

A SEVENTH APPLICATION OF ASPIRE FOR STATISTICS SWEDEN

Paul Biemer, Dennis Trewin and Dan Kasprzyk

June 21, 2017

TABLE OF CONTENTS

1. Executive Summary.....	3
2. Background and Introduction.....	6
3. Product Reviews.....	8
3.1. General Observations.....	8
3.2. ASPIRE Ratings.....	13
3.2.1. Labour Force Survey.....	13
3.2.2. Consumer Price Index.....	16
3.2.3. Producer and Import Price Index.....	18
3.2.4. Annual Municipal Accounts.....	20
3.2.5. Foreign Trade of Goods	22
3.2.6. Structural Business Statistics.....	25
3.2.7. Business Register.....	27
3.2.8. Total Population Register.....	29
3.2.9. Quarterly Gross Domestic Product.....	31
3.3. LCS/SILC Redesign.....	35
4. Crosscutting Issues and Recommendations.....	38
4.1. Mixed Mode Household Surveys	38
4.2. Integration and Coordination of Economic Statistics.....	39
4.3. Outsourcing Experiment with the LFS.....	42
4.4. Nonresponse in Household Surveys.....	43
4.5. Statistics Sweden's Revised Quality Concept.....	45
4.6. Household Consumption Data.....	47
4.7. Data Editing.....	48
5. Summary and Recommendations for Future Rounds.....	50
5.1. Summary.....	50
5.2. Highest Priority Recommendations.....	52
6. References.....	54
Annex 1 Checklists for Accuracy Dimension of Quality.....	55

1 EXECUTIVE SUMMARY

In 2011, the Ministry of Finance directed Statistics Sweden to develop a system of quality indicators for a number of key statistical products. This system was to include metrics that reflect current data quality as well as capture any changes in quality that occur over time. With the help of external consultants, Statistics Sweden developed a quality evaluation approach that is referred to as ASPIRE: A System for Product Improvement Review and Evaluation or ASPIRE (see Biemer and Trewin, 2013 and Biemer, Trewin, Japac and Bergdahl, 2014). The review process has been conducted annually since 2011 for essentially the same core set of statistical products.

This report summarizes the results from the seventh annual review (Round 7) of ASPIRE which was conducted in May 2017 by the ASPIRE team (viz., Biemer, Trewin, and Kasprzyk). The Round 7 report covers the following 9 products which were also reviewed in the previous round: Annual Municipal Accounts (RS), Consumer Price Index (CPI), Foreign Trade of Goods Survey (FTG), Labour Force Survey (LFS), Structural Business Survey (SBS), Business Register (BR), Total Population Register (TPR), Producer and Import Price Index (PPI), and the GDP component of the quarterly National Accounts, (GDP(Q)). The Survey of Living Conditions (LCS/SILC) was excluded from this round but there was a workshop with the LCS/SILC team to discuss the forthcoming redesign. Some recommendations are made in this report.

Although not part of the data product review process, the ASPIRE team participated in several other additional discussions and presentations including the coordination and integration of economic statistics, several projects that are part of the broader nonresponse project, Statistics Sweden's revised quality regulations, and the data editing project. Comments on these discussions are found in this report. The ASPIRE team presented a seminar on ASPIRE to staff in Stockholm and Örebro, in part to gain some input into how the ASPIRE process might be improved.

There were only three external reviewers available but we wanted to have some reviews conducted in parallel given the short time available. Two reviewers were involved in some reviews whereas a single reviewer conducted the other reviews. However, all three reviewers were involved in discussing the outcomes of the discussions and the determination of the final ratings and to improve the consistency across products. As in the prior rounds, the evaluation for each product involved a self-assessment, reviews of relevant documentation, presentations by staff, interviews of key staff, and a staff review of the preliminary evaluation results with feedback.

As in previous rounds, each product was scored (on a 10-point scale) against criteria that were standardized across error sources. The use of quality criteria guidelines and checklists greatly facilitate the application of the criteria and, we believe, provide more consistent ratings. Overall scores were tallied as a weighted average of the scores for each error source where the weights were 1, 2, or 3 corresponding respectively to low, medium, or high intrinsic risks associated with each error source. More details on the ASPIRE process may be found in Biemer, et al (2014).

With respect to the key results from this round, all but three of the nine survey and register products showed an improvement in ratings. With a maximum possible score of 100 percent (indicating perfect quality), the product scores ranged from 54.4 percent (for PPI) to 64.5 percent (for FTG) with an average rating of 59.4 percent. (Exhibits 2a and 2b provide the scores for each product by error source.) Although not in this report, we prepared a 'Change Matrix' for each product that provides explanations for any changes in ratings since the previous round. Drafts were reviewed by the product areas and they are available from Heather Bergdahl on request.

Changes in scores from Round 6 to Round 7 ranged from 1.4 (BR) to -0.4 (LFS) with an average of 0.5 percentage points. While an average increase of 0.5 percentage points is not a substantial improvement, it does indicate that overall, the quality of data products continues to steadily improve.

Some additional findings from the data product reviews found in Section 2 include the following:

- In prior rounds, model/estimation had the lowest mean rating although it was tied with frame error in the last round. The mean rating has improved slightly so it no longer is the lowest. Frame error and specification error had the lowest ratings.
- As in prior rounds, the error source with the highest quality score is sampling error.

In addition to the product reviews, the ASPIRE team had the opportunity to hear and discuss a number of current crosscutting projects and bring our perspective to these projects. There are also some new cross-cutting issues that we have identified.

- Mixed Mode Household Surveys (Section 4.1). Consistent findings have emerged from the research – response rates are higher, costs are lower, differences between response modes for the subject matter tested so far are negligible, but there is no evidence of nonresponse bias being reduced. Additional research is warranted, particularly on topical content more likely to have mode effects.
- Integration and Coordination of Economic Statistics (Section 4.2). This topic has been discussed in previous ASPIRE reports and is of major importance. The SAMSTAT program has been put in place to address these issues. It is a major program and the report makes some suggestions with respect to project management issues.
- Outsourcing Experiment with the LFS (Section 4.3). A successful experiment and analysis of the experimental results based on data quality and costs led to a decision to develop a new tender that increases the proportion of LFS sample outsourced. Continuous monitoring and comparisons of costs and data quality in the outsourced and Statistics Sweden samples should be an important activity.
- Nonresponse in Household Surveys (Section 4.4). Addressing nonresponse in household surveys continues to be a major area of attention. Significant resources, staff time, and research project effort continue to be expended to mitigate this problem. There appears to have been some important progress. Through increased knowledge of the causes of the problem, and mitigation efforts directed at these causes, the deterioration in response rates has been arrested for the household surveys we examined. Some of the most effective mitigation efforts have been the use of mixed mode data collection, the development of models to reduce both callbacks and noncontacts, and improvements in the management of data collection operations. Much more remains to be done in these areas including the consideration of the use of paper and pencil interviewing (PAPI) and other communication channels to mitigate web and telephone nonresponse and a greater use of propensity models to direct the data collection efforts. Better understanding of the different reasons for non-contact would help in this regard. We note the efforts to improve the accuracy of telephone numbers but there does not appear to be any quantitative knowledge of how important this is.
- Statistics Sweden's Revised Quality Concept (Section 4.5). During 2016, Statistics Sweden established a revised quality concept further elaborating legal standards for quality in

official statistics in Sweden. Unfortunately, as was true for the previous quality concept, the revised quality concept is somewhat inconsistent with the data quality framework that was adopted for ASPIRE in 2011 and is the foundation of the ASPIRE process. Perhaps the most important difference is the absence of a component of Accuracy in the new quality concept that is a key source of error in the ASPIRE framework – viz., Specification Error.

Specification Error reflects how well an observation measures the variable (i.e. the target variable) it is supposed to measure. The revised quality concept does not include the concept of Specification Error (or invalidity) as a component of Accuracy. We see this as a shortcoming of the revised quality concept.

- Household Consumption Data (Section 4.6). Largely because of the length of time since the last Household Budget Survey, the benchmarks used for the household consumption component in the National Accounts are out of date. This was highlighted by Eurostat in their recent GNI review of the National Accounts. The CPI and SILC are also affected by out of date household consumption data. We propose the establishment of a project team to address this important issue looking at alternative data sources, building on existing work, and a simplified Household Budget Survey.
- Data Editing (Section 4.7). We support the establishment of a Committee to look at editing undertaken by Statistics Sweden in a holistic way. A large proportion of the effort in data collections is devoted to micro and macro editing. There are plenty of opportunities to improve the cost-effectiveness of editing. We support the initial activities proposed for investigation by the Committee and have made some additional suggestions.

In Section 5, we have identified 10 recommendations that we consider highest priority for improving the quality of Statistics Sweden's data products. Priorities were assessed on the basis of impact and viability with cost being an important aspect of viability. They are listed below in approximate priority order and further discussed in Section 5.

1. Undertake an analysis of the reasons for non-contacts as a basis for targeting mitigation efforts.
2. Undertake a study of the cost and quality differences between EVRY and Statistics Sweden's LFS operations.
3. Review the timing and sequencing of the set of projects involved in the Integration and Co-ordination of Economic Statistics
4. Research alternative methods and data sources for household consumption data.
5. Research web data collection to better understand how to reduce nonresponse bias through alternative modes to reach hard to contact respondents; further examine potential mode effects.
6. Continue profiling of the largest and most complex enterprises.
7. Redesign of LCS/SILC.
8. Research on data editing.
9. Specify and develop a Statistical Business Register.
10. Greater clarification of the treatment of specification error in Statistics Sweden's quality concept.

2 BACKGROUND AND INTRODUCTION

This is the seventh round of ASPIRE, an annual evaluation program that began in 2011. As with the previous rounds, the focus of this ASPIRE round is on the Accuracy quality dimension. Nine of the ten products that were reviewed last year were reviewed again this year. The Living Conditions Survey (LCS/SILC) was excluded from the ASPIRE review this year because it is being redesigned and Statistics Sweden thought it would be more productive to review and comment on the plans for the redesign rather than to conduct another ASPIRE review for this product. Our comments on this redesign can be found in Section 3.3. The nine products that comprise the scope of our review are listed in Exhibit 1.

Exhibit 1. Sources of Error Considered by Product

Product	Error Sources
Survey Products 1. Labour Force Survey (LFS) 2. Consumer Price Index (CPI) 3. Producer and Import Price Index, (PPI) 4. Annual Municipal Accounts (RS) 5. Foreign Trade of Goods (FTG) 6. Structural Business Statistics (SBS)	Specification error Frame error Nonresponse error Measurement error Data processing error Sampling error Model/estimation error Revision error
Registers 7. Business Register (BR) 8. Total Population Register (TPR)	Specification error Frame: Overcoverage Undercoverage Duplication Missing Data Content Error
Compilations 9. Quarterly Gross Domestic Product (GDP(Q))	Input data error Compilation error Modelling error Data processing error Deflation/Reflation error Balancing error Revision error

As in prior follow up rounds, an objective of Round 7 was to identify areas where clear improvements were made or deteriorations were experienced since the previous evaluation. Another objective was to follow up on previous year's recommendations and to take note of other efforts to improve quality. Our report documents the key findings of our quality reviews and attempts to identify the highest priority areas for improvement at the product level. In addition, some general recommendations are made for crosscutting issues that we consider to be high priority for Statistics Sweden.

The ASPIRE process, error sources and evaluation criteria, that were applied in this review are essentially the same those used in Rounds 5 and 6 (see Biemer, Trewin, Kasprzyk, and Hansson, 2015, 2016), which differed somewhat from those applied in rounds prior to Round 5. The general

ASPIRE process is described in greater detail in two papers: Biemer, et al (2014) and Biemer, et al (2017).

Exhibit 2a provides the overall scores for the six survey products and the two registers for each product's relevant error sources. The applicable error sources are shown in the first column of these tables while the other columns refer to the products being evaluated. For each product, a dark shaded cell corresponds to a "High Risk" error source for that product. Medium shaded cells correspond to "Medium Risk" error sources and unshaded cells correspond to "Low Risk" error sources for a product.

As discussed in prior ASPIRE reports, the interpretation of the error sources and criteria may vary between surveys, registers and compilations. For example, for a survey, it may be appropriate to consider measures such as bias and variance because the products of surveys are estimates. For registers, the concepts of bias and variance are not particularly appropriate because they are essentially data sets, not estimates. Instead, it may be more appropriate to consider the validity and reliability of the register data because these quality concepts are applicable for any set of measurements of population characteristics.

Exhibit 2b provides the scores for the quarterly GDP; i.e., GDP(Q). As discussed in Biemer, et al (2014), the error structure used in the evaluation of this product has been customized to reflect the unique operations associated with compiling the data and generating quarterly estimates of GDP. For that reason, the accuracy of GDP is treated separately from the other nine products.

There are a number of other differences among the error models used for surveys, registers and compilations which are explained greater detail in Biemer, et al (2014).

Exhibit 2c summarizes the total scores for all nine products over all seven ASPIRE rounds in the form of a histogram. These exhibits will be discussed in greater detail in the next section.

3 PRODUCT REVIEWS

3.1 GENERAL OBSERVATIONS

Before discussing each product's detailed ratings, some general observations regarding the results in Exhibits 2a, 2b and 2c as well as a few caveats can be stated.

First, there is a natural tendency to compare the overall scores across the products or to rank the products by their total score. However, the ASPIRE model was not developed to facilitate such inter-product comparisons and there are some risks associated with ranking products in this manner. For one, the total score for a product reflects a weighting of the error sources by the risk levels, which can vary considerably across products. Products with many high-risk error sources, such as GDP, may be at somewhat of a disadvantage in such comparisons because they must perform well in many high-risk areas to achieve a high score.

Second, the assessment of low, medium, or high risk is done within a product, not across products. Thus, it is possible that a high-risk error source for one product could be of less importance to Statistics Sweden than a medium risk error source for another product if the latter product carries greater importance to Statistics Sweden or for official statistics. If resources devoted to quality improvements are greater for one product than another, this could also explain why some products can show greater improvements than others. Further, although we have attempted to achieve some degree of consistency in ratings among products, some inconsistencies surely remain.

Finally, the scores assigned to an error source for a product have an unknown level of uncertainty due to some element of subjectivity in the assignment of ratings as well as other imperfections in the rating process. A difference of 2 or 3 points in the overall product scores may not be meaningful because a reassessment of the product could reasonably produce an overall score that differs from the assigned score by that margin. Thus, any ranking of products would need to acknowledge these inevitable and unknown uncertainties in the ratings.

A more appropriate use of the product scores is to compare scores for the same product across review rounds as a way of assessing progress toward improvements. As noted in Biemer et al (2014), the ASPIRE review process focuses on process changes, new knowledge gained or communicated, new research conducted or planned since the prior round, and the demonstrated success of mitigation efforts that could alter the error risks and justify changes in the quality ratings. We believe this process assures a high level of reliability in the round-to-round changes scores for each product.

Before discussing the results in Exhibits 2a and 2b, it should be noted that a number of corrections to the ratings were made in Round 6 that may affect the Round 6 to Round 7 comparisons, viz.

- for the SBS, the Round 6 rating of 7 for Specification Error was corrected to 5 and
- for the BR, the Round 6 rating of 8 for Specification Error was corrected to 5.

Close inspection of scores in Exhibits 2a and 2b yield the following observations:

- The last row of Exhibits 2a and 2b shows the Round 6 to Round 7 changes in the overall quality ratings by product. Ratings increased for six out of nine products participating in both this and the previous round. The increases were quite small in that none of them exceeded 1 point. For Exhibit 2a, the average increase is 0.5 points for surveys and registers and 0.5 points for GDP (Exhibit 2b).

- The largest improvement in Exhibit 2a is the BR (1.4) which is a substantial improvement. The largest deterioration was recorded for the LFS (-0.4). However, this change only represents a small percentage of the LFS' overall score and may not be particularly meaningful. Regarding nonresponse, the progress the LFS is making is quite positive.
- The FTG also posted a small decrease in its overall scores, but this too should be interpreted as essentially no appreciable change in rating.
- Two error sources in Exhibit 2a are tied for the lowest mean rating: Frame Error (undercoverage) and Specification Error at 55. The former only applies to the two registers while the latter affects all products and thus may deserve more emphasis while noting that often specification error arises from a deliberate decision for practical reasons. In Round 6, Model/Estimation Error was the lowest rated error source with a score of 55. Its rating increased slightly this year, however, this error source consistently ranks among the lowest.
- Measurement/Content Error is of high risk for six out of eight products in Exhibit 2a – the most of any error source. It has also seen great improvement across the seven rounds; its rating has increased by more than 1 point per round on average.
- Likewise, Data Processing Error is high risk for five out of six products and it, too, has increased by more than 1 point per round on average.
- Not surprisingly, the error source with the highest quality score, and by a wide margin, is sampling error. This was also true in all the prior rounds.

To increase their ratings, products could concentrate on high risk areas having below average ratings if viable mitigation strategies can be identified. However, these areas may not be the highest priority areas for quality improvement as other factors need to be considered. For example, one should consider the feasibility and costs of the improvements, the needs and relative priorities of other products and improvement activities, the importance of improving Accuracy compared with other dimensions of quality and initiatives that are promoted by Eurostat and other external groups.

Exhibit 2a. Product Error-Level, Overall Level, and Error Source-Level Ratings with Risk-Levels Highlighted and Comparisons to Round 6 Overall Ratings

Error Source/Product	LFS	CPI	PPI	RS	FTG	SBS	BR	TPR	Mean rating
Specification error	62	68	48	48	58	52	55	52	55
Frame error	60	65	55	50	57	62	56	59	58
overcoverage							58	65	62
undercoverage							53	57	55
duplication							57	57	57
Nonresponse error /Missing data	60	55	57	58	62	70	47	63	59
Measurement error/Content	65	68	50	55	65	55	58	55	59
Data processing error	55	63	58	62	65	58	N/A	N/A	60
Sampling error	73	68	63	N/A	N/A	85	N/A	N/A	73
Model/estimation error	62	55	50	48	75	48	N/A	N/A	56
Revision error	N/A	N/A	N/A	60	65	55	N/A	N/A	60
Round 7 mean rating	62,6	64,2	54,4	55,6	64,5	59,3	55,6	59,3	59,4
Round 6 mean rating	63,0	63,2	53,7	54,9	64,6	59,3	54,2	58,7	58,9
Change (improvement/deterioration)	▼ -0,4	▲ 0,9	▲ 0,7	▲ 0,7	▼ -0,1	■ 0,0	▲ 1,4	▲ 0,7	▲ 0,5
HIGH RISK MEDIUM RISK LOW RISK N/A= Not Applicable									

Exhibit 2b. Product Error-Level, Overall Level, and Error Source-Level Rating with Risk-Levels Highlighted and Comparisons to Round 6 for the National Accounts

Error Source/Product	GDP(Q)
Input data sources Production Side (Average)	56
Index of Service Production (ISP)	63
Index of Industrial Productions (IIP)	52
Merchanting (including royalties, licensing, R&D)	52
Input data sources Expenditure Side (Average)	54
Turnover	53
Government	57
Investments	52
Inventories	55
Net Exports in Goods and Services	53
Compilation error (modelling)	47
Compilation error (data processing)	50
Deflation error	60
Balancing error	55
Revision error	57
Round 7 mean rating	54,5
Round 6 mean rating	54,0
Change (improvement/deterioration)	▲ 0,5
HIGH RISK MEDIUM RISK LOW RISK N/A= Not Applicable	

Exhibit 2c shows the overall ratings by product for the seven evaluation rounds. Recall that, in Round 5, a sixth criterion (viz., Effectiveness of Mitigation Measures) was introduced which substantially reduced the ratings for most products and error sources. This criterion was also applied in the last two rounds with essentially the same results although there were a few improvements. Some reasons for the low ratings for this criterion as well as some remediation measures were discussed in the Round 5 report. Because of this new criterion, the ratings for the current round and Round 5 are not directly comparable to the ratings for Rounds 1-4 which did not use the sixth criterion. For Round 5, we reported ratings under both the new (with Effectiveness) and old (without Effectiveness) criteria to bridge the Round 5 ratings with the prior rounds' ratings. For Exhibit 2c, we continue to use both sets of scores for Round 5; however, for the current round, only the scores under the new, revised criteria are shown.

In prior rounds, the LCS/SILC consistently the lowest rated product in the ASPIRE process after posting substantial improvement in Round 3. This product is being redesigned (discussed in Section 3.3) and was not rated this year. We believe this redesign will create substantial improvement opportunities for this product. As in prior rounds, the FTG tends to be the highest rated product although its ratings have been somewhat flat over the last three rounds.

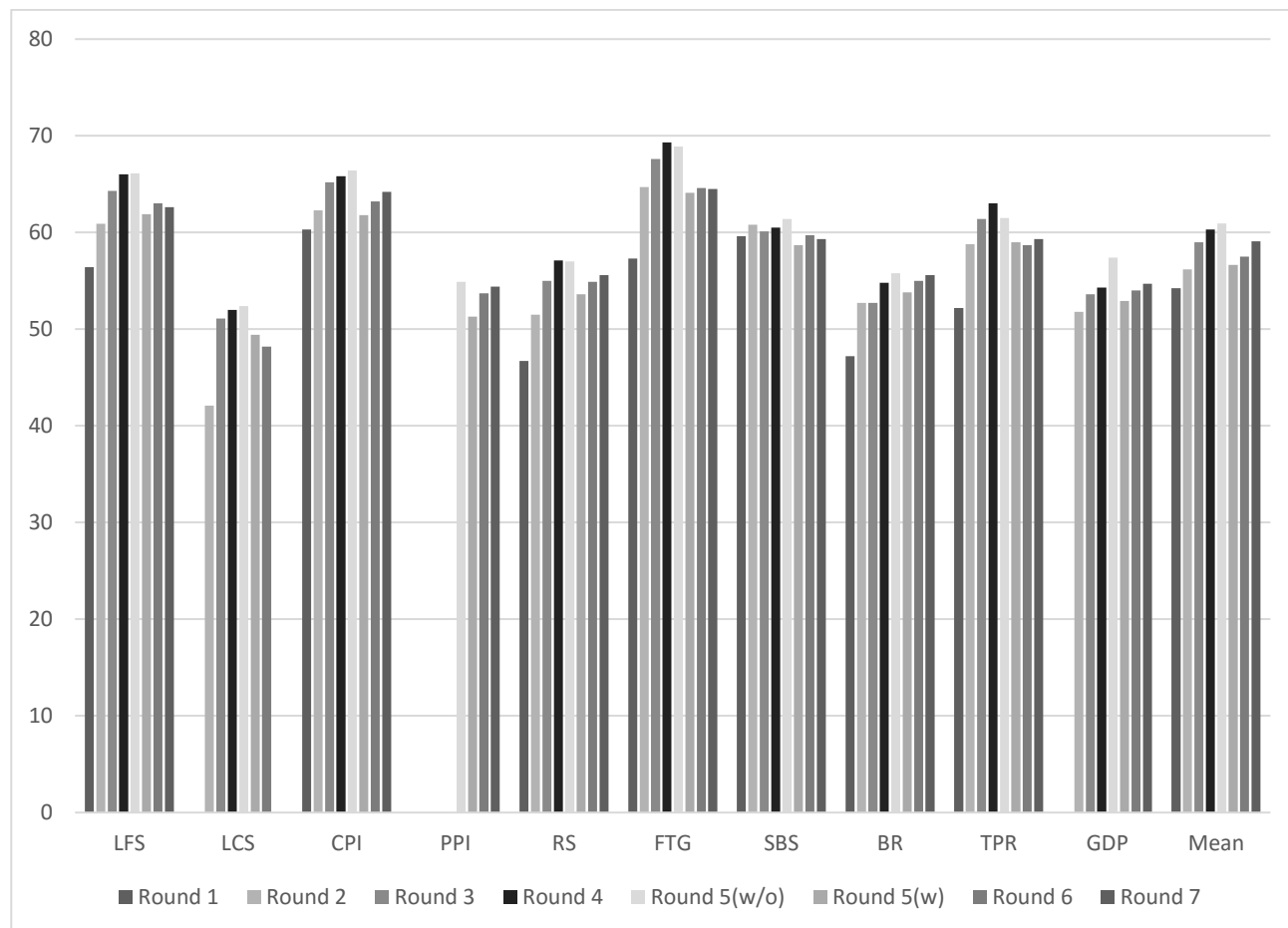
Although the mean quality ratings continue to increase year after year, the average increase for the last three rounds has been somewhat smaller than in earlier rounds, as can be seen from the "Mean" bars, the last set of bars in Exhibit 2c. Some possible explanations were also noted in our Rounds 5 and 6 reports. One is that the so-called "low hanging fruit" of quality improvement (i.e., improvements that can be more readily accomplished with low budgets and minimal activity) was picked up in early rounds. The achievement of further improvements will require a greater commitment of resources, personnel and innovative thinking.

In addition, we have noticed that some products, citing budgetary constraints and production demands, do not assign sufficiently high priority to continuous quality improvements. This can happen when management's attention is so focused on the routine production work that the objectives of continual quality improvement are given lower priority. There may be other distractions to quality effort such as the need to implement a new IT system.

Still, even products that assign a high priority to quality improvements find it difficult to maintain ratings of "Very Good" or "Excellent" for planning for mitigation for some error sources as they attend to quality improvements for other, needier error sources. This is acceptable performance because a product-area seldom has enough resources, personnel and time to do all that is needed to plan quality improvement across all error sources simultaneously. The challenge for the product manager is to ensure the focus is on the highest priority quality improvements at any point of time.

Finally, as we have cautioned in our prior reports, the results in Exhibit 2c do not necessarily represent the pace of data quality improvements for the 10 products. Although data quality improvement is the goal of ASPIRE, an improvement in ASPIRE ratings means that products have improved relative to the six ASPIRE criteria. We can only say that data quality has been improved to the extent the six criteria reflect actual reductions in the risks of product error. As an example, products may increase their ratings by developing plans designed to reduce the error. However, actual error reduction may not be realized until these plans have been implemented. This is one of the reasons the sixth criterion (viz., for Risk Mitigation Effectiveness) was introduced in Round 5.

Exhibit 2c. Overall Quality Ratings for All Products for all Seven Rounds including the Two Ratings for Round 5 Corresponding to “with (w)” and “without (w/o)” the Separate Assessments of Mitigation Effectiveness



Notes: ¹LCS/SILC was not evaluated in Round 1 and Round 7.

²PPI was evaluated for the first time in Round 5.

³The GDP(P) component of GDP(Q) was evaluated in Rounds 2-4. Thus, ratings for GDP(P/Q) reflect GDP(P) for Rounds 2-4 and GDP(Q) (i.e., GDP(P) and GDP(E)) for Round 5 and 6.

⁴New criterion added for Effectiveness of Mitigation Activities in Round 5

3.2 ASPIRE RATINGS

In this section, we review the progress over the past 12 months for the nine products shown in Exhibit 1 that were also reviewed in Round 7 using the checklists that appear in Annex 1. Customized versions of the checklists were used for the quarterly GDP to consider the unique error structure of the national accounts (see Biemer, et al, 2014). The ratings for each of the six criteria and applicable error sources are updated to reflect this progress. Then, we conclude the review of each product with our recommendations for the coming year.

A notable development is that almost all products have updated their quality declarations to better align them with the revised quality concept and quality declaration template. Many products used that opportunity to expand and clarify their quality declarations. The result is that there was general improvement of the documentation of product quality for these products. We believe the quality profiles should be widely available to the public, possibly accessible somewhere on Statistics Sweden's website.

3.2.1 LABOUR FORCE SURVEY (LFS)

CONTEXT

Notwithstanding some slight decreases in ratings, the LFS staff continue to make progress in several areas that impact data quality. Remarkably, response rates have remained at their Round 6 levels which suggest that they have finally stabilized. Only time will tell if this progress will hold. Still, the research on household survey nonresponse is continuing to yield important results for the LFS. Section 4.4 provides an extended discussion of these efforts. The outsourcing of data collection has continued and is being expanded. These efforts are discussed in greater detail in Section 4.3. Research on the use of web interviewing for possible use in the LFS is producing positive results so far (see also Section 4.1 for an expanded report on this research). A study of coverage error using population register data is reaching its finishing stages. Although the LFS recorded the largest deterioration across products (-0.4), this drop only represents a small percentage of the LFS' overall score and may not be particularly meaningful. We view the LFS's performance to be generally positive, particularly regarding the nonresponse issue.

SELECTED ACCOMPLISHMENTS

Progress toward Prior Recommendations

- *New Procurement.* Statistics Sweden is in the process of entering a new contract with an external supplier to collect up to 50 percent of the LFS sample. In the beginning, the planned workload for the external supplier will be incrementally increased to 13,000 six months later. The contract will stipulate that no more than 2 percent of the interviews should be by proxy. The interviewers with the supplier will be monitored at least once per month as they are at Statistics Sweden.
- *Web Data Collection Research.* LFS staff is preparing for a first test of web data collection in 2018 or 2019. The test will be limited to waves 2-7 and only the wave 1 permanently employed. Respondents will be offered the web option if they indicated interest in responding that way in wave 1. The web option may not have much impact on the nonresponse rate but could increase representativity of the sample and reduce data collection costs.
- *Frame Coverage.* To evaluate coverage error, LFS staff is comparing TPR and other register totals with the weighted LFS totals for the full sample by overall and by panel. Recent results

suggest that undercoverage is a somewhat bigger problem than overcoverage, particularly now that TPR provides an overcoverage indicator on the register. These results could lead to either more effective weighting adjustments or more frequent refreshment of the sample drawn from the register.

- *Nonresponse Bias Analysis.* LFS staff completed a research project that suggests that even though nonresponse increased over the past 5 years, the nonresponse bias does not appear to have. These results were based upon fully adjusted weighted data. It would be interesting to repeat some of this work with base weighted-only data to understand which demographic groups in the population were under-represented and how this changed over the five-year period. It is also important to determine whether the unequal weighting effect changed over the five-year period because of increasing weight adjustment variation.
- *Gap Analysis.* Progress was made on producing a “gap” analysis. This is a report that shows the number of interviewing hours requested by the LFS staff (by timeslot) (denoted by R_s) compared with the number of interviewing hours provided by the Data Collection Department (P_s). It can be used to monitor the ability of the Data Collection Department to meet the demands of the LFS. However, see recommendation 1 below for possible problems with this approach.

Other Accomplishments

- *Monitoring.* Behaviour coding is being carried out for both EVRY and Statistics Sweden interviewers. Comparisons of results between the two organizations should illuminate any differences in the way that interviewers approach the LFS interviewing task.
- *Recontact Improvements.* New questions were added to the questionnaire that will hopefully lead to a reduction in noncontacted sample members in subsequent waves. In addition, an agreement has been established with the Social Insurance Agency to merge their contact data onto the LFS sample. The Social Insurance Agency telephone numbers are thought to be more current. This will be further tested in the coming year.

RECOMMENDATIONS

Recommendations for Coming Year

1. *Revising the Gap Analysis.* Apparently, the gap analysis report does not reflect the number of interviewing hours that were requested by the LFS staff and the number of hours that were provided. Rather it reflects the Data Collection Department’s ability to staff to the levels they specify. We believe the gap analysis should consider both and would encourage this revision of the approach.
2. *Web Data Collection Research.* The mixed mode research that has been conducted so far has not found any systematic mode effects. However, that may be a consequence of the type of subject matter that has been tested. In the LFS, concepts such as looking for work or temporary (versus permanently) unemployed are difficult and comprehension may suffer without interviewer involvement. In addition, these problems will not surface if web interview testing is only carried out on the permanently employed. Therefore, we encourage the LFS staff to give some consideration to testing web interviewing on the unemployed as well.
3. *Analysis of Noncontacts.* Further analysis of the paradata for contacts, noncontacts, interviews and refusals is needed to better understand the causes of nonresponse, particularly regarding noncontacts. We encourage the LFS staff to continue their work in this area with emphasis on optimal calling strategies for noncontact cases.

4. *Cost and Quality Comparisons with EVRY.* Statistics Sweden has two interviewing facilities or components – one centralised (in Orebro) and another decentralized. These two components have very different costs and quality structures. Therefore, in considering costs and quality comparisons with EVRY, we believe it is important to breakout the two Statistics Sweden components separately as well as combined. For example, is the substantial cost/interview differential between Statistics Sweden and EVRY the same for both components? We suspect that decentralized costs/interview may be considerably higher than their centralized interviewing counterparts.

Other Areas for Consideration

1. *Updating the quality declaration.* We appreciate the recent revision of the quality declaration. However, two areas where the quality declaration could be improved are (a) some discussion of indirect (proxy) interviewing and its effects on measurement error and nonresponse, (b) a discussion and examination of the effects of dependent interviewing on response quality – both advantages and disadvantages, and (c) the addition of quantitative information as it becomes available from the ongoing studies. These elements should be considered in the next revision of the quality declaration.
2. *Industry and Occupation Coding.* The LFS staff is collaborating with the staff that does industry and occupation coding and that work has produced positive results. One area that should be explored is whether the information now being collected by interviewers on occupation is sufficient for the coders to assign accurate codes. A study conducted by Biemer and Caspar (1994) showed that if certain key elements of an occupation are not recorded by interviewers, occupations are coded much less accurately. Perhaps coding the portion of the interview that elicits the occupation and industry information could be a part of the behaviour monitoring study.

Exhibit 3. Labour Force Survey (LFS) Ratings for Round 7

	Error Source	Average Score Round 6	Average Score Round 7	Knowledge of risks	Communication	Available expertise	Compliance with standards & best practices	Plans towards risk mitigation	Effectiveness of mitigation measures	Risk to data quality
Accuracy (control over error sources)	Specification error	62	62	☑	○	☑	☑	☑	☑	L
	Frame error	57	60	☑	☑	☑	○	☑	☑	L
	Non-response /Missing data	58	60	☑	○	☑	○	☑	○	H
	Measurement /Content	65	65	☑	☑	☑	☑	☑	☑	H
	Data processing error	55	55	○	○	☑	☑	○	☑	M
	Sampling error	78	73	☑	○	☑	○	○	☑	M
	Model/estimation error	63	62	○	○	☑	☑	☑	☑	M
	Total score	63,0	62,6							

Scores					Levels of Risk			Changes from round 6	
●	◐	○	☑	○	H	M	L		
Poor	Fair	Good	Very good	Excellent	High	Medium	Low	Improvements	Deteriorations

CONTEXT

The Swedish CPI continues to be of a very high standard especially when compared to those of other countries. Many important initiatives have been taken over the last year. Nevertheless, there continues to be a range of initiatives that result in continuous improvements to the accuracy of the CPI taking advantage of new technology and new data sources.

SELECTED ACCOMPLISHMENTS

Progress toward Prior Recommendations

- *CPI Error Study.* The planned study did not take place in full but there was some work on the most important TSE components of the CPI. There has been work on creating an algorithm that optimally allocates the CPI sample.
- *Extended Use of Scanner and Internet Data.* Scanner data use has been extended to include prices for real estate fees, fresh fruits including berries, vegetables, cheese and meats. There has been no further extension of the use of web scraping but possibilities have been investigated. Much more of the price data collection is now being undertaken centrally where quality is easier to manage.
- *Quality Adjustments.* A very innovative Implicit Quality Index diagnostic tool was introduced in the year before last. This enables the impact of quality adjustments to be assessed and is an important macro-editing tool. It has also been used to help users to interpret the data – that is, how much of the change in a CPI component is due to quality change.
- *Monitoring the work of Price Collectors.* Updated tablet computers have been successfully introduced and paradata on time is being collected. The use of this data has not happened as fast as anticipated but there are plans to analyse this paradata as well as supporting the optimum allocation of the sample.
- *CPI Weights.* A review of the upper level weights of household consumption by product group, and data sources for these weights, has been carried out. It is recognised that more work is needed as the quality of many of the weights is uncertain because of the lack of a reliable and comprehensive data source.
- *Owner Occupied Housing Costs.* The methodology for condominiums has been improved and is now consistent with other types of owner occupied housing. Interest payments are no longer included as part of mortgage costs.

Other Accomplishments

- *Improved Commodity Coverage.* Some new commodity groups are now included in the CPI, namely real estate fees, care leasing, child care, and elderly care.
- *Other Quality Improvements.* There have been a range of important improvements such as centralised collection of some clothing items, investigation into how to incorporate dentistry costs, centralised quality adjustments for home electronics and an improved methodology for package holidays.

RECOMMENDATIONS

1. *CPI Error Study*. When the data become available, they should be used to support a TSE approach to improving the accuracy of the CPI because this would provide the evidence base for deciding where to best place the research and error mitigation efforts. It would also provide information in support of optimising resources across the whole CPI. In the interim, further work on the optimal allocation of the sample should take place.
2. *Extended Use of Scanner and Internet Data*. Continue to broaden the use of scanner data and “web scraping” to reduce sampling errors in the relevant components but, perhaps more importantly, to reduce the measurement errors, especially those associated with assessing price discounts. Extend this approach to problematic areas like international internet purchases. The work being undertaken in the Innovation Laboratory is important but there should be a less conservative approach to web scraping and the perceived barriers to the more extensive use should be resolved.
3. *Monitoring the Work of Price Collectors*. There is a lot of dependency on the work of the price collectors and their work should be routinely monitored. The newly introduced tablet technology has been used to collect ‘paradata’ as well as price data. This capability should be used to better monitor and evaluate the quality and effectiveness of the work of the price collectors. There would be merit in researching how other countries have used paradata in their CPI collections.
4. *CPI Weights*. Because it is unlikely that reliable Household Budget Survey (HBS) data will become available soon, there should be some investigations into data sources (e.g. scanner data), other than the HBS, that could be used for updating commodity weights used in the CPI. The focus should be on those products where the price movements might be quite different to the rest of the CPI. Sensitivity analysis of the weights would assist in this regard.

Exhibit 4. Consumer Price Index (CPI) Ratings for Round 7

	Error Source	Average Score Round 6	Average Score Round 7	Knowledge of risks	Communication	Available expertise	Compliance with standards & best practices	Plans towards risk mitigation	Effectiveness of mitigation measures	Risk to data quality
Accuracy (control over error sources)	Specification error	68	68	🟢	○	🟢	🟢	○	○	H
	Frame error	62	65	🟢	🟢	○	🟢	🟢	○	M
	Non-response /Missing data	55	55	○	○	🟢	🟢	○	🔴	L
	Measurement /Content	67	68	🟢	○	🟢	🟢	🟢	○	H
	Data processing error	67	63	🟢	○	🟢	🟢	○	🔴	H
	Sampling error	67	68	🟢	🟢	🟢	🟢	🟢	○	H
	Model/estimation error	52	55	○	○	○	○	🟢	○	H
Total score		63,2	64,2							

Scores					Levels of Risk			Changes from round 6	
🔴	🔴	○	🟢	🟢	H	M	L		
Poor	Fair	Good	Very good	Excellent	High	Medium	Low	Improvements	Deteriorations

CONTEXT

The PPI, among other things, provides very important input to the National Accounts when calculating GDP in constant prices. It is also an important statistic in its own right, used in the analysis of inflation by the Riksbank and others. It makes an important contribution to monetary policy. Analysis has shown that specification error and measurement error are the most important sources of error.

SELECTED ACCOMPLISHMENTS

Progress toward Prior Recommendations

- *Change in Data Collection Process:* The Data Collection Department took over data collection in January last year. After “teething” problems over the first few months, the product area believes that the quality of the data provided by the Data Collection Department is of a higher quality than that provided when collected by them.
- *Communication with Core Users:* Discussions continue to be held regularly with the National Accounts and different units in the Economic Statistics Department. This communication has given important insights to both producers and users of the PPI including on the relative importance of the different types of error.
- *Relative Risk Assessment Tool:* This evaluation method has been discussed with Statistics Denmark and has resulted in some adjustments to methods. This work will be presented at the Nordic meeting on producer price indexes and the Voorburg Group meetings on price indexes.
- *Benchmark with Another Country:* Three bilateral meetings have been held with Statistics Denmark exchanging thoughts and communicating information on risks for error in the PPI and how they might be addressed. Because of these discussions there have been some adjustments to the approach used by Statistics Sweden.
- *Measure the Price of Trade Margins:* A request for internal funding for a pre-study was not approved. The product area presented their ideas to the User Council for Economic Statistics and they were very positive.
- *Monitor Quality Adjustments:* An Implicit Quality Index similar to that used by the CPI is being studied. It was not possible because of the lack of information on the quality adjustments but the new coding system for the types of quality adjustments means that it will be possible to separate between different kinds of implicit and explicit quality adjustments. This means that it will be possible to produce an Implicit Quality Index in the future.

Other Accomplishments

- *New Product Groups.* A pilot survey for four new product groups (TV-broadcasting, radio, travel agency and licences) was conducted this year. They are in the experimental stage and it is planned that the resulting indexes will be first published in 2018. Including these products will increase coverage of the services prices, excluding health care and schools where no market prices exist, from about 80 to 86 per cent.
- *Web Questionnaire Improvements:* The number of questions in the web questionnaire has been reduced. Also, an English version of the questionnaire has been provided for those

businesses whose language preference is English. There are now controls in place in the web questionnaire comparing values with the previous period.

RECOMMENDATIONS

1. *Measure the Price of Trade Margins.* The Economics Statistics User Council regards this as high priority as do the National Accounts. A promising pre-study already exists that could be used as a base for a field study with experimental price collection. Funding of this next stage of this important work is necessary for it to progress.
2. *Monitor Quality Adjustments.* We continue to support a recommendation for a measure comparable to the Implicit Quality Index of the CPI for the PPI as this measure could be a valuable tool for both keeping track of quality adjustments as well as an illustration of their significance to key users (e.g. the NA). It may also provide measures of the effectiveness of the quality adjustment processes. The required information on quality adjustments in the PPI is now available to enable an Implicit Quality Index to be compiled.
3. *Evaluation of Measurement Error.* There have been several improvements to the web questionnaire. They should be evaluated to see whether they have been effective in mitigating measurement error.
4. *Expanding Coverage of Imports and Exports of Services:* The proposed work in this area is strongly supported particularly considering the high growth in trade in services. A pre-study should ensure that a reliable survey framework can be developed to support the expanded commodity coverage.

Exhibit 5. Producer and Import Price Index (PPI) Ratings for Round 7

	Error Source	Average Score Round 6	Average Score Round 7	Knowledge of risks	Communication	Available expertise	Compliance with standards & best practices	Plans towards risk mitigation	Effectiveness of mitigation measures	Risk to data quality
Accuracy (control over error sources)	Specification error	48	48	○	○	○	○	○	●	H
	Frame error	53	55	○	○	○	●	○	●	M
	Non-response /Missing data	57	57	○	○	●	○	○	●	M
	Measurement /Content	48	50	○	●	○	●	○	●	H
	Data processing error	58	58	●	○	●	●	○	●	H
	Sampling error	63	63	●	○	●	●	●	○	H
	Model/estimation error	48	50	○	●	○	●	○	●	H
	Total score	53,7	54,4							

Scores					Levels of Risk			Changes from round 6	
●	●	○	●	○	H	M	L		
Poor	Fair	Good	Very good	Excellent	High	Medium	Low	Improvements	Deteriorations

CONTEXT

Ratings for the Annual Municipal Accounts (RS) are slightly higher than the previous year. This is due to the plans RS staff have made to reduce error based on the research conducted this past year. The plans include improving instruction to municipalities on what elements should be considered as part of “common costs” and a pre- and post-edit analysis of the effects of changes to the edit system. During this past year, the focus has been on data transmission, understanding edit failures, and improved understanding of reported “common costs” data.

SELECTED ACCOMPLISHMENTS

Progress toward Prior Recommendations

- *Common Cost Data.* Last year, research with the common costs model found that common costs have been increasing significantly more than gross costs. Eighteen municipalities exhibited high trends in the change in common costs data. Staff followed up with these municipalities to obtain an understanding of their reporting. This research will result in revised instructions to the municipalities as to which areas are included in “common costs.”
- *Analysis of Changes at Decision Points.* Staff have just begun to work with the current data collection to produce and review statistics, such as edit failure rates and the number and type of changes at the decision points in the flow charts.
- *Quantifying Effectiveness.* Little has been done this past year on providing evidence of the effectiveness of edits and processes. Plans for next year include evaluating the effects of changes to edits implemented last year and evaluating the effect of implementing a few “hard” controls (mostly consistency checks).
- *Manual Editing.* The flow charts identified an important aspect of the RS edit and submission system – the manual editing phase. Staff reviewed the flow charts to identify differences in the editing processes for different parts of the survey; adjustments have been made to make the editing processes more similar across the surveys. Visual edit checks have been removed, resulting in less dependency on an individual editor. A team that benefits from the perspectives of different editors has been established to improve the consistency of micro-editing.
- *Analysis of Respondent Comments.* RS staff note that there were fewer respondent comments this past year than the previous year. Staff reviewed comments as to whether they have value or not. Comments are coded as accepted or not or if a followup with the respondent is necessary. Coding the substance of the comment is not done and it appears that little has been done to develop realistic substantive categories of comments to monitor the comments through the IT Interface; if implemented, this would allow the RS staff to understand or explain the data submitted.

Other Accomplishments

- *Editing and reduction in respondent burden.* RS staff reached an agreement with one of the larger county councils for an ongoing project during the year to better understand the difference between “raw” data, as submitted by the county council and Statistics Sweden edited data. The study will compare the raw data with processed data. The study goals are to improve the editing process, improve the quality of submitted data, and reduce respondent burden for the county council, with the latter being the primary goal.

RECOMMENDATIONS

1. *Measurements of impacts.* RS staff should develop measures of the effectiveness of their edits and processes, including statistics that show the sensitivity of the estimates to changes in edit specifications and processes. Many initiatives are taking place with the RS system. Staff need to produce quantitative evidence of the impact of the editing and processing initiatives.
2. *Analysis of changes at decision points.* As suggested previously, RS staff should identify areas within the system that are inefficient or produce little change in estimates. Statistics, such as edit failure rates and number and type changes at the decisions points in the flow charts, should be produced.
3. *Improved understanding of comments.* An improved understanding of respondent comments would likely help staff to improve the efficiency of their edits. RS staff should code the substance of comments; realistic substantive categories of comments should be developed to monitor the comments through the IT Interface, thus allowing the RS staff to understand or explain the data submitted.
4. *Methodological communication.* RS staff should communicate with staff in the Department for Development of Processes and Methods (PMU) who have a project on the use of controls from previous reports for business surveys. RS staff have real experience to contribute to this project.
5. *Technical documentation.* Several projects will look at the effects of edits and processes, RS staff should take the necessary time to describe the project, and document the impact on estimates of changes to the editing system.

Exhibit 6. Annual Municipal Accounts Ratings for Round 7

	Error Source	Average Score Round 6	Average Score Round 7	Knowledge of risks	Communication	Available expertise	Compliance with standards & best practices	Plans towards risk mitigation	Effectiveness of mitigation measures	Risk to data quality
Accuracy (control over error sources)	Specification error	48	48	○	○	○	○	●	●	L
	Frame error	50	50	○	○	●	●	●	●	L
	Non-response /Missing data	58	58	○	○	●	○	○	●	M
	Measurement /Content	53	55	○	○	●	●	○	●	M
	Data processing error	60	62	●	○	●	●	●	●	H
	Model/estimation error	48	48	○	●	●	○	○	●	M
	Revision error	60	60	○	○	●	●	○	○	L
	Total score	54,9	55,6							

Scores					Levels of Risk			Changes from round 6	
●	●	○	●	○	H	M	L		
Poor	Fair	Good	Very good	Excellent	High	Medium	Low	Improvements	Deteriorations

CONTEXT

The FTG area experienced quite a bit of staff turnover this year – two methodologists resigned and three IT experts left the unit, and thus the area was somewhat understaffed, particularly in the IT area where only one, shared person remains. Nevertheless, important progress in some areas was made. The staff provided two documents that summarize some of their progress this year: (a) Improvement of tools and methods for ITGS and (b) Distribution keys in foreign trade statistics. There were a few, small decreases in ratings which translated into a very small decrease in FTG's overall scores, but this is not an important change in the overall rating from last year.

SELECTED ACCOMPLISHMENTS

Progress toward Prior Recommendations

1. *Estimation Below the Cut-off*. Reference (a) above describes some of the work that was accomplished to improve below cut-off trade in Intrastat. This work produced more stable estimates for reduced costs and complexity. This work addressed the estimation of both the level of trade and the distribution of trade over commodities and countries.
2. *Statistical Value Survey*. We encouraged the FTG staff to continue to pursue redesign options for the Statistical Value Survey that will reduce respondent burden. It may be possible to use alternative sources of data to estimate statistical value rather than the survey. The FTG received a grant to study the options and hope to conduct a test of these methods in 2017. However, the full-scale data collection to update the statistical value adjustments to invoice value is not scheduled until 2018.
3. *Respondent Burden*. Although there was some progress made on reducing respondent burden, the work in the area continues. In particular, the staff has yet to evaluate the extent to which respondent burden has been reduced by the Intrastat Data Entry Package (IDEP.web).

Other Accomplishments

- *Other Work Related to Estimation Below the Cut-off Estimation*. A few traders are below the cut-off and are thus exempt from having to report their trade to Statistics Sweden. Their trade must still be estimated by commodity and country which can be quite complicated. A method based upon distribution keys (which is a method for mapping commodity codes to countries) was developed that seemed to produce mixed results at the 2-digit NACE level. Although not considered in the study, results at the three and higher digit NACE level are expected to be worse.
- *SIMSTAT*. FTG received additional funding from Eurostat to pursue additional research on SIMSTAT.
- *Data collection*. A survey for vessels/airplanes that was previously carried out by the subject matter unit with paper questionnaires has been moved to the data collection unit in order to conduct more effective and coherent data production.
- *Specific Movement of Goods*. The paper data collection process for Specific Movement of Goods (vessels and aircraft) was converted to an electronic collection, reducing burden and data processing error.
- *New Format for File Deliveries*. Because the delivery of some files must be sent under the SDMX format, the process for doing this with an Excel spreadsheet was quite burdensome for

FTG staff. This project developed a SAS program for producing this file automatically, thus reducing burden and data entry error.

- *New Methods for Validation.* This project implemented a dynamic editing schedule in the new Intrastat IT system which distributes the editing work continuously to FTG staff. This may increase the likelihood of obtaining a response from a respondent to correct an edit failure. It is not yet functional because the new IT system has not been activated.

RECOMMENDATIONS

Recommendations for Coming Year

1. *IT-Support.* Many high priority data quality improvements in FTG require IT support. Apparently, that support has been inadequate and many significant projects are, consequently, delayed, particularly those that would take advantage of the new IT system to improve data quality. We recommend that IT support be increased for FTG in the coming year if possible so that important quality improvement work can be accomplished.
2. *SIMSTAT.* It is important for FTG to explore the accuracy of their Intrastat import statistics through comparisons with other EU countries. This is particularly true if Eurostat requires that Intrastat import statistics agree with the exporting country's corresponding export statistics. This work should be given high priority next year.
3. *Documentation for Users.* FTG should continue their work to document their data collection and processing methodologies in furthering transparency with users. We believe this will increase their users' knowledge of FTG processes which may also lead to quality improvements of those processes.
4. *Estimation below the Cut-off.* As noted in our prior review, the sampling cut-off for Intrastat was doubled to SEK 9 million for imports in 2015 to reduce respondent burden measured in the aggregate. The effects of this change on model/estimation bias have not been assessed. Further, the process for estimating the value of shipments below the threshold is neither well documented nor well understood. We recommend that the FTG (a) evaluate the effect of raising the cut-off for Intrastat imports on estimator bias and (b) substantially improve the documentation of the estimation process for enterprises below the cut-off.
5. *Statistical Value Survey.* We encourage the FTG staff to continue to pursue redesign options for the statistical value study that do not require traditional survey methods, but rather would rely on administrative data sources. The FTG should pursue a research agenda in this area with possible funding from Eurostat and close consultation with the National Accounts.

Other Areas for Consideration

- *Separating Trade in Goods and Services.* The effects on FTG and FTS estimates due to the inability to accurately separate trade in goods from trade in services for some products.
- *Accuracy of Commodity Codes.* The accuracy of the FTG statistics is highly dependent on the accuracy of the commodity codes that enterprises assign to a good. There is a need for an evaluation of commodity coding error for the most problematic codes, particularly in the CN6 and higher granularity levels.

Exhibit 7. Foreign Trade of Goods (FTG) Ratings for Round 7

	Error Source	Average Score Round 6	Average Score Round 7	Knowledge of risks	Communication	Available expertise	Compliance with standards & best practices	Plans towards risk mitigation	Effectiveness of mitigation measures	Risk to data quality
Accuracy (control over error sources)	Specification error	58	58	○	○	☑	☑	☑	☑	M
	Frame error	57	57	○	○	☑	☑	○	○	L
	Non-response /Missing data	62	62	☑	☑	☑	○	○	○	M
	Measurement /Content	65	65	☑	○	☑	☑	☑	○	H
	Data processing error	65	65	☑	☑	☑	☑	☑	☑	H
	Model/estimation error	73	75	☑	☑	○	○	☑	○	M
	Revision error	67	65	☑	☑	☑	☑	○	☑	H
	Total score	64,6	64,5							

Scores					Levels of Risk			Changes from round 6	
●	☑	○	☑	○	H	M	L		
Poor	Fair	Good	Very good	Excellent	High	Medium	Low	Improvements	Deteriorations

CONTEXT

Improvements in the SBS were somewhat stalled in this round while staff await further development on several interdependent fronts, including: the new Business Register, the new IT-system, guidance on fining enterprises, and guidance on which profiling system should be used to delineate Kind of Activity Units (KAUs). For that reason, their overall score is not much changed from the last round.

SELECTED ACCOMPLISHMENTS

Progress toward Prior Recommendations

- *Processing Data on Editing.* The staff pursued improvements on macro-editing and looked at the cost-effectiveness of the various stages of micro-editing and macro-editing as part of the study. One important finding was that approximately 42 percent of the SBS budget is spent on editing of which about one third is on macro-editing. In addition, macro-editing appears to have small effects on the estimates. The staff is considering a more automated approach to macro-editing to reduce the cost but not editing quality. Given the high cost of editing, more selectivity in the editing process is warranted. A sensitivity analysis conducted by the NA could shed light on which areas of editing are the most important to retain.
- *Prepare for the New BR.* SBS has been preparing for the new BR and how to manage the likely discontinuities in the SBS data series. The current emphasis of the BR work seems to be on the reduction of burden for enterprises. One should also be vigilant that these burden reductions do not have important consequences on data quality.

Other Accomplishments

- *Embedded Edits.* The staff have been experimenting with embedding additional edits into the electronic questionnaire which will reduce cost and possibly improve accuracy. The work so far suggests that hard edits may be off-putting for respondents and may lead to greater burden. The current work is focusing on soft-embedded edits.
- *Electronic Reporting.* The number of businesses providing their data electronically continues to increase.
- *Merchanting.* Correctly classifying merchanting (for example, either as goods or services or possibly both) has become an important issue in light of the experiences in Ireland which showed that how merchanting is classified can have a substantial impact on GDP. The SBS has been studying the classification of merchanting and hopes to bring greater clarity to the issues in the near future.
- *Reduced Reporting Burden.* There was a decrease in number of surveyed companies from 600 to 500 in order to decrease respondents' burden and the survey cost. The plan was carefully executed with a design to minimize the risks for quality. However, some reduction in quality is expected.

RECOMMENDATIONS

Recommendations for Coming Year

1. *Inconsistencies with Final SBS and Corrected SBS at $t+1$.* In the discussion about Revision Error, SBS staff noted that the difference between the preliminary and final SBS estimates is

not as concerning as the difference between the final estimates and its revision one year later. The size of that difference is relatively large for some key statistics which suggests bias in the final SBS estimates. More work is needed to illuminate this finding and its impact on the NA and other important SBS users.

2. *Increase Profiling of Large Enterprises.* SBS should continue to promote efforts to increase the number of large enterprises that are profiled to ensure the NACE classifications are accurate in SBS and National Accounts (NA) statistics. The focus should be on the largest and more complex enterprises where the impact from profiling is likely to be greater.
3. *Macro and Micro Editing.* The costs of editing (42 percent of the total SBS budget) raises the question: “Is the money devoted to editing appropriately allocated from a total survey error perspective?” There may be opportunities to reduce the costs of editing substantially with only minimal deteriorations in data quality. That would allow resources to be more effectively allocated – for example, toward profiling or reducing overcoverage bias in the survey because of the inclusion of inactive businesses in the sampling frame. The SBS staff should pursue this goal in the coming year building on the work done to date.

Other Areas for Consideration

- *Virtual Statistical Business Register (SBR).* SBS staff should work with BR staff to make the SBR a reality. SBS staff can contribute to the development in a well-articulated plan describing the key characteristics of the SBR while maintaining its compatibility with the general BR. SBS staff also can contribute ideas as to how the SBR would be created and maintained, the costs and benefits of a SBR and how it would impact SBS data quality. We believe this would help clarify and promote the concept of the SBR and facilitate and expedite its development.

Exhibit 8. Structural Business Statistics (SBS) Ratings for Round 7

	Error Source	Average Score Round 6	Average Score Round 7	Knowledge of risks	Communication	Available expertise	Compliance with standards & best practices	Plans towards risk mitigation	Effectiveness of mitigation measures	Risk to data quality
Accuracy (control over error sources)	Specification error	53	52	○	○	☑	○	☑	☑	M
	Frame error	60	62	☑	☑	☑	○	○	☑	M
	Non-response /Missing data	70	70	☑	○	☑	☑	☑	○	M
	Measurement /Content	55	55	○	○	○	○	○	○	H
	Data processing error	58	58	☑	○	☑	○	☑	☑	H
	Sampling error	85	85	○	☑	☑	○	○	☑	M
	Model/estimation error	48	48	○	○	○	☑	○	☑	H
	Revision error	55	55	○	○	☑	○	☑	○	H
	Total score	59,3	59,3							

Scores					Levels of Risk			Changes from round 6	
●	☑	○	☑	○	H	M	L	Improvements	Deteriorations
Poor	Fair	Good	Very good	Excellent	High	Medium	Low		

CONTEXT

The Business Register (BR) has shown steady improvements in quality since the ASPIRE process was initiated. Last year was no exception, showing higher ratings from Round 6 of ASPIRE. BR staff have engaged in ongoing critical work with the re-engineering project and the coordination of economic statistics (SAMSTAT) project. Staff are also participating in a project about profiling to delineate statistical units, especially kind of activity units (KAUs). Research was documented on the delineation of enterprise units from related legal entities and the evaluation of NACE codes for companies in the construction and construction services industries by comparing business register data with occupational register data. Due to higher priorities and reduced resources, work on the Statistical Business Register (SBR) has been delayed, an unfortunate occurrence.

SELECTED ACCOMPLISHMENTS

Progress toward Prior Recommendations

- *Increase Profiling of Large Enterprises.* Several options for profiling are being explored: the Eurostat method (which includes business units located outside Sweden), the Swedish method and an automated approach to creating enterprises. The latter approach is being considered for less complex enterprises with only two or three legal entities. One objective of the research is to understand the effect of using the Eurostat versus the national approach and whether a single approach satisfying both Eurostat and Statistics Sweden's objectives is feasible. At present, only about 50 enterprises have been split into KAUs and the goal is to profile at least 200 enterprises.
- *Development of the new BR System.* This work continues to be part of the re-engineering project: an important project for Statistics Sweden with a target date to have the system ready for use by the SBS in November 2018. Two BR staff participate in the project.
- *Development of the Statistical BR System.* Little progress has been made on supporting the creation of a BR specifically for statistical purposes (i.e., the Statistical BR). The Objects and Population (OP) group prepared a discussion manuscript on this topic but the unavailability of staff and budget resources as well as higher priority projects has delayed any serious activity on the project.
- *Accuracy of NACE Coding.* A pilot study report that compares a unit's classification, its business description, and its employees' skills was completed for construction industries and it suggests a potential way to identify potential misclassification. It showed that inaccuracy of NACE coding in these industries may be a significant concern. Additional work on the accuracy of NACE codes is planned for this year.
- *Quality Indicators on Base Registers.* Last year the "Quality Indicators on Base Registers" project reached the proof-of-concept stage. Additional work has been done to prepare data for further testing.

Other Accomplishments

- *Data improvements.* Continued cooperation with the Tax Agency resulted in the high likelihood of monthly income statements being provided to Statistics Sweden. Staff are preparing for the possible availability of this information.

- *Communication.* Launching a new Sharepoint site provides an important communication channel about BR for staff and internal users of the BR.

RECOMMENDATIONS

Recommendations for Coming Year

1. *Coordination of economic statistics.* The re-engineering project and the SAMSTAT project are important high visibility projects for Statistics Sweden. BR staff should continue to place a high priority on active participation toward achieving project goals.
2. *Profiling Enterprises.* Eighty of the largest enterprises have been identified for profiling this year. Perhaps that should be rationalised to focus on the largest most complex enterprise groups where profiling will have the most impact on the accuracy of statistics. BR staff should establish good lines of communication with the Coordination and Corporate Affairs unit to establish responsibilities for the profiling. There should be no ambiguity about responsibilities as to the number of enterprises and which enterprises each unit should profile.
3. *Development of the Statistical Business Register System (SBR).* Even though the SBR has a lower priority in the larger project of coordinating economic statistics, harmonization, planning and development work should be initiated that emphasizes where research should be prioritized and quality improvements made. It should be possible to start early work on the functional specifications.
4. *Accuracy of NACE Coding.* As in previous years, steps should be taken to monitor the level of error in NACE coding on an ongoing basis through independent coding studies. The results of these studies should be made available to users, especially internal users. A strategy for addressing the most important inaccuracies in the NACE codes should be developed.
5. *Quality Declaration.* The BR quality declaration should be made available on Statistics Sweden's web site.

Exhibit 9. Business Register (BR) Ratings for Round 7

	Error Source	Average Score Round 6	Average Score Round 7	Knowledge of risks	Communication	Available expertise	Compliance with standards & best practices	Plans towards risk mitigation	Effectiveness of mitigation measures	Risk to data quality
Accuracy (control over error sources)	Specification error	53	55	○	○	☑	○	☑	☑	M
	Frame overcoverage	58	58	○	○	☑	○	○	○	H
	Frame undercoverage	52	53	○	☑	○	○	○	○	H
	Frame duplication	57	57	○	○	☑	☑	○	☑	L
	Non-response /Missing data	47	47	○	○	○	○	○	☑	L
	Measurement /Content	55	58	○	○	☑	○	○	☑	H
	Total score	54,2	55,6							

Scores					Levels of Risk			Changes from round 6	
●	☑	○	☑	○	H	M	L		
Poor	Fair	Good	Very good	Excellent	High	Medium	Low	Improvements	Deteriorations

CONTEXT

The TPR's ratings are slightly higher this year due to a successful project related to the improvement of the household register. The TPR staff continue to reach out to Statistics Sweden product staff, either directly or through methodological work groups on the use of the indicator of overcoverage variable. Good communication with the Swedish Tax Agency has fostered useful working relationships, resulting in lower overcoverage in the TPR.

SELECTED ACCOMPLISHMENTS

Progress toward Prior Recommendations

- *Overcoverage.* The TPR staff continued its successful outreach of the constructed overcoverage variable. A paper about the overcoverage variable was presented at the Nordic statisticians' conference last summer. Discussions with the Data Collection Department have taken place concerning the variable's use in developing better contact strategies. An internal user group has met and discussed the variable and data products, such as LCS/SILC, have participated.
- *Impact of overcoverage on survey estimates.* TPR staff have discussed with the LFS staff the importance of a study that quantifies the impact of the overcoverage on LFS estimates. Education statistics staff have used the overcoverage variable in its nonresponse analysis; however, results are not yet available.
- *Continue working with the Tax Agency.* The TPR staff maintains good and regular communication with the Tax Agency by meeting 3-4 times a year. Less overcoverage of the population occurred in 2015 when compared to 2014. A TPR initiative to study coverage at the regional level indicates that at least in one region's coverage can be off by 3 percent.
- *Overcoverage/undercoverage at the Regional Levels.* The TPR staff working with the Statistics Sweden education register validated the accuracy of address information and showed that for Uppsala the accuracy of the population register can differ by as much as 3 percent. This undercoverage is primarily because of students who do not register where they are living while attending university. In some sense, this may also be regarded as a content error in that their current addresses are not accurate. These findings were forwarded to the Tax Agency so that they would encourage students to register where they live.

Other Accomplishments

- *Improvements to the household variable.* Approximately 212,000 persons have missing data for their dwelling. TPR staff developed an imputation model that uses "register family" as an indicator of the formation of a household. Under this assumption, the model could assign 194,000 out of the 212,000 individuals to households, a significant improvement.

RECOMMENDATIONS

Recommendations for Coming Year

1. *Evaluation of the overcoverage model.* Following up on last year's recommendation, it is useful to evaluate the model misclassification error (false positive and false negative) rate. This study will inform uses of the model and will provide information necessary for this work to be published.

2. *Continue discussions with the Tax Agency.* Regularly scheduled discussions with the Tax Agency are necessary to maintain and improve the accuracy of the TPR. In particular, discussions are necessary to encourage the Tax Agency to improve outreach to students to register where they live, thereby improving the content of the TPR at the municipal level.
3. *Measure Impact of overcoverage on survey estimates.* TPR staff should continue to work with survey areas that use the TPR to study the impact of overcoverage on survey estimates. Staff should follow-up with LFS staff and the education statistics staff to learn the results of their studies that evaluate the impact on survey estimates. In addition, TPR staff and methodologists should collaborate on experiments designed to understand how TPR's overcoverage indicator can be used effectively in prioritising nonresponse follow-up.
4. *Improve understanding of the household variable.* As part of the LCS/SILC data collection, a household roster is developed during the interview. It is useful to collaborate with the LCS/SILC staff to measure differences in household membership between the TPR with the interview-reported household.
5. *Quality indicators for core variables.* TPR staff should continue to monitor the extent of missing data for critical register variables. While staff have addressed this issue in previous years, it is important to monitor these quality indicators annually to identify if missing data rates have changed.

Exhibit 10. Total Population Register (TPR) Ratings for Round 7

	Error Source	Average Score Round 6	Average Score Round 7	Knowledge of risks	Communication	Available expertise	Compliance with standards & best practices	Plans towards risk mitigation	Effectiveness of mitigation measures	Risk to data quality
Accuracy (control over error sources)	Specification error	52	52	○	○	○	○	○	●	M
	Frame overcoverage	65	65	●	●	●	●	○	○	H
	Frame undercoverage	57	57	○	○	●	●	●	●	L
	Frame duplication	57	57	○	○	●	●	●	●	L
	Non-response /Missing data	60	63	○	○	●	○	○	●	M
	Measurement /Content	55	55	○	○	●	○	●	●	L
	Total score	58,7	59,3							

Scores					Levels of Risk			Changes from round 6	
●	●	○	●	○	H	M	L		
Poor	Fair	Good	Very good	Excellent	High	Medium	Low	Improvements	Deteriorations

CONTEXT

The quarterly GDP estimates are produced from a very large and complex set of inputs from Statistics Sweden and other external sources. For our review, as with previous rounds, we could only look at a small number of the data sources that provided the greatest risk to the accuracy of GDP covering both the production and expenditure side. These are shown in Exhibit 11.

In 2015, we reviewed the input data sources for the expenditure-based estimates for the first time. We identified significant areas of potential improvement in collaboration with the National Accounts (NA) staff. The current situation with some of these data sources is:

Household consumption. Although the quarter-to-quarter data on turnover may appear reliable, there are problems with outdated annual benchmarks. There was criticism of this during the recent Eurostat review. The Household Budget Survey is potentially a very important source but the 2016 survey was cancelled because of significant nonresponse problems and other data quality issues. It is not clear when the next survey will be conducted and whether it will provide reliable data. National Accounts require aggregate data by commodity group for their purposes. This is less demanding than many other users who require micro level data. This brings into place the possibility of using alternative data sources or a simpler HBS.

Investments. The lack of data for the second quarter remains an issue. Forecasts are used instead. This is a volatile part of the accounts and there are large annual revisions when the annual SBS data is introduced. This problem should reduce when the quarterly business survey is introduced in 2018.

Research and Development. Quarterly estimates are model based using the trend in value added in the relevant industry as well as foreign trade of services in R&D. A survey on R&D is conducted every 2 years. It is an item where large revisions are expected and improvement of quarterly estimates is important. Since the last review there have been investigations into the quality of data on forecast R&D for the coming year. It appears that it will be quite useful for supporting the extrapolation models used for the quarterly estimates.

Foreign Trade. Companies in certain industries (for example, IT) are having increasing difficulty distinguishing between goods and services. This distinction is not easy and by having different surveys for goods and services, problems with the double reporting of transactions or missing transactions can occur. The work of the Coordination and Corporate Affairs unit has helped to reduce this problem.

Industrial Production Indexes. The newly introduced Production Value Index addresses this problem and other problems associated with distinguishing goods and services.

We note that a major five-year revision of the National Accounts is planned for 2019 which includes the updating of benchmarks. Other big issues will also be included in this revision e.g. trade margin estimates for the trading industries. The major motivation for the revision is the Eurostat recommendation to do this every 5 years where new sources and methods can be included and updates to structures and benchmarks. Statistics Sweden will also introduce the new quarterly business survey if it is not implemented earlier.

SELECTED ACCOMPLISHMENTS

Progress toward Prior Recommendations

- *Replacement IT System.* Work has continued a replacement of the Swedish NA IT-system scheduled for full implementation in 2019. Some parts have now been implemented, namely to store and receive input data. There are also plans for the considerable transition work that needs to take place including the availability of historical data within the new system.
- *Quarterly Business Statistics.* There are now definite plans to conduct a quarterly business survey for the largest enterprises to obtain estimates of intermediate consumption and to be supplemented by VAT data for the smaller enterprises. It will be introduced in 2018 one year later than previously anticipated. This will also allow investments data to be collected four times per year. The more accurate intermediate consumption data should reduce the size of the revisions when quarterly data is benchmarked to data from the annual accounts.
- *Sensitivity Studies.* Work on the sensitivity analysis has been continuing. Although the initial focus has been on annual GDP, there are implications for quarterly GDP. The most recent work has been to examine the automatic balancing of the annual national accounts using different starting points for some key component series. Several papers have been prepared on this work and have been presented in relevant conferences. One paper has been accepted for publication in the Statistical Journal of the IAOS.
- *Inventories.* Inventories now include sizeable military inventories such as military weapons, and spare parts for airplanes and other equipment.
- *Training.* There is no specific training course in national accounts but the new recruits have a study circle using *Understanding the National Accounts*, an OECD publication by François Lequiller and Derek Blades. National Accounts staff also participate in Eurostat's courses when they are offered at times that don't clash with the NA production timetable. Some National Accounts staff participate in the new course for Economic Statistics whilst others help with presentations thereby increasing knowledge of National Accounts in other parts of Statistics Sweden.
- *Consistency with Balance of Payments.* The working group of national accounts and balance of payments staff has been investigating data sources, conceptual and modelling differences between the two collections' estimates of the current account. Reconciliations of these differences should lead to more reliable data for both collections and some changes have already been made to balance of payments.

Other Accomplishments

- *Production Value Index.* The harmonization of the industrial and services production indexes has been completed with the harmonized survey introduced in 2015. This has enabled compilation of a Production Value Index from the second quarter in 2016. It is used for extrapolation of measures of production. This has enabled the capture of services and trade margins in the manufacturing industries as well as trade margins rather than total turnover from trade in the service industries. Adjustments for changes in inventories have also been made.
- *Merchanting.* With the support of the Coordination and Corporate Affairs unit, the quality of the company reporting of merchanting on a quarterly basis has improved.
- *Deflation.* Prices indexes for deflation are available for four new areas of services production.

RECOMMENDATIONS

Recommendations for Coming Year

1. *Household Consumption Data.* Largely because of the length of time since the last Household Budget Survey, the benchmarks used for the household consumption component are out of date. This was highlighted by Eurostat in their recent GNI review. The CPI and SILC are also affected by out of date consumption data. Furthermore, there is a need to identify the data sources to support the regular update of benchmarks. The full detail of the HBS is not required, only aggregated data at a relatively broad commodity group level. Some of these data may be available from alternative data sources. We recommend the establishment of a project team to look at effective means of developing household consumption benchmarks (see section 4.6). This should build on the existing project looking at alternative data sources for the HBS. If such a project team is established, National Accounts should be an active participant.
2. *Sensitivity Studies.* We strongly recommend the continuation of the sensitivity studies. There may be further useful work that could be done on the balancing process. However, a specific recommendation is to do sensitivity analysis on the household consumption benchmarks. The benchmarks need to be more accurate for those commodity groups with high growth rates, high volatility or rapid price change and the sensitivity analysis could focus on the impact of having out-of-date benchmarks for these commodity groups.
3. *Training of staff.* There needs to be more formality in the training, especially for new staff, while making greater use of new technologies to deliver that training. Self-paced training courses supplemented by coaching/tutoring by NA staff would be the ideal especially if use can be made of existing material. The Statistical Institute for Asia and the Pacific has a library of e-learning materials, some of which have been used by other Offices (e.g. UK Office of National Statistics) that might be able to be used by Statistics Sweden. Other advanced statistical offices should be approached to see what e-learning materials they hold that might be able to be provided to Statistics Sweden.
4. *Merchanting.* Improved profiling offers the best opportunity for improving estimates of merchanting. In the interim, there should be improved collaboration with the Coordination and Corporate Affairs unit to obtain increased knowledge of merchanting activity of the most significant Swedish enterprises.

Exhibit 11. Quarterly Gross Domestic Product (GDP(Q)) Ratings for Round 7

Error Source			Average Score Round 6	Average Score Round 7 with Effectiveness	Knowledge of risks	Communication	Available expertise	Compliance with standards & best practices	Plans towards risk mitigation	Effectiveness of mitigation measures	Risk to data quality
Accuracy (control over error sources)	Production Side	Input data, ISP	62	62	○	○	☑	☑	☑	☑	H
		Input data, IIP	62	63	○	○	☑	☑	☑	☑	H
		Input data, Merchanting (including royalties, licensing, R&D)	50	52	○	○	○	○	○	☑	H
	Expenditures Side	Input data, Turnover	55	53	○	○	☑	○	○	☑	H
		Input data, Government	57	57	○	○	☑	○	○	☑	M
		Input data, Investments	52	52	○	○	☑	○	○	☑	H
		Input data, Inventories	55	55	○	○	☑	☑	○	☑	H
		Input data, Exports in Goods and Services	53	53	○	○	☑	☑	○	☑	H
		Compilation (modelling)	47	47	○	○	○	○	☑	☑	H
		Compilation (data processing)	50	50	○	○	○	○	○	☑	H
		Deflation	60	60	☑	○	☑	☑	○	☑	H
		Balancing	52	55	○	○	☑	○	○	☑	H
		Revision	57	57	☑	○	☑	○	○	☑	M
		Total score		54,0	54,5						

Scores					Levels of Risk			Changes from round 6	
●	◐	○	☑	⦿	H	M	L		
Poor	Fair	Good	Very good	Excellent	High	Medium	Low	Improvements	Deteriorations

3.3 LCS/SILC REDESIGN

CONTEXT

The LCS/SILC surveys have a long history at Statistics Sweden. The history of the LCS dates from 1975 and has a strong national user base. The SILC began in 2005 and is conducted largely to satisfy Eurostat data requirements. The two surveys have competing demands made by data users, and while significant content is similar between the two surveys, a number of differences exist, resulting in data processing complications, multiple data products, and delays in products to the users. An additional complication to the data system is the Children's Survey, which uses as its sample a child in the selected LCS/SILC address, resulting in additional response burden and low response rates. The response burden on the respondents, the significant complications in processing multiple surveys, and new Eurostat requirements has made the continued conduct of the LCS/SILC data system unsustainable. A redesign of the survey is necessary to address these issues and bring the content up-to date and reduce complexity.

LCS/SILC WORKSHOP

The LCS/SILC Workshop allowed discussion of many issues. The LCS/SILC staff have initiated outreach to the Statistics Sweden Scientific Board and to the users of the surveys. Even though multiple user communities exist, more specific goals and objectives for the surveys are desirable. Staff described the user community, the current LCS/SILC survey design and the outreach project being undertaken to assess user data needs, and scheduled to end this fall. One topic that is not considered by data users is whether sufficient fiscal and staff resources are available to satisfy user demands; survey redesign efforts are constrained by the availability of budget and staff.

Staff also provided a succinct statement of redesign goals- namely, to reduce the complexity of the survey design, to avoid multiple estimates of similar variables, to satisfy the legal requirements for data, and to free up staff to focus on data quality issues in the surveys.

Other topics addressed by staff included a discussion of the problems associated with the current design. These include a number of definitional differences between LCS and SILC, including difference in definitions of the "income year," difference in the ages of the in-scope population, different weights, and different estimates for similar variables to name a few.

Contributing to the complexity and burden of the LCS/SILC is the Children's Survey. A household roster is developed as part of the LCS/SILC and a child is selected as the sample individual for the Children's Survey. Questions are administered to the sample child. As a supplement to the LCS/SILC, the survey provides a low-cost data collection approach but suffers from low response rates, particularly for some subgroups (e.g. migrants), and high response burden (in conjunction with LCS/SILC),

An updated Quality Declaration provided background information to supplement the workshop presentations.

RECOMMENDATIONS

1. *Integration of the LCS/SILC into one survey.* There are substantial resource inefficiencies in conducting 2 surveys in the same time frame with similar content. Staff should make the difficult decision to combine the two surveys into one survey. This recommendation, also advocated by the Statistics Sweden Scientific Board, will result in fewer data products and additional staff availability after implementation. Staff should begin to review the questions covering similar content in the two surveys and harmonize the questions so that a "core" set

of questions required by both surveys are defined for the redesigned survey. Decisions on how to address existing population and conceptual differences should be made. Timing of the implementation of the revised content modules should be identified for the redesigned survey. Plans for “bridging” the old surveys and the redesigned survey should be made. Survey redesigns generally require compromises among the various user communities. Consequently, the positive and negative implications of decisions and reasons for each decision should be described and available to the user communities.

2. *Develop a detailed schedule for the survey redesign project.* Survey redesigns have multiple components that lead to successful redesign goals and transition from one data system to the redesigned system. A detailed realistic time schedule should be developed that identifies the dates for implementation of the new system and all individual activities that lead to the new survey, including linkages between the activities. This includes planning time to harmonize questions common to LCS and SILC, identifying timing and content of supplementary modules, planning and implementation of experiments and tests to inform redesign decisions, identifying and designing data products, and plans for developing a bridge from current estimates to those based on the redesign.
3. *Ensure sufficient time exists for testing and experimentation.* Major issues will be identified in the redesign planning - issues that require experimentation and testing- for example, different design alternatives, cognitive testing of new and revised questions, and the use of the web as a mode of interview. Sufficient time to implement and analyse the results of methodological tests is necessary for making informed decisions. Planning for and implementing tests and experimentation must be done.
4. *Develop a draft survey redesign proposal.* There are many directions a survey redesign can take, given time and resource constraints. While the staff must be sensitive to user data requirements, it will be useful in the near term to develop a sufficiently detailed draft redesign proposal – “a strawman proposal” – that can be shared with the user communities and solicit feedback on the draft redesign proposal. The redesign proposal – the strawman – should be discussed with the user communities in the context of the available project staff and budget. This has the benefit of propelling the discussion forward and focussing on the new proposed design model and its ability to meet survey objectives with an understanding of resource limitations. It will also help develop ownership of the new design by some of the user community.
5. *Survey design, data delivery requirements, and data collection staffing.* Under the current design, Eurostat’s requirement for a faster release of data results has the potential to result in the need for many more data collection staff in the first half of the collection year than in the second half of the year. The lack of a standard and constant workload across the whole year can result in an interviewing staff that is part time and insufficiently trained. In considering survey redesign options, LCS/SILC staff should consider a continuous data collection design in which SILC based estimates are based on the 12 months prior to close out date.
6. *Identify new approaches to the Children’s Survey.* The current two stage design for identifying a sample child in a cooperating LCS/SILC household has problems that make the approach difficult and not sustainable. The Children’s Survey has very low response rates, especially in subgroups of non-mainstream households. And increasing the sample of households is only possible through sampling additional LCS/SILC households, a costly and inefficient process. Staff should bring together several external methodologists and internal methodologists to identify new and innovative approaches to the design of the survey. For

example, a member of the ASPIRE team, proposed the redesigned LCS/SILC have two sampling strata, defined by age (children/adults). Even though staff identified several difficulties with the approach, it, nevertheless, suggested another way to think about accomplishing the goals of the redesign.

Other Areas for Consideration

1. *Develop user training approaches for the redesigned survey.* Users will have many questions about the redesigned survey once it is implemented. Staff can help the process by developing good user documentation, user training materials and conducting user hands-on training (locally and at conferences). The training, materials, and documentation will mitigate the impact of the design changes for users.
2. *Develop staff analytic capabilities.* Eurostat's requirements for the SILC survey include the production of both cross-sectional and longitudinal data. This is true of the current SILC design as well as any redesigned SILC. Staff have not been able to take advantage of the analytic potential of the survey because of continuing operational responsibilities as well as insufficient preparation for analysing panel data. Investing in opportunities for staff to better understand the data collected and providing training opportunities in analysis methods should be an important and continuing component of a data product's program.
3. *Conduct research on nonresponse in the LCS-SILC and Children's Survey.* Nonresponse in the LCS/SILC and the Children's Survey is quite high by most standards. Staff should develop and implement plans to better understand the nonresponse rates and nonresponse bias. In particular, more should be learned about the noncontact rates- who was not contacted and for what reason, and apply what is learned to improving call scheduling strategies. Staff should collaborate with methodologists who have been involved with the broader nonresponse project to understand the research conducted on LFS and identify how LCS/SILC can benefit.

4 CROSSCUTTING ISSUES AND RECOMMENDATIONS

4.1 MIXED MODE HOUSEHOLD SURVEYS

Statistics Sweden has initiated several experimental projects to test the viability of mixed mode data collection, where the additional mode offering is the use of the web to complete the questionnaire. The research questions identified by staff are:

- Does offering the web as an alternative to the telephone mode increase response rates?
- Does the mode used affect the respondents' answers?
- Can costs be reduced by the offering of an alternative interview mode?
- Can nonresponse bias be reduced by offering an alternative mode for response?

The results of the experimental projects have been consistent – response rates are higher, costs are lower, differences in responses by mode are negligible, but there is no evidence that nonresponse bias is being reduced. While the latter result is disappointing, much has been learned from these experiments and confidence in offering two modes is warranted. The LFS is going to implement web as a response mode in waves 2-7 for those who hold a “steady job” and we await its successful implementation in the LFS.

RECOMMENDATIONS

1. Experimental data found no evidence that nonresponse bias has decreased through the offering of web data collection alternative, suggesting that other means are necessary to reach nonresponding subgroups such as youths and the elderly. In that regard, staff should study and test applications that collect web data via smart phones and test text messaging to increase response via smart phone.
2. Staff should continue to conduct research on the effects of respondent-reported data for respondents who choose to use the web. Even though mode effects have not been observed in the experiments thus far, additional research that focuses on the effect of mode on important subgroups as well as content that is more susceptible to mode effects should continue. If mode effects are found in specific content areas, staff should study how best to reduce these effects for that type of content. (See also Recommendation 1 in Section 4.4 for additional ideas for studying mode effects.)
3. The LFS should test the offer of a web collection alternative in Waves 2-7 on respondents other than those having a steady job. Testing the web collection alternative on those individuals who have more complicated job situations is desirable if the web alternative is to be offered across the entire sample.
4. Little discussion occurred concerning the testing of a self-administered paper and pencil response mode (PAPI). More research is needed on the offering of this mode in a mixed mode survey and its effect on the quality of reported data.

4.2 INTEGRATION AND COORDINATION OF ECONOMIC STATISTICS

For most rounds of ASPIRE we have made recommendations relating to the integration and co-ordination of economic statistics. In recent rounds, we have received presentations from Statistics Sweden on work they have done in this area. This progress has been most gratifying and we would like to acknowledge these very positive steps.

There is now a formal program known as SAMSTAT with a program leader and a high-level Steering Committee. It is aligned to Strategy 2020, Statistics Sweden's overarching strategy. We were provided with an Activity List of the various tasks involved in SAMSTAT. We did not see the underlying plan for SAMSTAT showing the various projects, key milestones, interdependencies, etc. but understand one is being developed. SAMSTAT has received a special internal funding allocation but it recently received a significant reduction because of more general reductions in the funding of Statistics Sweden.

It might be useful to reflect on progress on recommendations from the last round before looking at new recommendations. Our updated recommendation is shown after each of the six recommendations from the last round.

1. *Business Register*. When the current development of the Business Register (BR) is complete, work should commence on including the information within the BR, and the systems capabilities, to enable a Statistical Business Register to be extracted. For example, this would exclude businesses that are registered but inactive. More generally, the next version of the BR should be designed so that it can better address the major quality concerns in economic statistics.

Updated recommendation: Whilst we understand the emphasis on completing the current BR development, work should commence on the Statistical Business Register development as soon as possible. It will lead to significant improvements in the accuracy of economic statistics. The commencement has recently been deferred until spring 2018. It might be possible to start some of the preliminary specification work in parallel to current BR development work.

2. *Profiling of Large Enterprises*. There are plans in place to ensure that the largest and most complex enterprises are profiled into KAUs so that significant industry activities within the enterprise are identified. Where the full set of financial information is unavailable at the KAU level, it has been agreed that models utilising available information should be used to provide estimates at the KAU level. This work should be given a high priority. Furthermore, steps should be taken to ensure the KAUs are used uniformly across business surveys in Statistics Sweden through the application of common business frameworks.

Updated recommendation: Work has been taking place and some large complex units have been profiled but the recommendation remains active. It has been suggested that perhaps 200 enterprises should be profiled. We would caution against this large number. It would be best to focus on smaller number but including the most complex multi-industry enterprises where the impact from profiling would be greatest.

3. *Harmonisation of Business Units*. As far as possible, harmonize the selection of business units across the business surveys especially for those surveys that contribute to the National Accounts. It has been suggested that different units might be used for selection than for publication and this would be worth considering. Before a change in unit definition is

adopted, the transition issues should be carefully considered because they could be disruptive for some collections.

Updated recommendation: It is planned to do work on this during the current year. The recommendation remains active.

4. *Standard Classifications.* There should be a study to assess where more should be done to implement standard classifications in a way that would support national accounts and other users.

Updated recommendation: There do not appear to be any plans to work on this in SAMSTAT. The recommendation remains active although we acknowledge that it may be lower priority than some of the other SAMSTAT activities.

5. *Methodological Decisions.* Ensure that key methodological decisions, such as adjustments for nonresponse, are performed in a consistent way. A current methodological decision of interest is whether to use dependent survey feedback (see Section 3.2.8) and an agreed position should be reached on this.

Updated recommendation: This recommendation remains valid. It appears that no work on this aspect is planned for SAMSTAT. We strongly recommend that the scope of the SAMSTAT activities be expanded to include this work.

6. *Rationalisation of Collections.* Given the extensive uses of administrative data, Statistics Sweden still conducts a large number of data collections. There is also some duplication resulting in inconsistent or incoherent estimates. There is scope for rationalisation (that is, re-evaluating whether each collection is necessary), reducing respondent burden and possibly freeing resources for other activities. An example of inconsistent estimates is in the area of employment estimates and one early task might be to rationalise the multiple labour collections that are currently conducted, not always giving coherent results.

Updated recommendation: We note the plans to include the compilation of labour accounts within the scope of SAMSTAT. We support this. It should lead to improved knowledge of the strengths and weaknesses of the various labour collections and provide some insights into future rationalization of these collections. We also note the Quarterly Business Statistics project introduces a new quarterly business survey which will incorporate data collected in existing surveys including the investment and inventories surveys.

The presentations and discussions have identified some other issues we thought we should raise. This is mostly concerned with project management aspects. Of particular concern is the very ambitious program outlined for SAMSTAT especially in the first few years. The budget reductions make this even less viable. Some hard decisions need to be made on what should be done in these early years. Some activities will have to be deleted or deferred.

Of relevance to this is the three objectives of SAMSTAT. These are:

1. Simplify and facilitate for the data providers (i.e. reduce reporting burden),
2. Increase quality in statistics (includes integration aspects),
3. Make the production process more efficient.

All three are worthwhile objectives but there will be some tension between them as it will not generally be possible to attain all three objectives simultaneously. For example, there can often be trade-offs involved when looking at both accuracy and efficiency. It will be an important task of the Steering Committee to resolve these tensions. It may assist if there is corporate agreement on the

highest priority objective. Discussions with Statistics Sweden staff suggest this is reduced reporting burden but this should be confirmed.

The focus is on infrastructure and foundation activities first. We would support this. The intention is to cover most of these activities in 2017, although the Statistical Business Register was delayed by 12 months.

There appear to have only been selective discussions with users. This should be addressed especially for the main users. They will provide feedback which is important for determining priorities. Also, it will help get ownership from them which may be of use when seeking external funding.

RECOMMENDATIONS

In addition to the revised recommendations made above, we would like to issue some new recommendations.

1. *Timing*: There should be a review of overall timetable to assess what is achievable given the resources available. This may involve deferring or deleting some activities.
2. *Clarify main priority*: Three objectives are stated but there will be tension between these objectives. It would be helpful if there was agreement to the highest priority objective.
3. *Discussion with users*: There should be discussions with the main users of economic statistics to inform them of SAMSTAT and obtain guidance on what are the highest priorities. It may also assist with the next recommendation.
4. *Funding*: The funding level is inadequate. It is doubtful if the proposed activities within SAMSTAT can be achieved within the resources currently available in Statistics Sweden unless there is discontinuation of some significant collections. It is recommended that a case for external funding be developed given the importance of economic statistics to monetary and economic policy analysis.
5. *Project management advisor*: SAMSTAT is a very complex project involving many activities and interdependence between activities. Good project governance is essential. A project advisor who is expert in setting up project planning arrangements would be very useful in this respect including the identification of linkages between activities and the critical path.

4.3 OUTSOURCING EXPERIMENT WITH THE LFS

The experiment with outsourcing 20% of the LFS sample using EVRY resulted in a lot of information concerning the quality and efficiency of private sector data collection compared with government staff data collection. A detailed analysis was conducted last fall that showed higher response rates for EVRY, a higher proportion of sample units contacted on the first and second contact during the correct time window, more contacts per unit, and fewer deviations from the questionnaires by the EVRY interviewers. EVRY had a higher proportion of indirect (proxy) interviews, longer interview times, and significantly lower costs per interview. The analysis resulted in a favorable conclusion that outsourcing a portion of the LFS sample was not only feasible but desirable. Consequently, a new tender was issued to continue using a private sector company for a portion (50 per cent) of the LFS sample. The tender made explicit the need for reduced number of indirect interviews and mandated the contractor achieve a 65% response rate.

RECOMMENDATIONS

1. The difference in cost per sample unit between EVRY and Statistics Sweden is quite large and needs to be better understood. Statistics Sweden should continue to conduct a detailed cost analysis for the LFS data collection. The cost components for both centralized and decentralized interviewing staffs at Statistics Sweden should be compared.
2. Staff should monitor continuously the data quality characteristics of the EVRY sample and the Statistics Sweden sample. Because of the importance of the LFS and the split sample approach to data collection, staff must be vigilant that the LFS data do not deteriorate over time. Continuous monitoring of costs and data quality characteristics will become even more important since the new tender places additional data quality requirements on the contractor.

Other recommendations that are relevant to the EVRY work can be found in Section 3.2.1.

4.4 NONRESPONSE IN HOUSEHOLD SURVEYS

The substantial resources, time and attention that have been devoted to arresting the decline of response rates over the past six years appear to be bearing fruit this year. Response rates seem to have stabilized somewhat to about 59% in the Labour Force Survey which is where they were in Round 6. Other household surveys (such as the LCS/SILC) also seem to have benefitted from these long-term efforts with response rates generally stabilizing. Knowledge regarding the nonresponse problem continues to increase through reviews of the literature and nonresponse studies. Progress has been made to improve communication with respondents to help motivate response as well as with the data collection units to maintain appropriate staffing levels and initiate nonresponse mitigation efforts. In addition, the outsourcing experiment with EVRY seems to have delivered improved response rates both in the outsourced sample as well as the in-house sample. The latter may be the result of a transfer of knowledge between EVRY's and Statistics Sweden's Data Collection Department regarding efficient and effective interviewing methods. However, changes to the interview facilities (for example, the use of open interviewer cubicles) may have also contributed to these positive results.

The mixed mode data collection research which began in 2015 continues to show promise for reducing costs with no apparent degradation in data quality, particularly mode effects. Plans are progressing to offer the option of completing a web questionnaire to persons holding steady jobs in rotation groups 2-7 of the LFS with the promise of substantial cost savings.

The so-called Change Project also promises to further improve cost efficiency, data quality and the work environment surrounding data collection. This project seeks to improve communication among interviewers and supervisors regarding the daily goals for their work assignments. The project also proposes to reorganize the Data Collection Department so that production planning is more integrated with production operations. The hope is that this structure will allow the Data Collection Department to better meet the production and quality requirements leading to more stabilized and enhanced outputs. The results so far suggest that in-house data collection is now more efficient and effective and response rates have ceased declining. We applaud these efforts and remain optimistic that more positive results will come from the Change Project.

We were also made aware of a study that will determine the most efficient and effective strategy for contacting sample members when multiple telephone numbers are available for the sample member. In particular, this study will also evaluate the quality of telephone number obtained from the Social Insurance agency which is believed to possess telephone numbers having high contact probabilities.

RECOMMENDATIONS

The program on nonresponse in household surveys has progressed, but there are several high priority areas where concerns remain. These are:

1. *Mixed Mode Data Collection.* Although there have been several studies showing small and practically insignificant mode effects between web/phone and phone only interviewing, we believe the subject matter of these studies have not provided a rigorous test of mode effects. In that regard, a recent study by Pew Research (<http://www.pewresearch.org/2015/05/13/from-telephone-to-the-web-the-challenge-of-mode-of-interview-effects-in-public-opinion-polls/>) may be useful for identifying survey items that are more prone to mode effects. For example, assessing changes from permanent employment to unemployment or out of the labour force could be very sensitive to mode due to the complexity of the concepts. We encourage research comparing the differential bias effects of

telephone and web interviewing for more complex survey items than those that have been tested to date. See also Section 4.1 for additional discussion of mixed mode surveys.

2. *Better Understanding of Noncontacts.* As we have stated in prior reviews, information on noncontacts is sorely needed in order to appropriately address the issues. In particular, Statistics Sweden should investigate the following:
 - the proportion of noncontacted numbers that are cell phone vs. landline phone numbers.
 - the proportion of noncontacts that are due to wrong phone numbers – this is needed, among other reasons, to justify the work with the Social Insurance agency telephone numbers.
 - the proportion of noncontacts due to dialing correct numbers that are never answered vs answered but immediately discontinued vs answered but discontinued after the introduction, and so on.
 - characteristics of the sample members for each of the above types of noncontacts.
 - whether the overcoverage indicator on the TPR can be used to predict which noncontacts are nonresidents and how best to use this information.

We believe an ongoing research project is needed to thoroughly analyze the characteristics of noncontacts and then to develop mitigation strategies for dealing with each type of noncontact and/or noncontact population domain. For example, knowing what proportion of noncontacts are due to bad numbers would help determine whether the project described above to evaluate the Social Insurance agency's telephone number database could have any meaningful effect on the noncontact problem. Likewise, if a high proportion of noncontacts are mobile phone numbers, the use of text messaging could be an effective mitigation strategy.

3. *Reducing Nonresponse Bias at the Estimation Stage.* We encourage the continued emphasis in all household surveys to quantify and characterize nonresponse bias for critical survey variables. In this regard, bias-mitigation modeling and weighting research is extremely important since it is the last opportunity to reduce nonresponse bias after all efforts to increase response rates have been tried.
4. *Review of Data Collection Department.* As we noted in Round 6, there is a need for an evaluation of the costs and efficiency of the Data Collection Department. One idea is to conduct a time and motion study of both EVRY and Statistics Sweden's telephone operations and to compare the results. Of course, some provision should be made to include the decentralized interviewing operations in this study as well since we suspect their activities are very different from the centralized phone center.

4.5 STATISTICS SWEDEN REVISED QUALITY CONCEPT

During 2016, Statistics Sweden updated a previously well-established quality concept in the form of a regulation to further elaborate legal standards governing the quality of official statistics in Sweden. The provisions for this concept are documented in “Statistics Sweden’s Regulations on Quality in Official Statistics” (SCB-FS 2016:17). In addition, an updated template for documenting the quality of official statistics (which is referred to as a “Quality Declaration”) was also prescribed in an update of “Statistics Sweden’s Regulations and General Guidelines for the Official Release, etc., of Official Statistics.” (SCB-FS 2002:16). That it is required by law is an important advance for Statistics Sweden for the consistent and transparent reporting of product quality for official statistics.

Like its former version, the revised quality concept is somewhat inconsistent with the data quality framework that was adopted for ASPIRE in 2011. One issue is that there are significant differences in terminologies between the two systems which may cause confusion among staff who are trying to comply with both the 2016 quality concept regulations as well as with the ASPIRE recommendations in their work. This confusion may be ameliorated to some extent because both the revised quality concept and the ASPIRE documentation (see, for example, Biemer, et al, 2016) provide clear definitions of the key terms they use. Nevertheless, some of the terms in the revised quality concept (such as “sampling uncertainty” rather than “sampling error”) are non-standard in the survey literature and may generate confusion in publications for more statistically advanced audiences. However, we fully understand that for less statistically advanced data users, the changes in terminology are welcome. For example, it is possible for some data users to think the term “sampling error” refers to mistakes made in drawing the sample rather than the natural and acceptable result of selecting a subset of the population to survey rather than conducting an unnecessary and costly census.

Another important difference between the revised quality concept and ASPIRE is the absence of a component of Accuracy in Statistics Sweden’s revised quality concept that is a key source of error in the ASPIRE framework – viz., Specification Error. ASPIRE defines Specification Error essentially as the *invalidity* of an estimate or data item. Simply stated, validity reflects how well an observation measures the variable it is supposed to measure. The revised quality concept does not include the concept of Specification Error (or invalidity) as a component of Accuracy. We see this as a possible shortcoming of the revised quality concept. Indeed, the recent literature (see for example, Groves and Lyberg, 2010) explicitly includes validity as a component of total survey error (i.e., accuracy).

The revised quality concept includes some discussion of concepts that may include Specification Error as a special case under the Relevance dimension. For example, Section 5 in SCB-FS 2016:17 states the following:

The agreement between the target characteristics and the characteristics of interest is an important element in the assessment of the relevance of statistics.

However, this definition seems to be referring to a much broader concept of which Specification Error may only play a small role. This is borne out by the next sentence in Section 5 which states:

Before the statistical agency defines the target characteristics to be estimated, the agency makes considerations regarding user information needs, quality requirements associated with the statistics, the cost of producing the statistics, and the response burden.

This sentence seems to focus on the scope and contents of a data collection rather than the accuracy of the concepts that are to be measured. For example, a user community may wish that survey ABC would collect information on an extensive list of characteristics, say $X_1, X_2, \dots, X_P$. However, because of constraints on costs, respondent burden and other practicalities, the survey is only able to collect information of a subset of these characteristics, say $X_1, X_2, \dots, X_L$ for $L < P$. For users who find the characteristics $X_L, X_{L+1}, \dots, X_P$ highly relevant to their work, survey ABC lacks relevance. We agree with this definition of Relevance. However, it is important to note that all surveys lack Relevance to some extent and for some users in the survey's "constituency" or user community. However, it is important that the survey address the needs of its primary constituency.

Some references in Section 5 in SCB-FS 2016:17 could also include the much narrower situation where survey users have a need for a specific characteristic X which is among the characteristics $X_1, X_2, \dots, X_L$ that are presumably being measured by survey ABC. However, because the concepts underlying X either do not translate well to a survey questionnaire or because the survey designers may have misunderstood the concepts, the characteristic X^* is measured instead of X . Thus, X^* is delivered to the users in the guise of X . Either knowingly or otherwise, users then use X^* in their research and analysis rather than X . Thus, results based upon X^* will be inaccurate to the extent that X^* and X differ. In ASPIRE terms, the difference between X^* and X is called Specification Error and it can be shown that it is an important component of Accuracy when Accuracy is defined as how closely a collection measures X .

The argument that X^*-X should be a component of Relevance is debatable, in our view. Relevance has been defined broadly as the degree to which a survey "illuminates the issues that are of importance for users." Under this definition, any lack of accuracy can also be construed as a lack of Relevance because inaccuracies can obscure the issues that are of importance to users. This is true whether the source of the inaccuracy is Specification Error, Nonresponse, Measurement Error or other data errors. Clearly, it is not useful to subsume the Accuracy dimension under the Relevance dimension.

Possible confusion regarding what constitutes Relevance underscores the importance of refining the definition in the revised quality concept. We believe it is important to distinguish between survey scope versus survey validity. While a survey that purports to measure X but measures X^* instead may lack Relevance, more importantly it lacks Accuracy. Mitigation strategies that address Specification Error include refining the wordings of questions, using of multiple indicators of X and adjusting the estimates of X^* to better reflect X . These strategies fit more naturally under the Accuracy dimension, in our opinion. By contrast, strategies for mitigating the lack of Relevance due to reduced survey scope may be to simply begin collecting data on $X_L, X_{L+1}, \dots, X_P$ in the survey.

We believe that relegating Specification Error to the Relevance dimension and somewhat obscuring it in ambiguous language that could also include broader issues of survey scope diminishes its importance to data quality. Indeed, Specification Error is a medium to high risk error source for 8 out of 10 ASPIRE products and should be clarified and emphasized in the quality declaration.

RECOMMENDATION

Consider revising the 2016 Statistics Sweden Quality Concept so that it prominently and unambiguously acknowledges the term X^*-X as an important component of Accuracy, regardless of whether it is referred to as Specification Error.

4.6 HOUSEHOLD CONSUMPTION DATA

Largely because of the length of time since the last Household Budget Survey, the benchmarks used for the household consumption component in the National Accounts are out of date. This was highlighted by Eurostat in their recent GNI review of the National Accounts. The CPI and SILC are also affected by out of date consumption data. The CPI is largely dependent on the household consumption weights used by the National Accounts. SILC use household consumption data for estimating household equivalence scales which are used to support analysis of household income data.

In addition to benchmark data, there is a need to identify the data sources to support the regular update of the household consumption benchmarks. For National Accounts and CPI there are about 160 commodity groups. Therefore, the full detail of the HBS is not required, only aggregated data at that relatively broad commodity group level. It is also most important for those commodity groups where weights are changing significantly and price movements may be quite different to the CPI. Some of these data may be available from alternative data sources. Also, it may be possible to obtain more reliable data from a simplified HBS.

RECOMMENDATION

Establish a project team to look at effective means of developing household consumption benchmarks, looking in particular at alternative data sources. This should build on the existing work looking at alternative data sources. If such a project team is established, National Accounts and CPI should be active participants. One early step should be to assess what is being done, or planned to be done, in other national statistical offices.

4.7 DATA EDITING

All survey data requires some amount of editing to detect anomalies in the data because quite often such anomalies turn out to be errors that need to be corrected. For example, if an enterprise's total sales figure does not equal the sum of sales for each division in the enterprise (and it should), then an error condition is indicated. Such errors need to be repaired because they may significantly affect the accuracy of the data and subsequent data analysis. In addition, data users quickly lose confidence in the data when they discover inconsistencies that indicate obvious errors in the data.

Given the importance of editing to data accuracy, surveys invest considerable resources to editing. But data editing can be quite time consuming, labour intensive and costly. For example, in the SBS, data editing has been estimated to cost as much as 40% of the total survey budget. But is all that editing worth it? One might ask, "from a total survey error perspective, is it optimal to allocate 40% of the survey resources to editing or would some of that budget be better spent on profiling enterprises, reducing nonresponse, improving the questionnaire or on other high priority areas?" This is a question that every survey should consider at the survey design stage. How much editing is needed? It is possible that editing reaches the point of diminishing returns where some editing has a very positive effect on data quality, but beyond that point, data editing is much less effective and may even be detrimental to data quality.

Fortunately, an interdepartmental committee is investigating data editing practices at Statistics Sweden. This committee is taking a "holistic perspective on the data editing process." This includes micro editing, macro editing, and editing rules for electronic questionnaires, including embedded edits in web surveys. The latter can be very effective from both a cost and accuracy perspective although there appears to be little research in Sweden or elsewhere on the effectiveness of these embedded edits. The committee is investigating methods for measuring the cost and accuracy effectiveness of editing, including stopping criteria for the editing process. Another priority for the committee is the use of visual analytics for monitoring quality during the macro editing process.

We applaud the efforts of this committee which we believe are long overdue. There is a need to coordinate editing practices across surveys so that best practices are being consistently used across survey products. For example, we understand that considerable effort is being devoted to studying the editing processes in both the SBS and the RS. Although these two products are quite different, much could be learned by both staffs through information and resource sharing. We encourage the editing committee to pursue this work, considering the recommendations below.

RECOMMENDATIONS

The Editing Committee should consider the following questions as they pursue their work:

1. Are the product areas investing appropriately in data editing from a total survey error perspective? In this regard, consider expressing the goals of data editing in terms of its costs versus its benefits.
2. What are the trade-offs between micro-editing versus macro-editing? For example, should macro-editing be relied upon to a greater extent to reduce the costs of micro-editing while maintaining acceptable levels of outgoing data quality?
3. What are the best practices associated with data editing and to what extent are product areas following them. In particular,
 - a. What data editing techniques are most effective for detecting and repairing data errors?
 - b. What analysis techniques are best for evaluating editing processes? Consider techniques like before/after analysis and simulation studies.

4. Should guidelines and standards be developed for editing? In particular, what guidance can be given to surveys that use edits that are embedded in questionnaires regarding:
 - a. The burden on respondents of interactive embedded edits?
 - b. Wording of warnings and queries to respondents regarding inconsistencies in their responses?
 - c. The reduction of missing data due to “Don’t Know” or “Refused” responses using embedded edits?

We are not recommending that the committee attempt to address these questions within the next year. Rather, our recommendation is that the committee discuss them, prioritise them and then develop a plan for pursuing the key questions for the near term and a secondary set of issues to pursue as a longer-term strategy.

5. SUMMARY AND RECOMMENDATIONS FOR FUTURE ROUNDS

5.1 SUMMARY

As we have previously stated, Statistics Sweden is a world-class organisation and each ASPIRE round provides new evidence of this. We continue to see improvements with each review round with only a few, minor deteriorations. It is both satisfying and gratifying to see many of our prior recommendations being actively pursued by the product areas, most resulting in significant improvements in quality. A recent example of this is with the LCS/SILC which has received some pointed critiques in prior rounds. That this survey is now being redesigned to address these problems is a win, both for ASPIRE and Statistics Sweden.

Exhibits 2a and 2b show the current ratings, prior year ratings, and the improvements by product. Exhibit 2c provides a summary of the ratings since Round 1 in the form of a bar chart. Justifications for the rating changes are summarized to some extent in the product reviews whereas details of each change are provided in rating change tables for each product that are available separately upon request from Heather Bergdahl.

In the current round, all but three of the nine survey and register products showed an improvement in ratings. With a maximum possible score of 100 percent (indicating perfect quality), the product scores ranged from 54.4 percent (for PPI) to 64.5 percent (for FTG) with an average rating of 59.4 percent. (Exhibits 2a and 2b provide the scores for each product by error source.) Changes in scores from Round 6 to Round 7 ranged from 1.4 (BR) to -0.4 (LFS) with an average of 0.5 percentage points. While an average increase of 0.5 percentage points is not a substantial improvement, it does indicate that overall, the quality of data products continues to steadily improve.

After seven rounds of ASPIRE, it is likely that ratings for Knowledge, Communication, Expertise, Compliance with Standards and Best Practices have stabilised and are not likely to change appreciably from round to round. This means products should begin planning and implementing new evaluation studies to increase their knowledge of the error risks as well as to identify risk mitigation strategies that result in real, demonstrative improvements. Ratings could increase dramatically if staff routinely measured the effectiveness of their risk mitigation activities. This criterion, which was added to ASPIRE in 2015, draws the lowest ratings of all six criteria. Notwithstanding the relatively small increase in average scores for this round, there has still been a substantial increase in ratings for all products relative to the 2011 baseline year.

The ASPIRE process continues to be modified and improved in each round. This round, we reviewed and revised the Error Source Delineation Descriptions and shared these with the product areas for comment. These documents clarify what types of errors should be considered under each relevant error source for each product. These documents are an important aid in the review process to maintain inter-product and year-to-year ratings consistency.

In addition, the six criteria that were revised in Round 5 worked quite well and seemed to capture the information Statistics Sweden seeks regarding product quality. The new web interface developed by Statistics Sweden (described in our Round 6 report) facilitates product completions of checklists and seems to provide even better product inputs with less effort from the product areas. The system has allowed us to follow developments, ratings and comments for individual error

sources over time in a more efficient way. We are quite pleased with this system although we continue to look for ways to improve it.

In preparing for their Round 8 ASPIRE reviews, we hope staff will consider the product-specific recommendations we have made and make progress to the extent resources and time allow. In addition, as we proposed last year, we suggest greater consideration be given to demonstrating the effectiveness of the improvement efforts rather than simply relying on reasoning that an intervention that was designed to address some quality issue, achieved the desired effect.

In the discussion of the reviews for each of the products, we have identified the highest priority areas for improvement. In general, the highest priority should be given to error sources with high risk ratings combined with quality criteria with below average ratings. Some desired improvements are crosscutting or general in nature and we have discussed these in Section 4 of this report. These recommendations require consideration by top management rather than the individual product areas. This year, as requested by management, we have identified what we consider the highest priority general recommendations.

Some of the highest priority improvements for the products might require additional funding although products should be encouraged to do as much as possible from existing funds. As previously suggested, it may be worth considering a pool of funding for quality improvements. Bids could be made against this pool and funds allocated to those proposals that are judged to be the highest priority based upon their impacts on quality, costs, and probabilities of succeeding.

Finally, we would like to thank Statistics Sweden for enabling us to work on this important and interesting project. In particular, we would like to thank Heather Bergdahl and Mikaela Järnbert for their tireless and professional support and the excellent co-operation from all the Statistics Sweden staff with whom we had contact.

5.2 HIGHEST PRIORITY RECOMMENDATIONS FOR FUTURE ROUNDS

We have identified 10 recommendations that we consider highest priority for improving the quality of Statistics Sweden's data products. Priorities were assessed based on impact and viability with cost being an important aspect of viability. They are listed below in approximate priority order.

1. *Undertake an analysis of the reasons for non-contacts as a basis for targeting mitigation efforts* (see Section 4.4). Some of the areas needing investigation include the following:
 - a. the proportion of noncontacted numbers that are mobile phone vs. landline phone numbers.
 - b. the proportion of noncontacts that are due to wrong phone numbers – this is needed, among other reasons, to justify the work with the Social Insurance agency telephone numbers.
 - c. the proportion of noncontacts due to dialing correct numbers that are never answered vs answered but immediately discontinued vs answered but discontinued after the introduction, and so on.
 - d. characteristics of the sample members for each of the above types of noncontacts.
 - e. whether the overcoverage indicator on the TPR can be used to predict which noncontacts are nonresidents and how best to use this information.

An ongoing research project should commence to thoroughly analyze the characteristics of noncontacts and then to develop mitigation strategies for dealing with each type of noncontact and/or noncontact population domain.

2. *Undertake a study of the cost and quality differences between EVRY and Statistics Sweden's LFS operations* (see Section 4.3). The initial experimental study and evaluation of the EVRY and Statistics Sweden reported questionnaire data and cost data was very successful. With a new contract and increased sample size to be assigned to the private sector data collector, ongoing analysis of data quality and costs between the public and private sector collected data is necessary to assure that the high standards necessary for government statistics are maintained. Cost comparisons should review compare EVRY costs with Statistics Sweden's centralized and decentralized interviewing staff analyzed separately.
3. *Review the timing, sequencing, management and funding of the set of projects involved in the Integration and Co-ordination of Economic Statistics* (see Section 4.2). There should be a review of overall timetable to assess what is achievable given the resources available. This may involve deferring or deleting some activities. Three objectives are stated but there will be tension between these objectives. It would be helpful if there was agreement to the highest priority objective. There should be discussions with the main users of economic statistics to inform them of SAMSTAT and obtain guidance on what are the highest priorities. The funding level is inadequate. It is doubtful if the proposed activities within SAMSTAT can be achieved within the resources currently available in Statistics Sweden unless there is discontinuation of some significant collections. It is recommended that a case for external funding be developed given the importance of economic statistics to monetary and economic policy analysis. SAMSTAT is a very complex project involving many activities and interdependence between activities. Good project governance is essential. A project advisor who is expert in setting up project planning arrangements would be very useful in this respect including the identification of linkages between activities and the critical path.

4. *Research alternative methods and data sources for household consumption data.* (see Section 4.6) Establish a project team to look at effective means of developing household consumption benchmarks, building on existing work that is examining alternative data sources. If such a project team is established, National Accounts and CPI should be active participants. One early step should be to assess what is being done, or planned to be done, in other national statistical offices.
5. *Research web data collection to better understand how to reduce nonresponse bias through alternative modes to reach hard to contact respondents; further examine potential mode effects.* (see Section 4.1) The use of alternative modes, in particular, smart phones and text messaging, to both increase response rates and investigate whether mode effects are present in survey estimates for subgroups must continue to be a research priority. Additional work on the use of a paper and pencil response mode (PAPI) should be looked at as an alternative mode in a mixed mode survey.
6. *Continue profiling of the largest and most complex enterprises.* (see Section 3.2.7) Eighty of the largest enterprises have been identified for profiling this year. Perhaps that should be rationalised to focus on the largest most complex enterprise groups where profiling will have the most impact on the accuracy of statistics. BR staff should establish good lines of communication with the Coordination and Corporate Affairs unit to establish responsibilities for the profiling. There should be no ambiguity about responsibilities as to the number of enterprises and which enterprises each unit should profile.
7. *Redesign of LCS/SILC.* (See Section 3.3) The LCS-SILC should be combined into one survey, while at the same time satisfying Eurostat's requirements. Consideration should be given to a continuous data collection strategy that provides 12 months of SILC estimates for Eurostat. Significant planning and scheduling needs to be done to ensure redesign schedules are achieved.
8. *Research on Data Editing.* (see Section 4.7) Continue to investigate/coordinate data editing practices at Statistics Sweden taking into consideration the questions posed in Recommendations in Section 4.7.
9. *Specify and develop a Statistical Business Register.* (see Section 3.2.7 and Section 4.2). Even though the Statistical Business Register has a lower priority in the larger project of coordinating economic statistics, harmonization, planning and development work should be initiated that emphasizes where research should be prioritized and quality improvements made. Work could also start on the specification of a Statistical Business Register.
10. *Greater clarification of the treatment of specification error in Statistics Sweden's revised quality concept.* (see Section 4.5). Consider revising the 2016 Statistics Sweden Quality Concept so that it prominently and unambiguously acknowledges what ASPIRE refers to as "Specification Error."

6. REFERENCES

- Biemer, P. P., & Caspar, R. A. (1994). Continuous quality improvement for survey operations: Some general principles and applications. *Journal of Official Statistics*, 10(3), 307–326.
- Biemer, P. and Trewin, D. (2012). “Development of Quality Indicators at Statistic Sweden,” Internal Statistics Sweden report.
- Biemer, P. and Trewin, D. (2013). “A Second Application of the ASPIRE Quality Evaluation System for Statistics Sweden,” Internal Statistics Sweden report.
- Biemer, P., Trewin, D., Bergdahl, H., and Japac, L. (2014). “A System for Managing the Quality of Official Statistics,” *Journal of Official Statistics*, 30(3).
- Biemer, P.P., Trewin, D., Kasprzyk, D. and Hansson, J. (2015). “A Fifth Application of the ASPIRE Quality Evaluation System for Statistics Sweden,” Internal Statistics Sweden report.
- Biemer, P.P., Trewin, D., Kasprzyk, D. and Hansson, J. (2016). “A Sixth Application of the ASPIRE Quality Evaluation System for Statistics Sweden,” Internal Statistics Sweden report.
- Biemer, P. P., Trewin, D., Bergdahl, H., & Xie, Y. (2017). Chapter 17: ASPIRE: An approach for evaluating and reducing the total error in statistical products with application to registers and the national accounts. In Biemer, et al. (Eds.), *Total survey error in practice*. Hoboken, NJ: Wiley.
- Groves, R.M. and Lyberg, L.E. (2010). “Total Survey Error: Past, Present and Future,” *Public Opinion Quarterly*.
- SCB-FS 2002:16. Statistics Sweden’s Regulations and General Guidelines for the Official Release, etc., of Official Statistics. (Not officially translated at the date for this report)
- SCB-FS 2016:17. Statistics Sweden’s Regulations on Quality in Official Statistics.

ANNEX 1 – CHECKLISTS FOR ACCURACY DIMENSION OF QUALITY

Accuracy Dimension Checklist. *For each applicable error source, indicate either compliance or noncompliance with an item in the checklist by marking “Yes” or “No,” respectively. To achieve a higher rating for a criterion, all items for that higher rating must be checked. You may use the “Comments” field to provide comments you deem necessary to explain your response to an item.*

Knowledge of Risks	Check Box	Comments
1. Documentation exists that acknowledges this error source as a potential risk.	☐ Yes ☐ No Fair	
2. The documentation indicates that some work has been carried out to evaluate the effects of the error source on the key estimates from the survey.	☐ Yes ☐ No Good	
3. Reports exist that gauge the impact of the source of error on data quality using proxy measures (e.g., error rates, missing data rates, qualitative measures of error, etc.)	☐ Yes ☐ No Good	
4. At least one component of the total MSE (bias and variance) of key estimates that is most relevant for the error source has been estimated and is documented.	☐ Yes ☐ No Very Good	
5. Existing documentation on the error source is of high quality and explores the implications of errors on data analysis.	☐ Yes ☐ No Excellent	
6. There is an ongoing program of research to evaluate the components of the MSE that are relevant for this error source.	☐ Yes ☐ No Excellent	

Communication	Check Box	Comments
1. Data users have been informed of the risks from this error source to data quality through verbal communications, reports, websites and other formal and informal means.	☐ Yes ☐ No Fair	
2. Likewise, for data providers whose inputs pose some risk to data quality from this error source, there have been communications regarding these potential risks.	☐ Yes ☐ No Fair	
3. These communications to data users and providers have explained the risks in terms of the potential degradation to overall accuracy of the estimates.	☐ Yes ☐ No Good	
4. The potential impacts on users have been conveyed using sampling errors and/or proxy measures of bias and variance components. The measures have also been interpreted in a satisfactory way to facilitate the users' understanding of these risks.	☐ Yes ☐ No Good	
5. User documentation speaks clearly, comprehensively, and with appropriate detail on the size of the MSE components for the target audience.	☐ Yes ☐ No Very Good	
6. Provider communication is sufficiently detailed regarding the effects of errors including the quantification of impacts, and provides adequate information to enable the data providers to develop mitigation strategies that have real impacts on product quality.	☐ Yes ☐ No Very Good	
7. Based upon the communications they have received, users should be able to act appropriately regarding the risks from this error source when analyzing the data.	☐ Yes ☐ No Excellent	
8. There is evidence that data providers have been intimately involved in the process of mitigating the risks of error from this error source resulting in a significant reduction in the risk from this error source. Communication has been ongoing, positive, productive, and produced important changes in the inputs resulting in a significant reduction in the risk from this error source.	☐ Yes ☐ No Excellent	

Available Expertise	Check Box	Comments				
1. The product staff, or those areas servicing the product, include at least one person who is quite knowledgeable about methods for controlling or reducing the effects of the error source.	<table border="1"> <tr> <td>☐</td> <td>Yes</td> </tr> <tr> <td>☐</td> <td>No</td> </tr> </table> Fair	☐	Yes	☐	No	
☐	Yes					
☐	No					
2. Expertise for this error source is adequate in most areas that are relevant for this collection (design, data collection, estimation, analysis, and data dissemination).	<table border="1"> <tr> <td>☐</td> <td>Yes</td> </tr> <tr> <td>☐</td> <td>No</td> </tr> </table> Good	☐	Yes	☐	No	
☐	Yes					
☐	No					
3. At least some members of the product staff are adept at communicating risks for this error source to the both data users and providers clearly and concisely.	<table border="1"> <tr> <td>☐</td> <td>Yes</td> </tr> <tr> <td>☐</td> <td>No</td> </tr> </table> Good	☐	Yes	☐	No	
☐	Yes					
☐	No					
4. The expertise could be made available if required and Communication is good across the internal groups that need to coordinate to reduce the risks from this error source.	<table border="1"> <tr> <td>☐</td> <td>Yes</td> </tr> <tr> <td>☐</td> <td>No</td> </tr> </table> Very Good	☐	Yes	☐	No	
☐	Yes					
☐	No					
5. A good working relationship exists between the product staff and external groups who are key to reducing the error from this error source and their impact on SCB statistics.	<table border="1"> <tr> <td>☐</td> <td>Yes</td> </tr> <tr> <td>☐</td> <td>No</td> </tr> </table> Very Good	☐	Yes	☐	No	
☐	Yes					
☐	No					
6. The key experts frequently participate in conferences, workshops, and other venues where approaches for minimizing the risks of error from this error source are pursued.	<table border="1"> <tr> <td>☐</td> <td>Yes</td> </tr> <tr> <td>☐</td> <td>No</td> </tr> </table> Excellent	☐	Yes	☐	No	
☐	Yes					
☐	No					

Compliance with Standards and Best Practices	Check Box	Comments				
1. Staff are aware of internal and external standards that apply as they pertain to this error source.	<table border="1"> <tr> <td>☐</td> <td>Yes</td> </tr> <tr> <td>☐</td> <td>No</td> </tr> </table> Fair	☐	Yes	☐	No	
☐	Yes					
☐	No					
2. Key staff members are aware of best practices in the field that apply as they pertain to this error source.	<table border="1"> <tr> <td>☐</td> <td>Yes</td> </tr> <tr> <td>☐</td> <td>No</td> </tr> </table> Fair	☐	Yes	☐	No	
☐	Yes					
☐	No					
3. Current activities for controlling or minimizing data quality risks from this error source comply with all appropriate standards.	<table border="1"> <tr> <td>☐</td> <td>Yes</td> </tr> <tr> <td>☐</td> <td>No</td> </tr> </table> Good	☐	Yes	☐	No	
☐	Yes					
☐	No					
4. There are no serious violations of standards and best practices as they relate to this error source.	<table border="1"> <tr> <td>☐</td> <td>Yes</td> </tr> <tr> <td>☐</td> <td>No</td> </tr> </table> Very Good	☐	Yes	☐	No	
☐	Yes					
☐	No					
5. The steps that have been taken to comply with standards and to minimize the risk from this error source may be regarded as state of the art and represent current best practices. Compliance with best practices is routinely monitored.	<table border="1"> <tr> <td>☐</td> <td>Yes</td> </tr> <tr> <td>☐</td> <td>No</td> </tr> </table> Excellent	☐	Yes	☐	No	
☐	Yes					
☐	No					
6. Key staff actively read the literature as it pertains to this error source and some staff members are actively contributing to best practices in this area through conference presentations and publications.	<table border="1"> <tr> <td>☐</td> <td>Yes</td> </tr> <tr> <td>☐</td> <td>No</td> </tr> </table> Excellent	☐	Yes	☐	No	
☐	Yes					
☐	No					

Planning Towards Error Mitigation	Check Box	Comments
1. Documented discussions are being held with appropriate staff with the objective to control or reduce the risks from this error source.	☐ Yes ☐ No Fair	
2. A written plan has been drafted that lays out a clear and effective strategy for mitigating the risks to data quality from this error source.	☐ Yes ☐ No Fair	
3. If applicable, a Service Level Agreement (or its equivalent) with the source data providers is being drafted that specifically targets this error source.	☐ Yes ☐ No Fair	
4. The written plan with measurable objectives has been approved by management. The plan adequately addresses the work required for mitigating the risks of poor data quality for this error source.	☐ Yes ☐ No Good	
5. If applicable, a Service Level Agreement (or its equivalent) with the source data providers has also been approved by management that specifically targets this error source.	☐ Yes ☐ No Good	
6. Appropriate resources have been allocated and Progress toward achieving the goals of the risk mitigation plan is regularly reviewed and compliance with the plan is appropriately monitored.	☐ Yes ☐ No Very Good	
7. Considerable progress has been made and the plan and SLA (if applicable) are updated appropriately as work progresses and new knowledge is gained regarding the error source.	☐ Yes ☐ No Very Good	
8. Mitigation plans have been fully implemented or well underway. Information has been provided to users/providers regarding progress toward risk mitigation.	☐ Yes ☐ No Excellent	
9. Accountability measures are in place to ensure compliance with the plans.	☐ Yes ☐ No Excellent	

Effectiveness of Mitigation Measures	Check Box	Comments				
1. There have been some current efforts to mitigate the risk of error from this source.	<table><tr><td>☐</td><td>Yes</td></tr><tr><td>☐</td><td>No</td></tr></table> Fair	☐	Yes	☐	No	
☐	Yes					
☐	No					
2. Because of these efforts, current proxy measures of the error from this source suggest that the error risks were mitigated to some extent. Further, these efforts have been well-documented.	<table><tr><td>☐</td><td>Yes</td></tr><tr><td>☐</td><td>No</td></tr></table> Good	☐	Yes	☐	No	
☐	Yes					
☐	No					
3. The work undertaken to reduce the error from this source has resulted in significant reductions in the error risks based upon both proxy error measures as well as some direct measures of the MSE components. These improvements efforts have been well-documented.	<table><tr><td>☐</td><td>Yes</td></tr><tr><td>☐</td><td>No</td></tr></table> Very Good	☐	Yes	☐	No	
☐	Yes					
☐	No					
4. Direct estimates of the MSE components associated with this error source indicate that substantial reductions of the error risks were the result of current mitigation efforts. These accuracy improvements have been documented, have been discussed with key users and are publicly available.	<table><tr><td>☐</td><td>Yes</td></tr><tr><td>☐</td><td>No</td></tr></table> Very Good	☐	Yes	☐	No	
☐	Yes					
☐	No					
5. There is strong evidence based upon direct estimates of the MSE components that current mitigation efforts have substantially reduced the risks of error from this error source resulting in important improvements in accuracy. The evaluation has also considered the possibility that other errors sources may have been adversely affected by these mitigation efforts and no such unintended consequences were identified. These results have been thoroughly documented and are publicly available.	<table><tr><td>☐</td><td>Yes</td></tr><tr><td>☐</td><td>No</td></tr></table> Excellent	☐	Yes	☐	No	
☐	Yes					
☐	No					
6. In addition, key users have confirmed that the mitigation measures have succeeded in providing them with statistics that are more accurate and fit for purpose.	<table><tr><td>☐</td><td>Yes</td></tr><tr><td>☐</td><td>No</td></tr></table> Excellent	☐	Yes	☐	No	
☐	Yes					
☐	No					