



This module is part of the

# Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

# Method: Manual Editing

## Contents

General section .....	3
1. Summary .....	3
2. General description of the method .....	3
2.1 Introduction and historical notes .....	3
2.2 The use of recontacts .....	4
2.3 Potential problems .....	5
3. Preparatory phase .....	6
3.1 The editing staff.....	6
3.2 Editing instructions.....	6
3.3 Error messages .....	7
3.4 Efficient edit rules for manual editing.....	7
4. Examples – not tool specific.....	9
5. Examples – tool specific.....	10
6. Glossary.....	10
7. References .....	10
Specific section.....	12
Interconnections with other modules.....	14
Administrative section.....	15

## General section

### 1. Summary

In manual editing, records of microdata are checked for errors and, if necessary, adjusted by a human editor, using expert judgement. Nowadays, the editor is usually supported by a computer program in identifying data items that require closer inspection – in particular combinations of values that are inconsistent or suspicious. Moreover, the computer program enables the editor to change data items interactively, meaning that the automatic checks that identify inconsistent or suspicious values are immediately rerun whenever a value is changed. This modern form of manual editing is often referred to as ‘interactive editing’.

If organised properly, manual/interactive editing is expected to yield high quality data. However, it is also time-consuming and labour-intensive. Therefore, it should only be applied to that part of the data which cannot be edited safely by any other means, i.e., some form of selective editing should be applied (see “Statistical Data Editing – Selective Editing”). Furthermore, it is important to use efficient edit rules and to draw up detailed editing instructions in advance.

### 2. General description of the method

#### 2.1 Introduction and historical notes

*Manual editing* is the traditional way to perform data editing. Other data editing methods, in particular automatic editing techniques, did not emerge until the 1960s, and their application has only become widespread from the 1980s onward. Even today, practically all surveys at statistical offices and elsewhere include some form of manual editing. Manual editing is in fact widely viewed as an essential part of any data editing process.

Ideally, a person who performs manual editing – an *editor* – should be an expert who has extensive knowledge of the survey subject, the survey population, and the kind of errors that are likely to occur in the survey data. If necessary, he or she may recontact a respondent to check whether a suspicious value is correct, or to obtain a new value for a data item that was originally missing or incorrect. The editor may compare a survey unit’s data to reference data, such as data on the same unit from a previous survey or from an external register, or data on similar units. Finally, he or she may have access to other sources of information, for instance through internet searches.

In its ideal form, manual editing is expected to yield high quality data. In particular, it should lead to better results than automatic editing. However, it should be clear that the quality of manual editing depends strongly on the competence and training of the available editors. In certain less than ideal situations, the quality of manually edited data need not be significantly higher than that of automatically edited data, and it may even be lower (EDIMBUS, 2007).

Traditionally, manual editing was performed directly on the original paper questionnaires. Later, mainframe computers were used to check the data for inconsistencies and other violations of edit rules. To this end, the information on the questionnaires first had to be keyed in by typists. A list of edit failures identified by the computer was printed out on paper and used by the editors as a guide for making manual adjustments on the original questionnaires. When all questionnaires had been edited, the adjusted data were re-entered into the mainframe computer by typists and the edit checks were run

again, to see the effect of the proposed adjustments on the edit failures. Often, the automated checks revealed that the adjusted data still failed some of the edit rules, and another round of manual editing was required. It was not unusual that five, ten, or even more iterations of automatic checking and manual adjusting were needed before all questionnaires were considered sufficiently edited (Granquist, 1997; Van de Pol, 1995).

The advent of the microcomputer in the 1980s made it possible to integrate automatic checking and manual treatment of errors, thereby improving the data editing process in several ways (Bethlehem, 1987). From now on, the information on the questionnaires had to be keyed in only once.<sup>1</sup> After that, all adjustments could be made by the editors directly on the captured data. This obviously benefited the efficiency and timeliness of the editing process. A second improvement was that the editors could now get immediate feedback on the adjustments they made, because the automatic edit checks could be rerun instantaneously whenever the value of a data item was changed. This made it much easier for them to find adjustments that satisfied the edit rules. In addition, each record/questionnaire could now be edited separately, by one editor, until all violations of edit rules had been either removed or explained. This improved form of manual editing is called *interactive editing*.

Interactive editing requires a survey-processing system that provides the above-mentioned interaction between automated checks and manual adjustments. Well-known examples of survey-processing systems are *Blaise* (see, e.g., Blaise, 2002) and *CSPro* (see, e.g., CSPro, 2008). Pierzchala (1990) discusses general requirements of computer systems for interactive editing.

In today's statistical practice, interactive editing has effectively replaced all older forms of manual editing. Hence, the terms 'manual editing' and 'interactive editing' have become more or less interchangeable. In the remainder of this module, they shall be used as synonyms.

## 2.2 *The use of recontacts*

In the previous subsection, possible actions were listed that an editor may take when confronted with a record that requires review. One of these possible actions is recontacting the respondent. At first glance, a recontact may appear to be the natural way of obtaining better values for data items that were reported erroneously during the original field work, as well as items that were originally missing. Actually, depending on the survey, it may not be possible to contact the original respondents. For instance, if an external register is used as a data source and questions are raised about the quality of the incoming data, then the statistical office can usually only contact the supplier of the data set. Direct contact with the individual entities in the register is usually not possible in this case.

However, even when recontacts are possible, this approach can be considered problematic for several reasons. First of all, recontacts clearly increase the burden on respondents, whereas many statistical institutes are trying to reduce the response burden. In addition, recontacts tend to slow down the editing process and can therefore adversely influence the timeliness of statistics. Finally, if one considers that a respondent was not able to give a correct answer in the original survey – supposedly while filling in a meticulously designed questionnaire or talking to a highly qualified interviewer –,

---

<sup>1</sup> A more recent development is that data often arrive at the statistical office already in digital form, so that no keying is necessary at all. This is true for nearly all registers and for electronic questionnaires. For a discussion of the implications of electronic data collection for the editing process, see the theme module "Questionnaire Design – Editing During Data Collection".

then it is not at all obvious that he/she will give the correct response when talking to an editor. According to EDIMBUS (2007): "...respondents' ability to report should not be overestimated. In fact, if the structure of the questions does not fit their understanding, no amount of badgering will get the 'correct' answers out of them."

Following Granquist (1997), if recontacts are used during interactive editing, their main purpose should be to reveal problems that *cause* respondents to give erroneous answers, rather than merely correcting the individual errors that occurred. When used this way, recontacts can provide important insights into respondents' behaviour – in particular their ability to understand the concepts and definitions used in the survey. They may also reveal differences between what is asked in the survey and what kind of information is readily available in the survey units' accounting systems. These insights may be used as a basis for improvements at the data collection stage in subsequent surveys (see, e.g., Hartwig, 2009; Svensson, 2012).

### 2.3 Potential problems

There are several potential problems associated with interactive editing. The most important of these are the risks of *overediting* and *creative editing*.

According to Granquist (1995), overediting occurs when "the share of resources and time dedicated to editing is not justified by the resulting improvements in data quality." Manual editing is in fact a very labour-intensive and time-consuming activity, even in its modern, interactive form. Moreover, statistical output is typically affected by all kinds of errors (Bethlehem, 2009), including sampling error, selective unit non-response, coverage errors, measurement errors, etc. Only a subset of these can be treated during data editing: in particular, measurement and processing errors and, to a lesser extent, errors in the survey frame. Therefore, as soon as the data have been edited to a point where the influence of the latter types of errors on the statistical output is negligible compared to other sources of error (e.g., the sampling variance), manual editing should be stopped to prevent overediting. This notion – which was suggested already by Nordbotten (1955) – has received much attention since the 1980s. It has led to the development of methods for selective editing (see the theme module "Statistical Data Editing – Selective Editing") and macro-editing (see the theme module "Statistical Data Editing – Macro-Editing").

Another aspect of overediting is that if the editing process is continued too long, it may actually start to do more harm than good. In general, not all values that appear to be implausible are also incorrect. Hence, replacing all unusual combinations of values by more plausible ones would lead to a data set that does not reflect the natural variability of characteristics in the population. Overediting may therefore adversely influence the quality of the statistical output. An important part of the 'art' of manual editing is understanding which implausible values to adjust and which to leave as they are. This requires expert judgement and, in some cases, a recontact.

A second potential problem is the risk of creative editing: editors inventing their own, often highly subjective, editing procedures. Creative editing often involves complex adjustments of reported data items, done for the sole purpose of making the data consistent with a set of edit rules. Granquist (1995) remarks that creative editing may "hide serious data collection problems and give a false impression of respondents' reporting capacity."

To reduce the risk of overediting and creative editing, it is important to design efficient edit rules and to provide the editors with good editing instructions. These issues are discussed in the next section.

### **3. Preparatory phase**

In this section, several issues will be discussed that are related to the design of manual editing. These are: the desired characteristics of the editing staff (Section 3.1); the use of editing instructions to rationalise the manual editing process (Section 3.2); the design of error messages (Section 3.3); the design of efficient edit rules for manual editing (Section 3.4).

#### *3.1 The editing staff*

As mentioned in Section 2.1, the quality of manual editing strongly depends on the competence of the individual editors that are involved. A good editor should have the following characteristics:

- He/she has a large knowledge of the survey subject and of survey methodology. Since most of this knowledge is rather specialised, it has to be acquired through experience and training.
- He/she is communicative and responsive. This is particularly important if recontacts are used. Granquist (1995) remarked that if recontacts are done by telephone, “the editors also become telephone interviewers, needing adequate training and monitoring as in regular telephone interview surveys.”
- He/she is responsible and able to work accurately.
- Preferably, he/she should have an analytical mind, with an interest in problem-solving.

#### *3.2 Editing instructions<sup>2</sup>*

Editing instructions are an important aid in rationalising the manual editing process. They should contain at least the following components:

- A description of the purpose of the survey and the intended statistical output. In addition, the data collection phase and relevant data processing steps prior to editing should be briefly described.
- If relevant, instructions on the order in which the selected records should be treated. If manual editing is used in combination with selective editing (see “Statistical Data Editing – Selective Editing”), then an explanation is needed about the selection criteria and their interpretation. If manual editing is used in combination with macro-editing (see “Statistical Data Editing – Macro-Editing”), then detailed analysis instructions are needed regarding the selection of individual records that need further review.
- An overview of the types of errors that can occur in the data. Common errors in business surveys include classification errors with respect to NACE code or size class (i.e., errors in the survey frame), measurement errors, and processing errors.
- Suggestions about additional sources of information – such as auxiliary registers, sector organisations, and the internet – which should be consulted when following up data that have

---

<sup>2</sup> This subsection is to a large extent based on Hoogland et al. (2011).

been flagged by edit rules (see below). For example, many businesses nowadays have websites that contain relevant information for verifying potential NACE code errors.

- For each common type of error, an indication of how the error can be treated. Deterministic correction rules may often be specified for treating systematic errors (see also ‘Deductive Editing’). Clear instructions on this point can prevent the occurrence of creative editing.
- Instructions on how to log the editing actions taken during interactive editing. The survey-processing system should provide a comments field for this. Editors should be encouraged to provide details about the reasons for the adjustments they make. This information can be useful for improving the data collection process as well as the editing process itself.
- Instructions on specific follow-up actions that may be needed for certain types of errors. In particular, in case a NACE code or size class error is detected, it should be clear whether and how this must be communicated to the administrator of the survey frame.

### 3.3 *Error messages*

As mentioned in the main theme module, an important technique for finding errors in microdata is the inspection of data items that fail *edit rules*. Edit rules (edits for short) describe restrictions that should be satisfied by the data. Edits can be hard (meaning that they have to hold by definition, so that any failure corresponds to an error in the data) or soft (meaning that they are expected to hold for most survey units, but they can sometimes be failed by correct data items).

When edit rules are implemented in a computer system, an error message has to be associated with each edit rule. This message contains the information that the computer system gives to the editor about the unit and variables that are flagged by the edit rule as being (suspected to be) in error. The purpose of the error message is to give sufficient information for a rational follow-up of error flags. It also forms a basis for (process) data about the data collection and production processes.

The content of an error message generally consists of:

- Identifying properties of the flagged unit.
- The name of the flagged variable(s). For the purpose of manual editing, this should be a descriptive name rather than a technical one; e.g., not *TURNOVE100000* but *Total net turnover from domestic sales*.
- The code of the edit rule that was failed.
- A verbal description of the edit rule that was failed or, equivalently, a verbal description of the suspected error.
- If relevant and available, suggestions for auxiliary data that may be consulted in a follow-up of the error flag.

### 3.4 *Efficient edit rules for manual editing*

Typically, a large part of the work done during manual editing concerns the follow-up of soft edit failures. For this reason, it is important to formulate soft edit rules that are as efficient as possible. Here, an edit rule is considered efficient to the extent that it detects suspected errors that turn out to be

actual errors during manual follow-up, and inefficient to the extent that it detects suspected errors that turn out to be correct. (A measure of efficiency known as the hit rate will be introduced below.)

According to Norberg (2011), most edits that are used in practice consist of three components: an *edit group*, a *test variable*, and an *acceptance region*. The edit group defines the subset of the units to which the edit should be applied. The test variable is a known function of the observed variables that is evaluated by the edit. Finally, the acceptance region describes for which values of the test variable the edit will be satisfied. (Equivalently, one could define a *rejection region* that describes for which values of the test variable the edit will be failed.) Using these components, an edit may be written in one of the general forms

**if** ( *unit*  $\in$  *edit group* ) **then** ( *test variable*  $\in$  *acceptance region* )

or

**if** ( *unit*  $\in$  *edit group* **and** *test variable*  $\notin$  *acceptance region* ) **then** *error*.

Both formulations are equivalent. Human editors often find it slightly easier to work with the first formulation (Van de Pol, 1995). In a computer implementation, the second formulation can easily be extended to associate a unique error code and error message to each edit rule.

For a simple example, consider the following edit rule:

**if** *Size class* = 'small' **then**  $0 \leq \textit{Number of employees} < 10$ .

For this edit, the edit group can be defined as "all units for which *Size class* = 'small'". The test variable is identical to one of the observed variables, *Number of employees*. The acceptance region consists of the interval [0, 10). A computer implementation of this edit could further specify the following actions:

**if** ( *Size class* = 'small' **and** ( *Number of employees* < 0 **or** *Number of employees*  $\geq$  10 ) )  
**then** ( *error\_code\_E1* := "failed";  
*error\_message\_E1* := "The number of employees does not match the size class." )

The first statement in the then-part assigns the error code "failed" to the current record for this edit (identified here by E1). The second statement gives an error message describing the nature of the current edit failure to the human editor. Of course, the precise implementation of these actions will depend on the survey-processing system.

To give another example, consider the following conditional ratio edit:

**if** ( *Economic activity* = X **and** *Size class* = 'medium' )  
**then**  $a < \textit{Total turnover} / \textit{Number of employees} < b$ .

Here, the edit group consists of "all medium-sized units with *Economic activity* X", the test variable is defined as the ratio of the observed variables *Total turnover* and *Number of employees*, and the acceptance region is given by the interval (a,b).

Norberg (2011) notes that, for the editing to be efficient, one should choose edit groups that are homogeneous with respect to the test variable. In some cases, the choice of an edit group may be natural (e.g., the first example given above). If this is not the case, suitable edit groups may be derived from an analysis of previously edited data. Norberg (2012) suggests to use classification or regression trees for this. In addition, the acceptance region should reflect the natural variability of the test

variable within the edit group (Norberg, 2012). Again, previously edited data may be analysed (e.g., using box plots) to find suitable acceptance regions. It may be worthwhile to transform a test variable so that its distribution becomes more amenable to summary in the form of an acceptance region (e.g., so that the transformed test variable is approximately normally distributed, or at least symmetrical). Moreover, in repeated surveys, the acceptance regions should be regularly updated.

Outlier detection techniques are often used in the construction of soft edit rules. We refer to EDIMBUS (2007) for a discussion of outlier detection in the context of statistical data editing. Methods that may be used to construct soft edit rules in repeated surveys are discussed in the theme module “Statistical Data Editing – Editing for Longitudinal Data”.

At the design stage, it is useful to assess the efficiency and effectiveness of a proposed set of edits  $E$  by means of simulation. This requires historical data that have been fully edited, as well as the original, unedited version of the same data set. Interesting indicators for an edit  $e \in E$  include the *failure rate* (the proportion of records in the unedited data that fail edit  $e$ ) and the *hit rate* (the proportion of edit failures with respect to  $e$  in the unedited data that are associated with adjustments in the edited data). Note that for all hard edit rules, the hit rate should be 1. These indicators are local, i.e., defined for one edit at a time. Similar global indicators can be defined for the set of edits  $E$  as a whole. It is also interesting to assess to what extent the edits are ‘overlapping’, in the sense that the same error is often detected by multiple edits. Ideally, there should be as little overlap as possible between the edits.

Furthermore, making the assumption that the edited historical data do not contain any errors, one can evaluate the *missed error rate* (the proportion of errors in the original data that were not flagged by any edits in  $E$ ) and an estimate of the measurement bias due to untreated errors if editing were based on  $E$ . See EDIMBUS (2007) and Silva et al. (2008) for formal definitions of these and other indicators.

The Office for National Statistics in the United Kingdom and Southampton University have developed a tool called Snowdon-X which “can be used to understand how current edits are working within the survey and also the impact on quality of any changes to the edit rules” (Skentelbery et al., 2011). Snowdon-X evaluates the indicators mentioned above as well as many other indicators. See Silva et al. (2008) for more details on Snowdon-X.

Note that the failure rate and hit rate of edits can and should be evaluated also during regular production. On the other hand, evaluating the missed error rate requires edited historical data. For repeated surveys, suitable historical data sets are available in theory, if not always in practice (Lindgren, 2012). For a one-off survey, as well as the first cycle of a survey that will be repeated, the situation is different. Often in this case, a small pilot study is conducted beforehand. The data from this study can be used to test the effects of different editing approaches, including experiments with different formulations of edit rules. In addition, experts should be consulted that have had experience with similar surveys in the past.

#### **4. Examples – not tool specific**

## 5. Examples – tool specific

## 6. Glossary

For definitions of terms used in this module, please refer to the separate “Glossary” provided as part of the handbook.

## 7. References

Bethlehem, J. G. (1987), The Data Editing Research Project of the Netherlands Central Bureau of Statistics. Report 2967-87-M1, Statistics Netherlands, Voorburg.

Bethlehem, J. G. (2009), *Applied Survey Methods*. Wiley Series in Survey Methodology, John Wiley & Sons, New Jersey.

Blaise (2002), *Blaise for Windows 4.5 Developer’s Guide*. Statistics Netherlands, Heerlen.

CSPRO (2008), *CSPRO User’s Guide*, version 4.0. U.S. Census Bureau, Washington, D.C.

EDIMBUS (2007), *Recommended Practices for Editing and Imputation in Cross-Sectional Business Surveys*.

[http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/RPM\\_EDIMBUS.pdf](http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/RPM_EDIMBUS.pdf).

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

Granquist, L. (1997), The New View on Editing. *International Statistical Review* **65**, 381–387.

Hartwig, P. (2009), How to Use Edit Staff Debriefings in Questionnaire Design. Paper presented at the 2009 European Establishment Statistics Workshop, Stockholm.

Hoogland, J., van der Loo, M., Pannekoek, J., and Scholtus, S. (2011), *Data Editing: Detection and Correction of Errors*. Methods Series Theme, Statistics Netherlands, The Hague.

Lindgren, K. (2012), The Use of Evaluation Data Sets when Implementing Selective Editing. Working Paper, UN/ECE Work Session on Statistical Data Editing, Oslo.

Norberg, A. (2011), The Edit. Working Paper, UN/ECE Work Session on Statistical Data Editing, Ljubljana.

Norberg, A. (2012), Tree Analysis – A Method for Constructing Edit Groups. Working Paper, UN/ECE Work Session on Statistical Data Editing, Oslo.

Nordbotten, S. (1955), Measuring the Error of Editing the Questionnaires in a Census. *Journal of the American Statistical Association* **50**, 364–369.

Pierzchala, M. (1990), A Review of the State of the Art in Automated Data Editing and Imputation. *Journal of Official Statistics* **6**, 355–377.

Silva, P. L. N., Bucknall, R., Zong, P., and Al-Hamad, A. (2008), A Generic Tool to Assess Impact of Changing Edit Rules in a Business Survey – An Application to the UK Annual Business Inquiry Part 2. Working Paper, UN/ECE Work Session on Statistical Data Editing, Vienna.

- Skentelbery, R., Finselbach, H., and Dobbins, C. (2011), Improving the Efficiency of Editing for ONS Business Surveys. Working Paper, UN/ECE Work Session on Statistical Data Editing, Ljubljana.
- Svensson, J. (2012), Editing Staff Debriefings at Statistics Sweden. Working Paper, UN/ECE Work Session on Statistical Data Editing, Oslo.
- Van de Pol, F. (1995), Data Editing of Business Surveys: an Overview. Report 10718-95-RSM, Statistics Netherlands, Voorburg.

## **Specific section**

### **8. Purpose of the method**

Detecting and treating errors in microdata

### **9. Recommended use of the method**

1. Because of its expensive and time-consuming nature, it is best to apply manual editing only to that part of the data where expert judgement is really needed. In other words, one should always try to use this method as part of a strategy for selective editing or macro-editing (cf. “Statistical Data Editing – Main Module”). This usually means that manual editing is only applied to units that are either very large or complex, or for which the reported data are likely to contain many and/or influential errors.
2. A survey-processing system should be used that allows real-time interaction between manual adjustments and automated checks (i.e., manual editing should be interactive editing)
3. It is important to draw up editing instructions in advance, to guide the decisions made by the editors during manual editing. This lowers the risk of overediting or creative editing. It is also important to design efficient edit rules and informative error messages.

### **10. Possible disadvantages of the method**

1. If recontacts are used as part of manual editing, the method places additional burden on survey units that are recontacted. Recontacts may also affect the timeliness of statistical production.

### **11. Variants of the method**

1. n/a

### **12. Input data**

1. A data set containing unedited microdata.

### **13. Logical preconditions**

1. Missing values
  1. Allowed.
2. Erroneous values
  1. Allowed; in fact, the object of this method is to replace erroneous values with better values.
3. Other quality related preconditions
  1. n/a
4. Other types of preconditions
  1. n/a

**14. Tuning parameters**

1. A collection of edit rules for the microdata at hand.

**15. Recommended use of the individual variants of the method**

1. n/a

**16. Output data**

1. A data set containing edited microdata.

**17. Properties of the output data**

1. If manual editing has been performed correctly, the records in the output data set are consistent with all hard edit rules. In addition, all remaining soft edit failures have been explained and accepted by a subject-matter expert.

**18. Unit of input data suitable for the method**

Incremental processing

**19. User interaction - not tool specific**

1. As the term ‘interactive editing’ suggests, user interaction is needed throughout. In fact, all changes made to the data during manual/interactive editing are initiated by a human editor.

**20. Logging indicators**

1. Comments made by the editors to explain the adjustments they made to the data, as well as the soft edit failures that they left in.
2. If recontacts are used: comments made by the editors regarding identified problems that caused respondents to report erroneous values in the original survey.
3. Process indicators for the efficiency and effectiveness of the edit rules used in manual editing include: failure rate, hit rate, missed error rate, estimated measurement bias. See also Section 3.4 of this module, EDIMBUS (2007), and Silva et al. (2008).

**21. Quality indicators of the output data**

1. It is not straightforward to assess the quality of manually edited data, because in many applications the results of manual editing are actually taken as the standard by which other forms of editing are to be measured. Nordbotten (1955) suggests a way to measure the quality of regular manual editing, i.e., as it occurs in everyday statistical practice. This method takes a random sample of the original data and subjects it to a very refined form of manual editing (under ideal conditions, with near-unlimited resources). The quality of the regular editing process may then be measured in terms of the similarity of the data edited under regular conditions to the data edited under ideal conditions.

**22. Actual use of the method**

1. Interactive editing is used at Statistics Netherlands in many production processes, including that of the structural business statistics. The survey-processing system Blaise is used as a tool.

**Interconnections with other modules**

**23. Themes that refer explicitly to this module**

1. Questionnaire Design – Editing During Data Collection
2. Statistical Data Editing – Main Module
3. Statistical Data Editing – Selective Editing
4. Statistical Data Editing – Macro-Editing
5. Statistical Data Editing – Editing for Longitudinal Data

**24. Related methods described in other modules**

1. Statistical Data Editing – Automatic Editing

**25. Mathematical techniques used by the method described in this module**

1. n/a

**26. GSBPM phases where the method described in this module is used**

1. GSBPM Sub-process 5.3: Review, validate and edit

**27. Tools that implement the method described in this module**

1. Blaise
2. CSPro

Note: These tools support interactive editing, but – by its very nature – this method relies heavily on human interaction with the tool.

3. Snowdon-X

Note: This tool can be used to evaluate the efficiency of edit rules for manual editing.

**28. Process step performed by the method**

Statistical data editing

## Administrative section

### 29. Module code

Statistical Data Editing-M-Manual Editing

### 30. Version history

Version	Date	Description of changes	Author	Institute
0.1	05-06-2012	first version	Sander Scholtus	CBS (Netherlands)
0.2	01-03-2013	improvements based on Swedish review	Sander Scholtus	CBS (Netherlands)
0.3	12-04-2013	improvements based on second Swedish review	Sander Scholtus	CBS (Netherlands)
0.4	11-11-2013	minor improvements based on final Swedish review	Sander Scholtus	CBS (Netherlands)
0.4.1	26-11-2013	preliminary release		
1.0	26-03-2014	final version within the Memobust project		

### 31. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	21-3-2014 18:12