Working Paper
ENGLISH ONLY

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Paris, France, 28-30 April 2014)

Topic (iii): Getting the support of all people when implementing data editing

Applying Process Indicators to Monitor the Editing Process

Prepared by Karin Lindgren, Statistics Sweden and Martin Odencrants, Statistics Sweden

I. Introduction

1. A part of Statistics Sweden's business task is to refine data to statistical information through advanced methodological competence, broad knowledge of subject matter and modern techniques. The statistics should be impartial and of good quality, at the same time based on scientific principles. This task is complicated by the fact that collected data often contain errors. The process of correcting errors is an important part of a surveys production process and regards important aspects such as response burden, cost efficiency and timeliness. Records that are suspected to contain errors are usually discovered during the whole productions process. A challenge when designing the overall editing strategy is to allocate the resources marked for editing between different editing activities in order for the overall editing strategy to be efficient.

2. In many cases, a large number of edit rules and re-contacts with respondents are taken as a sign of good quality in final data. This is not necessary true. Instead of trying to correct all errors, one should concentrate on those that have a large impact on the final results (Granquist, 1997). Selective editing is a method that puts the focus on errors with a suspected large influence on the statistical output. Statistics Sweden has developed a generic tool for selective editing. SELEKT is a tool developed to signal an object as suspicious depending on the impact that the reported value will have on the published data and the probability that the reported value is incorrect. SELEKT was introduced in 2008 and is now used by a number of surveys at Statistics Sweden (Norberg, 2014).

3. Statistical products at Statistics Sweden are administrated by management teams that consist of coworkers with different areas of expertise. Each team consists of a methodological expert, subject area expert, IT expert and a product manager. All areas in the production process, from survey design to publication of final results, should be planned by the management team and experiences from the previous production round should be accounted for and documented. The iterative process of planning and evaluating each production round is an important part in the work of continuously improving efficiency in the overall production process. One important part in the continuous work is improvement of the editing process.

4. One main purpose of the editing process is to provide basic data for the improvement of the survey. By collecting information that describes the editing process of a survey it is possible to identify potential sources of error as well as identifying sub-processes that can be improved. Improvement of the editing process is an important task; internal studies as Statistics Sweden have shown that the editing process accounts for approximately 30 percent of the total survey budget. An essential question in the work with improving the production process is: "What have we learned during this production round and how can we use this information to perform better next time". One important source of information is data from the editing process. By producing process indicators it is possible to monitor the editing process and identify sub-processes that are not working as planned and therefore have to be adjusted.

5. Several process indicators designed for different purposes are available to monitor different aspects of the editing process. A large number of indicators for the editing and imputation processes are provided in EDIMBUS (2007). Process indicators can be used to monitor and identify problems with the measurement process and the editing process. In this paper we will not address the imputation process. Process indicators are not frequently used in practice at Statistics Sweden and it is hard to measure if the editing process shows signs of improvement over time. Almost only for the surveys where SELEKT has been implemented some process indicators have been calculated. Not many of the management teams are taking the necessary actions to improve the process. To be able to work with continuous improvement of the editing process, indicators need to be generated and analyzed continuously by the management team. In order for the indicators to be useful they need to be presented in an accessible way. Data created during the production process is usually not designed to measure quality and performance of the editing process, which complicates the task to generate indicators. There are currently no easy accessible standard reports that can be used to monitor the editing process as Statistics Sweden. Several initiatives are needed to make the management teams aware of the importance of improving the editing process. One way of increasing awareness of process indicators and their use is to provide the management teams with concrete examples, which is the purpose of this paper. Another way is to provide tools to easily generate process indicators. Statistics Sweden is currently developing a common platform for the production process. Surveys connected to the platform will have the possibility to generate standard reports containing process indicators in the future.

6. For the survey *International Trade in Services*, process indicators have been generated for the purpose of describing both performance and quality aspects of the editing process. Indicators have been calculated over three consecutive quarters, 2013Q1-2013Q3. The results are described in two different sections below; the first section describes the editing process from an overall perspective and the second section contains more detailed information on the editing process.

II. Indicators

7. The primary objective with computing indicators for the editing process is to monitor and optimize the process. Another important objective is to discover flaws in the measurement process and detect common causes for measurement errors e.g. a badly designed questionnaire. The chosen indicators have one or more of these purposes. Indicators come from EDIMBUS (2007) directly, or slightly altered, and are presented in table 1. It should be useful to study their behavior over time to detect changes in the data structure or to monitor the effect of improvements.

- n is number of observations
- e_{il} is the indicator variable related to the edit rule l and takes the value 1 if observation i fails the edit (detection of erroneous or suspicious values) and 0 if the edit is passed by observation i.
- f_{ij} = 1 if variable j of unit i is flagged for errors and 0 otherwise.
- o_{lj} =1 if variable j is involved in edit l and 0 otherwise
- y_{ij} is the raw value for observation i and variable j.
- $\hat{y}_{ij}$ is the edited value for observation i and variable j.
- $I(\cdot)$ =1 if its argument is true and 0 otherwise.

I_1	Overall failure rate. Rate of observations failing at least one edit rule.	$\frac{\sum_i(1-\prod_l(1-e_{il}))}{n}$
I_2	Overall hit rate. Rate of observations failing at least one edit rule and adjusted.	$\frac{\sum_{ij} f_{ij} I(\hat{y}_{ij} \neq y_{ij})}{\sum_{ij} f_{ij}}$
I_3	Weighted reject ratio.	$\frac{\sum_{ij} f_{ij} w_{ij} y_{ij}}{\sum_{ij} w_{ij} y_{ij}}$
I_4	Failure rate for each edit rule *l*.	$\frac{\sum_i e_{il}}{n}$
I_5	Hit rate for each edit rule *l*. Rate of edit failures which lead to adjustments	$\frac{\sum_i e_{il}[1-\prod_j(1-I(\hat{y}_{ij} \neq y_{ij})o_{lj})]}{\sum_i e_{il}}$
I_6	Unweighted editing rate for each variable *j*.	$\frac{\sum_i I(\hat{y}_{ij} \neq y_{ij})}{n}$

Table 1: Indicators

III. Brief Background of the Selected Survey and its Editing Process

A. *International Trade in Services*

8. *International Trade in Services* is a quarterly sample survey. Its target population is Swedish enterprises and organizations which trade in services, compensation of employees and transfers (in the following shortened services) with foreign countries. The sample consists of about 5000 enterprises and organizations. They report values concerning their income and expenditure as a subsequence of trade in services with foreign parties. The respondents are asked to provide data on the three following levels:

(a) Total expenditure and income originating from trade in services
(b) The expenditure and income specified by type of service (approximately 100 different types)
(c) The expenditure and income specified by country of the foreign trading partner for each type or service

When calculating the indicators the country specific values are not regarded.

9. Most respondents send in their information using a web based questionnaire but some still send in paper forms. In the web based questionnaire some edits are built in to prevent the respondents to submit erroneous data. At Statistics Sweden the data is later checked with a number of edit rules, both fatal edits and query edits. SELEKT is utilized to flag questionnaires for follow-up. In the following each enterprise and questionnaire is referred to as an observation.

10. If an observation is flagged by at least one fatal edit SELEKT always sends it to the editing staff for manual follow-up. If an observation is flagged by query edits alone, local scores for each service type are calculated based on the suspicion of error and its potential impact on the statistical output. The local scores are aggregated to a global score for each observation. A local threshold is used to only aggregate types of services with high local scores. Observations with high global scores are sent to follow-up. Observations with suspected errors which give low global scores are not investigated further and the data

are accepted and used in the estimation. A global threshold is used to separate observations in need of follow up. If an observation is sent to follow-up, all the suspected errors belonging to types of services with local scores above the local threshold are investigated. Suspicion of error is no longer a factor. This is based on an assumption that if a respondent is to be re-contacted the marginal cost of asking about an additional error is low. The large cost comes from each additional observation to follow up on.

B. Edit Rules

11. Since SELEKT was implemented in 2010 the survey has had 26 active edit rules: 21 query edits (Q01-Q21) and 5 fatal edits (F01-F05). Not all edit rules are activated each quarter. Only a few edit rules comprise the country level data. Some of the query edits flag an entire questionnaire and some only flag one service type as suspicious. The query edits can also be divided into edits which are meant to find deviation errors (Q08-Q14, Q19-Q21) or definition errors (inliers) (Q01-Q07, Q15-Q18). Different purposes mean different possibilities to utilize the indicators. Deviation errors imply that the respondent has submitted very deviant values, e.g. extremely high values, which might be erroneous or outliers. One common mistake made by respondents in the survey is submitting values in SEK instead of in thousands of SEK, which is the required unit. Definition errors can come from misunderstandings or a difference between what the question is intended to ask and what the respondent can provide. Definition errors are often harder to detect and to construct edit rules for, since the submitted values might be perceived as reasonable. One common mistake in *International Trade in Services* is reporting on the wrong kind of service. Definition errors become a potential problem when many respondents make the same type of error. When calculating the indicators each income/expenditure for a type of service is treated as a variable.

IV. Results

C. Overall Indicators on Performance and Quality

12. Some basic variables can be identified which give important descriptive information about the editing process. By combining the variables it is possible to generate overall process indicators. The indicators described below are limited to the information accessible from the collected and processed data. Table 2 and 3 show descriptive information from the editing process.

	2013Q1	2013Q2	2013Q3
Total number of observations	4041	3876	3890
Number of observations failing at least one edit rule	438	297	403
Number of observations failing at least one edit rule and adjusted	258	187	168

Table 2: Number of observations 2013Q1-2013Q3

	Expenditure			Income		
	2013Q1	2013Q2	2013Q3	2013Q1	2013Q2	2013Q3
Weighted raw value sum	373	433	302	457	694	730
Weighted raw value sum, failing at least one edit rule	319	373	194	368	616	538
Weighted edited value sum	129	130	151	216	214	297

Table 3: Total values, 2013Q1-2013Q3. Values are given in million SEK.

13. An essential indicator of overall performances of the editing process is the ratio between number of observations flagged for error and total number of observations (Overall failure rate, I_1). This measure illustrates how the work load varies over time but can also be used to identify deviations in the editing process. In order to make reliable conclusions much more than three time points would be preferred; the charts below are just a start. If one can show that the failure rate varies with a seasonal pattern the indicator could be used to reallocate resources throughout the year.

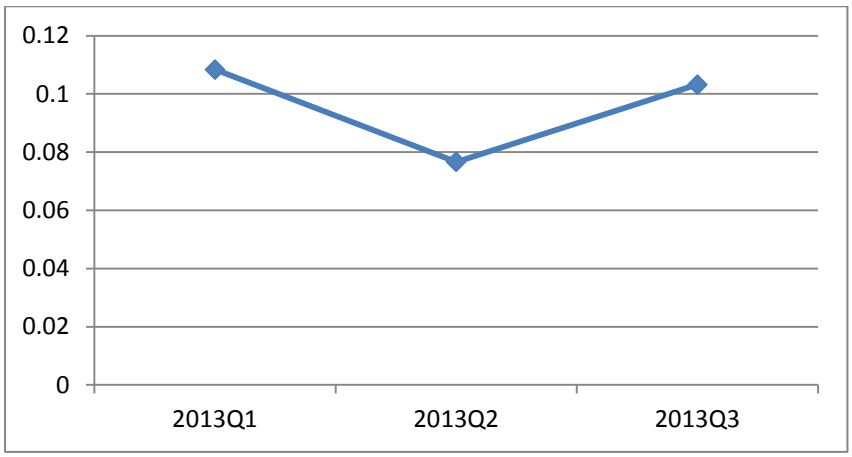

Figure 1: I_1 Overall failure rate, 2013Q1-2013Q3

14. The overall failure rate varies during the three quarters. A small drop in the ratio during the second quarter is followed by an increase during the third quarter. An increase in the overall failure rate between two consecutive quarters can indicate that the quality of raw data is decreasing, given that edit rules are the same. An increase can also be the result of a change of the edit rules. Tighter acceptance regions in the edit rules will result in more observations being flagged for error. This could lead to an increase in quality in the final results since more observations will pass to manual investigation and possible re-contact with the respondent. Without metadata on the editing process it is difficult to analyze the development of the flagged for error ratio.

15. The overall hit rate (I_2) gives information on effectiveness of the editing process. A rule of thumb at Statistics Sweden is a target hit rate of 60 percent (Granquist et. al, 2002). The figure below illustrates the overall hit rate for the three quarters.

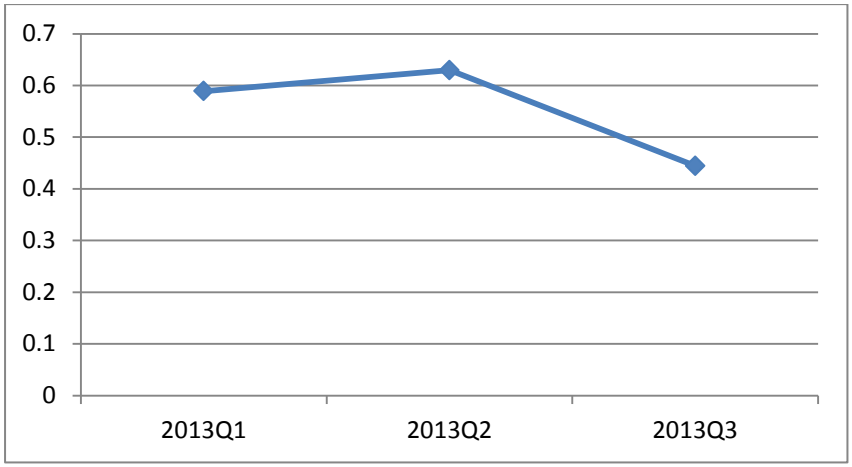

Figure 2: I_2. Overall hit rate, 2013Q1-2013Q3

16. The overall hit rate lies around 60 per cent which is our target. It shows a decrease during the third quarter. This can be an indication that one or more of the edit rules, F01-F05 or Q01-Q21, did not work properly during that quarter, maybe because of an irregularity in the editing process or because of the received data. When analyzing the hit rate of the individual edit rules, one discovers a large increase of the failure rate of edit rule Q02, the hit rate of Q02 was also very low the third quarter. Too much importance should not be made on only one quarter of course when having such a short time line.

17. By combining the weighted raw value sum of all observations flagged for errors with the weighted raw value sum of all observations, it is possible to construct the reject ratio, I_3 (Nordbotten, 2000). This ratio gives information on the impact of the editing process on the target variables. Figure 3 illustrates the reject ratio over the three quarters.

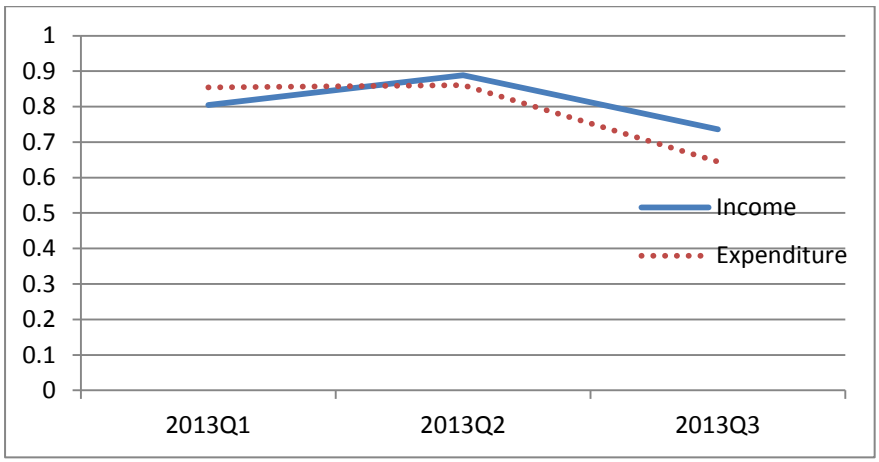

Figure 3: I_3. Weighted reject ratio, 2013Q1-2013Q3

18. Figure 3 shows that a large proportion of the reported values are flagged for error. One of the more common errors made by respondents is when the respondents have used SEK instead of thousands of SEK when reporting figures, an error which has a very large impact. The weighted reject ratio can be used as a quality measure of the editing process and its effectiveness.

D. Edit Rule and Variable specific Indicators

19. To evaluate the performance of the edit rules the failure rates (I_4) for each edit rule are computed for each quarter. Most edit rules have a very low failure rate. Table 4 shows the results. Some edit rules stand out and should be investigated; they are marked dark grey in the table. Some edit rules flag a higher amount of the observations constantly and some flag many observations just one quarter which looks a little suspicions. Three edit rules stand for the main part of error flags each quarter, they are marked light grey in the table.

Edit rule	2013Q1	2013Q2	2013Q3
Q01	0.02%	0.00%	0.00%
Q02	0.00%	0.03%	4.14%
Q03	5.05%	0.77%	1.93%
Q04	0.17%	0.03%	0.10%
Q05	0.05%	0.00%	0.03%
Q06	0.05%	0.03%	0.10%
Q07	0.12%	0.08%	0.08%
Q08	6.14%	6.22%	5.66%
Q09	5.32%	5.08%	4.50%
Q10	0.40%	0.34%	0.39%
Q11	0.89%	0.88%	1.05%
Q12	0.74%	0.80%	0.64%
Q13	0.07%	0.05%	0.00%
Q14	2.13%	2.66%	2.24%
Q15	0.20%	0.03%	0.10%
Q16	0.64%	1.86%	0.80%
Q17	0.00%	0.00%	0.03%
Q18	0.00%	0.03%	0.05%
Q19	0.47%	0.80%	0.72%
Q20	0.17%	0.18%	0.10%
Q21	1.19%	0.00%	0.00%
F01	0.02%	0.05%	0.08%
F02	0.94%	0.59%	0.95%
F03	0.00%	0.00%	0.03%
F04	0.10%	0.08%	0.00%
F05	0.12%	0.10%	0.08%

Table 4: I_4. Failure rate per edit rule, 2013Q1-2013Q3.

20. For the query edits, hit rate (I_5) is used to assess their efficiency, and presented in table 5. A hit rate of at least 60 % is a rule of thumb also for each edit rule (Granquist, et.al, 2002), but it can be hard to tell what the optimal hit rate should be for an individual edit rule. Low hit rates cause follow-up on values that are not erroneous which give unnecessary costs. Too high hit rates might imply too generous range edits and deviation errors that are not found. The table shows that some edit rules have low hit rates. If these edits also have flagged many objects they carry a high cost and might be in need of reconstruction. If the hit rate is low throughout all three quarters the edit rule might be poorly constructed. If the hit rate only shows a temporary drop one quarter something might have happened during that quarter. Edit rules with low hit rate and high failure rate are marked dark grey in the table. Q02, which been mentioned earlier, has a high failure rate and a low hit rate during the third quarter. The explanation lies in the way the edit rule is constructed in combination with the way the error lists were created that particular quarter, which led to the high failure rate. There was not an increase of confirmed errors. In the future this high failure rate can be adverted. Q21 (which is only in use the first quarter of the year) on the other hand, has a very high hit rate and a high failure rate. A reconstruction to make it tighter might be considered to capture errors within the range used in the edit today.

Edit rule	2013Q1	2013Q2	2013Q3
Q01	100%	.	.
Q02	.	0%	10%
Q03	54%	60%	91%
Q04	43%	100%	100%
Q05	100%	.	100%
Q06	50%	100%	75%
Q07	20%	67%	67%
Q08	67%	65%	53%
Q09	68%	64%	50%
Q10	63%	23%	60%
Q11	53%	24%	29%
Q12	77%	77%	76%
Q13	67%	100%	.
Q14	72%	67%	64%
Q15	100%	100%	100%
Q16	85%	89%	87%
Q17	.	.	0%
Q18	.	100%	100%
Q19	95%	84%	89%
Q20	57%	29%	50%
Q21	98%	.	.

Table 5: I_5. Hit rate per edit rule, 2013Q1-2013Q3.

21. Hit rate has its downsides as a quality measure. Most edit rules flag many variables and most variables are flagged by many edit rules. After the editing phase is over it is hard to determine which flag lead to which adjustment. The same value might have been flagged by many edit rules which will count as a hit for all of those.

22. Edit rules meant to find deviation errors are mostly range edits, where the value must lie in an acceptance interval. Range edits with poor hit rate need a correction of the accepted range. Often the values tend to increase over time, especially if it is economic information. Frequent monitoring of hit rates can give a cue on when the intervals need to be adjusted. Edit rules meant to find definition errors are harder to reconstruct if they show poor hit rate. They are based on known misinterpretations which the respondents tend to make. To detect systematic definition errors another type of indicator might be more useful. Editing rate per variable (I_6) can tell something about how common it is that the respondents make a lot of mistakes regarding the same variable. In table 6 the kinds of service with the most adjustments are shown. Editing rates are quite low for all kinds of services.

Income

Kind of service	2013Q1	2013Q2	2013Q3
463 *Other services between affiliated enterprises*	0.4%	0.4%	0.3%
143 *Goods freight by road*	0.3%	0.3%	0.3%
123 *Purchasing costs for merchanting goods sold during the period*	0.3%	0.2%	0.2%
442 *Architectural, engineering and other technical services*	0.3%	0.1%	0.2%
122 *Goods that don't cross Swedish borders (merchanting)*	0.3%	0.1%	0.1%

Expenditure

Kind of service	2013Q1	2013Q2	2013Q3
410 *Computer services*	0.8%	0.4%	0.4%
442 *Architectural, engineering and other technical services*	0.7%	0.5%	0.4%
143 *Goods freight by road*	0.6%	0.6%	0.1%
440 *Advertising and marketing services*	0.5%	0.4%	0.2%
222 *Other travel-related services*	0.5%	0.5%	0.0%

Table 6: I_6. Unweighted editing rate, highest rates for different kinds of services.

V. Discussion

23. Correction of errors is an important part of a surveys production process and accounts for approximately 30 percent of the total survey budget. With increasing demands on efficiency in the production process and higher expectations from users, there is a need to monitor and improve the editing process.

24. All surveys at Statistics Sweden are administrated by management teams that consist of coworkers with different areas of expertise. These management teams are responsible for monitoring and improving the editing process. Even though the editing process accounts for a large part of the total survey budget, small or no resources are devoted to improving the editing process. One reason is that there is no clear framework for which process indicators to use and no easy way to generate them from available data. Another reason may be that there is no clear connection between improving the editing process and freeing resources for other activities. There is no information on how increased efficiency in the editing process impacts the overall costs of the survey.

25. In this paper we present six process indicators that describe the different aspects of the editing process. A few other indicators, not presented in the paper, were considered during the project. One example is an indicator describing the direction and size of the adjustment by variable. Variable specific indicators could also be calculated with weights and an interesting idea would be to construct confidence intervals for many of the indicators. When comparing the indicator values over time confidence intervals could be useful to detect significant changes. Since the *International Trade in Services* survey use SELEKT, suitable indicators for selective editing are also very important. One measure that has been discussed is the correlation between the score each object has been ascribed during the editing and the size of the adjustment. In theory this correlation should be high and could be used as quality indicator of the selective editing. There was no information on the cost of each edit rule and therefore it is not possible to construct process cost indicators. Indicators which display the cost of the editing process would give the product manager a clear incentive to monitor and improve the editing process.

26. One of the objectives of the project was to establish indicators which could be used by the editing staff and subject matter experts in cooperation with methodologists. However, indicators easily become complicated and hard to interpret since they are based on complex dependencies between versions of raw data, edited data, error flags and edit rules. To compute the six indicators presented in this paper turned out to be time consuming and slightly complicated, much dependent on the structure of the data and the IT-platform. As stated, Statistics Sweden plan to implement the possibility to generate standard reports containing process indicators within a common platform for the overall production process. Our experience shows that this idea would be beneficial, since it would simplify the computing process.

However, the presented indicators would need to be completed with information regarding how the indicators could be interpreted. It is possible that the indicators should be interpreted in different ways for each survey.

27. Indicators regarding the editing process give valuable information on the editing process that can be used by management teams to identify areas in need of development or improvement. Generating process indicators on a regular basis give the management teams a foundation for working with continuous improvement of the editing process. It will probably take some time to establish standardized procedures within the organization and within each management team. By spreading information on process indicators within the organization we stress the importance of an effective data editing strategy and hopefully also get support for implementing better editing procedures.

28. The interpretation of the indicators for the *International Trade in Services* is that the measurement process and the editing process seem to work quite well. Failure rates are low, hit rates and weighted reject ratios are high. This indicates that most received data entries are accepted, and when an entry is flagged as erroneous, confirmed large errors are often found. However, the known problem with reporting in SEK instead of thousands of SEK has a very large impact. It would be informative to compute the indicators after adjusting for that particular error type to get information on the impact of remaining errors. Another problem when interpreting the indicators stems from the fact that the failure rates are low. As a result the amount of data to analyze is small, especially for the variable specific indicators. They depend too much on random fluctuations.

29. Process indicators describing the editing process are needed for several reasons; to give information on operational characteristics of the process, e.g. workload as well as describing quality aspects of editing. The type of indicators that can be generated depends on available data. Therefore the type of indicators that are needed to monitor a survey need to be considered when designing a new survey or planning for a new production round of an existing survey. The structure of data influences the difficulty to generate process indicators. A lot of work was needed to validate and reshape data before process indicators could be calculated. This difficulty is probably not restricted to the *International Trade in Services* survey and is probably one reason why indicators are not used in other surveys. A system that generates standard reports would make it easier for the management teams to monitor the editing process.

References

EDIMBUS (2007), Recommended Practices for Editing and Imputation in Cross-sectional Business Surveys, August 2007,
http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/RPM_EDIMBUS.pdf

Nordbotten S. (2000). Evaluating Efficiency of Statistical Data Editing: General Framework. United Nations, Geneva,
http://www.unece.org/fileadmin/DAM/stats/publications/editingefficiency.pdf

Granquist, L. et. al. (2002). Guide till granskning. CBM 2002:1. Handbook on editing. Statistics Sweden.
http://www.scb.se/statistik/_publikationer/OV9999_2003M00_BR_X97%c3%96P0201.pdf

Granquist, L. (1997). The new view on editing. *International Statistical Review*. 65. No.3. pp 381-387.
http://www.jstor.org/stable/1403378

Norberg, A. (2014), Experiences from Selective Editing at Statistics Sweden, paper presented at the Work Session on Statistical Data Editing in Paris, France, 28-30 April 2012.