight where the improvement efforts are succeeding (eg decrease in non-response), and where further work is necessary.

### **3.3.1.3 Interviewing activities (SCB)**

#### **Introduction**

This report considers applying our process quality approach to data collection by interviewing. The remainder of the introduction provides a short description of the process, followed by a summary of the example from SCB that will form the focus of the report.

#### **Description of the process**

This report focuses on processes involved in interviewing activities. Interviewing activities include - for example - face-to-face interviewing, telephone interviewing, tracing, interviewer training, as well as central activities such as those performed by supervisors and survey managers.

#### **Summary of the report**

This example describes ideas and some results of a project at Statistics Sweden. The work procedure mainly follows the ideas in section 2 of this handbook. The objective is to identify process variables that are related to nonresponse errors, measurement errors and productivity in interviewing activities.

#### **Report on process variables at Statistics Sweden, Interview Unit**

The Statistics Sweden Interviewing unit carries out the data collection in all telephone and face-to-face surveys conducted by Statistics Sweden. Quality in interviewing activities plays an important role for the final product quality. At Statistics Sweden we are in the final stages of a project that has the purpose to increase the quality and the productivity in interviewing activities.

The project has the following *mission statement*:

*"The objective is to identify and measure key process variables that are related to nonresponse errors, measurement errors and productivity in interviewing activities. Process data should be stored in a database and standard reports and guidelines on how to analyse and use process data should be developed"*

The main users or "customers" of process data in this case are the managers of the Interview Unit, survey managers, fieldwork staff, supervisors, subject-matter department staff (internal customers), and external customers. Except for external customers, each category is represented in the project team, which is an advantage when identifying key process variables.

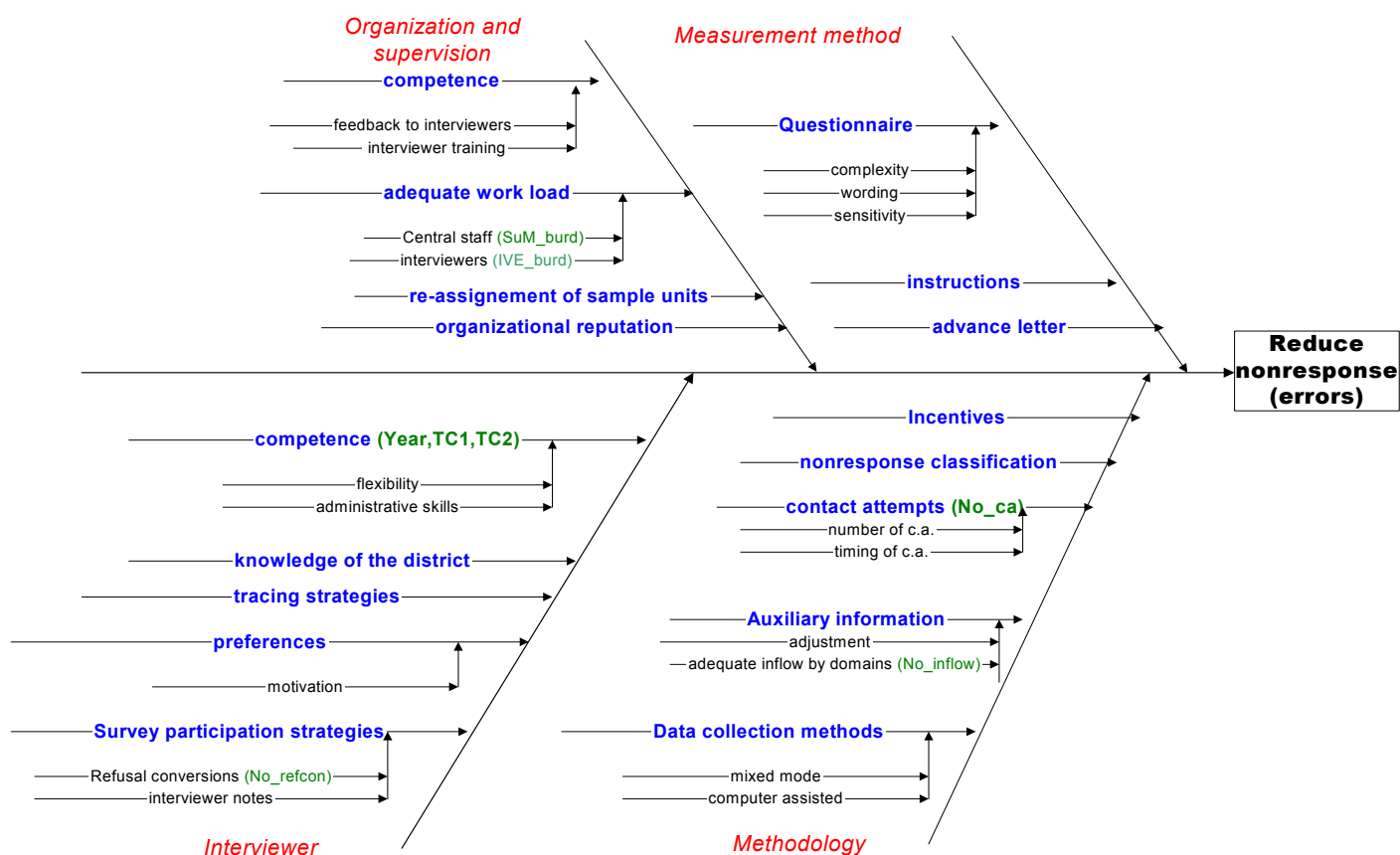
## Identifying key process variables

The objective is to identify and determine key process variables that are related to nonresponse errors, measurement errors and productivity. Cause-effect diagrams were used as tools for this procedure. The project team created one diagram for each area: nonresponse, measurement and productivity. For each area a diagram containing influential factors was created, irrespective of the potential to determine key process variables based on these factors. That is, the purpose of the procedure is:

- i. To try to identify all factors that might have influence on nonresponse errors, measurement errors and productivity in interviewing activities.
- ii. To define possible key process variables for each factor in the cause-effect diagram.

Cause-effect diagrams 7 to 9 below show that it is relatively easy to identify influential factors for each area, but considerably harder to define measurable key process variables for the factors. This is an indication of the difficulty in finding suitable process variables for some processes. However, the “two-step-procedure” above increases understanding of the factors that have an influence on the desired effects, even where it is not possible to identify and measure a key process variable.

**Cause-effect diagram 7. Reduce nonresponse errors**



The desired result of using cause-effect or ‘fish-bone’ diagram 7 is to identify factors that can be helpful in reducing nonresponse rates and eventually nonresponse errors. In this



example we limited our study to interviewing activities, where reducing nonresponse errors is seen as closely related to reducing nonresponse rates. From a total survey error perspective, reducing nonresponse errors involves applying an appropriate estimation procedure and nonresponse adjustment (see example 'SVC estimation').

The diagram is divided into four major "bones": Organisation and supervision, Measurement method, Interviewer, and Methodology. Every main bone is divided in "sub bones" which in turn are divided into to further "minor bones" (or factors). The measurable process variables defined are displayed within parentheses in cause-effect diagram 7. The following is a list of those variables with some comments as to their use:

*SuM\_burd = Number of planned hours/survey/survey manager.*

The main user is the management. The purpose is to gain control over the survey manager's workload. At Statistics Sweden, survey managers are usually in charge of more than one survey simultaneously. A survey manager with a heavy workload might not be able to control his or her fieldwork situation effectively.

*IVE\_burd = Number of planned hours/survey/interviewer.*

At Statistics Sweden most of the 150 field interviewers are working from their homes and have an allotment of sample units from different surveys. It is important for the fieldwork staff to control the *interviewer burden* for each interviewer. An interviewer with a heavy workload will probably have a negative influence on the response rate. *IVE\_burd* will serve as a guide for the fieldwork management staff in their decisions to re-assign sample units or not.

*Year = Number of years at work*

*TC1<sup>1</sup>: [1=participation Training course 1, 0=No participation Training course 1]*

*TC2: [1=participation Training course 2, 0=No participation Training course 2]*

Although these process variables may not be considered as key, it is meaningful to study correlations between them and individual interviewer response rates.

*No\_refcon = Number of successful refusal conversion attempts / Number of refusal conversion attempts.*

A refusal conversion attempt is a second (or later) contact attempt with a sample unit who refused to participate at the previous attempt. A refusal conversion attempt is assumed to be carried out by another interviewer specially trained for this purpose. *No\_refcon* is first of all of interest to study if refusal conversion attempts are successful. This key variable encourages analyses on interviewer level, as well as on a more overall level. Armed with knowledge about refusal conversion attempts it is also possible to present response rates split into refusal conversions and "normal" respondents. This is a recommendation at some NSIs, for example at Statistics Canada.

*No\_ca = Number of contact attempts / time period. (The real key variable is actually Number of contact attempts needed to achieve a contact related to different time periods.)*

---

<sup>1</sup> Statistics Sweden has two compulsory general training courses for interviewers

Such information is useful to create efficient call algorithms and other contact strategies. For more details concerning call algorithms we refer to Japoc et al. (1998).

$No\_inflow = \text{Number of final code units} / \text{time period} / \text{domain}.$

A basic purpose of inflow statistics is to have a means to evaluate reminder procedures.  $No\_inflow$  is a process variable for controlling response rates divided into important domains of interest. A continuous control of the inflow would help survey managers and field staff to gain control of the survey production and introduce necessary efforts in critical domains. For example, in Sweden we know by experience that non-Swedish citizens and big city citizens have a low response probability.  $No\_inflow$  will be a good tool when trying to avoid unpleasant surprises (in terms of response) for these domains.

Cause-effect diagram 8 follows the same principles as cause cause-effect diagram 7 and has the same main bones, with identified key variables within parentheses.

**Cause-effect diagram 8. Reduce measurement error situations**



Two of the process variables are identical to variables from cause-effect diagram 7:  $SuM\_burd$ ,  $IVE\_burd$ . This shows that some variables or factors might have influence on both nonresponse and measurement errors.

Under *Measurement method* only one process variable is defined, and is described below. It was found that measurement methods are usually dealt with through using an *evaluation process*, rather than studying continuous process variables.

*INT\_time* is simply interviewing time / respondent / survey. In continuing surveys the average *INT\_time* is a useful variable for survey resource calculations. Another area of interest is to study the interviewing time variation on interviewer level, for example with order statistics (median, percentiles).

Process variables identified under *Methodology* are defined below.

The first three variables listed should be viewed more as quality indicators of interviewing activity rather than key variables. The main reason why they are included in the Swedish project is that monitoring and re-interviewing procedures have been neglected during recent decades, and therefore require monitoring themselves.

*P\_monit* = *Proportion of monitored interviews / survey.*

*P\_reint* = *Proportion of re-interviews /survey.*

*P\_obs* = *Proportion of field observations /survey.*

The editing procedure aims to avoid or manage measurement errors. Editing is discussed in detail in chapter 3.3.3, but here we draw attention to one simple key variable:

*No\_errors* = *Number of error (messages) / item /survey.*

If there are large values of *No\_errors*, it is a clear indication of a bad question, and a revision of the question is needed.

Regarding data collection methods, mixed-mode is becoming increasingly common. With a mixed-mode design one mode is usually considered the 'main' mode. It is important to systematically follow response by collection method. If it differs significantly from what was expected by the design, one might lose control over the measurement situation. The following variable could be seen as a minimum in the case of a mixed-mode design:

*P\_diff* = *Proportion of responses obtained from modes other than the main one / survey*

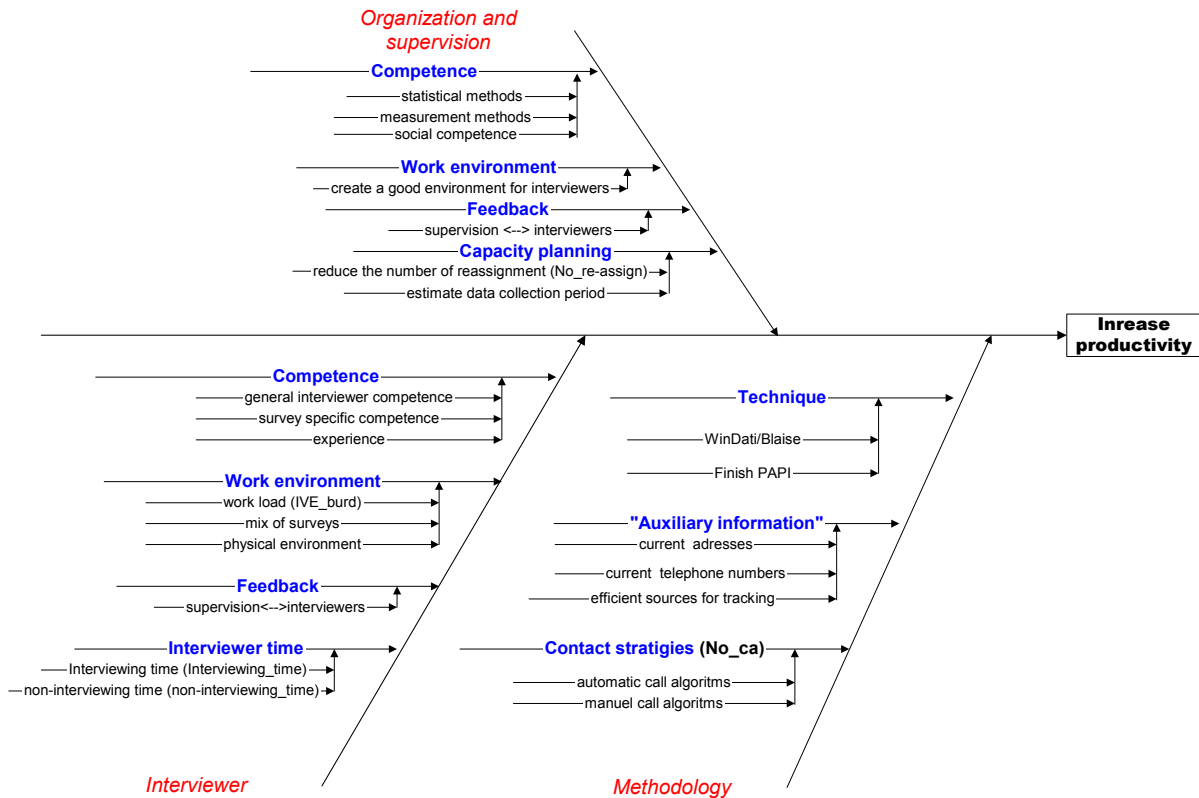
To increase the response rate, proxy interviews are often permitted. However, proxy interviews can increase measurement errors. For more detailed knowledge about the effect of proxy interviews on response rates and measurement errors, one must carry out special investigations. In continuing survey production it is at least necessary to have control over the proportion of proxy interviews over time:

*P\_proxy* = *Proportion proxy interviews / survey*

Under the *Interviewer* main bone, the factor "PC-slip-ups" potentially affects measurement errors. The aim is to define relevant process variables by means of audit trails and/or keystroke files, although at present this work is incomplete.

The structure of cause-and-effect diagram 9 is similar to diagrams 7 and 8 except that one of the main bones - *Measurement method* - is not included. The reason for the decision to exclude it is that, although the project team found measurement method to be closely related to nonresponse and even more so to measurement error situations, the links were not so strong for “pure” productivity.

Cause-effect diagram 9. Increase productivity



Considering for *Organisation and supervision* in the context of productivity, the assignment of sample units to interviewers is a very important process. As mentioned earlier, most interviewers are field interviewers, each receiving an assignment of sample units before every upcoming survey. That is, the interviewers are working simultaneously on several surveys. Due to various causes (illness, temporary leave of absence, etc) there are usually several re-assignments to carry out. The fundamental problem in this case is to find an optimal assignment procedure to work with. The key variable *No\_re-assign* is simply *the number of re-assignments by survey*. This is measured with the aim to reduce the number of re-assignments as they cost time and money. It is of special interest to study *No\_re-assign* by district. The Swedish field interviewers are divided into seven districts or regions.

One reason for re-assignments is the heavy workloads of some interviewers. That is the reason why *INT\_burd* (Interviewer burden) appears again in this diagram (as a “sub bone” to *Interviewer*).

Also under *Interviewer*, interviewer time is of great interest when one wants to increase productivity, for example by reducing the *non-interviewing time*. Interviewer time can be written as:

Interviewer time = Interviewing time + non-interviewing time

Similarly non-interviewing time can be written:

Non-interviewing time = travelling time + tracking time + residual time.

Residual time includes the reading of instructions, meeting time, etc.

To measure tracking time either interviewers must report time spent on tracking, or one may estimate tracking time using a model. This is not done in Sweden at present. *Standard report B* below includes the concepts of interviewing time, travelling time and "other time", where other time includes all other activities.

Under *Methodology*, contact strategies (*No\_ca*) is closely related to productivity as well as to nonresponse (see cause-effect diagram 7).

### ***Measuring and analysing key process variables***

In this report a subset of the project output will be presented. The desired total output of the project is:

- To create a process database with information from the Swedish CATI-system (WinDATI) and the Swedish interview reporting system (IDA).
- A set of standard reports. At present, eight standard reports are being prepared.

The standard reports are prepared in co-operation with statisticians, survey managers, field staff, supervisors and, above all, interviewers.

In this report two of the standard reports prepared are presented:

- a) Interviewer time by survey
- b) Survey progress report

#### ***Standard Report A: Interviewer time by survey***

An example report is provided in figure 14, with each component of Standard Report A described below.

The calculations are based on the final case coding of the sample units (WinDATI) and reported working hours (IDA). The purpose of the report is to:

- Illustrate reporting time for interviewer activities;
- Identify problems that might contribute to a decreased productivity;
- Generate basic data for cost follow-ups;
- Generate basic data for survey cost estimates.

**Figure 14:**

**Standard Report A. Interviewer time by Survey.**

Survey id	007	(A face-to-face survey)
Rounds(ar)	201, 202	
Stage(s)	1	
Unit of time: from	2002-08-01	
Until	2002-10-20	

Region	Number of sample units	Interviews		Interviewing time <sup>1</sup>		Travel time <sup>2</sup>		Other time <sup>3</sup>			Working time <sup>4</sup>		Total interviewer time <sup>5</sup>	
		Number	Pct	Hours	Min-utes/inter-view	Hours	Min-utes/inter-view	Hours	Min-utes/sample	Min-utes/inter-view	Hours	Min-utes/sample	Hours	Min-utes/sample
Sthlm	847	444	52%	481	65	464	63	771	55	104	1252	89	1715	122
East	411	210	51%	279	80	258	74	438	64	125	717	105	975	142
South	532	293	55%	379	78	270	55	493	56	101	871	98	1141	129
West	690	374	54%	431	69	344	55	775	67	124	1206	105	1550	135
Across	667	388	58%	417	64	386	60	660	59	102	1077	97	1463	132
Middle	384	197	51%	226	69	231	70	344	54	105	570	89	801	125
North	291	150	52%	173	69	255	102	360	74	144	533	110	787	162
CATI	0	0	0%	0	0	0	0	0	0	0	0	0	0	0
<b>Total</b>	<b>3822</b>	<b>2056</b>	<b>54%</b>	<b>2385</b>	<b>70</b>	<b>2207</b>	<b>64</b>	<b>3841</b>	<b>60</b>	<b>112</b>	<b>6226</b>	<b>98</b>	<b>8433</b>	<b>132</b>

1) Source: WinDati

2) Source: IDA

3) Other time = Working time – Interviewing time

4) Source: IDA

5) Total interviewer time = travel time + working time

### *Interviewing time by survey*

The main users are survey managers. The interviewing time is currently the most uncertain factor when estimating survey costs. This is partly because the number of test interviews is often very small, and generates very unreliable cost estimates. There is a requirement to follow up interviewing time and study its variations.

### *Travel time by survey*

This is a valuable input for survey cost estimates. Travel time contains trips to and from respondents in face-to-face surveys. Travel time also includes time for cases when the interviewer was “stood-up” by the respondent, i.e., when the respondent was not at home at the time agreed for an appointment.

### *Other time by survey*

*Other time = reported time – interviewing time – travel time.* ‘Other time’ is useful for survey planning and gaining control over ‘other time’ is important for supervisors. ‘Other time’ includes a lot of activities, for example tracking, work planning, contact attempts, and work with nonrespondents. Our ambition is to make it possible to divide other time into different sub-activities.

### *Working hours by survey*

This variable has the same purposes as other time.

### *Interviewer time by survey (total time)*

This variable is useful for assignment of sample units, for survey cost estimates and for cost follow-ups. It contributes to a comparison of survey cost estimates with the actual result.

It should be possible to generate standard report A anytime during the survey production. The result will, however, be most reliable after the data collection period has ended.

The user of the standard report states the following parameters:

- Survey;
- “Round” (every survey has one or more rounds);
- Stage (every round has one or more stages);
- Unit of time (minimum is one day);
- Region (define region or choose all).

The example in figure 14 presents process data from a face-to-face survey carried out by SCB during the time period August 1<sup>st</sup> 2002 to October 20<sup>th</sup> 2002. In this case unit of time was stated. Thus, all possible reported time for the survey was included. Since this is a face-to-face survey, the CATI-staff figures (= 0). Primarily, the results from standard report A are very useful to the next survey occasion and to similar surveys. In the last column (total interviewer time), region North has a very high average value: Minutes/sample unit. In this case we have a clear explanation: Region North is a sparsely populated area and interviewers need more travel time to do their work (see travel time in the same table).

Another standard report, described below, has the same layout and data sources, but the results are presented on interviewer level. This is useful for the regional supervisors.

**Standard Report B: Survey Progress Report**

This standard report - figure 15 below - contains two histograms based on data from WinDATI (finally case-coded sample units) and IDA (working hours). The objective is to give an overview of the inflow and survey costs for present and completed surveys.

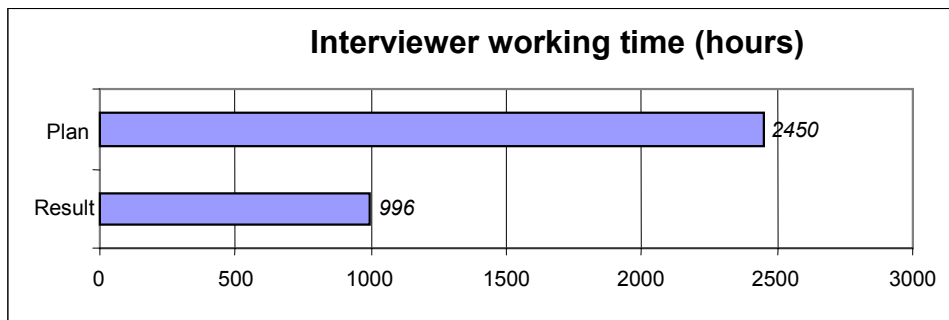
The report is intended mainly for the management. The management needs quick information about the status of one or several surveys. The report presents an overview, and for more detailed information one has to consult the survey manager in charge.

It will be possible to generate the report anytime during the fieldwork and, of course, after a survey is completed.

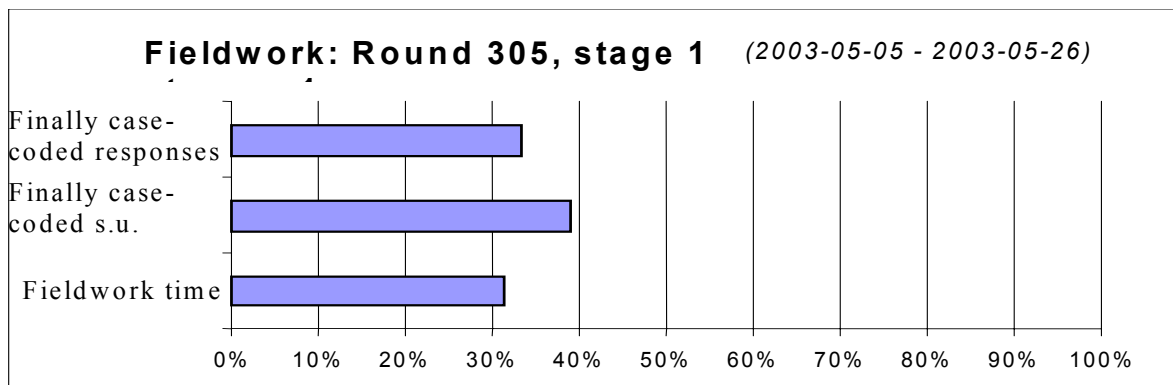
**Figure 15:**

**Standard Report B. Survey Progress report**

Survey id	006	<i>Party Preference Survey (PSU)</i>
Rounds(ar)	305	
Stage(s)	1	
Unit of time: from	2003-05-05	
Until	2003-05-26	
Survey Manager		Ingrid Fekkes



Key process variable: *Proportion of working hours (PWH) = Number of hours according to present result (Utfall) / Number of hours according to the survey planning (Plan) = 0.41*



Key process variable: *“continuous response rate” = finally case-coded responses / finally case-coded sample units (CCU) = 0.82*



Key process variable:  $CCU / PWH = \text{Proportion finally case-coded units} / \text{Proportion of working hours} = 0.80$

In figure 15 above the progress report gives an overview of the survey status after three weeks data collection, that is, just over 30 % of the (expected) data collection period. The user of this standard report can see that the proportion of 'finally case coded sample units (s.u.)' is almost 40 % and the proportion of finally case-coded responses more than 30 %. The facts that the proportion finally case-coded s.u. is larger than fieldwork time, and that the "continuous response rate" is over 80 % indicate a good survey status. On the other hand we can see that the proportion of working hours is just over 40% (more than the proportion of fieldwork time). The key variable CCU/PWH is 80%, which is an indication that the survey has used up to 20% more interviewer hours with respect to the number of finally case-coded units. We know by experience that the finally case-coded nonresponse units will increase towards the end of the fieldwork. Despite these facts the conclusion thus far will be that we have an adequate survey status. If the case should show a larger fieldwork time-chart compared with finally case-coded sample units and a high proportion of working hours together with a low continuous response rate, then we have an indication of serious problem with the fieldwork.

A set of Survey Progress reports that covers all ongoing surveys will be an excellent tool for managers of the interview unit.

### **Evaluation**

In this example some results from an ongoing project have been presented. The team succeeded with identifying measurable key process variables concerning nonresponse and productivity. The team found that measurement methods are usually dealt with through using an *evaluation process*, rather than studying continuous process variables.

An important output from the project is been a set of standard reports. Through team work 8 standard reports have been designed. Two of them are presented in this report. In the present phase of the project SCB has finished IT-solutions for 4 of 8 standard reports integrated in the WinDATI-system. A hindrance to the development of user-friendly IT-solutions for the standard reports has been restrictions with WinDATI. The next step of the continuing project (concerning process data on interviewing activities) is to create a process database beside the WinDATI system.

### 3.3.1.4 Literature review

Several papers describing quality work related to the process of data collection by paper self-completion questionnaire are listed below, along with a short summary of their content. These papers contain some further ideas for process variables and examples of their analysis, and will provide valuable information for those seeking to monitor and improve their data collection quality.

Bocci L, Brancato G and Signore M (2000) *The SIDI System: A Generalised Tool for Quality Control of ISTAT Surveys*, ISTAT publication.

This article describes a system used by ISTAT to store information about the surveys they carry out and calculate the quality measures for them. Some of the mentioned quality measures for data collection are the non-response rate and refusal rate.

Cevis T O and Peregoy A (1984) *An Overview of Quality Control in the Economic Area of the Bureau of the Census*, Proceedings of the Section on Survey Research Methods, American Statistical association.

This paper summarises the process the US Census Bureau use to print questionnaires, send them out and process them when respondents return them. The description includes details of the quality control checks they carry out on the mail packages and returned data. Some of the techniques used are acceptance sampling and code agreement between two separate coders. They also use an interactive processing system that automatically identifies obvious errors in the data being entered, gives the clerk a chance to correct data entry error and then, if necessary, sends the questionnaire to be checked by an analyst.

Filipucci C and Calia P (2001) *Towards Process Quality Control in the Italian Consumer Expenditure Surveys*, presented at the International Conference on Quality in Official Statistics, Stockholm, May 14-15, 2001.

Haselden L and White A (2001) *Developing New Quality Indicators in Social Surveys*, Paper for Quality in Official Statistics conference in Stockholm, Sweden May 2001.

These articles explain the ways ISTAT and the Office for National Statistics Social Survey Division map the survey process and identify quality indicators. The indicators identified for data collection fall into two main categories, namely the level of response and the implementation of the survey. The first category includes non-response rates, non-return rates, partial response rates and the number of 'don't know' responses. The second category includes reports on the question testing carried out, the punctuality of questionnaire despatch, the number of incomplete or unreadable mail packages reported and the average time spent completing the questionnaire.

The second paper discusses the effect of recording these process variables on the process itself. For example, if the number of replacement questionnaires sent out were monitored, then perhaps survey managers would be reluctant to issue replacements even when there is a problem with the questionnaire. One way of countering this would be to educate survey managers on the uses and limitations of the process variables. The authors have found quality indicators for many processes, including interviewing activities.

Some maps and indicators are presented in the paper. The aim is to be able to appraise changes to the processes as well as informing customers of quality.

INE-PT (2001) *Definition of a set of procedures for controlling process errors on surveys and their quantification (Definição de Procedimentos para Controlo dos erros no Processo de Produção Estatística e Respectiva Quantificação, relatório de estágio de Hélia Estevens, Abril de 2001)*, INE, Portugal.

This report constitutes an important achievement in survey quality control, in particular concerning face-to-face interviewing. The report was based on the quality control of the labour force survey and focuses on non-sampling errors. It is composed of three main chapters: error sources in the different stages of the statistical production process; how to deal with non-response; re-interviewing process as a quality control tool, for the evaluation of observation errors.

INE-PT (Quarterly report) *Quality analysis of the Labour Force Survey (Análise da qualidade da Informação do Inquérito ao Emprego, relatórios trimestrais)*, INE, Portugal.

This is a Quarterly quality control report on the Labour Force Survey, based on the procedures established in the above report.

INE-PT (2002) *Multivariable analysis on quality control (Análise multivariada no Controlo da Qualidade, relatório de Estágio de Ana Patrícia Martins, Dezembro 2002)*, INE, Portugal.

This is a report on multivariate data analysis of quality indicators at interviewer level, as mentioned in the above reports and on the process variables listed in the INE-Pt example.

Japac L, Ahtiainen A, Hörngren J, Lindén H, Lyberg L and Nilsson, P. (1997) *Minska bortfallet* (in Swedish), Statistics Sweden.

This book is one of a several handbooks: Current Best Methods (CBM) at Statistics Sweden. This CBM covers procedures to reduce nonresponse. It describes: advanced letters, tracking (of sample units), non response follow up strategies, incentives, sensitive questions, proxy, interviewer issues and combinations of these measures.

Japac L, Lundquist P and Wretman J (1998) *Interviewer strategies: How do interviewers schedule their call attempts?* Paper presented at the Ninth International Workshop on Household Survey Nonresponse in Bled, Slovenia, September 23-25, 1998.

This paper describes an attempt to identify the calling strategies that interviewers at SCB use when scheduling their call attempts. The authors have also carried out a study to find out about when (presumptive) respondents are at home.

Jones J (2003) A Framework for Reviewing Data Collection Instruments in Business Surveys, *Survey Methodology Bulletin*, Office for National Statistics, SMB 52 07/03, pp 4-9.

This paper describes the framework used within ONS to review data collection instruments, mainly paper self-completion questionnaires.

Marcelo C and Zilhão M J (1999) *Study on the quality of statistical operations that use face to face interviewing (Estudo sobre a Qualidade no âmbito das Operações de Recolha Directa)*, INE, Portugal.

This paper describes a study conducted with the aim of identifying non-quality issues related to the statistical operations that use face-to-face interviewing as a data collection modality. The final diagnosis clearly identifies areas of improvement and also recommends some proposal for quality improvements.

Mudryk W, Burgess M J and Xiao P (2000) *A quality control approach to CATI operations in Canada*, Statistics Canada.

This paper describes the application of quality control methods to CATI in Statistics Canada. The CATI process was analysed to establish desirable and undesirable quality characteristics associated with the interviewing aspects, and so to identify what to monitor and measure. Numerous interviewer behaviour quality characteristics with established definitions are recorded on a 'quality control monitoring form' by a quality monitor. Quality control monitoring forms input to a quality control feedback system that automatically generates statistical process control (SPC) charts. Each interviewer is considered as an individual process, and receives feedback at this level.

Office of Management and Budget, U.S.A. (2001) *Measuring and Reporting Sources of error in Surveys - Statistical Policy, Working Paper 31*, Statistical Policy Office, Office of information and Regulatory Affairs, Office of Management and Budget, U.S.A., July 2001.

This report takes the approach of studying "data quality" in terms of the measurement and reporting of various sources that affect data quality: sampling error, non response error, coverage error, measurement error, and processing error.

Saebo H V, Byfuglien J and Johannessen R (2001) *Quality Issues in Statistics Norway*, presented at the International Conference on Quality in Official Statistics, Stockholm, May 14-15, 2001.

Statistics Norway has adopted a systematic approach to quality, which is described in this paper. The authors describe how the production cycle is split into three main processes, one of which is data collection. Each process is examined in the context of the Consumer Price Index (CPI), where data are collected by paper questionnaires. Some process variables listed are the rates of missing price observations, the number of out of range items, and the number of inconsistent price observations.

During process mapping, the non-response rate was identified as a critical process variable. There is a need to analyse the distribution of non-response in a more systematic way. Such an analysis would enable a concentration of efforts on reducing non-response in areas with the largest impact.

## 3.3.2 Accessing Administrative Data

### 3.3.2.1 Accessing Administrative Data (INE-Pt)

#### Description of the process

Eurostat (2002) has defined administrative data sources as

*“sources containing information which are not primarily collected for statistical purpose”.*

Administrative data sources include: health registers, social security data, revenue authority data collected from tax forms; customs office data on imports and exports. Administrative data are primarily collected and used to inform policy, planning and resource allocation decisions.

#### Summary of the report

The importance of accessing administrative data is widely recognised:

- to increase the accuracy of information;
- to update statistical information;
- to reduce statistical burden;
- to reduce costs of statistical production.

The example given below is the result of a study conducted by INE-Portugal with the aim of improving the quality of the process of accessing administrative data, and dealing with administrative sources.

Finding key process variables for this process is not an easy task. However, a study to identify the issues that are important to address and to improve when accessing administrative data is described in detail in this example. Report 3.3.2.1 describes how INE-Portugal identified critical characteristics and improvement actions for the process, and summarises the findings.

Although few references for this topic were found, those identified are listed in sub-section 3.3.2.2.

## Identifying key process variables

### Identify critical product characteristics

With the aim of improving the process of accessing administrative data, INE-PT conducted an internal study during 2002. This started with a diagnosis of the “state of the art” of accessing administrative data, as well as future needs.

An internal request to all Statistical Production Departments asked for information through two questionnaires, shown in figures 16 and 17 below:

**Figure 16:** Questionnaire 1\_Inventory / situation in 2002

Identification of the Statistical Production Department:	
A	<ul style="list-style-type: none"> <li>• Identification of the Public Institution responsible for the Administrative Data (Administrative Source): _____</li> <li>• Identification of Administrative Data: _____</li> <li>• Statistical survey(s) using Administrative Data: _____</li> </ul>
B	<ul style="list-style-type: none"> <li>• Regularity in providing Information to NSI: _____</li> <li>• Type of format in providing information to NSI: (Diskette/CDROM/e-mail/Paper/Other) _____</li> <li>• Relationship between NSI and Administrative Source: Protocols / Other (describe) _____</li> </ul>
C	<ul style="list-style-type: none"> <li>• Strengths and weaknesses (synthetic evaluation regarding the following aspects: meeting data transmission deadlines; lag between data transmission and the reference period of data; quality of the information and quality of the relationship between the NSI and the Administrative Source) _____</li> </ul>

**Figure 17:** Questionnaire 2\_Evaluation of future needs of Administrative Data

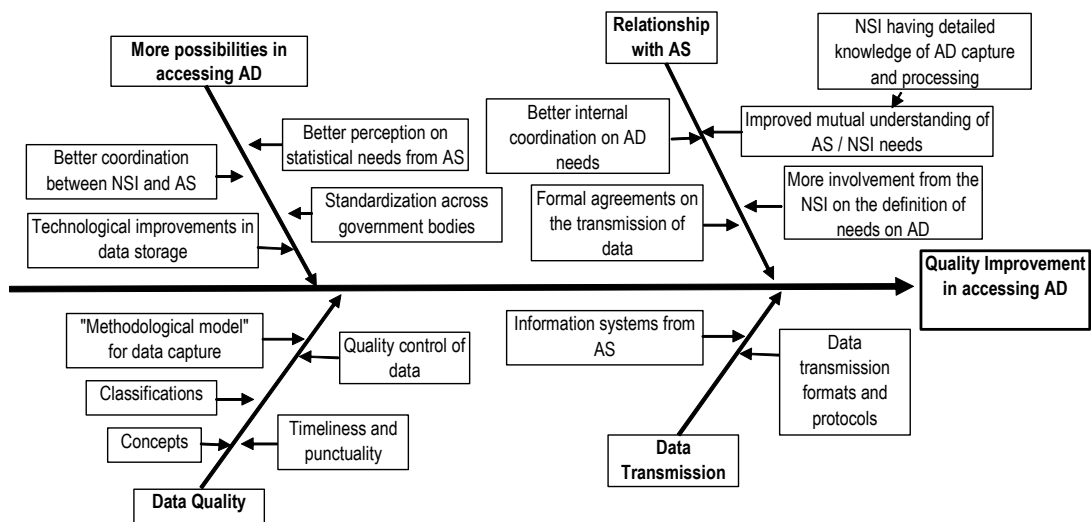
Identification of the Statistical Production Department:	
<ul style="list-style-type: none"> <li>• Identification of Administrative Data: _____</li> <li>• Identification of the Public Institution responsible for the Administrative Data (Administrative Source): _____</li> <li>• Statistical survey(s) using Administrative Data: _____</li> <li>• Regularity in providing information to NSI: _____</li> </ul>	
<ul style="list-style-type: none"> <li>• Observations (synthetic evaluation regarding the following aspects: procedures that were already implemented, major resistances to their implementation, and advantages for the statistical survey and to the NSI) _____</li> </ul>	

This diagnosis allowed the identification of all administrative sources used by INE-PT, and the most important limitations in accessing administrative data. The major conclusions of the diagnosis are limited to the fundamental aspects of the procedures of data transmission, as well as to the institutional involvement in this process.

The results also enabled us to draw cause-and-effect diagram 10 below, which shows the most critical aspects of this process.

**Cause and Effect Diagram 10:**

**Cause-effect Diagram - Quality Improvement in accessing to Administrative Data**



Based on the results of the questionnaires and on the cause-and-effect diagram, it was possible to identify the most important aspects for improvement in the process. The conclusions led to an action plan on the process, described below.

## ***Evaluation***

Our general conclusion is that the process of accessing administrative data should be promoted and developed.

Some important improvement actions were found for this process, listed below:

- To develop an accurately updated data base of the administrative data used in the statistical production;
- To co-ordinate the requests of administrative data and their regularity, when preparing the annual Plan for Statistical Production;
- To formalise, where possible, the relationship between the NSI and the administrative source, by protocols or contracts. This is essential to overcome some of the problems detected, and is an important instrument to improve the process, namely:
  - To ensure direct and regular access to administrative sources;
  - To facilitate the necessary methodological changes to improve administrative data;
  - To implement procedures to turn administrative data into statistical information;
  - To formalise data transmission deadlines, adjusted to statistical needs;
  - To implement procedures for solving and following up any constraints that might occur;
  - To enable the involvement of the NSI with the administrative sources in the conception of administrative data.



### 3.3.2.2 Literature Review

Ardal S and Ennis S (2001) *Data Detectives: Uncovering Systematic Errors in Administrative Databases*, from Proceedings of Statistics Canada Symposium 2001.

This paper focuses on health information and highlights the dangers of assuming that administrative data sources provide a sound basis on which to make policy decisions. The paper highlights stages within the data collection–policy cycle where errors may occur. A simple flow diagram is presented which highlights some of the errors which may occur in the cycle, including:

- event – may not be reported
- record – poor documentation
- transcription – inconsistent entry protocols employed
- file – structure inadequate to properly describe events
- master file – exclusion criteria
- derived files – recoding errors
- analysis – tabulations can compound errors
- policy – can introduce systematic reporting bias

Estimates of these systematic errors are important process variables to monitor.

Eurostat (2001) *Business Register Recommendations Manual. Chapter 10 Quality Policy. Second Draft.*

The paper focuses on business registers and highlights quality measures specific to the provision of such registers. The measures are not always applicable to administrative sources in general. The paper then discusses checks on the register themselves to ascertain their accuracy in representing the real life situation. These checks include control surveys to confirm the accuracy of classifications, and coverage checks of the register.

Eurostat (2002) *Business Register Recommendations Manual. Chapter 20 The Use of Administrative Sources, Fifth Draft.*

Harala R (1997) *Statistical Properties and Quality of Register Based Census Statistics in Finland*, Symposium 1997, New Directions in Surveys and Censuses, Ottawa, November 5-7, 1997

This paper describes Statistics Finland's advanced use of administrative registers in Official Statistics, and provides an overview of how administrative data are used in compiling statistics. The accuracy of the annual register based employment statistics is monitored through comparison with the results of the annual labour force survey (LFS). Results show that the register-based statistics are a good representation of the reality measured by the LFS.

Hoffmann E (1995) We Must Use Administrative Data for Official Statistics – But How Should We Use Them, *Statistical Journal of the United Nations* ECE 12, pp.41-48.

This paper provides a general perspective on issues of concern when using administrative data to substitute, supplement or support Official Statistics. Hoffmann emphasises the need for official statisticians to:

- get to know the data generating processes of administrative sources in detail;
- monitor and improve the data collection process;
- persuade the responsible agencies to make changes which lead to improvements in data quality and;
- calibrate observations generated by the administrative registrations.

### **3.3.3 Data Processing**

This chapter includes examples relating to data processing. The examples come under SVC Group 7: Editing and validation, derivation and coding. There are three reports on data editing and validation. We look at two instances where the quality of data validation on surveys in the UK has been improved using continuous quality improvement methods. There is a report on process variables relating to the data editing on the Greek Industrial Structure Survey. Finally there is a report on improving the quality of data editing in Sweden, incorporating an example from Canada. The data processing chapter closes with a description of the methods of quality control used when coding the 'occupation' question in the 2001 UK Census.

#### **Summary of the three Editing and Validation reports: 3.3.3.1, 3.3.3.2 and 3.3.3.3**

Data editing and validation is the process of detecting and adjusting errors resulting from data collection. The three reports below describe examples where continuous quality improvement methods have been used to improve the quality of the editing and validation process. The third report also includes a theoretical discussion of the continuous improvement approach to data editing.

The key message from these reports is that continuous quality improvement for the editing and validation process is best implemented by regular checking and improving of associated systems.

Quality can also be improved by redirecting resources to the most important survey units, by developing best practice for major processes and by monitoring the performance of staff involved in editing and validation.

### **3.3.3.1 Editing and Validation (NSSG)**

#### **Description of the process**

This report looks at editing and validation. Editing is a process aiming at improving the quality of the statistical information by detecting non-sampling errors in the survey data. International research indicates that in a typical statistical survey, the editing may consume up to 40% of all costs.

Our discussion for demonstrating the concepts will be based on the Industrial Structural Survey (ISS).

#### **Summary of the report**

The ISS is being conducted on an annual basis. NSSG collects data from about 5344 manufacturing enterprises all over Greece. The objective of the survey is to provide a comprehensive picture of the evolution, structure, and development of the industrial sector in terms of employment, sales, stocks, wages, assets and liabilities.

The ISS is carried out by a postal inquiry. Despite careful planning, many errors are found. The editing process has two main phases: the manual phase (field editing) and the automated phase. Manual editing is performed by industry experts and it is necessary in order to safeguard correctness of the reported figures. Automated editing follows the manual editing.

The manual editing phase takes about 8 months and has 30 individuals working on it. All collected questionnaires are carefully read and checked for consistency and accuracy. Once the data are converted to machine-readable form and the records have been machine edited, the data are ready for tabulation. Additional controls are applied to aggregated tabulated data.

Since editing is connected with other activities, such as data capture and coding, the description of the process includes these activities as well.

#### **Identifying key process variables**

##### *Develop a process flow map*

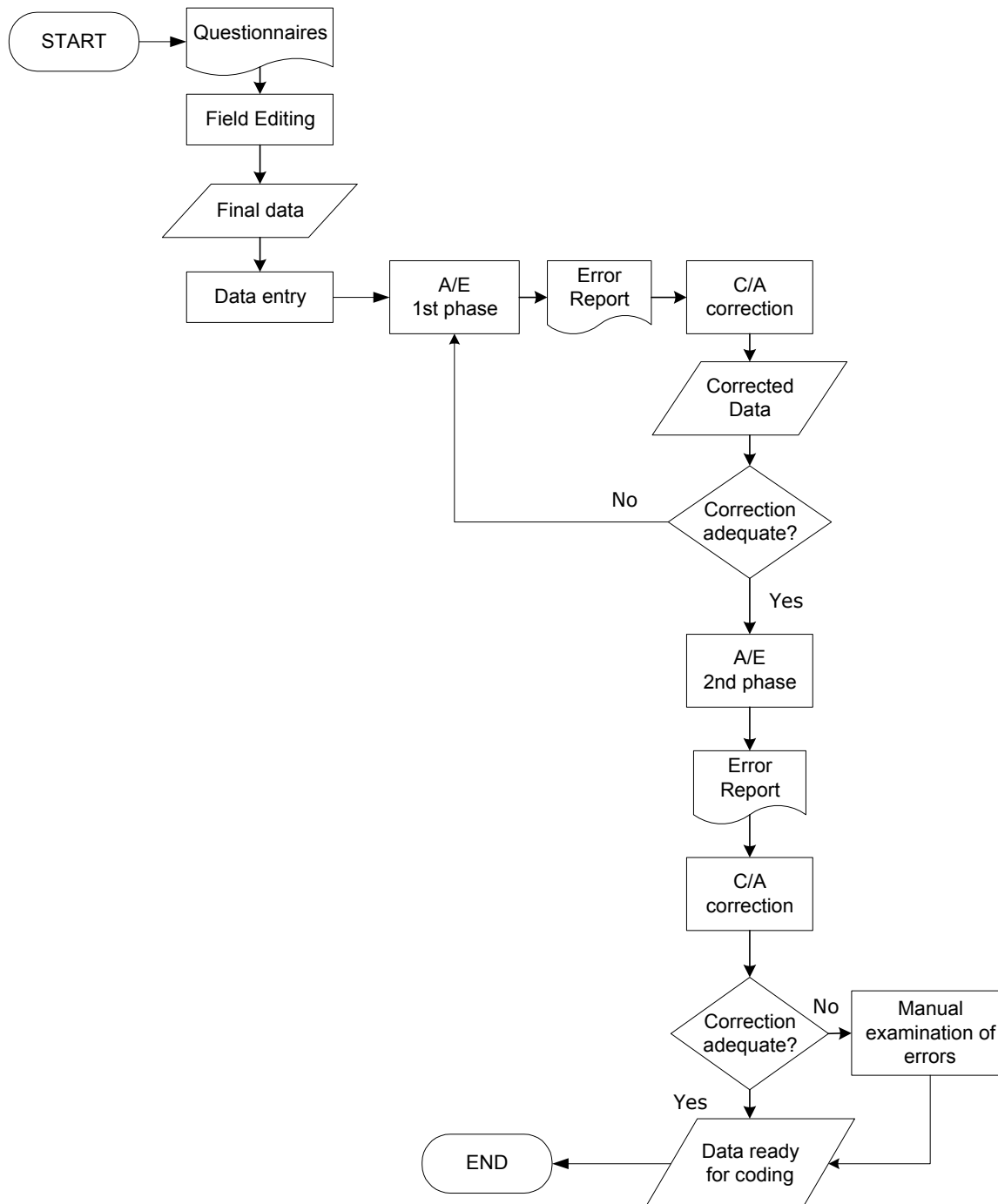
The following flow map (figure 18) examines the process of data editing.

The questionnaires are the input for the “Field Editing” sub-process. The output from this phase is data ready for data entry. “Field Editing” requires subject matter analysts, who review all questionnaires for consistency and accuracy before the data entry phase. In case ambiguities arise, enterprises are contacted by phone for clarification. In general, manual examination for surveys of establishments is more dependent on subject matter analysts than surveys of individuals and households.

Once the data entry phase has been completed, the automatic editing (A/E 1<sup>st</sup> phase-Automatic error recognition) will follow and an error report will be produced. Following this, a computer-assisted (C/A) correction is performed and if the correction is adequate the second editing phase commences. Otherwise, this operation is repeated until a predefined accuracy is reached. The second editing phase (A/E, 2<sup>nd</sup> phase) involves checks of aggregates. Similarly a correction phase will follow the error report.

If the output from this phase is considered unacceptable, it will be sent to analysts for the necessary corrections, otherwise the process ends. The final output is the corrected data set that will be used in the coding process.

**Figure 18:** Flow Chart of the Editing flow, NSSG



### *Determine key process variables*

Process variables are those factors that greatly affect the survey's **product characteristics** (e.g. *the accuracy, the budget or the time schedules*). It is possible to identify such variables by close inspection of the flow map. There are four sub-processes identified in figure 18:

1. *Field Editing,*
2. *Automatic Editing (1<sup>st</sup> & 2<sup>nd</sup> phase),*
3. *Computer-assisted correction,*
4. *Manual examination of errors,*

The process variables stemming from the sub-processes are:

1. *Field Editing,*
  - i. Time spent in manual examination of questionnaires for each group of economic activity,
  - ii. Number of analysts working in this process,
  - iii. Years (or months) of experience of the survey analyst in the specific survey, in the specific group of economic activity etc.
2. *Automatic Editing (1<sup>st</sup> & 2<sup>nd</sup> phase)*
  - i. The runtime of automatic editing adjusted by the size of the sample,
  - ii. Percentage of errors detected.
3. *Computer-assisted correction*
  - i. Percentage of errors corrected,
  - ii. Percentage of new errors.
4. *Manual examination of errors*
  - i. Reference material available to the analyst, i.e. collection of exogenous relevant information necessary for examining the accuracy of data,
  - ii. Timeliness of the external information,
  - iii. Years (or months) of experience of the survey analyst in the specific survey.

## ***Evaluation***

Time effort for editing should be calculated separately for each sub-process and for each activity sector. This will give detailed information for the allocation of resources. Note that the time spent in computations can be more easily determined than the human work.

It is quite helpful to have staff engaged many times in the same survey. Any information that will help the analysts during editing should be made available on time. Delays induced by late incoming of information will increase the total time spent on this process.

Since the effectiveness of edits should be continuously evaluated, the number of errors detected in each phase and the corrections made should be recorded. This periodic observation of errors is the best way to monitor the performance of editing and improve it in the future.

### **3.3.3.2 Validation (ONS)**

#### **Introduction**

This report looks at data validation. This is part of SVC group 7 – editing and validation, derivation and coding. Data validation is the method of detecting errors resulting from data collection. Validation tests or ‘gates’ are upper and lower limits used to test whether the incoming data are plausible. These gates are set based on information from previous cycles of the survey or from other surveys.

#### **Summary of the report**

We have identified two separate examples of testing validation in the ONS. The first of these involves reviewing the settings of the validation gates. The other is a range of quality checks built into the ONS results system to check the work of staff looking at data that fails validation.

#### **Examples from ONS of improving the quality of validation**

##### Reviewing validation gates

##### ***Identifying key process variables***

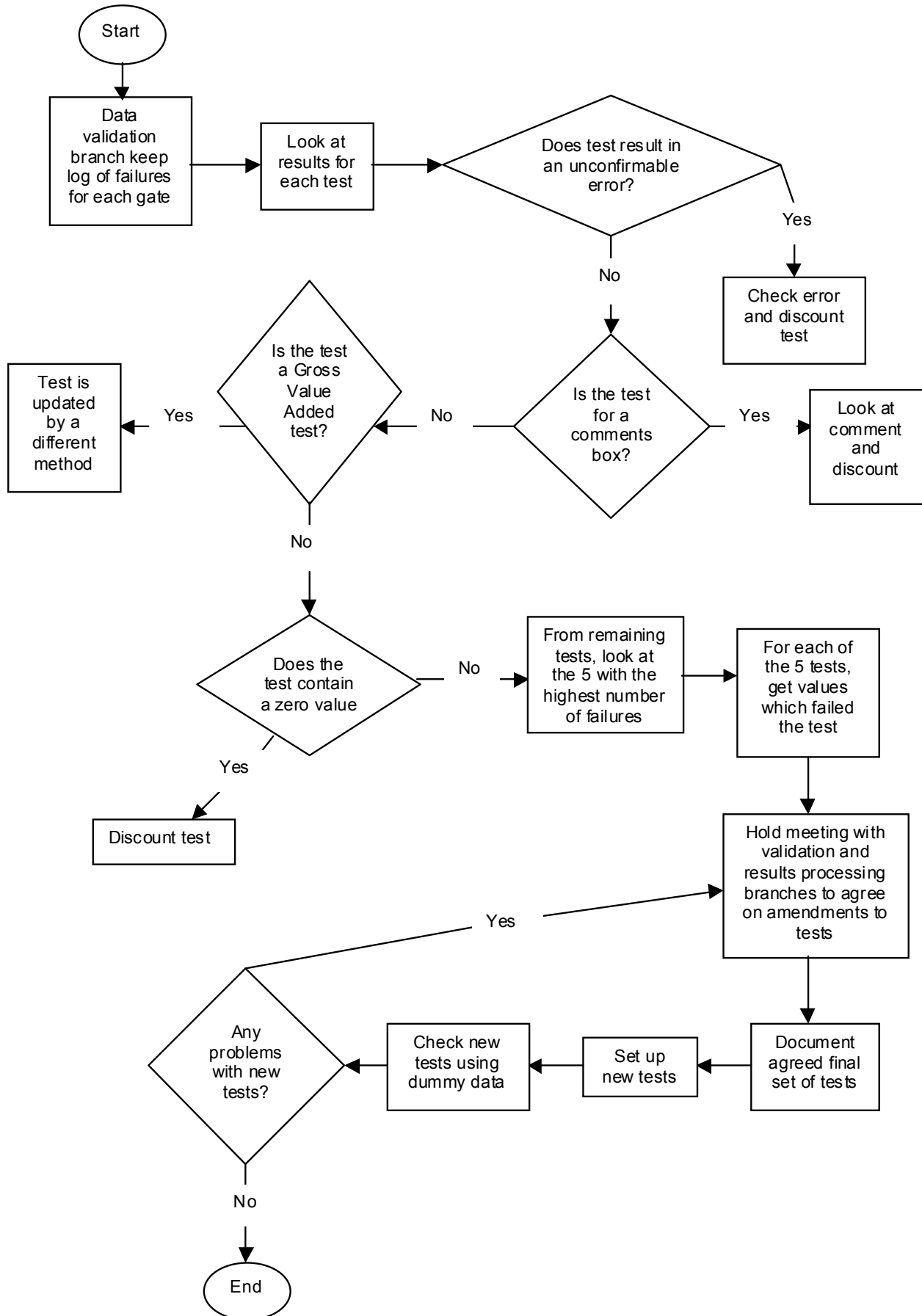
Validation tests and gates are used to identify errors and discrepancies in returned data. The gates are reviewed annually to see whether they are set at appropriate levels. If the gates are set too narrowly, then too little data passes through and a large amount of correct returns will fail the gates. On the other hand, if the gates are set too widely, then too much data passes through and there is a danger that some errors will not be picked up. In either of these cases the gates will need to be adjusted to more effective levels. The key process variable here is the number of failures for each gate, where a failure refers to data classified as an error by the gate.

##### ***Measuring key process variables***

The data validation branch keeps a log of the data failing each test. The gates with the highest numbers of failures are examined individually to check that they are identifying errors correctly. Some of the tests are excluded from this analysis, for example if the test relates to a comments box on the questionnaire. Once these have been discounted, the tests with the highest numbers of failures are examined individually. Usually around five tests are considered. The data validation branch and the results processing branch meet to decide whether any of these tests need amending. Any amendments are tested with the data before being implemented. The whole process is outlined on a process list. The process map shown below (figure 19) illustrates the steps involved.



**Figure 19:** Flow Chart of data validation using validation gates within the ONS



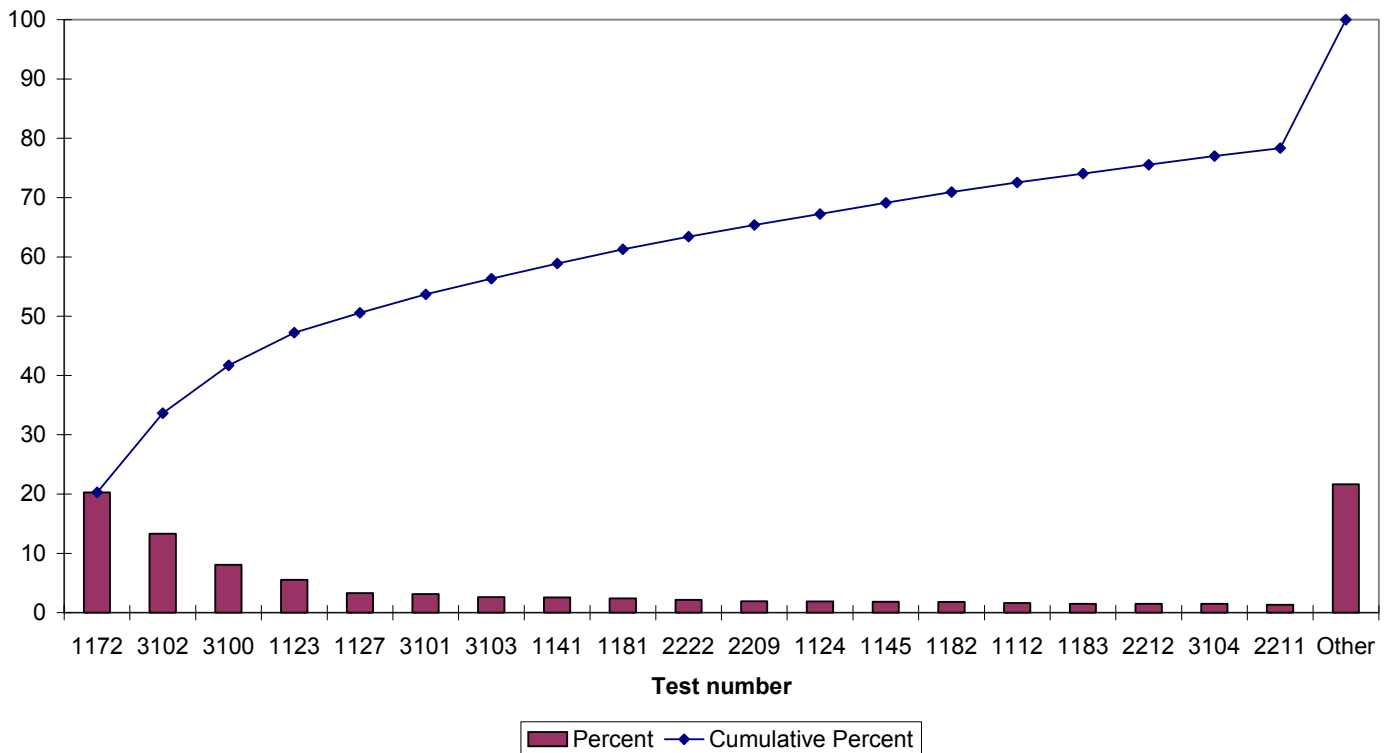
### Analysing key process variables

We have constructed a Pareto chart (figure 20) using data from the 2002 Annual Business Inquiry. There are 78 validation tests for this survey. The chart shows the tests with the highest numbers of failures.

The 'Other' category encompasses 59 tests. 78% of the failures are found in the 19 tests shown individually in the chart. The 5 highest tests account for just over 50% of all failures.

Figure 20:

Pareto chart for validation gate failures



### Evaluation

In this example we have seen how the quality of validation gates is maintained through regular checking. This is done efficiently by concentrating on the tests with the greatest numbers of failures. The use of a Pareto chart could further improve this process.

## Quality checks in Common Software

### ***Identifying key process variables***

This example looks at quality checks set up in the ONS results system, Common Software. There are seven quality checks in total. The checks allow line managers to quality assure the work of staff who check the data that has failed validation. The key process variable is staff performance.

### ***Measuring key process variables***

The manager usually looks at 1 in 10 of the data records that their staff have examined.

Five of the checks are used for monthly surveys. These look at errors resulting from validation checks that have been confirmed as correct, question values which have been changed, documentation of zero values, checks on out of scope units, and treatment of outliers and suspect values.

There is one check used for the Annual Business Inquiry. This checks that abnormal movements in year on year data have been properly examined and documented.

The final check is for the quarterly Capital Expenditure survey. This looks at the 50 contributors with the highest returned values and whether they have been correctly checked as being appropriate to the size of the company.

### ***Analysing key process variables***

The manager produces monthly highlight reports on the performance of each of their staff, and these are used to highlight training needs and thus seek to improve quality.

The information from these checks could be used to create control charts or Pareto charts, but this is not currently done. We have not been able to construct example charts, since the data required is confidential.

### ***Evaluation***

This example looks at quality checks that are used to monitor the performance of staff who check data failing validation. There is scope to use the information from the quality checks to improve overall quality by using control charts or Pareto charts.

### **3.3.3.3 Continuous Quality Improvement of surveys from an editing perspective (SCB)**

#### **Introduction**

This report has a more general feel than others. It is focused on the Continuous Improvement Approach (CIA) to editing and the underlying editing paradigm that advocates identifying and collecting data on errors, problem areas, and error causes to provide a basis for continuous improvement of the whole survey. The reason is that all readers of these guidelines may not be familiar with this approach to editing. Editing and different settings of editing and imputation are described, followed by a discussion of the aims of editing that lead to the CIA Approach.

#### **Summary of the report**

The CIA approach is first discussed in rather general terms, with references containing details provided. Then an example is given of how the approach has been implemented in practice. This focuses on increasing the effectiveness of the editing procedure.

#### **CQI from an editing perspective**

##### *What is Editing?*

In every survey there are errors in data that may distort estimates, complicate further processing or decrease user confidence in the data. *Data editing* is the procedure for detecting and adjusting individual errors resulting from data collection and data capture. The checks for identifying missing, erroneous or suspicious values are called *edit rules* or simply *edits*. The term for adjustment procedures is *imputation* (Granquist 1995). Timeliness and accessibility can also be improved through adoption of more efficient processing procedures. A commonly accepted definition of editing and imputation is the United Nations (2000) definition:

*“an editing procedure is the process of detecting and handling errors in data, including the definition of a consistent system of requirements, their verification on given data, and elimination or substitution of data which is in contradiction with the defined requirements”.*

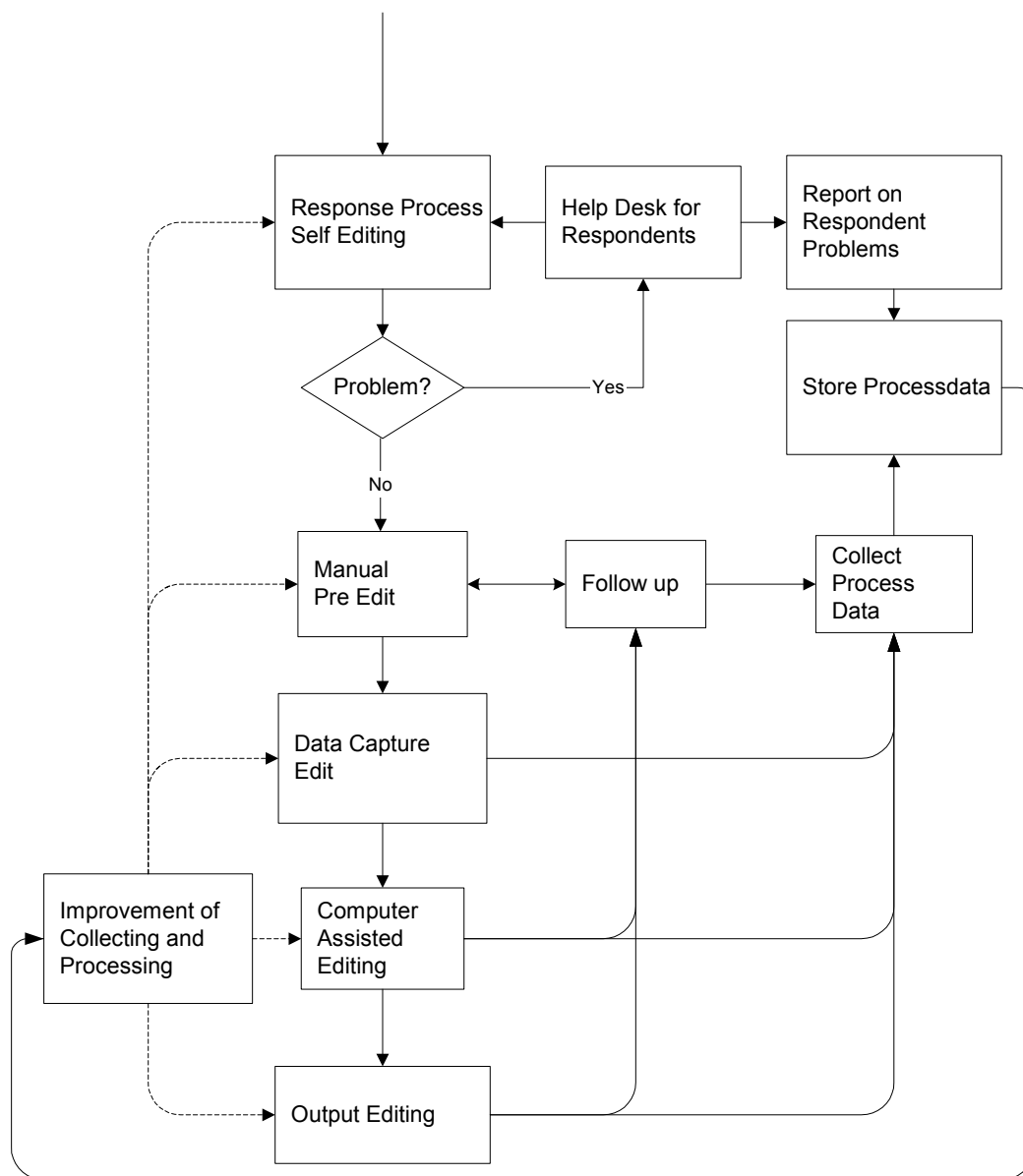
Editing can be carried out in many phases of the survey process in a number of settings comprising automatic and manual elements in various combinations, ranging from fully automatic editing and imputation to manual editing only. Furthermore, respondents may be involved in various ways, for example when suspicious data points are followed up.

The most common setting is: The computer identifies erroneous or suspicious data by means of a large number of edits provided by subject-matter specialists; the flagged records are manually reviewed, very often by follow-ups with respondents (see Granquist and Kovar 1997 for details). When an automatic editing and imputation system is used, for example based on the Fellegi-Holt method, edits are applied to data records during data capture or a machine-editing program. A score is given to each record depending of the magnitude of the record and the edits failed. All records exceeding a certain score are manually followed up and the others automatically imputed.

The editing approach can be illustrated in different ways. Figure 21 below shows a flow chart of the editing process for computer assisted editing of mail collections.

**Figure 21**

**Flow chart of computer assisted editing of mail collections**



### *Aims of Editing*

It is commonly accepted that the aim of editing should be: to provide information about the quality of the data, to provide the basis for the (future) improvement of the survey vehicle, and to tidy up the data. Because of this, the continuous improvement approach to editing and imputation was recognised early on. However, full realisations of the approach are still in their infancy, and experiences of a full application are not available.

Another more limited aim is to clean data files of records that are definitely errors and substitute those erroneous data with plausible data using data from approved records. That is the aim for completely automatic methods, e.g. those developed and evaluated by the now finished Euredit project, a research project funded by Eurostat.

### *The Continuous Improvement Approach (CIA) to Editing*

An outline of the approach is given in Granquist and Engström (1999). Key elements are collecting data on error sources, a process data system, developing Current Best Methods or guidelines for editing processes with manual follow-ups, and developing edits for detection of error sources.

This report presents extracts of the above mentioned paper relating to the continuous improvement approach, pointing out what has been done so far concerning the key elements. In the second chapter an example will be given that illustrates how useful CIA might be.

It should be noted that the principles have been completely applied in the Swedish Current Best Methods for editing (Granquist et al. (2002)) that is intended to serve as a guideline for developing editing systems for business surveys.

The Continuous Improvement Approach is a direct consequence of the editing paradigm that emphasises identifying and eliminating error sources ahead of cleaning up data. Related issues discussed are collection of data on error causes, the need for and the requirements of a high qualitative Process Data Subsystem, and standardisation of the editing process by developing and implementing Current Best Methods.

#### *Why we need a CIA to editing*

Practically all published studies of traditional editing processes indicate that the hit-rate (the proportion of flags that result in changes) is low. Many reported values are being changed by insignificant amounts, and just a few errors are responsible for the majority of the total change.

The studies present data such as: 10 to 15 percent of the changes contribute to more than 90 percent of the total change; 5 to 10 percent of the changes bring the estimate within 1 percent of the final estimate. The hit-rate lies between 20-30 percent in the few studies where hit-rates are estimated. These facts suggest two things. Firstly, the entire set of edits should be designed to identify errors more efficiently, and secondly many errors could be left unattended or subject to automatic treatments. Many statistical agencies are aware of these problems and devote considerable efforts to raising the productivity of editing systems.

During the last decade a number of selective editing methods have been developed. These methods can decrease the number of unnecessary flags and order the errors (or the suspect data) with respect to their (potential) impact on estimates either prior to or during survey processing, without having examined all the cases. Selective editing includes any approach which focuses the editor's attention on only a subset of the potentially erroneous micro-data items that would be identified by traditional editing methods. It is empirically shown that selective editing methods put together in a system, where rework and recontacts to respondents are minimised can increase productivity by 50 percent and more (Granquist and Kovar (1997)). But to improve quality we have to go further!

There is ongoing research on refining these kinds of edits and on new types of edits and editing procedures. Furthermore, interactive editing procedures at the data entry stage or when data are collected and captured have proven to be good ways of rationalising the editing of survey data.

The key objective of the new paradigm is that quality should be built into the processes to prevent errors, rather than identify errors once they have occurred and replace them with more accurate data. A successful way of doing this is to apply the concept of continuous

quality improvement to the whole survey process, where editing is but one process (Linacre (1991)). Note that editing under this paradigm is a key process, in that it will furnish data on errors as a basis for measures to eliminate root causes of errors from the survey.

### *Inlier edits*

Query edits are usually outlier checks. However, they cannot identify data that are affected by small but systematic errors reported consistently by a number of respondents in repeated surveys. Such errors are termed inliers and are defined as faulty data which lie within any reasonable bounds of ordinarily used edits. Inlier methods are probably of greater importance for quality than outlier methods, irrespective of how efficient they are in detecting erroneous outliers. Inliers occur whenever there are discrepancies between the survey definitions and the definitions used in the firms' accounting systems. Werking et al. (1988) present an illustrative example. In an ongoing Response Analysis Survey (RAS) designed to focus on differences in definitions and how to get firms to apply the survey's definitions, they found that the estimate of the main item "production worker earnings" for the RAS units became 10.7 (with a standard error of 3.2) in contrast with 1.6 for the control group. One method to cope with that type of inlier is to add some questions into the questionnaire, asking the respondent whether he or she included/excluded certain item components in the answer.

Research on inlier methods is fairly new. Winkler (1997) presents a description of the problem and suggests a number of methods for converting inliers to outliers using additional information that may be available in the files being edited.

Exploratory Data Analysis (EDA) methods using SAS/INSIGHT or JMP are probably the best way to identify the presence of inlier problems as they are focused on discerning patterns in data. The root causes to an inlier error have to be discovered and then adequate changes in the survey have to be made. The best source for finding root causes is discussed in the following paragraph.

### *Collecting data on error sources*

The paradigm imposes a new and probably rather heavy and difficult task to the editors. They have not only to verify flagged data and find acceptable values, but also they have to identify and register quality indications of the new data, error causes, respondent problems, and possible problem areas of the survey. It will require deep subject matter knowledge of and insight into the survey design. Furthermore, editors have to understand that this task is substantially more important in recontacting respondents than verifying suspicious data. This contrasts to the common comprehension that flagged data should be changed to pass the edits (creative editing). To build quality into the survey also means that recontacting respondents includes educating the respondents in answering the questions in continuing surveys.

Engström (1997) shows that it is feasible. This is the only published paper on this theme.

The paper presents a study from the 1995 Swedish European Structure of Earnings Survey (SESES), where data collection on error causes was integrated into the survey process. The editors had to identify and code error causes like misunderstanding, questionnaire problems, typing problems etc. Furthermore, they had to indicate whether respondents were contacted to solve flagged items. Engström found that the edits were rather efficient. The error cause data for the most erroneous item (4000 cases out of 16000) showed that 90 percent of the errors were due to respondent misunderstanding. It

was judged that most of these errors could be avoided by changing the wording of the question and improving the instructions to the respondent. However, the coding was burdensome and the editors had problems in referring the error cause to the erroneous item. Engström (1997) concludes that error cause data are extremely useful and that the system has to be carefully designed to facilitate the work of the reviewers.

### *The importance of efficient questionnaire design*

Linacre and Trewin (1989) indicate that rates for item nonresponse and form/system design errors are both about 30 percent of the errors in business surveys. They conclude that improving questionnaire design would improve the quality of incoming data. The example given by Engström (1997) emphasises that improving the questionnaires can prevent a significant number of errors and that a tight co-operation between questionnaire designers and survey managers would be extremely beneficial for the organisation. Linacre (1991) mentions that the Australian Bureau of Statistics established a forms design unit following the results of a number of evaluations of editing processes. The paper states that the quality of statistical estimates is largely influenced by the respondent's ability to understand questions unambiguously and to have relevant data available. If respondents do not have data for a particular item in their accounting systems, the strategy of collecting data for that variable has to be revised. Note that respondents are likely to deliver the data they have irrespective of any difference in definitions.

### *Process Data System (PDS)*

In an October 2003 work session in Madrid, evaluations of editing processes were an important issue, signifying the start of developing a handbook on evaluations of editing processes. At least two systems are developed: the Statistics Canada UES system that will be discussed in the example given in the second chapter, and the ISTAT IDEA system for calculating standard quality indicators on editing and imputation (Della Rocca et al (2003)).

Engström (1996) gives a rough sketch on how to monitor an editing process and outlines a number of basic indicators for studying the outcome of, in particular, the edits, that could serve as key process variables.

Nordbotten (1998) presents a Process Data Project outline for systematic collection and storing of data on editing architecture, quality and performance for individual surveys. Combined with other metadata, this provides a basis for survey design improvements. The paper has been carefully reviewed by the UN-ECE group on statistical data editing in two work sessions. Many of the statements are reproduced below.

In addition to final product quality indicators, the continuous quality improvement approach requires data on the applied editing system architecture as background data, and on the performance of the process, including interactions with respondents and others, to evaluate the process. The editing architecture data are by-products of the design of the system, while performance data have to be collected and stored during the editing. The product quality, the editing system architecture and the performance data have to be collected and stored in a well-designed PDS. Cost and timeliness constraints, particularly for short period surveys, exclude post evaluations for this purpose. The data have to be analysed and measures have to be taken to improve the current editing procedure.

A PDS has many purposes. Performance measures are needed during the editing process for monitoring and regulating the process while maintaining quality goals, and for



improving future system designs with regards to quality and performance objectives (Weir (1997)).

A PDS should give data on quality for both the user and the survey manager. The users want data about quality to evaluate whether the supplied statistics are suitable for their needs, while the survey manager need data on quality to analyse alternative production strategies.

Editing processes have to be described in a uniform way, making it possible for a statistical agency to compare the effectiveness of the editing between surveys. The top level managers need data in order to allocate their methodological resources, select surveys for revisions, and see the effect of research and development efforts.

Research is needed on designing a PDS which is more than simply a means for improving individual surveys. The PDS must become an integrated part of a general metadata system permitting research into the origins of errors, improved survey design in general, and of improved editing systems in particular.

## *Current Best Methods*

Lyberg et al. (1998) state that probably the most effective way to improve quality is to develop Current Best Methods (CBM) for its major recurring processes, to have them implemented and continuously updated as new knowledge is generated. The role of CBMs in the improvement of survey quality is discussed in detail in Morganstein and Marker (1997).

Agency manuals on editing, papers on editing strategies, and generalised software may have similar effects to CBMs in getting sound, recommended practices communicated and used within the agency. The advantage of CBMs is that they are supported by the top level management and developed by the agency's experts together with a number of carefully selected users (in Sweden statisticians responsible for editing processes) to assure that each CBM will reflect the organisation's apprehension of what are best practices.

## **An example from Canada**

The following is a shortened version of Martin and Poirier (2002). It is a good illustration of how process data combined with data from an interview survey among editors can be used as a basis for taking measures to improve data collection and the editing process. The outcome of the measures taken, and the subsequent analysis is missing to make it a perfect illustration of the Continuous Improvement Approach to editing.

Survey processing, especially components that are highly labour intensive, can be expensive. It is therefore important to find ways to gain efficiencies. Managers of Statistics Canada's largest multisector business survey have access to versions of data and to additional processing metadata that describe how the data were transformed from collection through post-imputation correction. They were able to use this information to detect inefficient or inappropriate methods and to replace them in favour of more efficient and appropriate methods. The paper presents the findings of a study involving a few years of survey data

The Unified Enterprise Survey (UES), initiated at Statistics Canada (STC) in 1997 with seven industries, now integrates just under 50 annual business surveys into one centralised survey system. Businesses of all sizes are in scope for the UES. Large firms are always in sample; the smaller businesses are randomly selected each year.

For the first few years of the UES, most industries use two questionnaires – a long questionnaire that asks for all the variables of interest to the industry and, in an attempt to ease response burden, a shorter questionnaire directed to smaller firms. The intention was that the details for these smaller firms could be derived from their tax data.

Editing and/or imputation is carried out in each of the first four phases of the survey process: (1) Data collection, (2) Post-collection review and correction, (3) Automated imputation and (4) Post imputation review and correction.

The data and metadata resulting from each of these phases are housed in a central data repository such that four versions of data with accompanying metadata (one for each of the four processes above) are available for analysis. In addition, for mail-back units, the raw captured data are available for research, but reside outside the central repository.

Only recently have Statistics Canada finally had the opportunity to assess the data available for the years 1997 to 2000. With this data, they wanted to:

- quantify the effect that manual intervention has on the data since the costs, in terms of people and time, are high;

- determine whether these costs could be decreased by directing interventions more efficiently;
- determine whether pre-specified automated edit and imputation procedures are the most appropriate or whether they are in fact leading to a need for more manual intervention.

Mail-back questionnaires are captured using a Quick Data Entry system with virtually no editing. Captured data are then batch edited and categorised according to the severity of their edit failures.

Two slightly different follow-up strategies are applied for mail-back units that fail capture edits. For the non-manufacturing sector, questionnaires categorised as having severe edit failures are flagged for follow-up. For the manufacturing sector, “critical” units categorised as having severe edit failures are flagged for follow-up. Mail-back questionnaires having only non-severe edit failures and manufacturing “non-critical” units are not flagged for follow-up.

The main concern during this phase of processing is the follow-up cost. It is estimated that a telephone follow-up takes on average 15 minutes. This estimated time does not account for the unsuccessful follow-up attempts that often precede a final contact.

The study concentrated on reference year 2000 and on units that mailed back their questionnaires. The path followed by mail-back units provides us with a fuller set of data and metadata. For these units, there are two versions of data - raw captured data and the data resulting from follow-up. In addition, there are the edit flags resulting from the batch edit and flags resulting from the follow-up process. Less information is recorded for units whose data are collected entirely by telephone.

For one specific industry, they identified the ten edits with the highest percentage of failing cases. All ten edits were severe, the category requiring follow-up. One edit was failing 27% of the time. The smallest of the ten was failing 13% of the time. Of these edits, they noted that six were query edits – not highlighting mathematical impossibilities, but rather potential unusual relationships.

Across all industries, Statistics Canada detected that:

- Less than 3% of units passed all edits;
- Units which failed any edits tended to fail severe edits;
- All units failing severe edits in the non-manufacturing industries were flagged for follow-up and all critical units failing severe edits in the manufacturing industries were flagged for follow-up;
- The rate of follow-up was lower for manufacturing industries than for non-manufacturing since manufacturing did not follow-up non-critical units;
- In some industries, 100% of units were flagged for follow-up;
- Short questionnaires were flagged for follow-up at a lower rate than long questionnaires, since they had fewer variables and therefore fewer edits.

Conclusions drawn identified a need to:

- find a way to encourage respondents to use our mail-back questionnaire in order to minimise the cost associated with telephone data collection;

- re-visit their edits, paying stricter attention to what should constitute an edit follow-up, so that those units responding by mail would not be contacted by telephone simply to have their reported data confirmed;
- find a way to prioritise individual units for follow-up in an even stricter fashion than was currently employed for the manufacturing sector.

### *Impact of Follow-up*

By comparing the raw data with the post-follow-up data, Statistics Canada split changes resulting from follow-up into 3 categories to determine what is the effect of each on the data:

- Missing data becomes available through follow-up - This brought about an increase of up to 20% for the weighted variable in its industry. On average, it produced a 2% increase.
- Value (properly captured) changes through follow-up - This brought about a change of up to 4% for the weighted variable in its industry. On average, it produced a 1% change. A high percentage of queried responses were confirmed by the respondent to be correct.
- Value (improperly captured) changes through follow-up - This brought about a change of up to 62% for the weighted variable in its industry. On average, it produced a 10% change. A very small number of changed records produced a very large actual change.

Statistics Canada concluded that they needed to direct their attention to the problem of edit failure caused by improper capture.

The primary concern during post-collection processing is the cost of review and correction conducted by the survey analysts. Statistics Canada learned early in the study that most analysts do little manual correction before automated edit and imputation, so they concentrated their efforts on the changes made after automated imputation. At this stage of processing they have the automatically imputed data together with the metadata that identify where each data value came from – reported, imputed by method-A, imputed by method-B etc., and they have the data after manual intervention, with each changed value easily identified.

Total Operating Revenue variable was manually changed over 15% of the time.

- 67% of these changes were for units that had been imputed through mass imputation;
- 24% of the changes were changes to “reported” data;
- 4% of the changes provided data for units that could not be imputed through the automated process and so still had missing data.

The Total Operating Expense variable was manually changed over 21% of the time.

- 55% of these changes were for units that had been imputed through mass imputation;
- 35% of the changes were changes to “reported” data;
- 5% provided data for units that could not be imputed through the automated process and so still had missing data.

These findings strongly indicated that Statistics Canada should be concerned with mass imputation and changes to reported data. They conducted a rather subjective survey amongst subject matter analysts to try and determine why they felt the need to change “reported” data. Two answers emerged - data that had been badly captured and the respondent had misunderstood the question.

The findings suggested several areas of concern, and a need to:

- find a substitute for tax data or improve the processes leading to the particular version of tax data so that there would be a consistent correlation between auxiliary data and survey data;
- address the issue of badly captured data, so that analysts could feel confident that “reported” data were truly reported;
- revisit the content/wording of our questionnaires, so that respondents would not misunderstand.

While 15% to 21% of these variables had been changed, when Statistics Canada looked at what would have happened had they made only the top 50 changes, these results were within a very small margin of the final estimates. Estimates resulting from the top 10 changes were also quite similar, especially in some industries.

With this information Statistics Canada felt that:

- They needed to find a way to identify fewer, large impact units that would yield the greatest improvement in the estimates.
  - For reference year 2001, for the non-manufacturing sectors, they now have only one questionnaire per industry. The length of each new questionnaire has been greatly reduced, compared to the equivalent long questionnaire of the reference years 1997 to 2000. The questions have been reworded in an attempt to avoid misinterpretation by respondents.
  - They expect to have a more successful mail-back response rate now that the respondent is faced with a much shorter questionnaire;
  - The number of collection edits has been greatly reduced as a result of the shorter questionnaire;
  - Follow-up rates should be lowered, with fewer edits involved.

For the reference year 2000, for mass imputation, Statistics Canada no longer use tax data to find a nearest neighbour. They have turned instead to using historical response for units previously in sample and the size measure that resides on the Business Register for new units in sample. The imputed values will improve, assuming we now have a tighter correlation between these auxiliary data and the reported data of the donors.

For the manufacturing industries, identifying critical versus non-critical units succeeds in reducing the follow-up rate. A technique to prioritise all UES units for follow-up must be found, such that unimportant, easily imputed units are not given the same attention as large, difficult to impute units. The method must recognise the variation across industries and geographic region – a small unit for one industry/province is not necessarily a small unit for another. Additionally, a large unit is not always difficult to impute, especially as they introduce more appropriate imputation techniques. A method that was devised for the pilot year was never fully implemented. That method and others will be evaluated.

The capture edits themselves must be revisited. Query edits in particular must be reviewed and perhaps dropped altogether. As was evident in the study, most of the queries were later confirmed by the respondents. A study is now underway in the area of Data Collection to determine how best to minimise the number and type of edits that trigger follow-up. The results should be ready to be implemented for reference year 2002.

The data capture errors must be addressed. Originally, the capture of mail-back questionnaires was done using the same system as telephone capture, with interactive editing. For seasoned keyers this proved to be a very time-consuming process and thus the process was changed. More thought needs to be given to finding the happy medium that assists keyers in finding and correcting their own keying errors, without frustrating them.

Statistics Canada need to find a way to help subject matter officers direct their attention to those units where change will have the most impact. While it was simple to find the top 50 units after the fact, it is difficult to find a way to identify these units before the change is actually made.

For post-collection imputation and manual correction, they will investigate at a more detailed level which imputation methods result in values with which analysts are dissatisfied, with the intention of adapting their methods, and if necessary their systems, to bring better results.

There continues to be at Statistics Canada a culture that insists on micro data correction beyond the point where the influence on the estimates is worthwhile. Statistics Canada feel that since they now have the capacity to obtain good information concerning the effect of corrections, they will be able to adapt to more appropriate procedures.

### **3.3.3.4 Coding (ONS)**

#### **Introduction**

This example focuses on applying our approach to process quality to coding. It may be argued that exercises examining the quality of coding are expensive to carry out and cannot be done routinely. Instead, typically studies are carried out from time to time and the results are quoted as quality indicators. But for continuous quality improvement there is a need for continuous monitoring of variables throughout the process, and not just publication of overall quality indicators.

This example, along with several of the references in the literature review in sub-section 3.3.3.5, describes such continuous monitoring, where process variables have been identified, measured, and analysed. The results of this quality improvement effort demonstrate that the coding process is amenable to the process quality approach.

The remainder of the introduction provides a definition of coding as a process, followed by an overview of the content of the report.

#### **Description of the process**

Typically, data are collected as qualitative or quantitative responses, which are coded into a specified set of classifications according to given rules. The data we are coding could be in several forms: paper questionnaire; scanned form; telephone or face-to-face interview (with an interviewer coding responses as they are given.) Coding may be automatic, usually built around a matching algorithm and parsing of words to improve match rates, or manual.

#### **Summary of the report**

The report examines a fishbone diagram, which sketches some sources of quality in the coding process. Such diagrams are proposed for use in identifying key process variables.

Following this, a real example of quality control of coding for the 2001 UK Census is described. The report will focus on the coding of the 'occupation' question. A rate reflecting the accuracy of the coding process was calculated by verifying samples of codes, and analysis of errors helped identify problem areas to address. Overall, the verification suggests continuous improvement in the sampled batches, as the accuracy rose from 87.30% in the first 10 batches checked, to 90.35% in the last 11 checked, out of 43 batches checked over nine months.

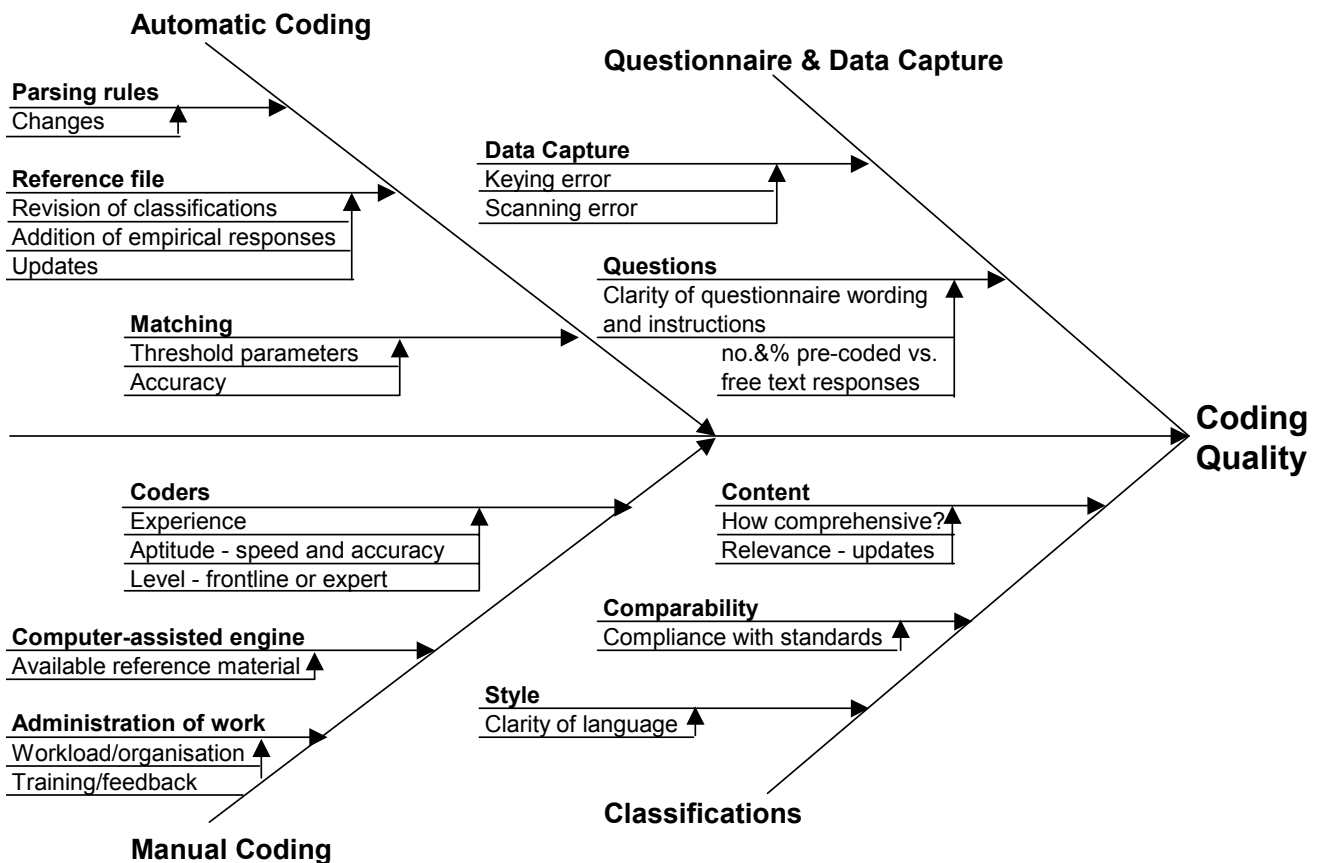
## Sources of Quality in the Coding Process

One of the main tools recommended for use in identifying key process variables is the 'fishbone' diagram, also called the 'Ishikawa' or 'cause-and-effect' diagram. Such a diagram shows the main causes and sub-causes of a certain effect. In this case the effect is the quality of the coding process, and we wish to use the map of causes to help identify possible variables to monitor.

A diagram listing four causes and associated sub-causes is provided in figure 22 below, followed by some explanation and potential variables for each cause. To be consistent with the report on 2001 UK Census below, we have assumed that:

- data are collected via a paper questionnaire;
- an automatic coding system similar to ACTR (Automated Coding by Text Recognition) developed by Statistics Canada is used;
- the automatic system is followed by computer assisted manual coding.

**Figure 22:** Cause-and-Effect diagram for coding quality



### Questionnaire & Data Capture

This part of the diagram aims to represent the effect of prior processes on the coding process. Data are collected in the form of returned questionnaires, and is entered onto a computer by a combination of scanning and keying. Therefore the input to (and output of) the coding process is affected by the quality of these processes. Some editing and validation may also occur before coding, but for simplicity these are omitted from the diagram.



Clearly worded questions will help respondents to respond accurately, and pre-coded questions allow the respondent to choose the most appropriate code and so reduce burden on the coding system. Errors made in data capture may follow through to coding, or lead to complications when coding the results. Therefore, the percentage of questions that are pre-coded, and keying and scanning errors are a few process variables that may be useful.

### *Classifications*

In general, a standard classifications manual for characteristics such as industry and occupation will be used in a given survey. A particular code is typically associated with one or more classifications, and so the classification manual in use may affect coding quality.

For example, coding may be difficult if a particular occupation is not represented in the manual. Therefore the content or coverage of the classifications manual appears on the diagram. The comparability (for example with different countries or across time) of the classification system will affect that of the codes in use. Also, if manual coders use the classifications to code then it helps if the classifications use a clear language that is easy to understand.

### *Automatic Coding*

The ACTR system follows the general procedure of first 'parsing' a response, then comparing the parsed response to a reference file, using a matching system to allocate a code or to express doubt and refer the code to manual operators.

Parsing is a sophisticated technique for text standardisation, including functions such as deletion of trivial words, definition of synonyms, suffixes removal, and character mapping. Parsing rules are pre-defined and may be updated throughout the coding process, and affect the resulting codes.

The reference file is a dictionary of parsed descriptions, each associated with a unique code. The file is constructed by revising the classification manual (to enable processing by a computerised system), and adding descriptions derived from empirical responses (given in previous surveys and coded by experts). Therefore we have identified the revisions of the classifications manual and the addition of empirical responses as sub-causes of coding quality. We could measure simple process variables such as the percentage of reference file descriptions derived from empirical responses; or the number of descriptions in the reference file (compared with the classifications manual).

When matching, either an exact match is found, or an algorithm is used to match parsed text to the most suitable partial match. The algorithm uses user-defined threshold parameters, which directly affect the accuracy of coding. The matching may lead to several outcomes: a unique 'winner' code, multiple winners, possible winners or no winners. A tally of each type of outcome provides some interesting process variables to monitor. Where the automatic system succeeds in allocating a code, this code might be wrong. A measure of the accuracy of the system is an important process variable – and is examined in detail in the 2001 UK Census example below.

## *Manual Coding*

The coding operators themselves have a big impact on the quality of the coding. Their experience and aptitude are shown as sub-causes. A simple process variable for coding is the experience of the coder (in months etc). To reflect their aptitude, process variables such as the rate of coding or the accuracy of coding could be measured (again, the accuracy rate is addressed in the 2001 UK Census example below). Typically there will be two levels of coders: frontline and expert. Codes too difficult for frontline coders are sent to experts who are thought to code more accurately. The number of expert coders and the threshold for referring a code to an expert may be important process variables.

Aspects of the computer-assisted engine used by manual coders will affect quality, for example the availability of reference material.

Finally the administration of the manual coders is important. Heavy workloads may have adverse effects on the coding quality. On the other hand, good training and feedback should improve the quality of the coders' work.

## **Report on Quality Control of Coding for the 2001 UK Census**

This report on the quality control of coding for the 2001 UK Census brings together information from several unpublished papers by both Lockheed Martin and the Office for National Statistics.

### ***Identifying key process variables***

#### *Identify critical product characteristics*

The 2001 UK Census provides key information on population counts. For any meaningful breakdown of this single 'total population' figure, it is critical to be able to classify the responses, for example by sex, age, religion, occupation etc. This is achieved by including relevant questions on the Census forms, and then coding the responses into the classification system required. Thus, effective coding is a key process in producing good quality Census statistics. The effectiveness or 'accuracy' of the coding system is therefore a critical product characteristic.

The coding system used a combination of automatic coding and manual coding. Automatic coding used inexact matching, which accepted a percentage of errors. When the automatic coding failed to assign a code, the data would be given to frontline manual coders. If the frontline coder could not assign a code, it was sent to an expert manual coder. Both frontline and expert coders utilised computer-assisted technology.

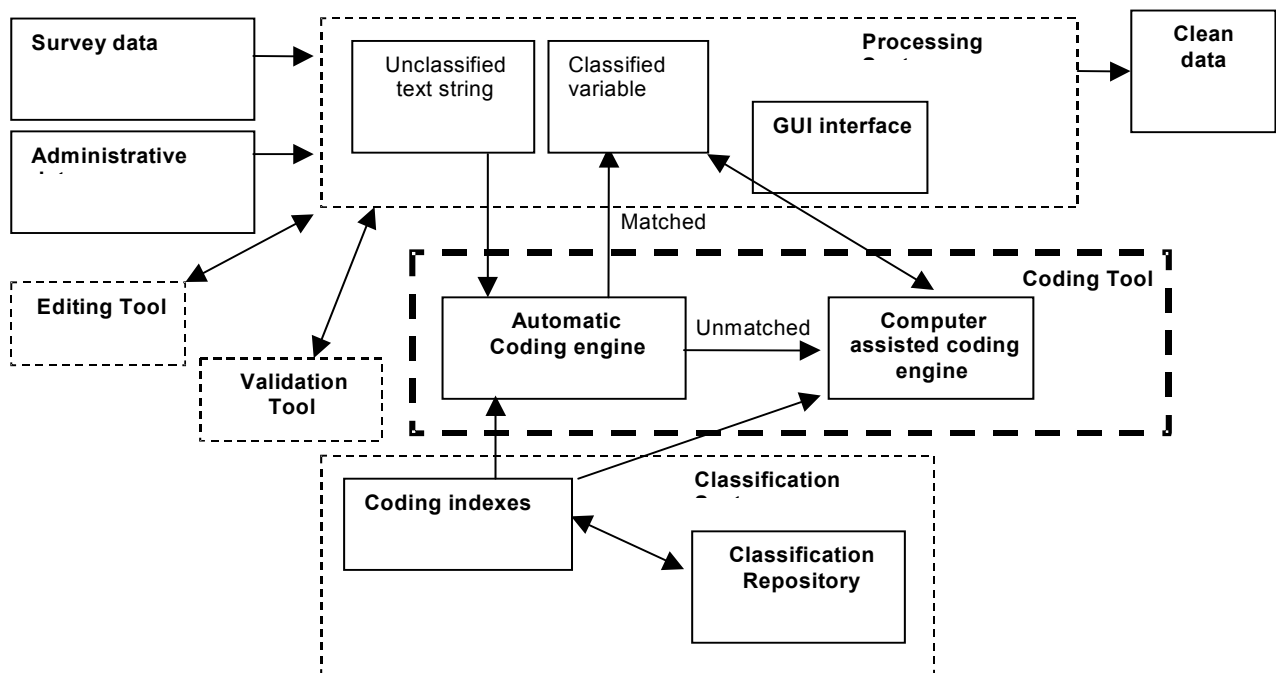
Since the coding is difficult for both automatic and manual systems, it becomes even more important to measure and monitor the effectiveness of the process.

### Develop a process flow map

Coding work for the 2001 UK Census was contracted to Lockheed Martin (LM), who undertook the automatic and subsequent manual (computer-assisted) coding. The Office for National Statistics (ONS) provided the data, as outlined below. It provided the capability to continually improve the databases by modifying and adding new data at any time.

Figure 23 below illustrates the typical relationship between a coding tool (with automatic and computer-assisted components) and the classification and processing systems. This is taken from documentation for the ONS Statistical Infrastructure Project on classification coding tools. The data are provided as input to the processing system, which contains the coding tool. In this case, the coding tool has two components: an automatic coding engine and a computer assisted coding engine. The former accepts an unclassified text string, and using coding indexes linked to the classification repository it will either find a matching code which will output as a classified variable, or fail to match a code. In the latter case the unmatched text is sent to the computer assisted coding engine, where manual coders use coding indexes to classify the variable. The classified variables then contribute to the target output of a clean dataset. Editing and validation tools are also part of processing, and are considered in reports 3.3.3.1 to 3.3.3.3.

**Figure 23:** Map of a coding tool in relation to classification and processing systems.



### The Classification System:

The data provided by ONS consisted of modifications of official indexes including the UK Standard Occupation Classification 2000 (SOC). In addition to the indexes, cleansed data from the 1991 Census and the 1997 Test responses were developed into databases. During operations, the success of matching was affected by increasing the size of the databases, and by further tuning of the systems.

### The Coding Tool:

Automatic coding was used with the aim to code a lot of data to an acceptable level, quickly, and at less cost than using expensive manual coders. The product selected for coding was a generic product for string matching, ACTR (Automated Coding by Text Recognition) developed by Statistics Canada. If the confidence level of the automatic coding was below a specified threshold, the code was sent for frontline manual coding. The aim was for these coders to code quickly, performing text edit and search, using computer-assisted technology. If they failed to find a code in good time, it was sent to an expert coder, who had more aptitude or experience, and had more time and reference material for coding.

### The Codes

The codes themselves are four digits long. The first digit level is referred to as the occupation 'major group'. There are nine major groups, given in detail in table 9 below.

**Table 9:** Table of Major Group descriptions

Major Group Number	Major Group Description
1	Managers and Senior Officials
2	Professional Occupations
3	Associate Professional/Technical
4	Admin/Secretarial
5	Skilled Trades
6	Personal Service
7	Sales and Customer Service
8	Process, Plant, Machine Operatives
9	Elementary Occupations

### *Determine key process variables*

Both LM and the ONS identified process variables to monitor for the coding process. The main process variable is an overall coding accuracy rate, measured by sampling and verification. After taking a sample of codes for verification, an estimate of the rate is obtained by dividing the number of correct codes by the total number verified.

LM outlined how data would be sampled and checked to calculate accuracy rates. LM and the ONS agreed Service Level Agreements (SLAs) for the accuracy rates: for occupation they were 88%. LM checks could show that coding was generally consistent, however it was possible that LM could be consistently coding incorrectly. Therefore, the ONS devised additional checks for systematic error (the non-random distribution of error across some relevant categorisation of the results). Errors that are not evenly distributed may bias the results, and so should be addressed.

From now on, to avoid confusion, the LM variable 1 will be referred to as the consistency rate, and the ONS variable 4 as accuracy or error rate. These and other process variables identified and measured by LM or ONS are listed in table 10 below. This report will focus on variables 1, 4, and 6.

## **Measuring key process variables**

The remainder of the report will refer to Estimation Areas (EAs) and Census Districts (CDs) of the UK, used in the 2001 Census. There are 112 EAs, usually containing around 500,000 people, although some larger EAs have up to 900,000 people. Each EA contains several CDs.

As described above, occupation coding underwent two main quality checks to measure errors being introduced that could affect the integrity of data. The first was carried out by LM (variable 1), and the second by the ONS (variables 4 and 6). The first two unnumbered sections below describe how these variables were measured, the last covers some additional issues in measurement.

**Table 10:** Table of process variables for the coding process

No.	Process Variable	Measured by	Details
1	Accuracy/Consistency (or conversely error) rate	LM	Sampling and independent verification of codes
2	Accept rate	LM	Rate of coding records, regardless of correctness
3	Number and percentage coded by mode	LM	The numbers and percentages coded automatically, frontline manually and expertly
4	Accuracy (or conversely error) rate	ONS	Sampling and verification: analysed overall and at major group level
5	Rate of incorrectly assigned 'uncodeables'	ONS	ONS verification of codes deemed 'uncodeable' by LM
6	Frequencies of types of error in coding	ONS	For the codes in error, details of the correct code and the incorrect code assigned allow detailed analysis of systematic errors

### *Lockheed Martin Quality Checks, variable 1*

LM undertook approximately two per cent sampling from each of the 112 EAs, giving a total of 748,385 codes checked. ONS found that on balance the LM proposed samples were acceptable, ensuring measurement capability. Where possible, sampling errors for estimates derived from the samples are included in this report.

The verification method agreed for LM is called 'the two of three rule', with a verification operator and possibly an arbitration operator blindly coding sampled descriptions. In the terminology of 2.3.4, this is independent inspection. For both automatic and manual coding, an Occupation code was selected at random at a rate of two per cent. The original code and source code for the data was stored. The code was marked as unknown and the source description was sent through correction steps, as illustrated in table 11 below in simple terms.

**Table 11:** Details of the ‘two of three’ verification method used by LM

Autocoding	Coded value			
	A	A	A	A
Frontline Coder 1	A	B	B	B
Frontline Coder 2		A	B	C
Coded value marked	Correct	Correct	Wrong	Wrong

Frontline Coding	Coded value			
	A	A	A	A
Frontline Coder 1	A	B	B	B
Expert Coder		A	B	C
Coded value marked	Correct	Correct	Wrong	Wrong

LM then derived consistency levels from the errors found during code verification. As we will see later, results show that consistency is a good indicator of accuracy, provided there is no systematic error.

#### *Office for National Statistics Quality Checks, variables 4 and 6*

As mentioned previously, the ONS devised additional checks to identify possible systematic errors.

The number of EAs checked was dictated by the delivery schedule, and the overall ONS assessments for coding are based on results from 43 EAs. For each EA checked, on delivery of coded data the ONS selected samples of varying sizes (dictated by the workload) from each CD. The default sample size was two percent for the Occupation question. Sampling errors for estimates derived from these samples have not been included in this report, due to insufficient information available for their calculation.

ONS coders then carried out manual checks using standard coding rules with all possible information: text response, details from other questions, whether coding was automatic or manual, the code allocated and access to the image of the form, if needed. Staff could differentiate between errors at the major group level and at finer levels.

All major problems in coded data were identified in comprehensive checks early on in the process. One CD was delivered well in advance and was 100% sampled by ONS. The first EA to fully arrive was also 100% sampled and checked for systematic error. The ONS checked all relevant records in the first EA to be delivered after the first major set of changes were made to the system.

#### *Other issues in measurement*

In general, as well as measuring a process variable it is important to record the time at which the process is undertaken, to enable the construction of reliable charts. In this case we would like to note the time as each EA enters the coding process. The information actually available is the scanning end date, a proxy for the time at coding.

The measurement capability of the ONS quality checks was assessed to some extent by LM. On performing analysis of ONS assessed rates, LM found some errors made by ONS

coders. These lead to the conclusion that the corresponding original LM assigned codes were correct. ONS assessed rates were not adjusted to take these findings into account, due to the small numbers involved.

### ***Analysing key process variables***

#### *Determine system capability*

Table 10 shows that the 2001 UK Census identified and measured several process variables. This report will focus on the analysis of systematic errors carried out by the ONS, and also examine the difference between the LM consistency rate and the ONS accuracy rate.

#### Systematic errors

On receipt of coded data from LM, ONS carried out sampling and verification as outlined in the previous section on measuring key process variables. Where LM had coded incorrectly, the data were marked to indicate this. The ONS calculated percentages of incorrect coding by each allocated code (variable 4). Where 15% or more of records within an individual code were marked as incorrect, and this produced 20 or more cases, data were flagged as identifying a potential source of systematic error. An extract from one of these tables is given in table 12, where we see code 2129 highlighted, with 31 codes and a 16.13% error rate. This information was fed back to LM, who investigated causes for the errors.

**Table 12:** An extract from a spreadsheet recording details of incorrect codings

<b>CODE</b>	<b>TOTAL OF EACH CODE IN NS</b>	<b>OF WHICH HAD STATUS X</b>	<b>WHICH % OF INCORRE CT ERRORS</b>	<b>HOW MANY AUTO</b>	<b>HOW MANY MANUAL</b>
2126	20	1	5.00	0	1
2127	13	2	15.38	1	1
2128	18	5	27.78	2	3
2129	31	5	16.13	3	2
2131	16	2	12.50	1	1
2132	54	3	5.56	0	3

LM determined if the errors could be corrected by a change to the system parameters, or by an update or addition to the databases. These adjustments were made to correct a high level of error within a single code, and so did not always produce an increase in overall accuracy, although they did eliminate future systematic errors. Manual coding procedures were also modified to improve the accuracy of coding operators. LM documented information on each code, including its number of errors, along with reasons and resolutions. An extract from the documentation is provided in table 13. Here we see software and parsing correction to the automatic system, as well as discussions with coders being implemented to address various errors.

**Table 13:** An extract from a spreadsheet detailing errors, reasons and resolutions

Code	Occupation text	Reason	Resolution	Implementation	Total Auto	Total Manual	Combined Auto and Manual Totals	Number of EA's this appeared in
2128	QUALITY CONTROLLER	Random logic error causing errors in matching on qualified professions index	LM to correct software	Software fix - patch 449 - Oct 21, 01	30	15	45	5
2129	MAINTENAUCE/ PRECISION/etc. ENGINEER	Close match on "Packaging Engineer" in survey, due to incorrect parsing	LM to modify parsing rule	Survey version 7.0 - patch 505 Nov 11, 01	113	40	153	14
2131	CONSULTANT	These no longer gets autocoded	no longer a problem	before version 5	11	5	16	1
2132	TECHNICAL SUPPORT, TEAM LEADER, etc.	Using industry information. These look correct, except for manager.	ICL.to review error list and discuss with coders.		9	12	21	3

In total, twelve sets of changes were made to the system during the processing operation, each covering many individual codes. The changes were identified from investigating apparent systematic errors in 133 individual codes. Some of the errors identified were not resolved, as their frequency was too small to warrant a fix. System changes incorporated additions, deletions and amendments to the indexes, tuning data, parsing rules, and coding instructions.

As explained in the above section on measuring key process variables, all major problems in coded data were identified in comprehensive checks early on in the process. As well as looking at the error rate (variable 4), ONS analysed the distribution of errors at major group level. This used frequencies of types of error at major group level (variable 6). Colour-coded bar charts highlighted common mistakes for investigation.

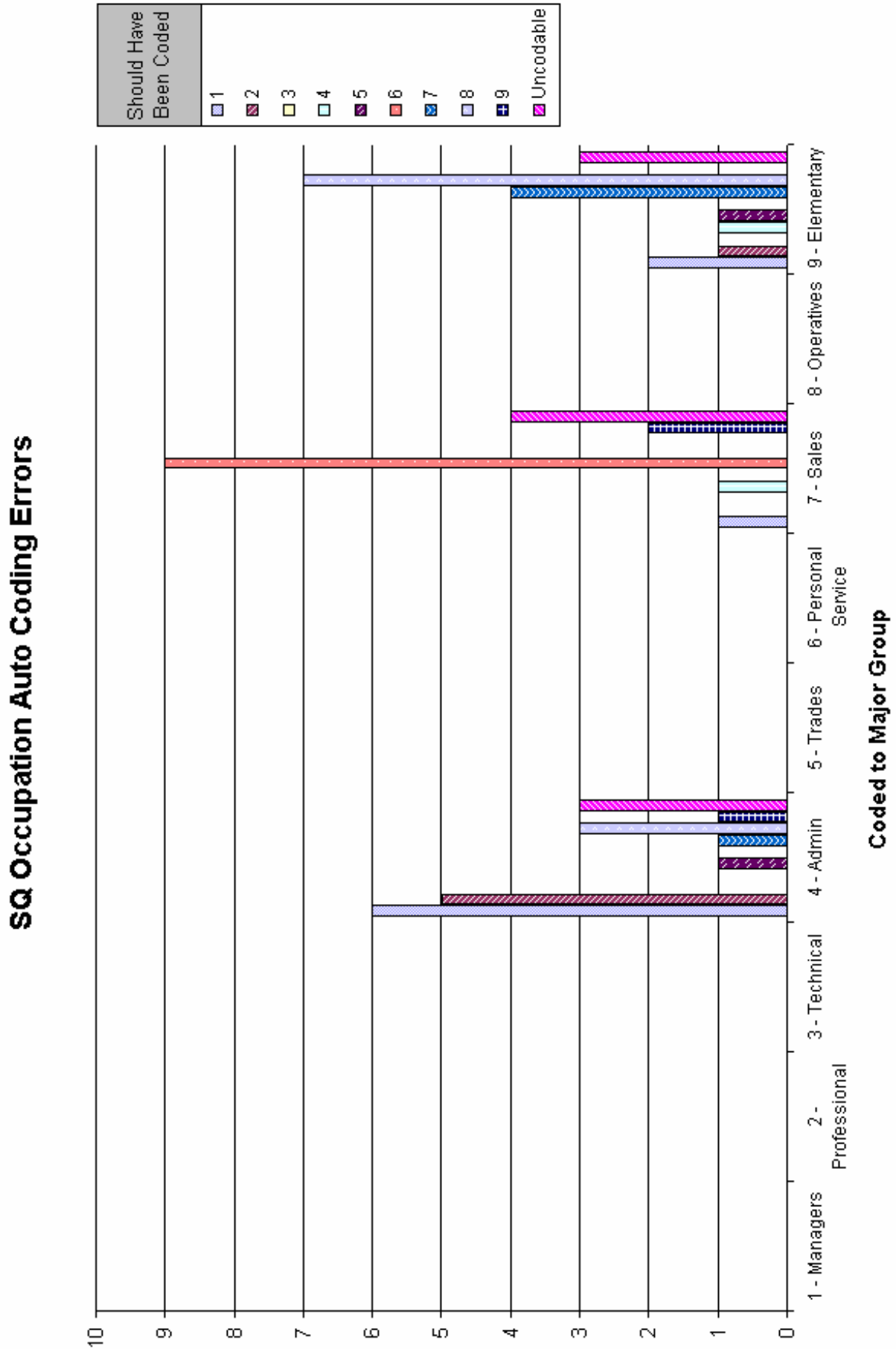
An example of the charts examined is given in figure 24, which shows automatic coding error types for the first EA delivered. Systematic errors were identified in the first EA checked, but by then there were 33 EAs in processing. Ideally, the first EA's data should have arrived earlier, enabling changes to be made before any further EAs were processed.

One of the problems identified involved qualified occupations. The coding system was allocating 'qualified codes' to unqualified respondents, and codes from the lower end of the classification for respondents with professional qualifications. There were four occupation major groups affected by this error: groups 1 (Managers and Senior Officials), 2 (Professional Occupations), 3 (Associate Professional/Technical) and 8 (Process, Plant, Machine Operatives). Analysing figure 24 confirms that one of the main problems was incorrect allocation of groups 1 and 2 to group 4 (unqualified, Admin/Secretarial) and of group 8 to group 9 (unqualified, Elementary Occupations).

As the skew in the data caused by this error was considered severe, 7,500 errors were identified and corrected after delivery. After applying a fix to the system, the ONS checked all records that could have been affected in one EA processed following the fix. The results show that none of these records were affected by the error.



**Figure 24:** Bar chart of automatic coding errors for the 'occupation' question in estimation area 'SQ', by major group



## Analysis and comparison of overall rates

As well as analysis of systematic errors, ONS computed an assessment of overall accuracy from the original data and compared this against the SLA and LM values. As noted earlier, the SLA for occupation coding accuracy is 88%.

### A control chart for the LM rate

LM sampled from each of 112 EAs, and achieved an overall estimated average of 91.32% consistency, with a 95% confidence interval of (91.26%, 91.38%). The estimated average for automatic coding (72.20% of codes) was 91.80%, for frontline manual was 89.40%, and for expert was 87.10%. This pattern in rates is to be expected, as the more difficult codes are those passed to manual coders, with the most difficult of all coded by experts.

A control chart of the LM consistency rate by scanning end date (for the 112 EAs checked) is provided in figure 25. Scanning end date is used on our x-axis as it is the best indicator (or 'proxy') available for the time when an EA went into the coding system. Ideally, the time the coding process began for each EA would be used on the x-axis, for a plot reflecting progress over time.

As can be seen from the chart, all EAs safely met the SLA according to LM's consistency rates. There is an apparent improvement in the reported consistency of the occupation field towards the end of coding. The averages, along with 95% confidence intervals for each of 4 consecutive groups of 28 batches are as follows:

- 91.41% (91.29%, 91.52%) in the first 28 EAs;
- 91.12% (90.99%, 91.26%) in the next 28;
- 90.62% (90.49%, 90.75%) in the next 28;
- 92.10% (91.98%, 92.23%) for the final 28.

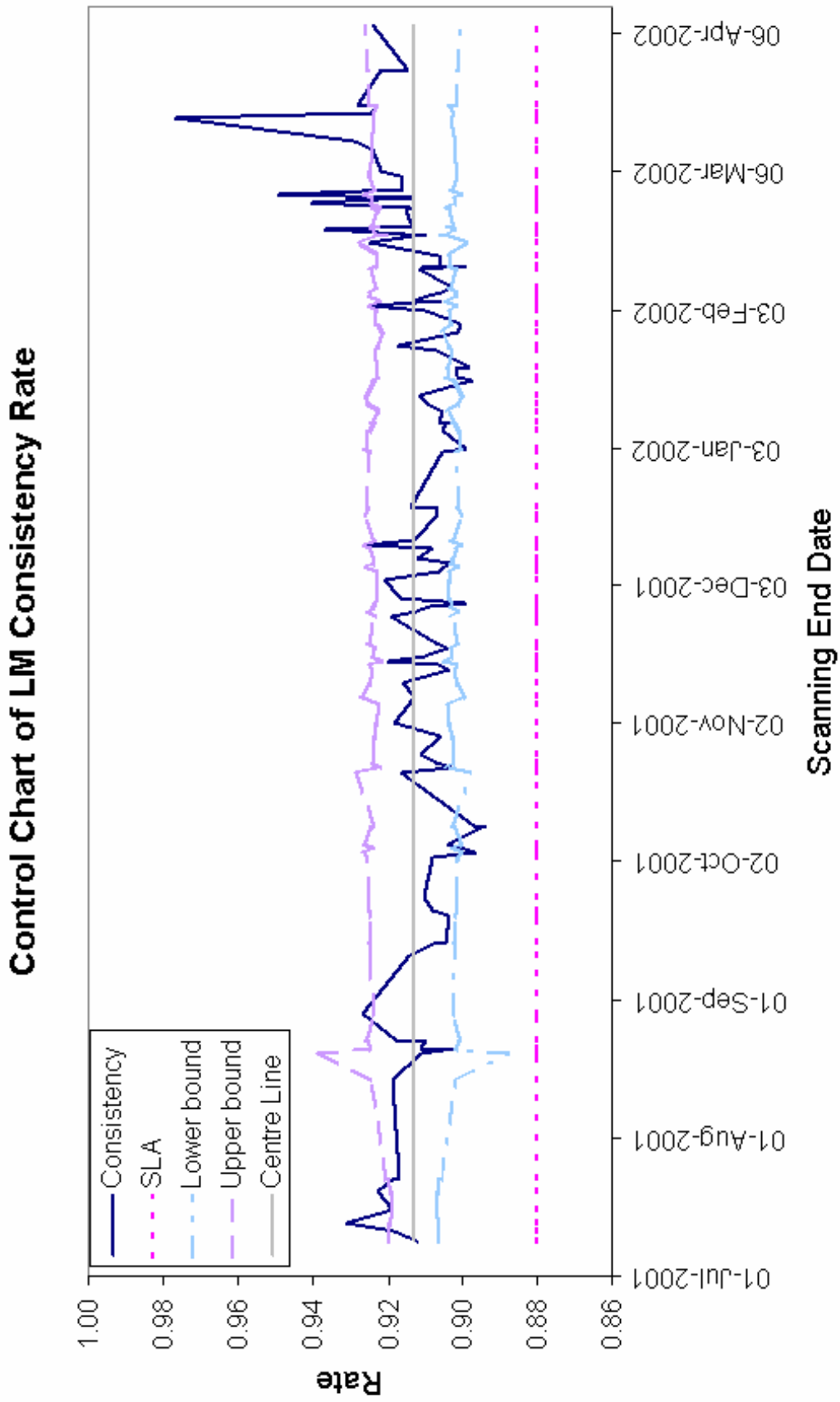
Examining the behaviour of the chart suggests that the process became fairly stable and remained so until mid-February 2002. During that period, a few points lay slightly above or below the control limits. In an operational environment any inconsistent result should be investigated to ascertain whether there is any special cause for the variation that could be eliminated.

The last month and a half of processing exhibited considerable 'out of control' behaviour. However these points lay above the upper limit and so indicated a desirably high consistency between coders. If the process and the upward 'out of control' behaviour had continued, this would suggest a permanent shift and improvement in the coding behaviour. In that situation we should recompute limits to enable fair assessment of process stability and behaviour.

It is interesting to consider whether any significant changes to the process occurred during this time that could explain this apparent improvement. There were a variety of factors that could have caused the fluctuations, including:

- the processing contract was coming to an end;
- coders were being given incentives to stay on;
- increased competition between the shifts;
- some coders were to be kept on after the main processing task was finished.

Figure 25: Control chart of Lockheed Martin's coding consistency rate, by scanning end date.



### ONS rate compared to LM rate

ONS sampled from 43 EAs, and found that accuracy rates gave an 89.04% overall estimated average. Hence there is an estimated difference of 2.28% between the LM and ONS rates, which may to some extent represent systematic errors identified by ONS in the first delivery of data, but not corrected until much of data had been coded (see further discussion below).

A chart comparing ONS accuracy and LM consistency rates by scanning end date is provided in figure 26. Note that points are plotted for the 43 EAs checked by both parties only. Overall, the ONS rates ranged from 83.23% to 93.38 % (with no outliers). In this case, 62.79% of sampled EAs meet the SLA: a marked difference from the LM assessment, partly due to systematic errors not identified by LM checks.

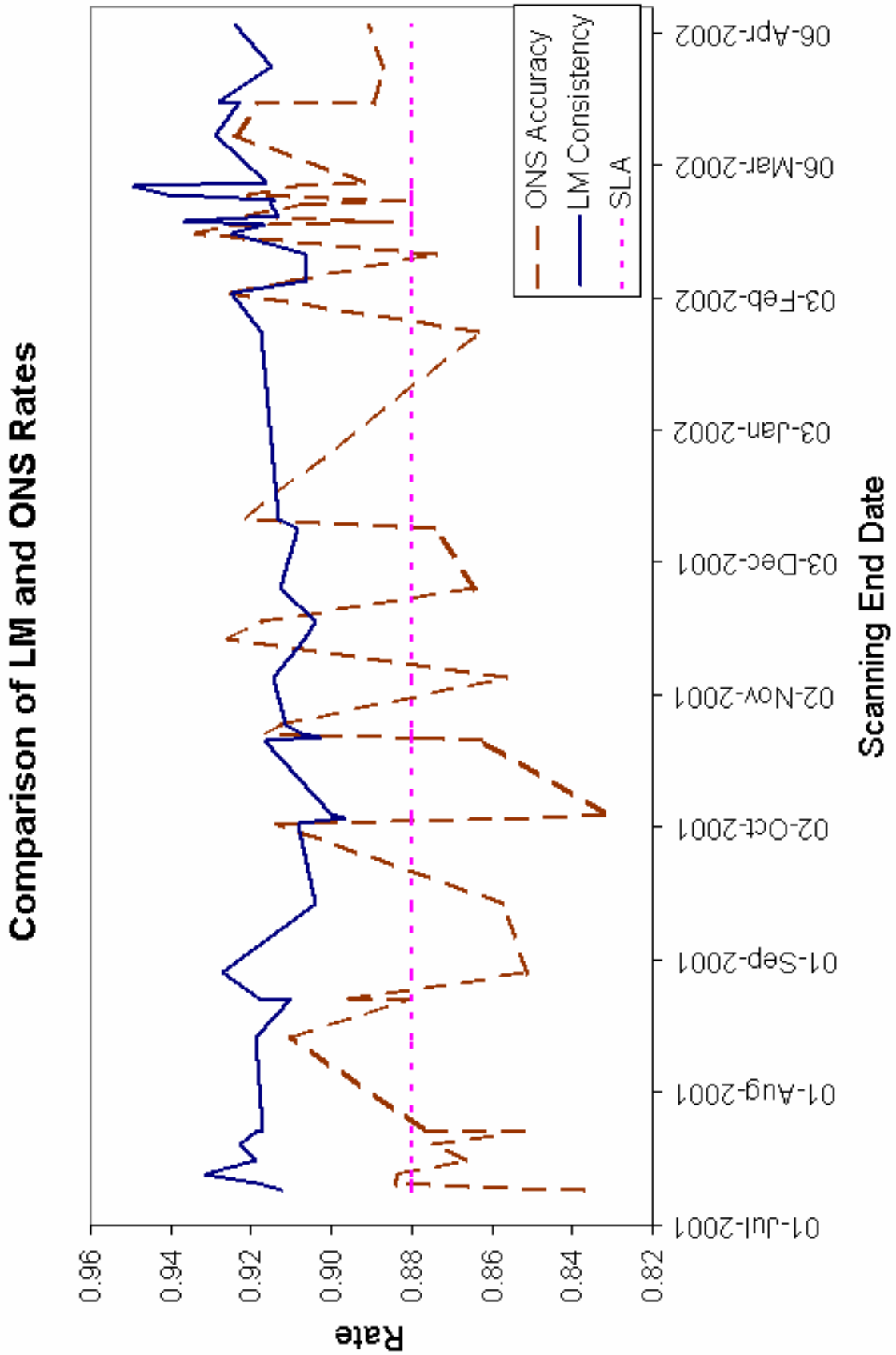
However, the ONS data also suggest improvement in reported accuracy: from

- 87.30% for the first 10 EAs, to
- 87.53% in the next 11, to
- 89.99% in the next 11, to
- 90.35% in the last 11.

The ONS rate clearly has a greater variance than the LM rate. The standard deviations are 2.71% and 1.13% respectively. Due to the high variability in the ONS data, a control chart does not provide any useful information and so is omitted from this report.

There seems to be little similarity in the movement of the rates up until around February 2002. For the last two months the lines follow a similar movement and are in general closer together than previously. Indeed, as systematic errors are identified and resolved over the nine months of processing, we would expect the accuracy rate to increase and become closer to the consistency rate. The discussion below expands on this issue.

**Figure 26:** Chart comparing the ONS coding accuracy and LM coding consistency rates, by scanning end date



### Further discussion of errors and accuracy

As shown above, there is some difference between LM consistency and ONS accuracy rates. For example, the overall estimated averages differ by 2.28%. If ONS assessments of accuracy were exact and if systematic error was the only type of error present then this difference would be a close indicator of systematic error. That is, the difference in the rates would represent the systematic errors identified by ONS in the first delivery of data, but not corrected until much of the data had been coded. However, the above assumptions do not necessarily hold, as discussed below. We can only assume that the difference in rates is mostly explained by the presence of systematic errors. And so we may tentatively conclude from figure 26 that, as the coding process progresses, systematic errors are reduced, causing the consistency and accuracy rates to move closer together.

### Consistency vs. Accuracy

In this report, LM consistency rates are compared with ONS accuracy rates, but we may ask: what is accuracy? Accuracy is consistency with the truth. This leads to another question of: what is the truth? In this particular example, we assume that consistency with better-trained (ONS) experts can be considered as accuracy. But when it comes to a process requiring judgement, such as coding, experts do not always agree. Therefore, it is important to note that the ONS measures are just considered better approximations of accuracy. Real accuracy should be based on verifying the results against a trusted source. In this context, that means either using truth decks (samples that have been verified to be accurate) or interviewing the respondent, which is generally not practical. In reality, we don't have a real accuracy measure for coding. Instead, we have better consistency measures that still have inherent error.

### Systematic vs. Random Errors

Systematic error is defined as error with a non-random distribution. These result from processes that are consistently applied incorrectly. While systematic errors are a component of the difference between consistency and accuracy, they are not the only component and are not necessarily a good characterisation of the other components.

Non-systematic or random error is the other major component of the difference between consistency and accuracy. In any complex task that involves approximate reasoning or human beings there will be random error. The random error in processes is reduced as the process is refined, which is what happened during coding for the UK Census. The tuning of the automated recognition systems was refined over time so that it was more accurate. Some of this was through the elimination of systematic error, and some was through improvements to the data and software strategies, which reduced random error. Similarly, random error would decrease as the skill of the manual coders evolved through experience and on-the-job training.

In addition to reduction of systematic error, the reduction in random error outlined above seems to have contributed to a reduction in differences between consistency and accuracy and better correlation between LM and ONS results later in the process, as shown in figure 26.

### *Establish a System for Continuous Monitoring of Processes*

In addition to tuning in the development environment and the analysis of accuracy rates during processing, live operations were also observed. Operator feedback was gathered during the Rehearsal for the Census, and a study was performed upon completion, to identify the most beneficial changes that could be made to the system. One such change was adding capability for frontline coders to pass their final text changes and short notes on to the expert coders. This helped reduce duplication of effort between the two modes of coding, and was seen as a great benefit in increasing the speed of expert coders.

### ***Evaluation***

#### *Conclusions*

Checking a sample of coded data was an effective method of identifying systematic error, and assessing reported accuracy. LM rates of consistency are good measures of the accuracy, except where there is systematic error. All major problems in the coded data were identified in a comprehensive check of the first EA.

#### *Recommendations*

Future processing systems should again include stringent quality measures. Future quality assurance procedures should include a process to identify systematic error at code level at source, with sufficient resource to analyse and correct data at source. A large block of coded data should be analysed early enough to allow corrections to be made before too much data passes through the system.

### 3.3.3.5 Literature review

Several papers describing quality control of the editing, validation or coding processes are listed below. These papers contain some further ideas for process variables and examples of their analysis, and will provide valuable information for those seeking to monitor and improve their data processing quality.

Biemer P and Caspar R (1994) Continuous Quality Improvement for survey operations, some general principles and applications, *Journal of Official Statistics*, Vol. 10. No. 3, 1994, pp. 307-326

The authors discuss sample inspection (or acceptance sampling) as a method of quality control. The limitations of the method lead them to consider a different methodology, which is the CQI approach described in section 2. This approach is applied to industry and occupation coding in the Research Triangle Institute.

A measure of coding accuracy, the coder error rate (CER), was developed. Using 1991 rates as a starting point, in 1992 they began implementing changes in the coding operation. Changes to the shift arrangements for coders were made, as well as enhancements to the on-line coding system following comments raised in team meetings.

However, CQI was not achieved until a feedback loop from adjudicators to coders was introduced. Weekly team meetings with a quality advisor were set up; during which Pareto charts of misassigned codes (for the group and individuals) facilitated discussions, explanations, and highlighted needs for retraining. Future research includes investigating the use of process control charts to identify special causes, however focusing on an individual operator as an assignable cause may affect morale, whereas the current approach has been received unanimously positively by staff.

Della Rocca G, Luzi O, Scavalli E, Signore M and Simeoni G (2003) *Evaluating, Monitoring and Documenting the Effects of Editing and Imputation in Istat Surveys*, Work Session on Statistical Data Editing, Madrid, Working Paper number 3.

Diskin D H, Karcher P, Mersch M and Pavnica P (1987) *Developing Process Controls for Surveys and Censuses*, Proceedings of the Section on Survey Research Methods, American Statistical association

The authors describe approaches used in the US Bureau of the Census to control the quality of their processes. Previous attempts used inspection and repair techniques, which proved unsatisfactory due to their cost, and not preventing errors before they occur. Inspired by the work of W. Edwards Deming, Census Bureau personnel 'decided to focus on improving processes to assure quality rather than control it after the fact.'

They followed this approach by developing a software system that generates charts and statistical information on process quality. The system provides supervisors with timely feedback and information to identify special and common causes of error. In addition to this, tasks were simplified and documented, and training improved.

Future plans outlined in the paper include a controlled experiment to compare traditional inspection and repair methods with an approach emphasising continuous monitoring and adjustment. A split panel design for the assignment of death certificates for coding is planned, with half the coders using each of the two approaches to quality control. Plans for



analysis include examining the variability, error rate reduction, time, cost and quality of data resulting from the two approaches.

Engström P and Granquist L (1999) *On improving quality by modern editing*, Work Session on Statistical Data Editing, Working Paper No.23, Rome 1999.

Engström P (1996) *Monitoring the Editing Process*, Work Session on Statistical Data Editing, Working Paper No. 9, Voorburg 1996.

Engström P (1997) *A Small Study on Using Editing Process Data for Evaluation of the European Structure of Earnings Survey*, ECE, Work Session on Statistical Data Editing, Prague, Working Paper No. 19.

Granquist L (1995) Improving the Traditional Editing Process, in B.G. Cox, D.A. Binder, N.Chinnappa, A.Christianson, M.J. Colledge, and P.S. Kott (eds.) *Business Survey Methods*, New York: Wiley, pp. 385-401.

Granquist L and Kovar J G (1997) Editing of Survey Data: How much is enough? in Lyberg L, Biemer P, Collins M, De Leeuw E, Dippo C, Schwarz N, and Trewin D (eds.) *Survey Measurement and Process Quality*, New York: Wiley, pp. 415-435.

Granquist L, Arvidson G, Elffors C, Norberg A, and Lundell L G (2002) *Guide till granskning*, CBM 2002:1 (in Swedish).

Granquist L (1998) *Efficient Editing- Improving Data Quality by Modern Editing*, presented to the New Techniques and Technologies for Statistics (NTTS '98) conference. <http://europa.eu.int/en/comm/eurostat/research/conferences/ntts-98/papers/sp/012s.pdf>.

The author discusses two ways of improving data quality by means of editing: continuous improvement of the survey so as to prevent errors and the design of edits. Other issues discussed are the identification of error causes, the requirements of a well-designed system called process data subsystem and the standardisation of the editing processes.

Harris K W (1984) *An overview of the Quality Control Program for the National Centre for Health Statistics*, Proceedings of the Section on Survey Research Methods, American Statistical association.

The National Centre for Health Statistics provide data on the health of the United States population. Data processing is a significant part of this work, and coding is the most important function of the processing.

The author describes the system of acceptance sampling for quality control of coding. The paper contains a detailed discussion of various types of verification, including two and three way independent sample verification, and sequential sampling.

The conclusion is that the two way independent verification system with lower sample rates employed has reduced the required resources of some coding operations by 75%, whilst continuing to provide unbiased estimates of quality.

Linacre S J (1991) *Approaches to Quality Assurance in the Australian Bureau of Statistics Business Surveys*, Bulletin of the International Statistical Institute: Proceedings of the 48th Session, Cairo, Egypt, Book 2, pp. 297-321.

Linacre S J and Trewin D J (1989) *Evaluation of Errors and Appropriate Resource Allocation in Economic Collections*, Proceedings of the Annual Research Conference, Washington, DC: U.S. Bureau of the Census, pp. 197-209.

Luzi O (2003) *Evaluating, monitoring and documenting the effects and imputation in ISTAT surveys*, Work Session on Statistical Data Editing Madrid 2003, Working Paper no. 3.

Lyberg L, Biemer P and Japac L (1998) *Quality Improvement in Surveys – A Process Perspective*, Proceedings of American Statistical Association, Dallas 1998.

Macchia S, Mastroluca S, and Reale A (2001), *Planning the quality of the automatic coding process for the next Italian General Population Census*, presented at the International Conference on Quality in Official Statistics, Stockholm, May 14-15, 2001

This paper examines the automatic coding process and its impact in terms of quality. The authors describe plans for measuring the quality of automatic coding for the 2001 Italian General Population Census, and present the results of two pilot surveys. The software used for automatic coding is ACTR (developed by Statistics Canada), as used for the 2001 UK Census example described above. Section 2 of the paper provides a detailed description of the software.

In the pilots, automatic and manual coding systems were compared by measuring several process variables. These include:

- precision – the percentage of correct codes automatically assigned;
- recall – the percentage of codes automatically assigned.

Manzari A and Della Rocca G (1999) *A Generalised System based on a Simulation Approach to Test The Quality Of Editing And Imputation Procedures*, UN/ECE Work Session on Statistical Data Editing, Working Paper No.13.

This paper evaluates the quality of an editing and imputation procedure in terms of its capability to recognize errors and adequately replace them with the true values. Generalized software (ESSE), based on a simulation approach, developed by ISTAT is also described along with an application.

Martin C and Poirier C (2002) *Analysis of Data Slices and Metadata to Improve Survey Processing*, Work Session on Statistical Data Editing, Working Paper No. 2, Helsinki 2002.

Mersch M, Gbuz P and Russell C (1993) *Statistical process Control in Decennial Census Industry and Occupation Coding*, Proceedings of the Section on Survey Research Methods, American Statistical association

This paper has links to the above paper by Diskin et al. (1987). The latter describes dissatisfaction with the QC approach in the Census Bureau, and the former is written after the 1990 Census when a new approach was taken for quality control of the occupation and industry coding.

Many process variables are mentioned in the paper, including a production rate (code per hour), referral rate, and an error rate derived from three way independent coding verification. A computer system automatically generated reports for each coder and unit of coders, so that supervisors could monitor performance. Some analysis of overall rates is provided in the paper, as well as recommendations for future changes and improvements in the quality monitoring system.

Morganstein D and Marker D A (1997) Continuous Quality Improvement in Statistical Agencies, in Lyberg L, Biemer P, Collins M, De Leeuw E, Dippo C, Schwarz N, and Trewin D (eds.), *Survey Measurement and Process Quality*, New York: Wiley, pp. 475-500.

Nordbotten S (1998) *Improving Editing Strategies*, Proceedings of the third International Conference on Methodological Issues in Official Statistics in Stockholm, October 1998.

Walker C, Brown A and Veevers R (1996) *Data Validation in the CAI World*, Proceedings of the Survey Research Methods Section of the American Statistical Association, 1996, pp. 620-625.

This paper describes data validation on a Statistics Canada social experiment using computer-assisted interview (CAI). The paper outlines the improvements to quality that can be gained from the data validation process by using CAI. The paper closes by stating that using CAI in this way can improve data quality by using a continuous improvement cycle to measure and validate data quality, and improve the survey taking process.

Wein E (2000) *The Planning of Data Editing*, UN/ECE Work Session on Statistical Data Editing, Working Paper No.3.

The author discusses the planning of data editing. He provides the various stages of editing and describes in detail three of the stages considered more crucial for a successful planning; namely, collection and judgement of relevant information, overall planning, specification of data editing activities.

Wein E (2002) *The Planning of Data Editing Processes*, UN/ECE Work Session on Statistical Data Editing, Working Paper No.3.

This paper examines the design of data editing processes so as to improve the efficiency of editing, that is to disseminate statistical results with low process costs, improved accuracy, timeliness and user-friendly documentation.

Weir P (1997) *Data Editing Performance Measures*, ECE, Work Session on Statistical Data Editing, Prague, Working Paper No. 38.

Werking G, Tupek A and Clayton R (1988) CATI and Touchtone Self-Response Applications for Establishment Surveys, *Journal of Official Statistics*, 4, pp. 349-362.

Winkler W (1997) *Problems With Inliers*, ECE, Work Session on Statistical Data Editing, Prague, Working Paper No. 22.

### 3.3.4 Weighting and Estimation

#### 3.3.4.1 Measuring nonresponse bias (SCB)

##### Introduction

This example focuses on the use of auxiliary information to measure survey nonresponse bias. It is an investigation of a survey process rather than an example of the continuous study of key process variables. This approach is necessary due to the nature of weighting and estimation. For an NSI to achieve improvements in the quality of estimation, it is often not sufficient to analyse key process variables. It is necessary to undertake survey specific investigations. It is also for this reason that we have chosen a post-survey procedure as an example for this process.

##### Description of the process

What is estimation? Every textbook that deals with survey sampling has a description of estimation. Särndal et al. (1992) describes estimation in the following way:

*“This phase entails the calculation of survey estimates according to the specific point estimator formula, with appropriate use of auxiliary information and adjustment for nonresponse, as well a calculation of measure of precision in estimates (variance estimate, coefficient of variation of estimate, confidence interval).”*

Later in the book the authors state that all errors from sample selection, data collection (see 3.3.1) and data processing (see 3.3.3) will affect the point estimates and should ideally be accounted for in the measures of precision.

##### Summary of the report

The two main parts of this report are:

- A short general section on estimation and key process variables;
- An example from a nonresponse study at SCB where nonresponse bias was measured by an empirical approach with use of auxiliary information.

## Estimation and key process variables

Fundamental key process variables concerning estimation are variance, sampling error and standard error. Other relevant key process variables are: relative standard error, confidence interval, mean square error, and nonresponse error/nonresponse bias. In the Eurostat working group (see Eurostat (2003)) the concepts above have following definitions<sup>2</sup>:

*Variance*: The variance is the mean square deviation of the variable

*Sampling error*: The part of the difference between a population value and an estimate there of, derived from a random sample, which is due to the fact that only a sample of values is observed; as distinct from errors due to imperfect selection, bias in response or estimation, errors of observation and recording etc.

*Standard error*: The positive square root of the *variance* of the sampling distribution of a statistic.

*Relative standard error*: The relative standard error (RSE) is a measure of the variability of estimates. The RSE of an estimate is obtained by dividing the standard error of the estimate (SE[r]) by the estimate itself [r]. This quantity is expressed as follows:  $RSE=100 \times (SE[r])/r$ .

*Confidence interval*: A *a*% confidence interval for an unknown population parameter \*, is an interval, calculated from sample values by a procedure such that, if a large number of independent samples is taken, a percent of the intervals obtained will contain \*.

*Mean square error*: The expected value of the square of the difference between an estimator and the true value of a parameter. If the estimator is unbiased then the mean square error is simply the variance of the estimator. For a biased estimator the mean square error is equal to the sum of variance and the square of the bias.

*Nonresponse error*: Nonresponse errors occur when the survey fails to get a response to one, or possibly all, of the questions. Nonresponse causes both an increase in *variance*, due to the decrease in the effective sample size and/or due to the use of imputation, and may cause a **bias**<sup>3</sup> if the nonrespondents and respondents differ with respect to the characteristic of interest.”

In survey practice it is often relatively easy to calculate and analyse variances/ standard errors, relative standard errors and confidence intervals. It is more difficult to measure nonresponse errors/nonresponse bias and coverage errors. Adjustment for nonresponse and coverage errors plays an important part in the estimation procedure. For a complete description on estimation in the presence of nonresponse and coverage errors refer to Lunström and Särndal (2002). This is a handbook (a Current Best Method or CBM) that offers effective methods to reduce nonresponse errors. It also covers sampling errors and coverage errors.

---

<sup>2</sup> The source of each definition is the Methodological Documents Glossary, Oct 2003

<sup>3</sup> Nonresponse bias (Notation by the authors of the Handbook)

## The use of registers to measure nonresponse bias in a Swedish survey

As mentioned above, it is often hard to measure nonresponse bias. A theoretical expression for nonresponse bias for an estimate of a total ( $t$ ) can be written:

$$B_{NR} = (1 - a_R) \cdot (\hat{t}_R - \hat{t}_{NR}),$$

where  $(\hat{t}_R)$  is an estimate of the total based on the respondents and  $(\hat{t}_{NR})$  is an estimate based on the nonrespondents.

We generally know the nonresponse rate ( $a_R$ ) but know very little about the difference in the characteristics of the study variable between the respondents and the nonrespondents, that is  $(\hat{t}_R - \hat{t}_{NR})$ . There are many textbooks and papers that discuss *nonresponse bias* and associated formulas, but there are fewer examples of quantitative estimates of nonresponse bias.

The following example is from SCB's survey: Activity after Graduation (AAG). For a complete description of the study refer to Hörngren (1999).

### *Description of the Activity after Graduation survey (AAG)*

The purpose of AAG is to describe students' activities after undergraduate or postgraduate exams in respect of a certain academic year. The main aim is to look at various aspects of their employment situation at the time of the study (reference period). This example is based on data from the 1994 survey, based on a sample of the graduate population. The sample size is about 7,800 individuals. The main data collection tool is a mail questionnaire, followed-up by telephone interviews among a subsample of nonrespondents according to the method of Hansen and Hurwitz (1946). In the estimation phase we assume that there exists a response distribution, which divides the population into a response stratum and a nonresponse stratum. We can view this example as a special case of response homogeneity groups: a "mail group" and a "telephone group". If we have response from every individual in the sub sample (telephone group) we will have an unbiased estimator, despite the nonrespondents in the initial sample (mail group).

### *Identifying key process variables*

The weighted nonresponse rate of the mail questionnaires, i.e., after the first phase in the survey, was 23.5%. Despite special efforts (telephone interviews) to make the individuals in the subsample respond, the weighted nonresponse rate in the second phase was about 50%. As mentioned above the requirement that makes unbiased estimation possible is full response in the subsample. A nonresponse rate of 50% in the subsample must be seen as a very serious problem.

The nonresponse rate is of course a process variable. But the key process variable for this "post-survey example" is *nonresponse bias*. Measuring nonresponse bias requires relevant information on the nonrespondents. In SCB's *Register of Employment (RE)* we have information on employment for the total population in Sweden. RE is in turn based on (at least) six other registers. The main source is the *statement of income from the employer*. In RE, employed people are defined as those who are 16 years old and who did an average of at least one hour's paid work per week in November (reference period). Consequently, employed people are defined by income.

The aim is that the definition used in RE will correspond as far as possible to the definition in the Swedish Labour Force Survey (LFS). In AAG we receive answers on the respondents' main activity during the reference week. The *correlation coefficient* of "AAG employed" and "RE-employed" is in this case 0.42. The correlation coefficient can in this case be seen as a process variable for the relevance of the auxiliary information.

### **Measuring key process variables**

RE-employed is related to AAG-employed. In a post-survey procedure it is possible to match AAG-data with RE-data on an individual level. Thus we can estimate the number of RE-employed and the RE-employment ratio on the entire sample. In this situation we also have "observations" for nonrespondents. We calculate the following estimates (for a complete description of the formulas see Hörngren (1999)):

a1) The RE-employment ratio based on the entire sample,

$$\hat{R}_{as} = \frac{\hat{t}_{as}}{N}$$

where N is the known population total and  $\hat{t}_{as}$  is an unbiased estimate.

a2) The RE-employment ratio based only on the respondents r,

$$\hat{R}_{ar} = \frac{\hat{t}_{ar}}{N}$$

where N is known. It is also of interest to estimate RE-employment ratio with an alternative estimator. If we treat the telephone interviews as a successful reminder, the RE-employment ratio can be calculated with a Horvitz-Thompson (HT) estimator based on three different "response sets":

b1) The RE-employment ratio based on the entire sample with a HT-estimator

$$\hat{R}_{bs} = \frac{\hat{t}_{bs}}{N}$$

b2) The RE-employment ratio based on the respondents r with a HT-estimator:

$$\hat{R}_{br} = \frac{\hat{t}_{br}}{N}$$

b3) The RE-employment ratio based **only** on the mail questionnaire respondents ( $r_1$ ) with a HT-estimator that is treating the telephone interviews as "nonrespondents". These estimates are of more interest than the "b2-estimates":

$$\hat{R}_{br1} = \frac{\hat{t}_{br1}}{N}$$



By virtue of the point estimates using the parameters above it is possible to estimate three measures of the nonresponse bias concerning the RE-employment ratio:

- i. The bias  $B$  in percentage of  $\hat{R}_{as}$  that occurs when using the idea of Hansen and Hurwitz is estimated by:

$$\hat{B}(\hat{R}_a) = 100 \cdot \left( \frac{\hat{R}_{ar} - \hat{R}_{as}}{\hat{R}_{as}} \right)$$

- ii. The bias in percentage that occurs when using a HT-estimator **including** the telephone interviews is estimated by:

$$\hat{B}(\hat{R}_b) = 100 \cdot \left( \frac{\hat{R}_{br} - \hat{R}_{bs}}{\hat{R}_{bs}} \right)$$

- iii. The bias in percentage that occurs when using a HT-estimator **excluding** the telephone interviews is estimated by:

$$\hat{B}(\hat{R}_{b1}) = 100 \cdot \left( \frac{\hat{R}_{br1} - \hat{R}_{bs1}}{\hat{R}_{bs1}} \right)$$

### **Analysing key process variables**

In table 14 post-survey estimates of RE-employment ratio are shown. All point and standard error estimates are calculated using CLAN 97 (Andersson-Nordberg).

**Table 14:** A table of post-survey Estimates of RE-employment ratio according to AAG with different Estimation Procedures. Point estimates (p.e.) and Standard Errors (s.e.)

Domain	Estimation Procedure									
	a1		a2		b1		b2		b3	
	p.e.	s.e.	p.e.	s.e.	p.e.	s.e.	p.e.	s.e.	p.e.	s.e.
<b>Overall</b>	<b>85.4</b>	<b>0.7</b>	<b>88.2</b>	<b>0.7</b>	<b>85.2</b>	<b>0.7</b>	<b>88.4</b>	<b>0.7</b>	<b>88.4</b>	<b>0.7</b>
Engineering (M Sc.)	87.6	0.7	92.6	0.6	87.6	0.7	93.0	0.7	92.9	0.6
Mathematics and										
Natural Science	82.3	1.4	85.8	1.4	82.3	1.4	85.6	1.4	85.4	1.6
Tech.gy and Data	72.0	1.0	74.5	1.2	72.1	1.0	74.3	1.1	73.8	1.2
Economy	84.3	2.1	87.0	2.2	82.7	2.1	87.8	2.1	88.4	2.1
Social and Beh. Sc.	84.6	1.1	87.4	1.2	84.1	1.1	87.4	1.1	87.6	1.2
Medicine and Dent.	86.3	1.3	90.5	1.4	85.7	1.3	90.4	1.3	89.9	1.3
Recreation/Nursery	86.1	1.7	87.0	1.7	86.8	1.6	87.0	1.7	87.1	1.8
Bach of Arts. New	71.2	3.0	77.4	3.2	70.7	3.0	77.4	3.1	77.1	3.3
Others	89.3	1.8	92.0	1.7	88.9	1.7	92.0	1.7	92.1	1.7

Estimates according to a1 and b1 are unbiased. The marginal differences that occur between point estimates in a1 and b1 are random errors. Estimates according to a2 and b2 are based on *all* respondents in the sample. In a2 the nonresponse model follows the idea of Hansen and Hurwitz although we failed to obtain response from 50% of the individuals in the subsample. In b2 (HT-estimator) we use a nonresponse model, which assumes independent responses with equal response probabilities within strata (and we treat a mail-respondent and a telephone respondent equally).

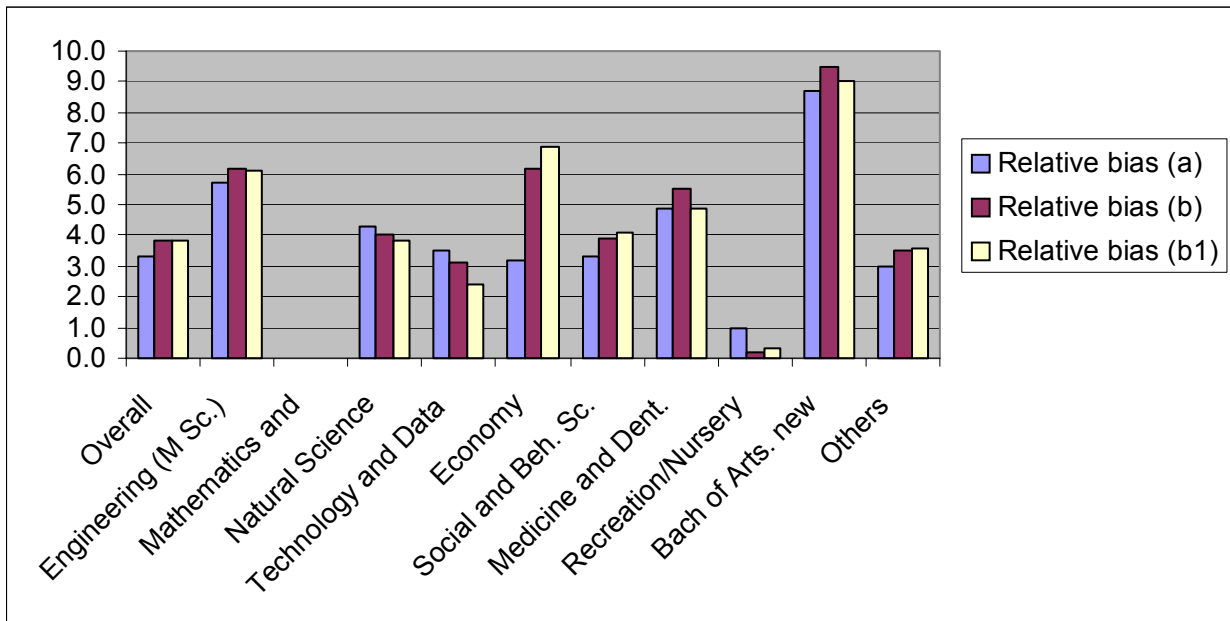
The estimates of the RE-employment ratio according to a2 and b2 suffer from bias. Table 14 shows that a2 overestimates the RE-employment ratio by 2.8% units and b2 also overestimates the same parameter by 2.8% units. Estimates using b3 are generated in a similar way as estimates by b2, with the difference being that telephone interviews are excluded. A comparison between b2 and b3 gives an indicator of the significance of the telephone interviews. Estimates b2 and b3 have approximately the same levels. The variance in b3-estimates is larger, particularly in some domains, since these estimates use fewer observations than b2.

In table 15 and figure 27 we get a more general view concerning the effects of the bias.

**Table 15:** The Bias in Percentage of Unbiased Estimates of RE-employment Ratio

Domain	Estimated Bias in Percentage of Unbiased Estimates of RE-employment Ratio		
	$\hat{B}(\hat{R}_a)$	$\hat{B}(\hat{R}_b)$	$\hat{B}(\hat{R}_{b1})$
<b>Overall</b>	<b>3.3</b>	<b>3.8</b>	<b>3.8</b>
Engineering (M Sc.)	5.7	6.2	6.1
Mathematics and Natural Science	4.3	4.0	3.8
Technology and Data	3.5	3.1	2.4
Economy	3.2	6.2	6.9
Social and Beh. Sc.	3.3	3.9	4.1
Medicine and Dent.	4.9	5.5	4.9
Recreation/Nursery	1.0	0.2	0.3
Bach of Arts. new	8.7	9.5	9.0
Others	3.0	3.5	3.6

**Figure 27:** Bar chart of the Bias in Percentage of Unbiased Estimates of RE-employment Ratio



As seen in table 15 and figure 27 there is no doubt that the RE-employment ratio is overestimated. The bias is more serious in some domains of study, especially *Bachelor of Arts (new types)*, which has a high nonresponse rate of 14%. However the domain *Engineering* also has a nonresponse rate of 14%, but a relative bias lower in comparison with *Bachelor of Arts*. The explanation is that the difference in RE-employment between respondents (who have higher employment) and nonrespondents is greater within *Bachelor of Arts*, in comparison with *Engineering*.

In the *Recreation and Nursery Teachers Education* domain of study the bias is negligible. This is the domain of study with the lowest nonresponse rate, of 6.9%. When we compare the three different measures of the bias (each with different estimation assumptions), it is shown that estimates according to Hansen and Hurwitz are slightly better. But we must keep in mind that these estimates give a larger variance, and that a telephone interview is about ten times more expensive than a mail questionnaire.

### **Evaluation**

In this example we have, in an empirical way, measured nonresponse bias in estimates of the RE-employment ratio. Since RE-employed and employed according to the survey (AAG) are closely related, the results are clear indications of the degree of nonresponse bias in the survey.

The study indicates that the nonresponse causes a bias of over 3% for estimates of the employment ratio. It is very clear that the response model according to Hansen and Hurwitz does not work with a high nonresponse rate in the second phase (the telephone follow-up). The obvious reason is that individuals who respond in the telephone follow-up do not differ in employment status from individuals who respond on the mail questionnaire. The negligible difference between estimates by b2 and b3 emphasises this statement.

The results of this example lead to two main recommendations for AAG-surveys:

- I. Stop using telephone follow-ups according to the idea of Hansen and Hurwitz. But SCB continue to use telephone interviews (as a reminder and for the interview) in strata/domains of study where we know we have serious nonresponse bias. That leads to a new estimation procedure:
- II. Using data from RE as auxiliary information in the sampling and estimation phase with the purpose of adjusting for nonresponse bias (and sampling errors).

The result of this example was presented in the "*Bakgrundsfakta till arbetsmarknads och utbildningsstatistiken*" series, which is background material presented for the statistics produced by the department for Labour and Educational Statistics at SCB. The publications of this series consist of product descriptions, accounts of methods used and compilations of various information that may be of help in gaining an overview of the statistics and facilitate their use. The publications are intended mainly for the users of Labour and Educational statistics.

The main benefit of this example is that AAG (and similar) surveys now use register information as auxiliary information with the purpose of reducing nonresponse errors and sample errors. However this makes it difficult to evaluate the new estimation procedure, as it is not possible to use the same auxiliary information in evaluation as used in estimation. New auxiliary information must be identified for the evaluation procedure.

### 3.3.4.2 Literature review

This literature review summarises information from relevant papers relating to estimation, and particularly nonresponse bias.

Andersson C and Nordberg L (1998) *A User's Guide to CLAN 97 - a SAS-program for computation of point and standard error estimates in sample surveys*, Statistics Sweden.

Clan is a SAS-program designed to compute point and standard error estimates in sample surveys. CLAN covers most of the common estimators, for example: The Horvitz-Thompson estimator, the Generalised Regression Estimator, estimators based on two-phase sampling and response homogeneity groups. CLAN has also the advantage of being an excellent tool for evaluating the estimation procedure and comparing different estimators with the same set of observations.

Eurostat (2003) *Methodological Documents Glossary* (Eurostat/A4/Quality/03/Glossary), Sixth meeting of Assessment of quality in statistics working group, Luxembourg 2-3 October 2003.

Hansen M H and Hurwitz W N (1946) The Problem of Nonresponse in sample surveys, *Journal of American Statistical Association* 41, 517-529.

A classical paper in survey sampling. Hansen and Hurwitz introduce the idea of using a subsample among the nonrespondents to produce unbiased estimates. It did not work in the example above – but the theoretical idea has been successful in many survey situations.

Hörngren J (1999) *An example of the Use of Registers to Measure Nonresponse bias in Swedish Surveys*, contributed Paper for the International Conference in Survey Nonresponse in Portland, Oregon, USA, October 1999. Statistics Sweden.

The example in this report is based on this paper.

Lundström S and Särndal C-E (2002) *Estimation in the presence of Nonresponse and Frame Imperfections*, Statistics Sweden (SCB).

This book has been prepared as a Current Best Methods (CBM) manual within the framework of quality improvement work at SCB. It offers a review of effective methods to reduce the influence of nonresponse in surveys. It also covers issues closely related to nonresponse, namely, coverage errors. The examples in this CBM are from SCB, but the recommendations are general and suitable for every NSI within the EU.

Särndal C-E, Swensson B, and Wretman J (1992) *Model Assisted Survey Sampling*, New York: Springer-Verlag.

The back cover of this widely used book explains that: "This book provides a comprehensive account of survey sampling theory and methodology which will be suitable for students and researchers across a variety of disciplines".

### 3.3.5 Analysis of Primary Outputs

#### 3.3.5.1 Tabulation (NSSG)

##### Description of the process

*The analysis of primary outputs* is part of SVC group 9. It refers to the process of summarising raw data and investigating any data discrepancies by means of exploratory data analysis or macro-editing methods before further analysis. All primary survey or census outputs should be analysed in this respect, in order to determine data consistency or to direct further analysis.

##### Summary of the report

Our discussion will be based on the survey for the construction of the Retail Sales Value Index (RSVI). The objective of the survey is to measure the variation of the value of total retail sales. This is achieved by continuous monitoring of sales values provided by a panel of enterprises.

The survey is conducted by a postal inquiry on a monthly basis using one-stage stratified random sampling all over Greece. NSSG collects data from enterprises (reference units), including all their branches in the country.

Tabulation is used for the identification of discrepancies in the estimates of the longitudinal series.

##### Identifying key process variables

###### *Develop a process flow map*

Tabulation is used for assessing the temporal picture of the aggregates according to sector of activity and groups of turnover. Prior to the aggregates analysis, micro level controls using the same logic (ratios across time) are performed on the enterprises that have values for the previous month, as well as for the same month with one-year lag. The checking process is described in the following paragraphs, and the corresponding flow map is provided in figure 28.

As mentioned earlier, a sample is used for the construction of the index. One single variable, the sales value, is requested from each enterprise. During tabulation NSSG filters the enterprises, which have data at time  $t$ ,  $t-1$ ,  $t-12$ , and calculates two indicators: the change related to last month's data and to the same month in the previous year.

We denote as  $SV_{j,t}$  the sales value for enterprise  $j$  at month  $t$  and the two indicators are:

$$R^1_{j,t} = \frac{SV_{j,t} - SV_{j,t-1}}{SV_{j,t-1}} \quad \text{and} \quad R^2_{j,t} = \frac{SV_{j,t} - SV_{j,t-12}}{SV_{j,t-12}}$$

If large deviations are found then a correction phase follows, where 2 options exist. The user can

- I. correct  $SV_t$  if it is incorrect (perhaps wrong entry),
- II. modify  $t-1$  (but not  $t-12$ ) if it is considered incorrect.

As is apparent in this phase, the Index does not include the new entries (enterprises that first appear at time  $t$ ).

If minor deviations are found then the next steps are the introduction of weights and the calculation of aggregates (sales values for each sector of activity or for groups of turnover). For these aggregated values the corresponding  $R^1$  and  $R^2$  indicators are also calculated.

If these deviations are considered acceptable, the Index is calculated. Otherwise the weights are checked for consistency and, if found correct, a re-examination of data is performed. Errors in the weights lead to a new estimation of totals without examination of the data. This process is repeated until acceptable deviations are found.

#### *Determine key process variables*

For achieving maximum efficiency of the process, the following two sub-processes are considered most important:

1. *Checking and correction of weights,*
2. *Correction, modification of data*

These sub-processes can be monitored with the following measurable indicators:

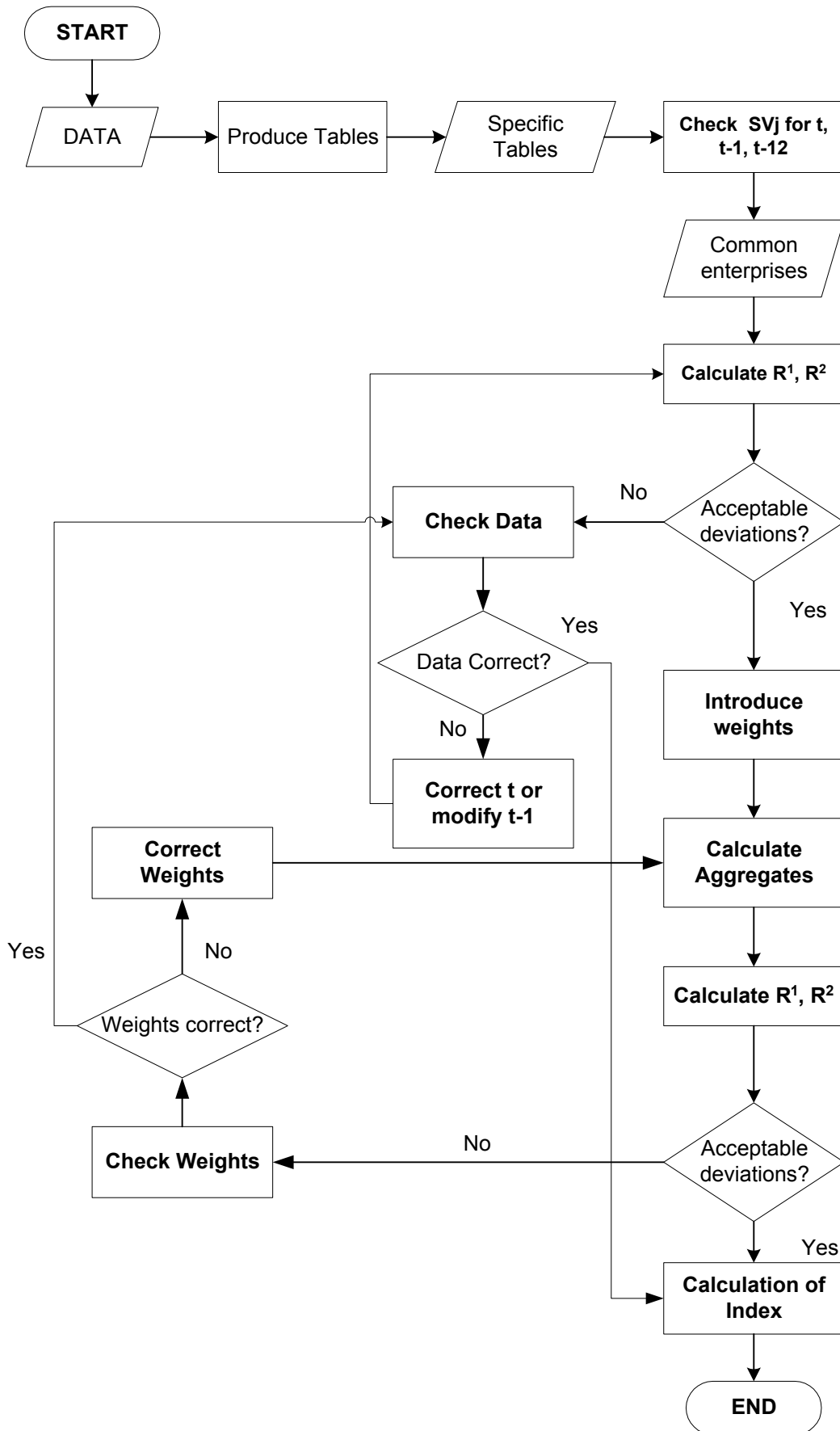
##### *1. Checking and correction of weights*

- i. Time spent in manual examination of weights
- ii. Percentage of errors due to wrong weights with respect to the total number of records.

##### *2. Correction, modification of data*

- i. Percentage of errors detected (at enterprise level) with respect to the total number of records.
- ii. Percentage of enterprises not modified at all,
- iii. Percentage of modifications at time  $t$
- iv. Percentage of modifications at time  $t-1$ .

**Figure 28:** Flow chart of the tabulation process, NSSG





## ***Evaluation***

The dimension of time is very important in this process. With the derived key process variables the analyst is able to produce clusters of error sources. Errors at time  $t-1$  are more important than at time  $t$  since they correspond to higher costs of verification and processing. Similarly, errors in the weights may reveal frame imperfections unknown even at time  $t-12$ .

Finally, we should note that time effort for this process and number of employees engaged many times in the same survey are potential process variables that could also be investigated. However, their monitoring depends on the institutional management policies and thus do not fall into clearly defined frameworks (eg if employee mobility rates between departments is low the latter potential process variable does not apply).

### 3.3.5.2 Literature review

Though the literature on quality improvement methods for Analysis of Primary Outputs is scarce, the following papers provided guidance for writing this report.

Banim J (2000) *An Assessment Of Macro Editing Methods*, UN/ECE Work Session on Statistical Data Editing:

<http://www.unece.org/stats/documents/2000/10/sde/7.e.pdf>

This paper makes a short introduction to various macro-editing techniques and reports on the application of the Hidioglou-Berthelot and Aggregate methods to Ireland's Annual Services Inquiry.

Revilla P (2002) *An E&I method based on time series modelling designed to improve timeliness*, UNECE Work Session on Statistical Data Editing:

<http://www.unece.org/stats/documents/2002/05/sde/13.e.pdf> .

The author describes an edit and imputation method, based on time series modelling, that improves timeliness of public statistics. Two examples of using the method are also presented.

### **3.3.6 Time Series Analysis**

#### **3.3.6.1 Reviewing Seasonal Adjustment (ONS)**

##### **Introduction**

This report focuses on time series analysis. There are four component activities in this SVC group: interpolation, seasonal adjustment, trend analysis and extrapolation, and documentation and reporting of quality.

The literature on applying Continuous Quality Improvement methods to time series analysis is very scarce. We have been unable to find any references on this topic; therefore there is no literature review in this chapter.

##### **Summary of the report**

This example relates to the seasonal adjustment aspect of time series analysis. We consider the annual seasonal adjustment review program carried out by the Time Series Analysis Branch in ONS Methodology Group. This review program re-analyses the use of seasonal adjustment across series published by ONS, and investigates whether seasonal adjustment is appropriate for series that are not currently seasonally adjusted.

##### ***Identifying key process variables***

The annual seasonal adjustment review program grew out of a recognition that there was a wide variety in the quality of seasonal adjustment in ONS outputs. In 1999 the Time Series Analysis Branch decided to undertake a comprehensive quality audit of all seasonal adjustment in ONS as a first step towards improving quality. The objectives of the audit were to assess the quality of seasonal adjustment in each branch, to identify how and where improvements were needed and to re-assess the approach to seasonal adjustment in ONS.

The key process variable here is closeness to optimal seasonal adjustment. A good quality seasonal adjustment allows for valid comparisons of movements over time.

##### ***Measuring key process variables***

To assess the quality of seasonal adjustment, each branch was given a grade from A – E, defined as follows:

A – Very Good

B – Good

C – Acceptable

D – Poor

E – Very Poor

The grades measure how close each branch got to optimally seasonally adjusting their data, and therefore give an indication of how much scope there is for improvement. They are assigned by members of the Time Series Analysis Branch according to the following criteria (note that in the below X11 ARIMA refers to the standard software used for seasonal adjustment in ONS during the time of the review):

Grade A: Very Good. Have made good technical use of X11ARIMA to achieve very good seasonal adjustments. Program options have been used appropriately and reviewed during the last year. There is clear evidence that knowledge of individual series has been used to enhance the quality of seasonal adjustment.

Grade B: Good. Evidence of some scrutiny of the data, that key features have been recognised and appropriate action taken. No significant errors.

Grade C: Acceptable. Nothing major that has been done is in itself wrong, but the seasonal adjustment is characterised by an over dependence on a standard set of options with limited knowledge or use of other options. Limited use of knowledge of the series to improve quality.

Grade D: Poor. Major features of the data or the X11ARIMA output have gone unrecognised and not been acted upon, or inappropriate action has been taken.

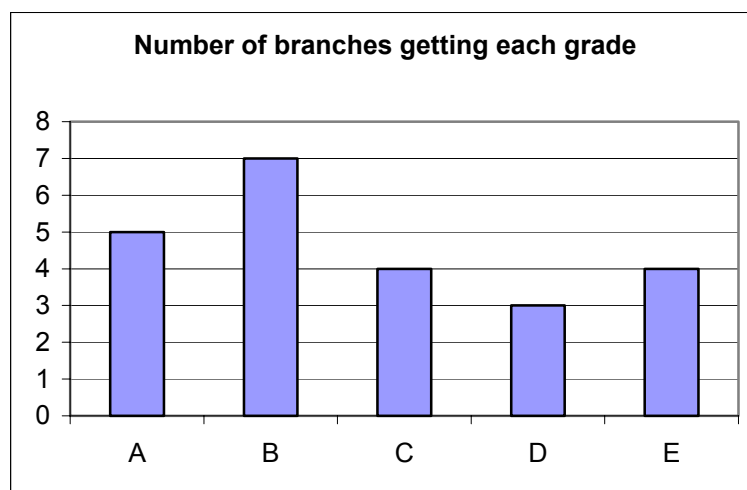
Grade E: Very Poor. Incorrect or inappropriate use of X11ARIMA. The misuse of the package is serious enough that the process of seasonal adjustment detracts rather than adds to the value of the data.

The definitions above come from an internal ONS report describing the audit carried out in 1999. Although assigning grades in this way is subjective, the report states that assigning grades is usually easy, as the distinction between good and bad is very obvious. In the 1999 audit, none of the branches questioned the grades they were assigned.

### ***Analysing key process variables***

The results from the initial audit of the quality of seasonal adjustment in ONS are given in figure 29.

**Figure 29:** Bar Chart of the number of branches achieving each grade in the audit of seasonal adjustment



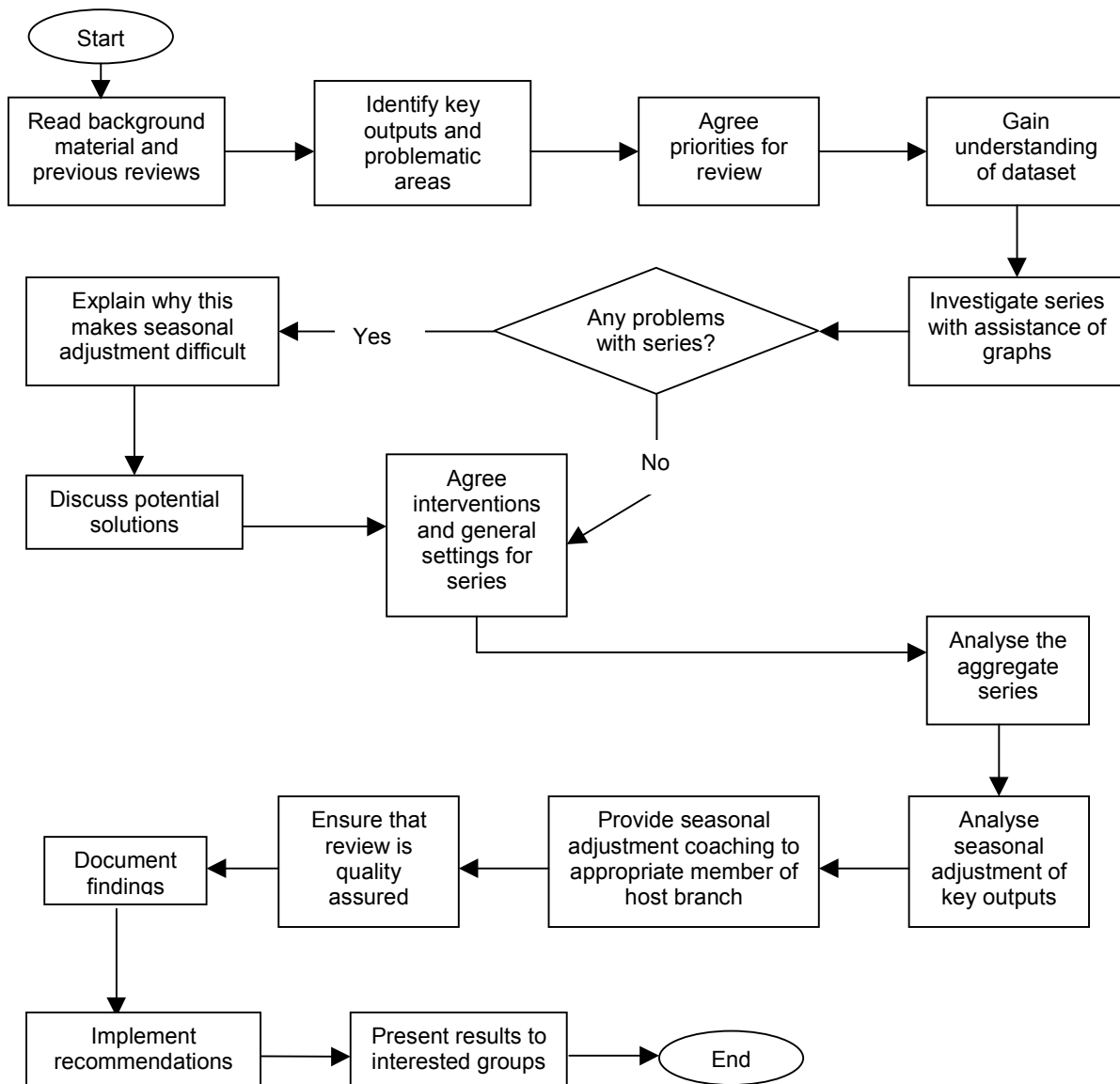
The conclusion drawn from this was that the quality of seasonal adjustment in ONS was too variable. In many branches the quality was unacceptably low.

At the time of the initial audit, ONS had a decentralised approach to seasonal adjustment. The areas that compiled statistics were also responsible for the seasonal adjustment of those statistics. The audit found that very few of the branches in ONS had sufficient expertise in seasonal adjustment or knowledge of X11ARIMA to carry out this role effectively. Given the findings of the audit, a more centralised approach to seasonal adjustment in ONS was proposed and subsequently introduced. Seasonal adjustment

remained part of branch production systems, but Time Series Analysis Branch took on the responsibility for carrying out annual reviews of seasonal adjustment in each branch.

In April 2000 Time Series Analysis Branch started the first round of seasonal adjustment reviews. The reviews involve re-analysing published seasonally adjusted series by running them through X11ARIMA and X12ARIMA to check their parameter settings. Attention is given to the treatment of any calendar effects present and the quality of the seasonal adjustments being performed. Seasonal series that are not currently seasonally adjusted are also examined, and their suitability for seasonal adjustment is assessed. Figure 30 contains a process map, which explains the review program in more detail.

**Figure 30:** Flow chart of the seasonal adjustment review process, ONS



## ***Evaluation***

The change of approach to seasonal adjustment in ONS has led to an improvement in quality. Table 16 shows the number of branches assigned each grade at the time of the audit and in July 2002.

**Table 16:** Table of the number of branches achieving each grade in the seasonal adjustment review, during 1999 and 2002

<b>Grade</b>	<b>Number of branches (August 1999)</b>	<b>Number of branches (July 2002)</b>
A	5	15
B	7	1
C	4	7
D	3	1
E	4	1

After the success of the review of seasonal adjustment in ONS, Time Series Analysis Branch recently undertook an audit of seasonal adjustment in government departments outside of ONS. The results showed large variability in the quality of seasonal adjustment between the different departments. Time Series Analysis Branch have proposed that they should strengthen their advisory role in the field of seasonal adjustment across these departments.

### **3.3.7 Confidentiality and Disclosure**

#### **3.3.7.1 Disclosure (NSSG)**

##### **Description of the process**

Statistical agencies routinely publish data in microdata files for public use and tables. The agencies implement statistical methods before dissemination, which provide protection to the respondents or simply restrict data access to authorised persons only. The aim of this report is to identify the process variables of the disclosure control process.

##### **Summary of the report**

A real life example could not be provided since, at present, NSSG does not implement any statistical disclosure control method, but simply avoids publishing data that may reveal the respondent's identity. (A special working group named Disclosure Review Panel is in charge). Therefore, a theoretical description of the process is provided and, through this example, potential key process variables are proposed.

##### ***Identifying key process variables***

###### *Identify critical product characteristics*

National Statistical Institutes, which play a dominant role in the dissemination of statistical data, need to be very careful in releasing statistical data for use outside the Institutes. They should take technical measures in order to prevent the identification of individual responses. The efficiency of the disclosure control methods is therefore a critical product characteristic.

The critical product characteristic and the assessment of its quality are directly related to the variation of the key process variables. However, in the case of disclosure control, the assessment of quality is not directly related to a measurable statistical asset. To clarify the above concepts a process map illustrating the steps made by statistical offices is presented below. This process map will be used for delineating the potential process variables.

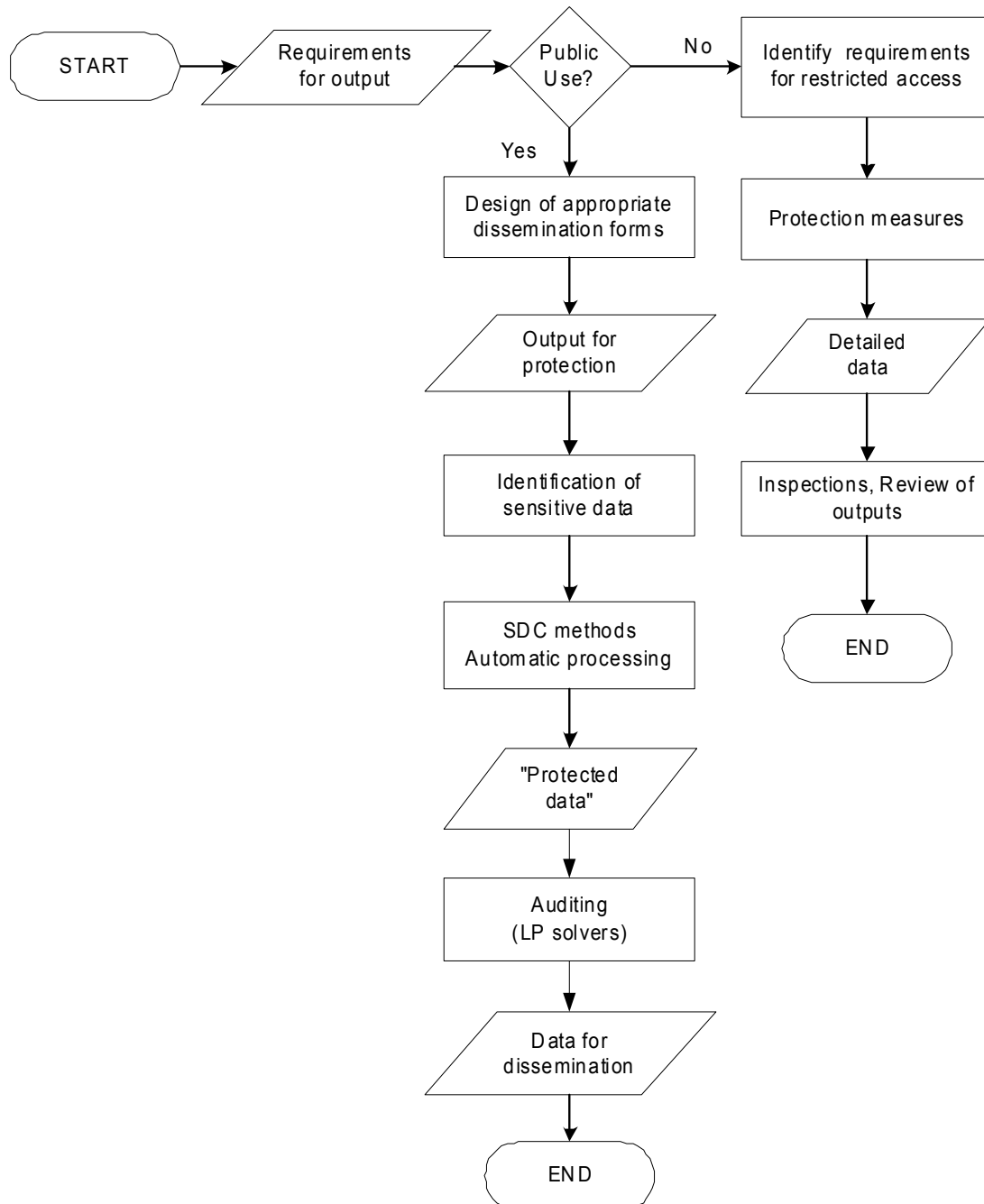
###### *Develop a process flow map*

Data (microdata or tables) can be protected either by restricting information before dissemination or by releasing it only to authorized persons. Though we try to describe both approaches the emphasis will be placed on the former. Figure 31 depicts a generic series of disclosure control activities, and can be useful in deriving process variables.

As can be seen the flow map comprises of two distinct groups of activities. The column on the left regards the sub-process related to producing data that are available to the general public. The actual final user is unknown to the statistician. The column on the right concerns the sub-processes related to provision of access under specific terms and conditions to a clearly defined group of users.

Overall, the whole sequence of disclosure activities, from the specification of the requirements to the generation of the output, can take two different paths. The description of these paths is provided below, along with a discussion on the identification of potential process variables.

**Figure 31:** Flow chart of disclosure control activities, NSSG





## *Determine key process variables*

As a first step statistical offices should identify the user requirements for the output. These depend on the expected usage of the data, i.e. data either for public or research needs. A distinction is made between these two categories since they usually correspond to different levels of security. Data for public use are public-use microdata files and, more often, tables. Data for research are usually detailed data with a low level of distortion.

Regarding data intended for public use, the first step is the “Design of the appropriate dissemination forms”. This is a tabulation of the data or a design of the requested microdata files.

The next sub-process is the “Identification of sensitive data”. In this process the methods used to identify sensitive information must be determined. These methods depend on the type of data, eg frequency counts or magnitudes. Subsequently, the appropriate disclosure control methods are applied to the data. Statistical offices usually limit their practices to the more easily applicable disclosure control methods. Such preferred methods are cell suppression, rounding, table redesign etc.

Quite often there are cases where the disclosure control methods fail to guarantee the required level of protection. For example, cell suppression patterns, which are often determined through heuristics methods, cannot always guarantee protection. In such cases the NSI ought to examine the data to assess its “degree of protection”. This process is called Disclosure Auditing and is intended to evaluate the results of the disclosure activities. Unfortunately, there is currently no consensus on this issue. In the literature, there are potentially useful techniques for testing the level of protection based on linear programming.

Towards this goal the CDAC (Confidentiality and Data Access Committee of the Federal Committee on Statistical Methodology) launched a project in 1999. Specifically, the project aimed to develop a user-friendly suppression audit program written in SAS. In order to measure the quality of published data, the software was designed to calculate protection ranges for primary and secondary suppressions. The comparison of the ranges with the true cell values provides the user with an audit “tool” (see [3]). **A potential process variable for this sub-process could be the number of cases where the SDC methods failed to protect the data.**

Another tool proposed by the Federal Committee of Statistical Methodology is the “Checklist”. It consists of a series of questions that are designed to assist an agency to *determine the suitability of releasing disclosure-limited data* products. The Checklist cannot provide all the information needed in order to decide whether a data product is adequately protected or not, but it can save time and money if it used early enough in the pre-dissemination phase. **Potential process variables could be indicators derived by the analysis of the answers to the checklist, eg the ratio of “negative” over “positive” responses.**

Regarding data for research purposes, methods called “restricted access methods” are employed. They are divided into 2 categories: licensing and research data centers. In licensing, the statistical office disseminates confidential data to researchers (for use in their private centers). Research data centers provide in their premises the facility to work on detailed data. First the requirements of the restricted access method should be identified. These include establishing procedures, legal agreements and all necessary preparation for each method. The next step is to determine the protection measures to ensure the protection of the respondent. These involve data security plans, research project proposals etc. Finally the data providers should inspect whether the researchers

have complied with the agreements made and whether there is a breach of confidentiality. **A potential process variable is the number of impingements to the total number of inspections made.**

### ***Evaluation***

We have shown that derivation of process variables in this abstract concept is possible. Further adaptations to specific survey instances are necessary under this general framework.

### **3.3.7.2 Literature Review**

Doyle P, Lane J, Theeuwes J and Zayatz L (2001) *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*, Elsevier.

This book provides a review of new research in the area of confidentiality and statistical disclosure techniques. It presents information on the different approaches taken by statistical agencies in disseminating data and provides a survey of what statistical disclosure techniques are used by statistical agencies. There is also a series of chapters on public perceptions of statistical agency actions.

Federal Committee on Statistical Methodology (1999) *Checklist on Disclosure Potential of Proposed Data Releases*, Office of Management and Budget, USA.  
<http://www.fcsm.gov/committees/cdac/resources.html>

The Checklist consists of a series of questions that are designed to determine the suitability for release of data (microdata and tabular data).

Zarate A, Bournazian J and Wolf V (2000) *Integrating Federal Statistical Information and Processes*, paper presented at the Federal Committee for Statistical Methods Statistical Policy Seminar November 2000.

This paper describes activities of the Confidentiality and Data Access Committee (CDAC), for example the “Checklist” and the development of an auditing software.

### 3.4 Conclusion

The preceding sub-section contains 12 reports on applying the Continuous Quality Improvement approach to statistical processes. This work has led to interesting findings and improvements in statistical process quality. These findings are summarised below. We consider the use of Statistical Process Control techniques, process variables in the context of the ESS Quality Dimensions, and some possible further work.

The processes examined fall into two categories:

- those amenable to the whole quality approach, including identifying measurable process variables;
- those where mapping the process and identifying key factors may lead to valuable quality improvement actions.

These are described in more detail below.

#### **Areas where measurable process variables are easily identified**

The process quality approach seems particularly appropriate for data collection and data processing. This may be due to the fact that these processes deal with individuals or 'sample units' as opposed to aggregates. This gives rise to a large volume of data from which we can derive useful process variables.

To identify problems and to improve quality, special 'one-off' studies examining these processes (eg for different types of errors) are often carried out. However the repetitive nature of the processes creates scope for continuous improvement throughout the production cycle. Evidence of this is seen in example 3.4 on coding quality, where continuous measurement and analysis of process variables seems to lead to continuous improvement in the consistency and accuracy of coding.

In many cases the process variables identified could also be considered as quality indicators (or 'product variables'). That is, they are indicative of the overall quality of the statistic, and of interest to users. For example the data collection reports of chapter 3.3.1 suggest that non-response, a commonly published quality indicator, is a useful process variable. In chapter 3.3.4 on weighting and estimation, a measure of non-response bias was used as a process variable to identify problem domains and to assess the effectiveness of 'follow-up' telephone interviews.

#### **Areas where more general quality improvement ideas and tools are useful**

In contrast, the process variable approach proved to be less straightforward to apply in other areas of the Statistical Value Chain. Tabulation, seasonal adjustment and disclosure deal with aggregates of data, and are performed less frequently during the production cycle. Although one or more variables are defined for each process in this handbook, the examples are not fully developed, reflecting the difficulties found in applying the approach. However process maps were found to be extremely useful for describing the detail involved in these complicated processes.

Accessing administrative data is an area of current interest to NSIs, and was examined in example 2.1. Identifying critical product characteristics using a survey, and representing

these on cause-and-effect diagrams led to the identification of several valuable improvement actions. This is in spite of the fact that no measurable process variables were identified.

### **Statistical Process Control (SPC)**

SPC concepts involve process stability and capability (as described in sections 2.3.5 and 2.3.6), and make use of control charts. Even in the areas where measurable process variables exist, it is difficult to find an example from within an NSI where SPC concepts have been applied. This may be due to a number of reasons, for example:

- insufficient data have been collected, as in the case of a new process or a long production cycle;
- unacceptable variation and its causes are self-evident to survey managers and identifiable without using SPC techniques;
- awareness of SPC techniques and their uses is low.

However there is one example in the handbook (see report 3.3.3.4) which uses a control chart to examine the behaviour of the process. This shows that there is scope to employ the SPC approach for statistical processes.

### **ESS Quality Dimensions**

It is important to relate process variables to the ESS Quality Dimensions of:

- Relevance;
- Accuracy;
- Timeliness;
- Accessibility and Clarity;
- Comparability;
- Coherence.

Timeliness and accuracy are relatively easy to consider, as they can vary throughout the production process. The reports in the previous sub-section contain several examples of process variables related to accuracy. For example coding 'accuracy' rates, non-response bias, and closeness to optimal seasonal adjustment all reflect the accuracy of the resultant statistics to some extent. Timeliness is considered in report 3.3.1.3, in terms of the time taken up by different components of an interviewer's work.

'Accessibility and clarity' has not been considered in depth in this handbook. In theory, at the dissemination stage it should be possible to derive process variables related to this dimension. For example the number of queries about where to obtain data, and the number of 'hits' for data available on the internet indicate how successful we are in providing access to statistics.

The remaining dimensions of relevance, comparability and coherence are perhaps the most difficult to relate to specific processes and therefore process variables. This is because they involve higher-level concepts such as the definitions, classifications, and methods underpinning the statistics. These are not likely to vary throughout (or between) production cycles. Overarching *product* variables are more appropriate here, for example adherence to international standards is a useful indicator of comparability.

There are significant differences in the quality improvement actions we can implement for different dimensions. For example, consider how we may improve the accuracy of coding by making changes to the parameters of an automated tool, compared to how we may improve the comparability of a statistic by adjusting the classifications used. The former quality improvement action may take place on a day-to-day basis at the discretion of a survey manager, whereas the latter may require the involvement of higher management and even the revision of a classifications manual. The use of process variables to inform *continuous* improvement is limited in the latter case.

### **Future work**

Future work could include addressing the gaps in this handbook in terms of applying the methods to some areas of the Statistical Value Chain, as described in sub-section 3.1. For example, it would be interesting to see how processes involved in sample design, and in the dissemination of statistics could be improved by identifying and monitoring process variables. More extensive use of SPC techniques could prove beneficial to quality improvement efforts.

Future work could also address the difficulties regarding the ESS Quality Dimensions, by considering in more detail whether the approach is useful for dimensions other than timeliness and accuracy.

## **Annex 1 – Glossary of Key Concepts in Process Quality**

There are an abundance of ideas and approaches to process quality and quality in general. Colledge and March (1993) stress that, although there are 'several prophets' of Quality Management, the basic ideas are all the same. This glossary will define some of the main concepts and definitions relevant to process quality, and some of the history of this theory. Further information on different Quality Management approaches is provided in the seventh chapter of the LEG report (Eurostat (2002)).

### **Key process variables**

Key process variables are those factors that can vary with each repetition of the process and have the largest effect on critical product characteristics, i.e. those characteristics that best indicate the quality of the product.

As is clear from the handbook, it is not as easy to identify and measure process variables in statistical production as in industrial production. But process quality concepts (some described below) and associated methods are of equal importance in both areas.

### **Quality Indicators**

Quality indicators are statistical measures that give an indication of output quality. Examples are estimated standard errors and response rates, which relate specifically to the accuracy of the output. Quality indicators differ from process variables, which give an indication of the quality of the process. However, some quality indicators can also give an indication of process quality. Response rates are an example of this.

### **Quality Assurance**

Quality Assurance (QA) is an organisation's guarantee that the product or service it offers meets the accepted quality standards. QA is achieved by identifying what 'quality' means in context; specifying methods by which its presence can be ensured; and specifying ways in which it can be measured to ensure conformance. This leads to the next definition of Quality Control (QC).

### **Quality Control - QC**

Quality Control (QC) aims to achieve a specified average outgoing quality limit. In other words, it is a technique used to check quality against a set standard or specification. QC ensures that a specified measurement from a component in a process stays within accepted tolerances (acceptable variations which stray from the optimum).

QC requires constant inspection throughout the process in order to detect components that are not up to the required standard. Often, these inspections are also carried out on the completion of the process or product by trained inspectors.

## **Total Quality Management - TQM**

Total Quality Management (TQM) is a management philosophy that is driven by customer needs and expectations. TQM aims to create a Quality Culture, and is based on a number of core values such as: customer orientation; leadership; participation of all staff; process orientation; teamwork; staff development; and continuous improvement.

Quality management did of course exist beforehand, including the QA and QC concepts above, but the TQM approach was a revolution in the field. This revolution was sparked by the teachings of Deming and subsequent events in Japan in the mid-20<sup>th</sup> century. As Deming (1991) describes, before 1950 Japanese consumer goods had earned a world-wide reputation for being 'shoddy and cheap'. In 1948 and 1949, Japanese engineers studying literature on quality control observed that improvement in quality leads to an improvement of productivity. Top management in numerous companies came to realise that quality is vital for export. One of these organisations, the Union of Japanese Science and Engineering (JUSE) brought in Deming as an expert. His visit was followed by several conferences with higher management in Japan. As a result, from 1950 the quality of Japanese goods improved markedly, and by 1954 they had captured global markets.

Although TQM was first applied in manufacturing, the concepts are equally valuable in a service organisation such as an NSI. Deming explains how 'the principles and methods for improvement are the same for service as for manufacturing. The actual application differs, of course, from one product to another, and from one type of service to another, just as all manufacturing concerns differ from one to another.'

The objective of TQM is to enable the organisation to deliver products with continuously improving quality, which leads us to the next definition.

## **Continuous Quality Improvement - CQI**

Continuous Quality Improvement (CQI) is a part of TQM theory, and its history began in industry as explained above.

CQI theory differs from QC in the sense that the latter aims to achieve a specified average outgoing quality limit, whereas CQI also aims to achieve the smallest error rate possible through continually improving quality for the duration of an operation. This is achieved by adopting new activities and eliminating those that are found to add little or no value. The goal is to increase effectiveness by reducing inefficiencies, frustrations, and waste. CQI makes extensive use of the methods in Statistical Process Control (SPC), described below.

## **Statistical Process Control - SPC**

Statistical Process Control (SPC) is a methodology that uses a set of tools to identify and analyse causes of variation in a process. Its aims are to establish process stability (a state where the process variation consists entirely of random components), and subsequently process capability (ability to meet specifications).

Some of the tools used are described by Colledge and March (1993). They explain that, 'basic to the notion of continuous improvement is the capacity to monitor processes and to measure the quality of products and the effects of changes on quality. Examples of measurement techniques are Shewart charts for monitoring and detecting when processes go "out of control", cause-effect fish-bone diagrams, and Pareto analysis for analysing and prioritising problems.'



## **Six Sigma**

More recently, in the early and mid-1980s Motorola engineers decided to measure defects per million opportunities (an opportunity meaning a chance for a defect) rather than the traditional quality levels of defects per thousands of opportunities. Motorola developed this new standard and created the methodology and the cultural change associated with it. They named the approach 'Six Sigma'. Since then, many companies around the world have adopted the Six Sigma approach to their business.

Six Sigma is a disciplined methodology for eliminating defects in a process, aiming for six standard deviations or 'Sigmas' between the mean and the nearest specification limit. This involves a measurement-based strategy that focuses on process improvement and variation reduction.

## **The European Foundation for Quality Management - EFQM**

The European Foundation for Quality Management (EFQM) helps organisations throughout Europe to participate in quality improvement activities and accelerate the process of Total Quality Management. The foundation has developed the EFQM Excellence Model for assessing organisational excellence. They use this model to judge the leading organisations in the annual European Quality Award.

Two of the fundamental concepts of organisational excellence defined by the EFQM are 'management by processes and facts' and 'continuous learning, innovation and improvement'.

Jeskanen-Sundström (2003) has described the use of the EFQM (and other management systems) in Statistics Finland, which has had encouraging results.

## **International Organisation for Standardization (ISO)**

The International Organisation for Standardization (ISO) is a network of national standards institutes from 148 countries working in partnership with international organisations, governments, industry, business and consumer representatives. National delegations agree on ISO standards, which help improve the efficiency and safety of products and services. ISO also produces guideline documents on specific standards, one of which covers statistical methods for quality control (see ISO (2000)).

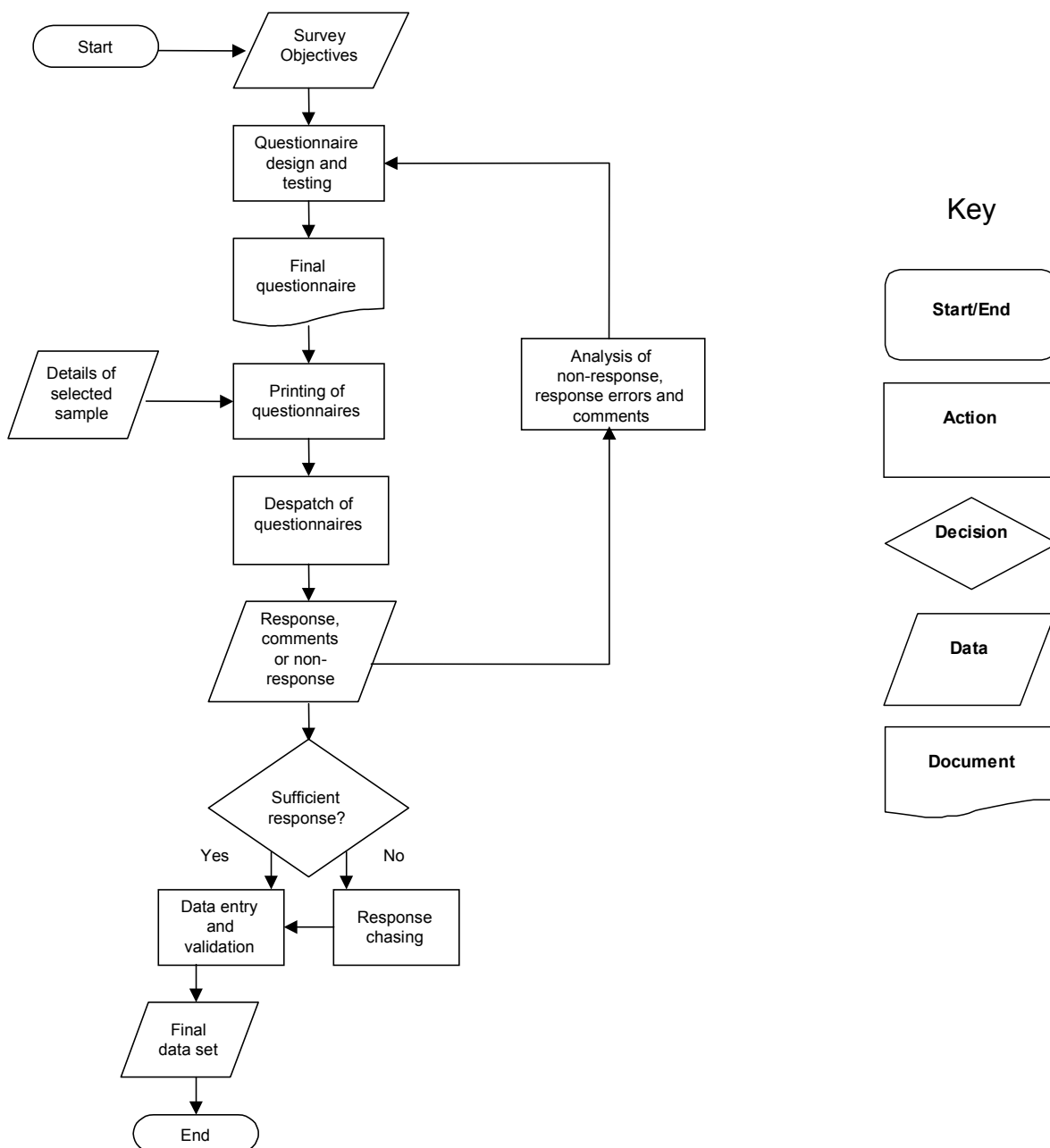
## Annex 2 – Flow Charts

**Use** - To allow a team to identify the actual flow or sequence of events in a process that any product or service follows.

The flow chart lists the order of activities. Different symbols have specified meanings. Figure 32 below shows an example, taken from report 3.3.1.1. It presents a fairly high-level flow chart, which aims to follow the typical procedure for data collection used within the ONS. The key on the right indicates the meaning of the symbols, and follows the standard set of flowchart symbols. These symbols are used in section 3 with a consistent meaning.

Other useful information to add to a flow chart are details of who 'owns' each step, and what the associated process variables are.

**Figure 32:** Flow Chart of part of the Data Collection process, with key to symbols



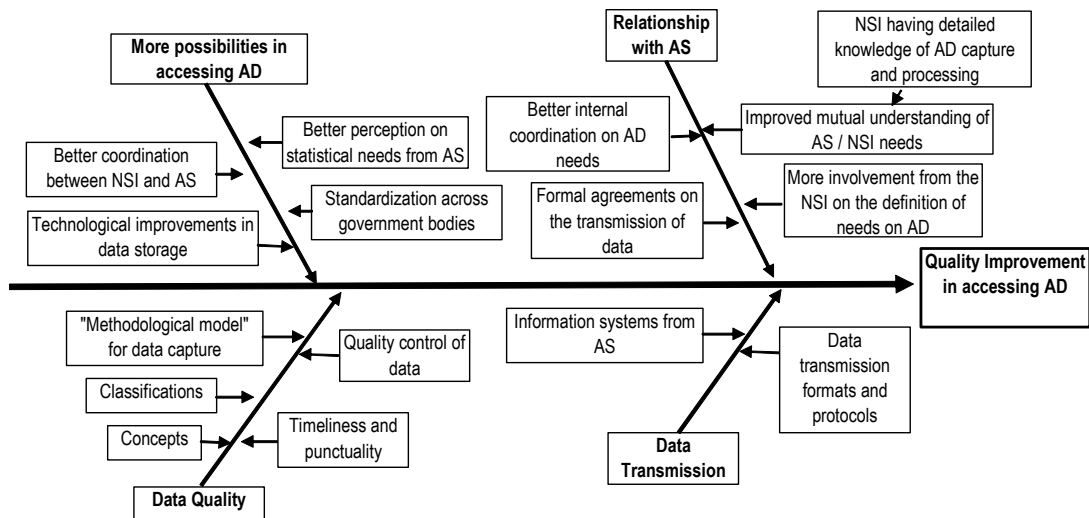
## Annex 3 – Cause and Effect Diagrams

**Use** - To allow a team to identify, explore, and graphically display, in increasing detail, all of the possible causes related to a problem or condition to discover its root cause(s). The diagram facilitates the identification of key process variables when frequency or other types of data are not available.

Because of its appearance, the cause and effect diagram is also called the fishbone chart. Another common term used is the Ishikawa chart, after Kaoru Ishikawa, who popularised the use of the chart in Japan (see Ishikawa (1976)). Its most frequent use is to list the cause of particular problems. The lines coming off the core horizontal line are the main causes and the lines coming off those are sub causes. From all of the factors on the fishbone, the five or six believed to be most important are the factors to measure and whose variability should be reduced.

The example in figure 33 is taken from report 3.3.2.1. It is a fishbone diagram describing the factors that affect the process of accessing administrative data (AD) from administrative sources (AS).

**Figure 33:**  
Cause-effect Diagram - Quality Improvement in accessing to Administrative Data



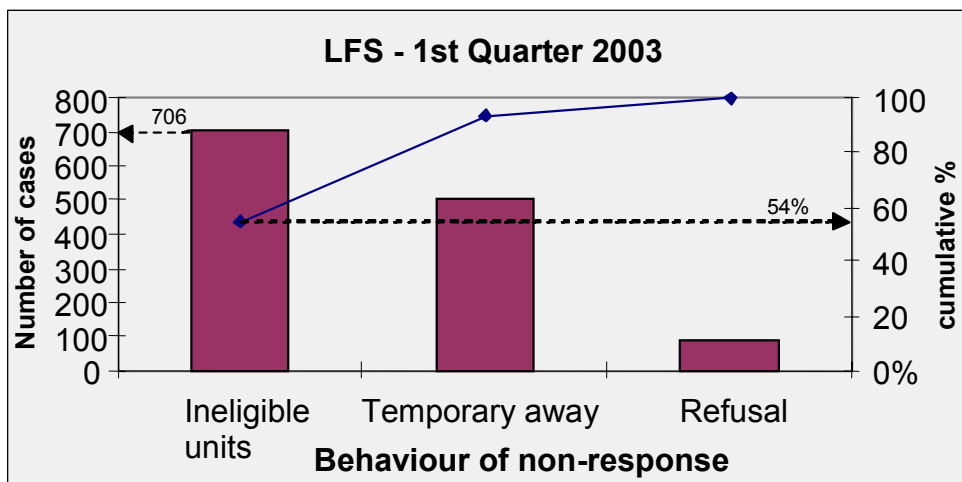
## Annex 4 - Pareto Charts

**Use** - To focus efforts on the problems that offer the greatest potential for improvement by showing their relative frequency or size in a descending bar graph.

The Pareto chart shows the distribution of items and arranges them from the most frequent to the least frequent, often with a final bar grouping miscellaneous items. The tool is named after Wilfredo Pareto, the Italian economist who determined that wealth is not evenly distributed: some of the people have most of the money. This tool is a graphical picture of the most frequent causes of a particular problem. It shows where to put your initial effort to get the most gain, hence improving effectiveness in deciding how to allocate resources.

The example below is taken from report 3.3.1.2. The chart displays the number of cases of different types of non-response for the Labour Force Survey (LFS) in a particular region of Portugal. The Pareto chart enables the easy identification of the vital non-response type: 'lost' units account for 54% of non-response. It is important to investigate why this rate is so high and take action to improve it.

**Figure 34:** Non-response behaviour for the Labour Force Survey, Lisboa e Vale do Tejo Region, first quarter of 2003.



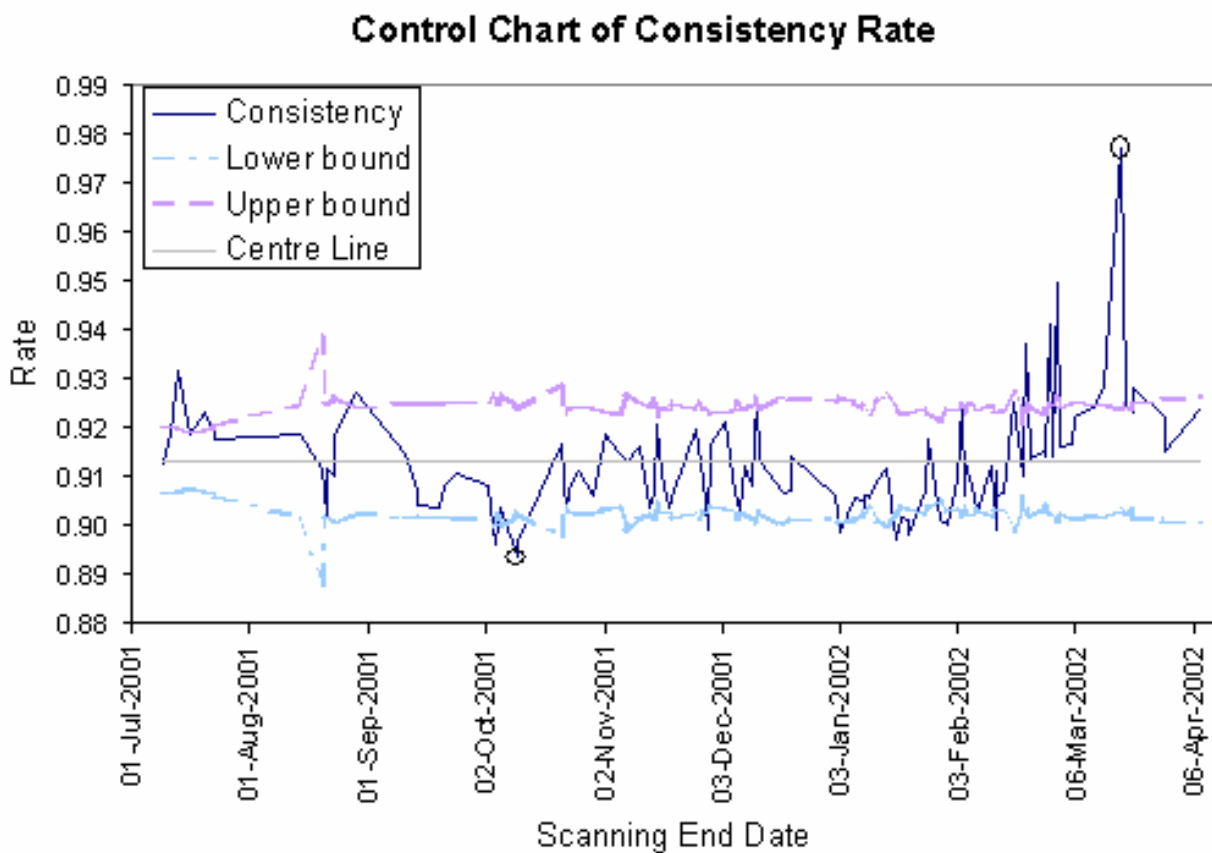
## Annex 5 - Control Charts

**Use** - To monitor, control, and improve process performance over time by studying variation and its source.

### Methodology of control charts

As described in sub-sections 2.3.5 and 2.3.6, the control chart is a line chart with mathematically constructed control limits. An example of a control chart is given below in Figure 35, adapted from report 3.3.3.4.

Figure 35:



By mathematically constructing control limits at three standard deviations above and below the average, we can then see which points fall outside the limits (i.e. are out of statistical control). Some such points are circled in Figure 35. Action should then be initiated to look for possible causes for this behaviour. We can determine what variation is due to normal ongoing causes (common causes) and what variation is produced by unique events (special causes), which should then be eliminated. A feature of this particular control chart is that there appears to have been a shift in the data around February 2002. It is important to examine the causes of such changes as they may affect the analysis.

After establishing statistical control, we can examine the process capability. The aim is to reduce common causes of variability, thus improving the quality of the process.

Further details on how to construct and interpret control charts are given in a technical annex.

# Technical Annex: Control Chart Methodology

## Introduction

Control charts are important tools, proven to be essential to quality control plans by highlighting unusual or unsatisfactory performances, detecting process instabilities, and providing a basis on which to reach decisions about a process.

Previous references to control charts are found in sub-sections 2.3.5, 2.3.6 and Annex 5. These mention the uses and limitations of the tool. This technical annex provides basic detail and references for the construction and interpretation of control charts, in sections A and B respectively.

Control charts are formed from simpler charts known as run charts. A run chart is a visual display of data collected at different time points. It is simply the raw data, for each time interval, plotted in time-order on a chart. They are used to show trends and shifts in a process over time, variation over time, or to identify decline or improvement in a process over time.

To create a control chart, a centre line (CL) and upper and lower control limits (UCL and LCL respectively) must be added. We can use historical 'baseline' data to calculate the CL, LCL and UCL. Baseline data are used to test the null hypothesis that the process is in control against the alternative that the process is out of control.

## A. Constructing Control Charts

### Types of Control Charts

Different types of control charts apply to different types of data. Data can be split into 'variable' and 'attribute' data. Calculation of the CL is similar for all types of chart, whilst calculation of the LCL and UCL depends on the type of chart.

'Variable' data are usually measurements such as length, time, or some count of quantity produced (eg number of forms processed in a specific time). 'Variable' data are analysed using variable control charts. These include the:  $\bar{X}$  chart to control individual values;  $\bar{\bar{X}}$  chart to control the process average;  $R$  chart to control the range;  $S$  chart to control standard deviations; and  $S^2$  chart to control the variance.

'Attribute' charts are used when the data are measured to meet certain conditions. That is items are compared with some standard and are then classified as to whether they meet the standard or not. Any item that does not meet the standard is considered to be a 'defective' or 'non-conforming' unit. This type of chart could be used for error rates and response rates. There may be situations where a single unit has more than one defect. Control charts for attribute data include the:  $p$  chart to control fraction defectives in a sample;  $np$  to control the number of defective units;  $u$  chart to control the number of defects per unit produced; and  $c$  chart to control the number of defects.

This section of the annex provides brief details on how to construct the CL, LCL and UCL for each of the above types of chart. Other considerations not covered in this annex include the process capability index, cumulative sum and moving average charts. Detailed explanations of these are given in Tapiero (1995).

The methodology described below was informed by the references: Burr (1976); Bowerman and O'Connell (1989); Hapuarachchi, March and Wronski (1997); Reed and Reed (1997); Sytsma and Manley (1999); and Tapiero (1995), which is the main reference.

### **General notes on construction of the centre line and control limits**

The first considerations for constructing a control chart are the sample size and sampling frequency of observations to input to the chart. At each time-point, a subgroup (a set of observations obtained while process conditions are not allowed to change substantially) of one or more observations is obtained. Subgroups should be chosen to allow comparison between possible sources of variation, for example, municipalities, interviewers, surveys, etc. Time-points for sampling should be frequent enough to prevent the process from being out of control for too long. A large overall sample size improves the statistical quality of estimates, and justifies use of the normal distribution. Tapiero (1995) provides further details.

Once we know the type of data and sampling, it is important to construct the chart appropriately, using sensible baseline data and the right formulae for the CL and control limits. Bowerman and O'Connell (1989) advise having at least 25 data points in the baseline time period. If there are less than 25 points available then we should use the available data with caution. The baseline data (which influence the CL and control limits) may need to be updated if a statistically significant change occurs (see Section B on the interpretation of control charts for more details).

In general, the CL is plotted as the mean of all data points from the specified baseline period. Control limits are plotted at say  $d$  standard deviations (sds) above and below the CL. The parameter  $d$  is based on the risk used in constructing the chart and an appropriate statistical distribution. In general,  $d=3$  is used. But calculation of the sd varies between types of control chart, with details given in sub-sections by chart type below. In some cases only the baseline data contributes to the sd, whilst for other types of chart the sd is calculated separately for each time-point.

After the baseline period, at each time-point new data will become available to plot. In general the CL is not re-calculated, but is extended to this new time-point. The LCL and UCL are also extended, or in the case of  $p$  and  $u$  charts re-calculated using the new sample size.

### **Control charts for variable data**

Suppose that a random sample or subgroup of size  $n$  of process variable observations is taken at each of  $k$  time periods. If the sample is sufficiently large, the law of large numbers tells us that the overall average, say  $\bar{X}$ , is normally distributed, say with mean  $\mu$  and standard deviation  $\sigma/\sqrt{n}$ . Then there is a probability of over 0.99 that the average of a random sample of size  $n$  from the population will be in the interval

$$\left[ \bar{X} - 3 \frac{\sigma}{\sqrt{n}}, \bar{X} + 3 \frac{\sigma}{\sqrt{n}} \right].$$

As explained above, the general approach is to first estimate the mean or other parameter of interest and its sd, and then to set the LCL and UCL at three sds from the parameter estimate.

### *X* or 'individuals' chart

An individuals chart is used when only a single observation per time period is taken (i.e.  $n=1$ ). This may occur when it is difficult or impossible to group measurements into subgroups so that an estimate of the process variation can be determined. The solution is to artificially create subgroups from adjacent data (most often pairs), and use the 'moving range' to estimate the standard deviation. Tapiero (1995) provides further details on the calculation of the control limits. Note that normality cannot be assumed with individual values, so interpretation should be cautious and the assumption tested.

### $\bar{X}$ chart

Suppose the independent samples at each time-point have multiple observations ( $n > 1$ : some guidelines suggest  $n > 3$ ), and that each sample of  $n_i$  observations is a rational subgroup. Then the  $\bar{X}$  chart monitors changes in the process average, that is, it detects variations between samples.

Suppose  $x_{ij}$  is the  $j^{\text{th}}$  observation from the sample drawn at time  $i$ . Let the sample mean at time  $i$  be  $\bar{x}_i = \sum_j x_{ij} / n_i$  and overall average  $\bar{x}_\bullet = \sum_i \bar{x}_i / k$ . The estimated standard deviation is  $s_i = \sqrt{\sum_j (x_{ij} - \bar{x}_i)^2 / (n_i - 1)}$ . The CL is plotted at  $\bar{x}_\bullet$ , and control limits for the  $\bar{X}$  chart are set at:

$$\left[ \bar{x}_\bullet - 3s_i, \bar{x}_\bullet + 3s_i \right].$$

In practice, the  $\bar{X}$  chart is used together with an  $R$ ,  $S$  or  $S^2$  chart. Alternative methods for estimating the sd when using an  $R$  chart are provided in Bowerman and O'Connell (1989), and when using an  $R$  or an  $S$  chart in Sytsma and Manley (1999).

### $R$ chart

The  $R$  chart provides an indication of the meaningfulness of the  $\bar{X}$  measurements, that is, it detects variations within samples.  $R$  charts use an alternative estimate of the variance, which may also be used in an  $\bar{X}$  chart. Some details on this estimate are provided below, but full details on  $R$  charts are provided in Tapiero (1995) or in Bowerman and O'Connell (1989).

The alternative estimate uses the average range of the samples. Let the range at time  $i$  be  $r_i = \max_j(x_{ij}) - \min_j(x_{ij})$ . The sample ranges  $r_i$  are plotted on an  $R$ -chart, with its CL at the average, denoted by  $\bar{r} = \sum_i r_i / k$ . The LCL and UCL are computed using this average, based on a framework provided by order statistics. Further explanation is provided in Tapiero (1995) or Sytsma and Manley (1999).

### $S$ and $S^2$ charts

When the sample range information is not satisfactory, we use  $S$  and  $S^2$  charts to control the standard deviation and the variance respectively. Tapiero (1995) contains full details. In summary, when the population mean is unknown, the standard sample variance estimate has a chi-squared distribution, which can be used to construct the LCL and UCL for both the  $S$  and  $S^2$  charts.



## Control charts for attribute data

'Attribute' charts are used when the data are measured to meet certain conditions, for example a correct or incorrect classification, or a response or non-response. The control chart is used to determine if the rate of defective product is stable and detect when a deviation from stability has occurred. The argument can be made that an LCL should not exist, since a rate of defective product outside the LCL is in fact a good thing; we want low rates of defective product. However, if we treat these LCL violations as simply another search for an assignable cause, we may learn from the drop in defects rate and be able to permanently improve the process. There are several types of charts available for different types of attributes and sampling schemes. The main four types are described below.

### *p* chart

The *p* chart can be used when the subgroups are not of equal size. The chart measures the ratio of defective items per unit sampled. At time  $i$ , let  $x_i$  be the number of defective units (units with one or more defect) and  $n_i$  be the sample size. Then the  $\bar{p}_i = x_i / n_i$  (estimates of the probability that a unit is defective) are the points plotted on the chart. The

CL is calculated as  $\bar{p} = \frac{\sum_i x_i}{\sum_i n_i}$ .

However the sd (and hence control limits) vary with the sample size, and must be calculated separately at each time point. The variance at time  $i$  is estimated by  $\bar{p}(1 - \bar{p})/n_i$ . Therefore, the LCL and UCL at time  $i$  are given by:

$$\left[ \bar{p} - 3\sqrt{\bar{p}(1 - \bar{p})/n_i}, \bar{p} + 3\sqrt{\bar{p}(1 - \bar{p})/n_i} \right],$$

to provide a 99% confidence interval.

### *np* chart

The *np* chart is used in the special case of equal subgroups, when it is not necessary to convert defective counts into the proportions  $\bar{p}_i$ . Here we can directly plot the counts  $x_i$  against the subgroup number  $i$ .

In this case the LCL and UCL are:

$$\left[ n\bar{p} - 3\sqrt{n\bar{p}(1 - \bar{p})}, n\bar{p} + 3\sqrt{n\bar{p}(1 - \bar{p})} \right],$$

Tapiero (1995) explains how the Binomial or Poisson distributions are used to compute these intervals.

### *c* charts

The *c* chart measures the number of defects per inspection unit, for example per day or per form. This chart can take account of a situation where there are several types of defects with different probabilities. Each defect type can also be weighted by its severity, namely its cost. (Note that, unlike the *p* and *np* charts, the *c* and *u* charts allow for multiple defects per unit).

Assume we have one type of defect only. We plot the points  $c_i$ , which are the number of defects at time  $i$  (as opposed to  $x_i$ , the number of defectives). Then the average number of defects per sample gives the CL,  $\bar{c} = \sum_i c_i / \sum_i n_i$ . Using the Poisson distribution, the LCL and UCL are:

$$\left[ \bar{c} - 3\sqrt{\bar{c}}, \bar{c} + 3\sqrt{\bar{c}} \right].$$

Further details on the use of  $c$  charts are provided in Tapiero (1995).

### *u charts*

The  $u$  chart is used when it is not possible to have an inspection unit of a fixed size of interest, for example a batch of 100 forms or a time-scale of one person-month. When the number of defects per inspection unit is converted into a ratio per standard unit, it may be controlled with a  $u$  chart. Notice that the number no longer has to be integer as with the  $c$  chart.

Using the notation above, the points plotted on the chart are the  $u_i = c_i / n_i$ . The CL is given by the average  $\bar{u} = \sum_i c_i / \sum_i n_i$ , and the LCL and UCL by:

$$\left[ \bar{u} - 3\sqrt{\bar{u}/n_i}, \bar{u} + 3\sqrt{\bar{u}/n_i} \right].$$

### **Control charts for dependent observations**

In some situations, observations may be serially correlated, autocorrelated or display cyclical behaviour. In this case, standard control charts are not applicable. Hapuarachchi and Wronski (1994), and Hapuarachchi, March and Wronski (1997) cite several references for constructing control charts in these various situations. The authors also recommend a method for dealing with data that have autocorrelation structure, seasonality and trend. Their specific dataset contains non-response rates for a survey using a rotation panel design. In summary, their approach is to:

- (a) identify the appropriate time series model that best describes the data;
- (b) estimate the parameters of the model in (a);
- (c) incorporate the model in (a) to construct an appropriate control chart for residuals generated from this model.

There are, however, further complications to analysing the resulting control chart. The authors advise that this method is appropriate for more sophisticated users and not for statistical output managers or operational staff in general.

## ***B. Interpreting Control Charts***

This section explains process variability and capability, providing some guidance and references on how to interpret control charts using these concepts. The last part of the section looks at the particular case of the  $\bar{X}$  chart, which is used together with an  $R$ ,  $S$  or  $S^2$  chart.

### ***Process Variability***

The concept of process variability is central to methods for continuous improvement. As explained in sub-section 2.3.5, process variation can be partitioned into two components. Natural process (or 'common cause') variation is the naturally occurring fluctuation inherent in all processes. Special cause variation is typically caused by some problem or extraordinary occurrence in the system.

The primary use of control charts is to quickly determine when a process is 'out of control' or unstable, that is when a special cause of variation is present due to an unusual occurrence in the process. The process is then investigated to determine the root cause of the out of control condition, and to identify a strategy to address it.

In a typical situation where control charts are employed as part of a continuous improvement effort, at first a process will be highly variable and out of statistical control. Then, as special causes of variation are found, the process comes into statistical control. Finally, through process improvement, variation is reduced.

Identification of out of control conditions is dependent on the control limits on the control chart. If all points lie within the LCL and UCL, then the process is in apparent statistical control, exhibiting only common cause variation. This does not mean that no special causes are present, only that investigating such causes will not result in an improved statistical control of the process. It is important to check for special causes and trends.

Types of out of control situations include: a point lying outside the control limits; numerous consecutive points lying in constricted areas within the limits; runs above or below the CL; linear or cyclical trends. The first type – a point lying below the LCL or above the UCL – is the most obvious out of control condition, which applies to all control charts. It indicates the existence of a special cause of variation, which should be isolated and dealt with. On the other hand, points that are too close to the CL may imply that limits are improperly drawn. Further information on interpretation of these different situations is provided in Tapiero (1995).

Investigation and correction of processes that are out of control is typically a team effort, as described in 3.5.2. Methods of reducing common cause and special cause variation include process improvement techniques, investing in new technology, or reengineering the process to have fewer steps and therefore less variation. Reduced variation makes the process more predictable, with output closer to the desired level.

## **Process Capability**

In mathematical terms, the capability of a process is defined as the inherent variability of a process in the absence of any undesirable special causes. That is the smallest variability of which the process is capable when variability is due to common causes only. As explained in sub-section 2.3.6, we consider capability in terms of customer requirements. Process capability studies are conducted to compare the performance of a controlled process to its requirements. Even if a process is in statistical control, the common cause variation may cause the process to fail in meeting individual product specifications, and so become 'incapable'. The remainder of this section will explain some of the background to assessing capability.

We can create a frequency distribution for the data, and construct our specification limits on the same plot. The specification limits are fundamentally different to the control limits, as they explicitly define requirements imposed on the process by its customers. We may assess capability (meaning whether specifications are being met) by comparing the process spread with the specification spread.

Other methods for assessing capability include: the use of an  $r$  chart to determine inherent variability; and the use of the capability index as a simplified measure describing the relationship between process variability and specification spread. These methods are described in Sytsma and Manley (1999) and Mendenhall and Beaver (1989).

## **Interpreting $\bar{X}$ charts**

First the within sample variation chart ( $R$ ,  $S$  or  $S^2$  chart) is constructed. If the observations are inside the control limits, the process is showing a consistent amount of variability. If not, appropriate action should be taken to stabilise the variability. The  $\bar{X}$  chart limits assume that the variance remains constant over time. Therefore, once there is a consistent amount of variability, we can construct the  $\bar{X}$  chart to check that the process is in statistical control.

If the amount of variability is inconsistent, we should be careful to look at both charts before drawing conclusions. For example, a large observation on the  $\bar{X}$  chart accompanied by a large observation on the within sample variation chart does not necessarily indicate a change in the population mean level. This could signify an increase in population variability, or both mean and variability.

When special causes are eliminated, both charts can be recomputed by omitting the points that are out of control. We monitor future process variable observations using the new limits. When all special cause variation has been removed (which is signified by all observations being in control), what remains is usual process variation (common cause variation, as described in 3.5.2).