A System for Product Improvement, Review and Evaluation, ASPIRE

- a manual **Version 1.0**

SCB

2021
Registration number A2021/0104

A System for Product Improvement, Review and Evaluation, ASPIRE - a manual, version 1.0

Producent
SCB, Statistiska centralbyrån
Avdelning för Process och
Metodutveckling, PMU
701 89 Örebro
+46 10 479 40 00

Förfrågningar
Heather Bergdahl
+46 10 479 67 82
heather.bergdahl@scb.se

A System for Product Improvement, Review and Evaluation,

Producer
Statistics Sweden, Department for
Development of Processes and
Methods
SE-701 89 Örebro, Sweden
+46 10 479 40 00

Enquiries
Heather Bergdahl
+46 10 479 67 82
heather.bergdahl@scb.se

URN:NBN:SE:SCB-2021-x99br2101_pdf

Denna publikation finns enbart i elektronisk form på www.scb.se
This publication is only available in electronic form on www.scb.se

Foreword

A System for Product Improvement, Review and Evaluation – ASPIRE, is an approach used to quantitatively evaluate the accuracy of important statistical products at Statistics Sweden. The agency engages a team of external reviewers to ensure objective assessments. An important starting point is that accuracy is set in relation to the purpose of the statistics. External reviewers provide valuable recommendations to staff for improvements from a fitness for purpose perspective. The results are reported annually to the Swedish government.

Statistics Sweden is very grateful for the dedication and expertise of the various reviewers that over the past 10 years have made up the review team, performed the reviews and submitted reports to Statistics Sweden. Paul Biemer and Dennis Trewin were the co-founders of the approach and led the reviews between 2011 and 2017. Jesper Hansson (2015-2016) and Dan Kasprzyk (2015-2017) along with Johanna Laiho-Kauranne (2018-2019) gave valuable contributions during their years with the review team. The present reviewers, Stephen Penneck and Susan Linacre, joined the review team in 2018 and are an important influence for the promotion of the fitness for purpose perspective at Statistics Sweden through ASPIRE.

All ASPIRE reports that have been submitted to Statistics Sweden – from 2011 until 2020 – are available on request.

This manual is written as a guide for new reviewers regarding evaluation activities and a reference for trained reviewers. It also provides a general description for any who are interested in learning more about how ASPIRE works at Statistics Sweden. Our warm thanks to Paul Biemer who suggested the idea for this manual and who produced a first draft in 2018. We are also most grateful to Dennis Trewin, Dan Kasprzyk, Susan Linacre and Stephen Penneck who gave valuable comments at various stages of the work. Our Statistics Sweden colleagues, Eva Elvers and Martin Axelson, co-authored section 2.4 *Considerations regarding the Swedish quality concept*. Eva Elvers also gave much appreciated support to Heather Bergdahl who edited the manual.

Statistics Sweden, February 3, 2021

Joakim Malmdin
Director of Quality Management

Innehåll

List of Abbreviations

ASPIRE	A System for Product Improvement, Review and Evaluation
$COSO_{IC}$	Committee of Sponsoring Organisations of the Treadway Commission
ES CoP	European Statistics Code of Practice
ESS	European Statistical System
GSBPM	Generic Statistical Business Process Model
INSPO	Evaluation method for the statistical production process [in Swedish, Intern statistikproduktionsutvärdering]
NSI	National Statistical Institute
ONA	Other National Authority
PSS	Process Support System [in Swedish, Verksamhetsstöd, VST]
SAM	National statistical agency (or authority) [in Swedish, Statistikansvariga myndigheter]
SOS	System of Official Statistics

List of Selected Swedish terms

Föreskrift	Regulations issued by the government or a government agency
Förordning	Regulations issued by the Swedish government
Lag	Legal Act
Regeringskansliet	The Swedish government offices
Riktlinjer	Guidelines
Sveriges Riksdag	The Swedish Parliament
Årsredovisning	Annual Report

1. Introduction

1.1. Background

In 2011, following some incidents with official statistics, the Swedish government directed Statistics Sweden to pursue quality improvements for a number of key statistical products and to report progress in these areas annually. Specifically, an annual report was requested in the form of metrics that reflect current quality in statistics as well as capture any changes in quality over time. Up until 2008, Statistics Sweden monitored the quality of statistical products by way of an annual self-assessment questionnaire, the results of which were included in the agency's annual report to the government. However, the process did not yield the objective measures of accuracy needed for this reporting, nor did it give a sufficient basis for effective quality management and continuous improvement. The self-assessment process was thus discontinued and Statistics Sweden did not quantify progress on product quality for its annual report again until 2012, at which time Statistics Sweden began to use A System for Product Improvement, Review and Evaluation, ASPIRE, (Biemer et al 2014).

Statistics Sweden commissioned top international expertise to develop an assessment process in partnership. The process was implemented in 2011 when a first evaluation was carried out of eight important statistical products providing a baseline measurement for coming annual reviews (Biemer and Trewin 2012). The results have been reported annually to the Swedish government. At the time of this writing, ASPIRE has run for nine rounds and have included 14 products.

The review process helps Statistics Sweden in a concrete way to

- identify important accuracy components of the statistics based on the purpose of the statistics,
- apply a structured, comprehensive approach for rating efforts to reduce uncertainties (i.e. improve accuracy), and
- distinguish between areas of priority from a fitness for purpose perspective.

The process is aided by objectively and consistently rating products against well-specified criteria. With recurring reviews, the process encourages continuous improvements. It also provides numerical scores per accuracy component and criterion, for comparisons across time.

1.2. Summary of changes since 2011

The ASPIRE approach has evolved in several ways since its start. The number of criteria has gone from five to six and the so-called checklists,

which indicate requirements for the ratings, have been clarified in successive rounds. The most significant development, however, took place in 2018 when adaptions were made to align to the quality framework of the System of Official Statistics (SOS) in Sweden. This affected the list of accuracy components to be evaluated for each product as well as the perspective of the evaluation that the products should be fit for purpose. In conjunction with this, the requirements for the ratings were further clarified in 2018-2019. These latest developments have led to a break in the time series making comparisons of the scores with previous rounds difficult.

The members of the review team have also varied over time. Due to the renewal of the review team to be trained in 2018 and the adaptions made to the ASPIRE approach, 2018 constituted a transition year when no reviews were carried out.

The product mix has been adjusted through the years. Common to all the products is that they are regarded to be among the most important products under Statistics Sweden's direct responsibility which are separate from products the agency produces on commission to other national statistical agencies (SAM). The present product mix comprises seven statistical products and one statistical register (see Table 5 in Appendix 1).

The changes in the approach, review team, and product mix are given in more detail in Appendix 1.

1.3. Key terminology

In some cases, Statistics Sweden's use of terminology is distinctive compared to other agencies. In this section, explanations are provided for some generally used terms. Further explanations of more specific terms are provided later on in their appropriate contexts.

The term *survey* – often the longer term *statistical survey* – is used in a broad sense, meaning not only sample surveys, which utilise direct data collection, but also surveys that utilise administrative data, so-called new data sources, and/or multiple sources (micro-data or already existing statistics). Statistics Sweden's use of the term *statistical survey* corresponds to the term *statistical process* used in the European Statistical System, ESS. (European Union 2020)

It follows that the term *survey design* refers to the design of any survey in a broad sense.

The term *statistical product* is used to describe the statistics that are produced. Normally a statistical product corresponds to one single statistical survey. Occasionally though, one single statistical survey can result in several statistical products, and several statistical surveys may contribute to a single statistical product.

The term *statistical register* is used for a register that is designed for statistical purposes, i.e. to be used in production of statistics. A statistical register may, for example, comprise several sets of administrative data that have been integrated and further developed when it comes to variables, possibly also units. This is described in several places below and especially in Section 2.4.4.

Some further specific concepts and terms are listed below. Most of them are explained in Section 2.4.

- *Statistical characteristic*, described in Section 2.4.1.
- The triplet – *interest*, *target*, and *observation* – described in Section 2.4.1.
- Avoidance of the term *error* (Section 2.1.2); emphasising instead *uncertainty* (Section 2.4.1).
- Explicit use of *model assumptions* as a source of uncertainty (especially Section 2.4.3).

1.4. Disposition

The manual contains six chapters including the introduction. Chapter 2 describes the Swedish quality framework which is fundamental to all evaluations of quality of statistics in Sweden. Here the Swedish quality concept along with its adaption to statistical registers is explained further providing relevant information for the products selected to ASPIRE. Chapter 3 describes the key elements of the review such as the evaluation criteria and the rating process with checklists. Chapter 4 explains the review process from preparations to the writing of the report. Chapter 5 discusses the reporting to Statistics Sweden and to the Swedish government in more detail. Chapter 6 summarises the strengths and weaknesses and some benefits of ASPIRE. Appendix 1 explains more on the changes made to ASPIRE, Appendix 2 gives a detailed description of the evaluation criteria and the checklists are provided in Appendices 3, 4 and 5.

2. The Swedish quality framework

2.1. Quality in the Official Statistics of Sweden

In working with ASPIRE, it is important to see the approach in light of the broader context of

- SOS with the legal quality framework that applies,
- Statistics Sweden's role in SOS,
- the European Statistical System (ESS), which has a non-ignorable impact on the statistics of the member states of the European Union (EU).

It is crucial to understand the quality work done at Statistics Sweden which supplements the legal framework governing quality in SOS. Both of these contexts will be described below. The purpose is to explain the most important aspects that govern quality in statistics in Sweden.

2.1.1. Official Statistics of Sweden and the legal framework governing quality

Statistics Sweden is, according to its directives from the government (Regeringskansliet 2016), responsible for

- developing, producing and disseminating official statistics and other government statistics,
- coordinating the Official Statistics of Sweden,
- promoting cooperation among the statistical agencies.

In addition, the government instructs Statistics Sweden to act as the National Statistical Institute (NSI), as defined by EU Regulation on European Statistics (European Union 2009).

Statistics Sweden, together with the other 27 national statistical agencies (SAM) as well as the so-called other national authorities (ONA) who produce European Statistics have committed themselves to adhere to the European Statistics Code of Practice, ES CoP, which is a self-regulatory instrument based on 16 principles covering the institutional environment, statistical processes and statistical outputs (European Commission, 2017).

Official statistics are regulated by the Official Statistics Act and Ordinance, the Ordinance with directives for Statistics Sweden, as well as Regulations from Statistics Sweden (Sveriges Riksdag 2019 and 2020;

Regeringskansliet 2016; Statistics Sweden 2002; 2016; 2017; 2018b; 2018c). The provisions communicated in this legal framework apply to all SAMs, who are appointed by the government in the Official Statistics Ordinance, including Statistics Sweden.

Following the EU-Regulation for European Statistics, the Official Statistics Act dictates provisions on quality.

The Official Statistics Ordinance lays out additional stipulations on quality for the SAMs that they provide documentation and quality reports for official statistics, and that they evaluate, each in their own area of statistics, the quality of official statistics. The Ordinance also gives Statistics Sweden the mandate to issue executional regulations on quality in these areas. (Sveriges Riksdag 2020).

The government has also charges Statistics Sweden to submit a report with an analysis of the above-mentioned evaluations of the quality of official statistics that the SAMs perform each year. (Statistics Sweden 2019a)

Consequently, Statistics Sweden has issued the following regulations and guidelines for the SOS:

- Regulations on quality in official statistics with a quality concept that maps to the quality provisions in the Official Statistics Act (Statistics Sweden 2016).
- Regulations and General guidelines for the official release of official statistics, with a template for quality reports (Statistics Sweden 2002 and 2020c)
- A supporting Handbook on Quality for Official Statistics of Sweden to support the implementation of the two regulations above (Statistics Sweden 2020a).
- Regulations for the annual evaluation of quality of official statistics (Statistics Sweden 2017) and a Handbook to support its implementation (Statistics Sweden 2018a).
- Guidelines for what constitutes Official Statistics (Statistics Sweden 2018c) to support a uniform application of the guidelines so that users of statistics understand the distinction between official and other government statistics.

2.1.2. The Quality concept, Quality reports, and Evaluation of the quality of official statistics

The quality concept communicated in Statistics Sweden's regulations (Statistics Sweden 2016) has five main quality components:

1. Relevance,
2. Accuracy,
3. Timeliness and punctuality,
4. Accessibility and clarity,
5. Comparability and coherence.

These are linked to the seven criteria the Official Statistics Act and are similar to the five principles for statistical output used in the ES CoP. Somewhat different wording is used however when the Swedish components are defined.

The five main quality components are further subdivided into 16 quality components on the next level. Four of these subcomponents are further subdivided.

The quality concept is to be used when developing, producing, and disseminating statistics. The quality concept is user-oriented, as was the case with previous Swedish quality concept, and most subcomponents are in line with the previous concept. As in previous versions of the quality concept, a guiding principle in the choice of terminology is to avoid the word *error*, as this may give users an impression of mistakes and low quality even when this is not the case.

The quality concept provides the basis for descriptions of official statistics. Thus, it is vital for the *quality reports* of statistics whose target audience is the prioritised users. Based on the information provided by the statistical agency, users assess the level of quality relative to the purpose of the statistics and to their specific uses. In addition, a well-written quality report may also serve to facilitate the communication between the user and the producer on user needs.

The structure of the quality report follows the quality concept with some added administrative information. The same approach and template are used for all statistics, regardless of the design of the survey, for e.g. sample surveys, total population surveys, surveys that utilise administrative data, further developed statistics based on multi-sources, etc.

The quality concept is fundamental for the annual *self-evaluations of the quality of official statistics*, which each statistical agency is obliged to carry out for all official statistics that have been officially released during a given year. The reported results from these evaluations are a

means for the Swedish government to follow up and assess the quality of official statistics at an overall level. They also give important input to statistical agencies in their dialogue with users and stakeholders, which can give valuable input for improving the quality of statistics relative to the purpose of the statistics in coming productions rounds.

2.2. Quality at Statistics Sweden

Statistics Sweden has a strategy for reaching its long-term goals (Statistics Sweden n.d.). One of the main goals is to continuously meet the new and changing needs of users for statistics that are fit for purpose. Production is based on scientific principles and follows international and Swedish regulations and guidelines for quality as well as recognised quality standards. The quality framework and the work invested to uphold and improve quality are of strategic importance for Statistics Sweden. This section describes some aspects of Statistics Sweden's work with quality over and above the requirements regarding quality given in the legal framework for Official Statistics of Sweden.

The quality policy (Statistics Sweden 2020b) conveys the overall level of ambition of the agency's work with quality as well as principles and values to characterise the whole organisation. The policy is a starting point for the efforts to continuously improve statistical quality, process quality, and quality management.

Regarding *statistical quality*, Statistics Sweden applies the framework described in section 2.1.2 with the quality concept as a basis for descriptions of quality, for example in quality reports, and with the evaluation of the quality of official statistics where the purpose of the statistics is a starting point. ASPIRE complements this evaluation with its focus on accuracy for a selection of Statistics Sweden's important products and the feature of external reviewers.

Statistics Sweden studies the level of public confidence with the agency as well as customer satisfaction for its fee-financed operations. Customer satisfaction is monitored by way of:

1. a Customer confidence survey, which is directed towards all customers who have been invoiced for SEK 10 000 or more for an assignment delivered by Statistics Sweden;
2. an intermittent Client Confidence survey, which involves in-depth and qualitative interviews with Statistics Sweden's 20 largest and most important clients.

Process quality, according to the quality policy, means that processes are transparent and documented. They are intended to prevent errors through a systematic approach and are continually evaluated and improved. Also, the agency's data collection process facilitates respondents in their provision of data and ensures its quality.

Statistics Sweden also applies an overarching process-based view on statistical production. The structure of the process which defines and describes the business processes used to produce official statistics, is similar to the Generic Statistical Business Process Model, GSBPM. (UNECE 2013). The GSBPM, as the name indicates, is generic, meaning that it applies to all types of statistical surveys, regardless of the design. This includes, e.g., sample surveys, total population surveys, surveys using administrative data and/or multiple sources. The collected data may be micro data or statistics. An example of the latter case is the National Accounts. A statistical survey generally involves all the processes (Figure 1), from the specification of information needs to the dissemination and communication of the statistics, and evaluation and feed back.

The GSBPM can also be used for the production of statistical registers designed for statistical purposes (Section 2.4.4).

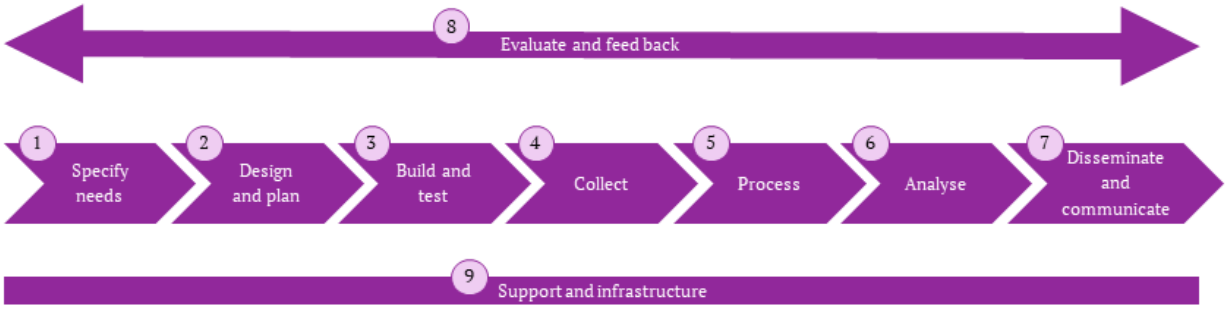

Figure 1. Level one of the GSBPM version in use at Statistics Sweden

The agency applies standardised operational procedures, methods, and tools to support statistical and process quality. These are described in the *Process Support System* (PSS) – a repository where staff can find documentation, instructions, templates, checklists etc., to support the application of standards and best practices.

An evaluation method called *INSPO* is used for the statistical production process, which puts emphasis on production, teamwork and improvements. The purpose of the evaluation is to study the previous production process in order to improve it in the next round, prioritise improvement possibilities, and to suggest measures.

Quality management involves systematic risk management and internal governance and control guided by the $COSO_{IC}$ framework (Committee of Sponsoring Organizations of the Treadway Commission 2013). The agency strives to have a suitable division of tasks with clear responsibilities and continuously to improve the work environment. Security management safeguards the collected data and the integrity of respondents.

Statistics Sweden has a Director of Quality Management who leads a central quality group. Together they support various central quality initiatives at Statistics Sweden.

2.3. The cyclical procedure and fitness for purpose

Statistics Sweden's cyclical procedure for evaluations supports the fitness for purpose perspective and is straightforward and consistent with the general principle of continuous improvements i.e. PDCA: *Plan, Do, Check, and Act* (Figure 2). (Statistics Sweden 2018a)

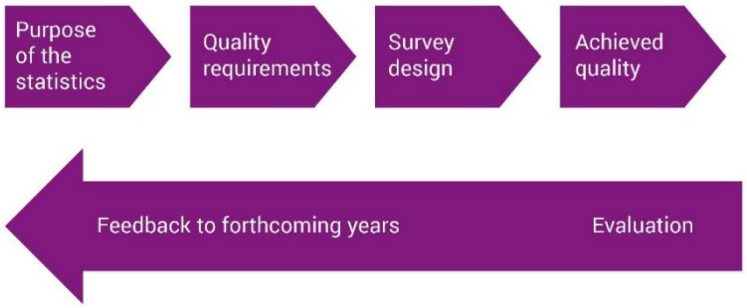

Figure 2. Cyclical procedure of the annual self-evaluation of the quality of official statistics.

As seen in Figure 2, the *purpose of the statistics*, which is stated in the quality report, is an important starting point. As previously, the term *survey* is used in a broad sense. The *purpose of the statistics*, *quality requirements*, *survey design* and the *achieved quality* comprise four links in a chain where the achieved quality builds on the earlier links. The purpose gives rise to quality requirements in terms of the quality components. The survey design in principle should be such that the quality requirements are satisfied. The achieved quality should reflect the objectives of the survey design. The achieved quality is evaluated regularly. Findings and conclusions are used in coming production rounds, for example to adjust the survey design. In this way, quality is set in relation to the purpose of the statistics reflecting its fitness for purpose.

Even though the cyclical procedure is quite straightforward, the transitions between the different links of the chain in Figure 2 involve challenges – for example, in the discussions with users, in determining possible data sources, and in measuring or assessing the achieved quality.

The cyclical procedure applies for all evaluations of quality, both of statistics and processes, at Statistics Sweden. The procedure is clearly evident in the sequence of questions in the form for the self-evaluation of the quality of official statistics (explained in Section 2.1.2). It is also

clearly reflected in the ASPIRE approach to evaluate *Accuracy* particularly in the evaluation criteria (see Section 3.2). A similar cyclical approach applies also for the evaluation of the statistical production process, INSPO (explained in Section 2.2).

In the end, it is up to the key users to assess whether the statistics are fit for purpose or not. Information provided by the producer in the quality report and from various evaluations can give important input for them to make such an assessment. It is crucial that insights from evaluations guide producers in the important and often challenging dialogue with users regarding the quality requirements of the statistics and the extent to which the resulting quality meets these requirements. A dialogue covering these issues can guide producers in fine-tuning the quality of the statistics making them more suited to key uses. In due course, producers of statistics can then provide better conditions for well-informed decision-making, debate and research.

2.4. Considerations regarding the Swedish quality concept

In the context of this manual, there is reason to further elaborate the main quality components of *Relevance* and *Accuracy*, along with some fundamental principles and terminology specific to the Swedish setting (Statistics Sweden 2016 and 2020a).

The last three main quality components – *Timeliness and punctuality*, *Accessibility and clarity*, and *Comparability and coherence* – will not be further discussed. Although relevant for any discussions on trade-offs with the components, *Accuracy* and/or *Relevance*, they do not need explanation here as they are defined similarly to international standards.

2.4.1. Fundamental principles and terminology

1. The generic statistical production process
Basic and common to all statistical products at Statistics Sweden, is the *generic statistical production process*. It is used for all statistical surveys regardless of the design (see Section 2.2).

2. Statistical characteristics
The quality concept builds on the concept of a *statistical characteristic,* which is defined as the numerical value obtained by summarising individual variable values for units in a population or subpopulation using a statistical measure. Both units and variables are associated with a specific reference time. They are often, but not always, associated with the same point or period of time. Reference times are sometimes regarded as separate from variables and populations, and sometimes as an integrated part.

The users' needs regarding the statistical characteristics are called *characteristics of interest*. When a statistical agency sets the *purpose of the*

statistics, it adopts the statistical characteristics to be estimated – the *statistical target characteristics*. This also involves consideration of the quality of the statistics, costs, and response burden. The target characteristics comprise target variables, target units, target populations, etc. The terms, *interest* and *target*, are useful when distinguishing between on the one hand, requests on the part of the users and on the other hand, the choices that the statistical agency makes.

In the Labour Force Survey (LFS) a population of interest is the *resident* population in a given country i.e. persons who are 15 years and older who reside in the country regardless of residency status or citizenship. The target population in the Swedish LFS consists of persons who are *registered* in Sweden and who have turned 15 years but not yet 75 years. The differences between these populations are fairly small, since the majority of the population of interest is contained in the population register, and since persons older than 74 years participate in the labour market only to a limited degree.

It is common that cells and margins in a statistical table correspond to statistical target characteristics. Any breakdowns show divisions into study domains or reference times.

In conjunction with the agency's decisions on the target characteristics, *observation variables* are selected i.e. the variables for which values are collected, either from registers or directly from respondents. The observation variables can be the same as or different from the *target variables*. There are several conceivable reasons for possible differences.

- It may be easier for the respondent to respond to several relatively simple questions instead of one single complex question. The questionnaire is constructed accordingly. The target variable is in this case derived from the observation variables;
- There are instances when it is deemed impossible to directly collect the target variable or where its use would cause significant measurement uncertainty or non-response. Similarly, and necessary to handle, the intended target variable may not exist in the case of administrative data. In both cases, an alternative could be to make use of a model that describes the target variable by means of other variables that *can* be observed. The model must be considered to be sufficiently good for use in statistics and better than any of the alternatives.

It is generally clear from the context whether the term *unit* refers to a certain type of unit – such as *individual* or *enterprise* – or to instances, such as the particular units in a population or study domain.

The statistical agency also chooses observation units. When using register or administrative data, the observation variables, observation units and reference times are already set.

3. Estimation
The objective of a statistical survey is to obtain numerical information for a set of statistical characteristics. The step from data to statistical values has two aspects: statistical inference and the actual calculation. This applies to every survey regardless of whether it is based on directly collected data or on existing data. A statistical value is typically an estimate of a target characteristic and therefore associated with uncertainty. An *estimation procedure* that is explicitly based on random characteristics is often called an estimator. When producing statistics based on registers or administrative data, the statistical inference may be less obvious compared to producing statistics from a sample survey, depending on the extent of the knowledge of uncertainty.

4. Frame procedure and frame coverage
The *target population* is the population referred to in the statistics. It is important that this is clearly defined in the survey design and in the presentation of the statistics. The reference time for the target population is an example; the specification can be particularly important for statistics based on registers (administrative or other data). A statistical survey may have several target populations, for instance, if there are several types of target units. For example, a survey may have both individuals and households as target units.

A *frame* is a tool used for identifying and delimiting units and as a source of information. For many surveys, the type of unit is the same in the frame as in the target population. This is the simplest case, where there is a *one-to-one* relationship. However, differences may exist – one example is that the type of unit in the frame is enterprise and the type of unit in the target population is goods. The unit in the frame then becomes a *channel* to collect data. It is important to distinguish between the frame with its units (called *frame elements*), and the *frame population* which has the same type of unit as the target population.

5. Exclusion from direct data collection (cut-off)
It is quite common in enterprise surveys to exclude the smallest companies from direct data collection in order to reduce the response burden and the costs. The fact that a subset of the target population is excluded from data collection is typically handled by using model-based estimation procedures for the contributions from these companies to the target characteristics. The purpose of the statistics, with its accuracy requirements, affects the choice of the *cut-off* threshold when balancing other quality components, costs, and response burden.

At Statistics Sweden, the expression in Swedish corresponding to *exclusion from direct data collection* is used increasingly and preferably instead of the somewhat improper term, cut-off-sample.

2.4.2. Relevance, an overview

Relevance refers to how well statistics illuminate the issues of importance for users of the statistics. The quality component *Relevance* has two subcomponents, both of which are further subdivided into sub-subcomponents:

- Purpose and information needs
    - Purpose of the statistics
    - User information needs
- Content of the statistics
    - Unit and population
    - Variables
    - Statistical measures
    - Study domains
    - Reference times

The subcomponent *Purpose and information needs* refers to the purpose, the information needs that the statistics are intended to meet, and the knowledge about the needs for statistical information. The subcomponent *Content of the statistics* refers to the target characteristics. Its sub-subcomponents follow naturally, considering the definition of a statistical characteristic.

The quality component *Purpose of the statistics* is different in character from the other components. It is set early – when developing the statistics – and it is a prerequisite. The other quality components are taken into account in the design, and subsequently measured or assessed for the resulting statistics. The design of the statistical survey should be based on the stated purpose (re-considered, if needed) while simultaneously taking into account user needs, the quality of the statistics, the costs, and the response burden. Of course, this does not preclude that other usages are possible. User information needs are normally many and often diverse. There may be conflicts. Hence, some prioritisation is necessary. In the ESS, there are many instances where the statistics are regulated by EU-regulations. These have a direct effect on the priorities of the statistical agency regarding user needs and a direct influence on the statistical agency in setting the purpose of the statistics, in its choice of the target characteristics and consideration of the quality requirements to be satisfied.

Clear descriptions of the content of the statistics for users is an important section of the quality report for official statistics. In this section, the target characteristics are described and also delineated in relation to characteristics of interest. Also, if there are observation units or

observation variables differing to the target units or target variables, these differences are described in this section as well.

The *Purpose of the statistics* reflects the objectives of the statistics and the information needs that are to be satisfied by the statistical survey and the resulting statistics. It follows that the purpose should be the starting point for any evaluation of the quality of statistics.

2.4.3. Accuracy, an overview

Accuracy refers to how well a statistical value estimates its target characteristic. Statistics may be disseminated using preliminary statistical values, once or more, followed by final statistical values disseminated at a later stage.

The quality component, *Accuracy*, has three subcomponents. One subcomponent is further subdivided.

- Overall accuracy
- Sources of uncertainty
    - Sampling
    - Frame coverage
    - Measurement
    - Non-response
    - Data processing
    - Model assumptions
- Preliminary statistics compared with final statistics

The subcomponent, *Overall accuracy,* refers to the expected deviation of the statistical value from the target characteristic.

If the overall accuracy is well known – which is rarely the case – the single sources of uncertainty become of less interest. However, it is often easier, or less difficult, for the producer to assess the sources of uncertainty one by one than to provide an overall measure.

It is desirable to be able to indicate some kind of limitation for the expected size of the deviation between the estimate and its target. One way of providing a measure of overall accuracy is in terms of an uncertainty interval. If possible, the uncertainty interval should refer to total uncertainty. Uncertainty intervals can be presented as objective confidence intervals (for instance based on the survey design and the chosen inference principle) or as subjective assessment intervals (for example based on subject matter knowledge and previous assessments). Depending on the situation, it may be possible and reasonable to use models to construct uncertainty intervals.

In the quality report, a brief description of the estimation procedure and its principles should be provided, such as whether the estimation procedure is based on the design of the statistical survey and if it depends on model assumptions. It is desirable to include uncertainty intervals or other uncertainty measures, if possible, for overall accuracy. (If there are such measures for some single sources of uncertainty, they should be placed under the relevant headings.)

Regarding the subcomponent, *Sources of uncertainty*, some calculations and assessments for single sources may be possible. Sometimes studies can be carried out as more information becomes available, for example using more up-to-date information, from registers or administrative sources, or comparing the short-term data with more detailed annual data. If possible, an assessment of which source(s) of uncertainty has (have) the largest impact on the accuracy of the statistical values should be given in the quality report, before presenting each of the sources of uncertainty.

There is no need to describe the first five sources in this context. The sixth source requires some explanation. *Model assumptions* refers to uncertainty caused by one or more factors: the model being a simplified description of reality, model parameters being estimated (normally the case), and random ingredients in the model (possibly). Estimation for small enterprises with no directly collected data is an example. When uncertainty is induced by the use of models and model assumptions pertaining to one of the first five sources of uncertainty, it is primarily assessed with that particular source. Imputation to compensate for non-response is an example.

The subcomponent *Preliminary statistics compared with final statistics* refers to information on the size and direction of the revisions made to the preliminary statistics before they become final. This component refers to routine revisions (planned and recurring) (Statistics Sweden 2018b). Routine revisions would for example include the planned and recurring annual rebase of GDP. Other types of revisions such as planned revisions due to changes in methods or definitions and unplanned revisions involving corrections are not included here.

Preliminary statistics are revised according to a pre-announced schedule, one or more times, until the statistics are final. For annual statistics, there may typically be two different releases: one with preliminary statistics and the other with final, more detailed statistics. For short-term economic statistics, the tabulation plans and the release times are often the same for the successive versions of the statistics; the chosen number of revisions may for instance be two preliminary versions before the final version.

The size and directions of the revisions of preliminary statistics compared to the final statistics provides some limited information on the accuracy of the preliminary statistics. It is limited since the size refers

to a difference between estimations, while accuracy refers to the statistics' closeness to the estimated target characteristics.

Both preliminary and final statistics can and should be of the quality required such that they are fit for their respective purposes.

2.4.4. Adaption of the quality concept for statistical registers

Sweden, along with the other Nordic countries, has a long tradition of making use of administrative data for statistical purposes. Administrative data contain several types of units, and possibilities exist to make linkages between several of these. Statistics Sweden has created a register system with three so-called base registers, for Population, Business, and Real Estate. Many of the agency's statistical registers spring from these base registers which themselves constitute statistical registers.

In 2019, a quality concept for statistical registers was established at Statistics Sweden that is quite close to the quality concept for official statistics. (The approach has been used for many years, but it is now further developed.) An internal handbook explains concepts and provides a template to describe the production and the quality of a statistical register (Statistics Sweden 2019b).

The quality components take on somewhat different meanings in the quality concept for statistical registers compared to the definitions that apply for statistics; the reason being that a register comprises data on a micro level whereas statistics comprise aggregated data on a macro level.

A statistical register is designed to be used in the production of statistics. It is often based on several data sources, typically administrative data that have been collected and processed by another government agency for purposes other than for statistics. The statistical register may have additional variables and unit types compared to the collected data. A statistical register is the final product of a process that follows the GSBPM with the exception of the aggregation to statistics.

The statistical register may have one or more of these four major areas of use:

1. comprise the basic source for register-based statistics,
2. provide a frame for statistical surveys,
3. provide auxiliary information in estimation procedures, and
4. provide support for data editing on a micro level.

Most statistical registers are produced on a regular basis. There may be different register variants for which different register versions are produced. A register variant may, for instance, comprise a certain subset of the register for which both preliminary and final versions are produced.

The Relevance component for statistical registers with its subcomponents and sub-subcomponents are as follows:

- Purpose and information needs
    - Purpose of the register
    - User information needs
- Content of the register
    - Unit and population
    - Variables
    - Reference times

Often delays are an important factor to consider regarding statistical registers, i.e. the time it takes until real changes are included in the register.

The Accuracy component for statistical registers with its subcomponents and sub-subcomponents is as follows:

- Overall accuracy
- Sources of uncertainty
    - Coverage
    - Measurement
    - Non-response
    - Data processing
    - Model assumptions
- Preliminary register compared with final register

The subcomponent, *Overall accuracy*, is considered in relation to the purpose of the statistical register i.e. the most important uses – whether external or internal. The accuracy of the collected data and the data processing are considered for each set of objects as well as the variables. Consideration is given to the consequences of the various sources of uncertainty. There are no self-evident measures of uncertainty here. If, for instance, the register is used for the production of statistics the user will be interested in information that affects the estimation procedure for the statistics as well as information that can be used to assess the quality of the register. To assess the overall accuracy, the data from the register can be compared to other data sources, if possible.

The subcomponent, *Sources of uncertainty,* is a heading for the five sub-subcomponents – C*overage, Measurement, Non-response, Data processing*, and *Model assumptions*. For instance, information may exist

about coverage deficiencies due to delays and as to how the source data affects measurement errors.

The subcomponent, *Preliminary register compared with final register,* is relevant for registers producing one or more preliminary versions before the final version of the register. Assessments are made of the revisions – i.e. the differences between any preliminary version of the register and the final version – and their consequences in different respects, for instance coverage.

3. Key Elements

This chapter gives a step-by-step description of the key elements of ASPIRE i.e. the accuracy components, the evaluation criteria, the rating process which is aided by the so-called checklists, the computation of average scores and the tabular presentation.

3.1. Accuracy components

Each of the accuracy components for statistics (Section 2.4.3) or for registers (Section 2.4.4) are evaluated from a fitness for purpose perspective (Section 2.3). As earlier established, the accuracy components apply in principle to all statistical products, regardless of the design.

Overall accuracy, which is of primary interest for the key users, is evaluated explicitly from a fitness for purpose perspective.

Sources of uncertainty are each evaluated for the respective products as applicable. In addition, an assessment is made of the importance of single sources of uncertainty to Overall accuracy.

Preliminary statistics compared with final statistics or, in the case of registers, *Preliminary register compared with final register* is evaluated as applicable.

Formulating the purpose of the statistics involves an iterative dialog with the users based on the intended uses and the prioritised information needs. The dialog includes drawing conclusions about the accuracy requirements and any trade-offs to be considered between different quality components.

Accuracy requirements for the prioritised uses refer commonly to *Overall accuracy* and to *Preliminary statistics compared with final statistics,* or, *Preliminary register compared with final register*. One requirement for *Overall accuracy* could be the length of an uncertainty interval. Single sources of uncertainty could be of interest to users if they have strong influence on overall accuracy, also if they are unstable over time with a varying and unforeseeable influence.

A significant part of the review process is the assessment of the relative importance of each single source of uncertainty relative to *Overall accuracy*. Consideration is given to the stability of the source of uncertainty. For example, the uncertainty from a particular source may decrease and/or become more stable over time in step with continuous improvements to accuracy and the implementation of new or improved procedures with successive production rounds. Improvements could lead to a reduced influence on *Overall accuracy* for a single source of uncertainty

compared to previous production and evaluation rounds, all other things being equal.

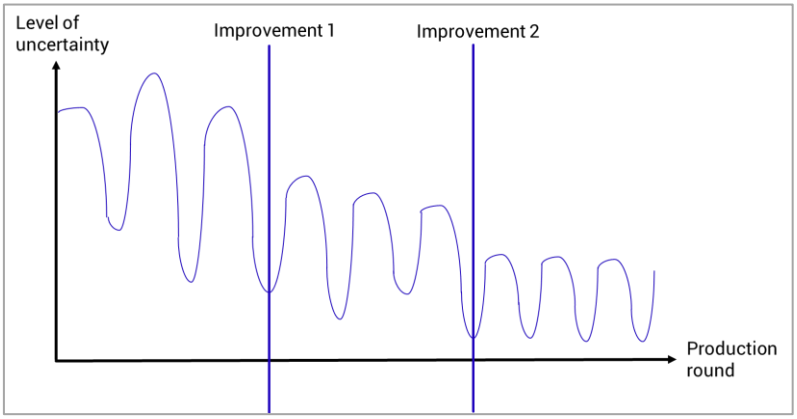

Figure 3. Uncertainty levels with successive improvements to accuracy

The opposite case with increased instability and a greater influence on *Overall accuracy* may also occur in a case where a study, an examination of process data, or an unexpected external factor revealed previously unknown uncertainties.

The assessment of the importance of single sources of uncertainty to *Overall accuracy* is therefore affected by improvements or deteriorations to accuracy as well as by other internal and external factors. An assessment of the importance of the source of uncertainty should be evident in the product's quality report.

3.2. Evaluation criteria

The accuracy components are evaluated based on six evaluation criteria (C1-C6).

C1 Available expertise (members of the production team or other internal expertise)
C2 Compliance with standards and best practices
C3 Knowledge of requirements, achievements, and improvement needs
C4 Plans for improvement activities
C5 Results of improvement activities and findings from other evaluations
C6 Communication with users and data suppliers

The criteria have been developed with several objectives in mind. First, the criteria should represent conditions that Statistics Sweden acknowledges as key and important for quality improvement. The ultimate goal being to achieve a level of accuracy in the statistical products such that they are fit for purpose. The criteria should also be easy to understand, apply and measure. They should also be as few as needed to keep the

process simple and transparent. The six criteria are explained briefly below and further elaborated in Appendix 2.

Criterion one, *Available expertise* (members of the production team or other internal expertise), refers to the existence and sufficient availability to support the production team in their work to

- apply standards and best practices (C2),
- pursue knowledge of accuracy requirements, achievements and improvement needs (C3),
- plan and implement improvement activities (C4),
- observe results of improvement activities and findings from other evaluations (C5), and
- communicate in a suitable way with users and data suppliers (C6).

Criterion two, *Compliance with standards and best practices,* is basic condition for quality in statistics. The legal frameworks for European Statistics, as for Official Statistics in Sweden, state statistics are to be developed, produced and disseminated on the basis of uniform standards and harmonised methods. Standards, whether internal or external, are obligatory. Best practice is a procedure that has been shown by research and experience to be a very good, or best, procedure. This is based on judgment given the current level of knowledge. They are often recognised or recommended in the statistical community.

Criterion three, *Knowledge of requirements, achievements, and improvement needs,* with the production team provides an important basis for work with quality improvement. The production team that oversees a production process should be knowledgeable of

- the accuracy requirements for their product,
- the accuracy attained in the design,
- the accuracy achieved in the production process, and
- improvement needs that should be addressed in future production rounds.

This knowledge proceeds from the knowledge of the fitness for purpose of the statistics and their key uses.

Criterion four, *Plans for improvement activities*, draws attention to the need to plan improvement projects before any effective results can be yielded according to objectives. There may be a wide range of plans that can include changes in the design or production process, as well as the conducting of experiments or launching of studies. The objective for

such plans is generally to acquire better knowledge, to improve accuracy (i.e. reduce uncertainties) or to attain a more effective use of resources (increased efficiency).

Criterion five, *Results of improvement activities and findings from other evaluations* has two aspects. It follows up the results and the effectiveness of any improvement activities referred to in criterion four, if there has been any implementation since the previous ASPIRE round. Also, it always acknowledges findings from the most recent regular evaluations of the statistical production process (according to the cyclical procedure) and the accuracy achieved.

Criterion six, *Communication with users and data suppliers* refers not only to the communication with key users and the suppliers of micro data or statistics, but also to their involvement with the production team in improving accuracy. Dialogue with the key users should focus on their needs relative to the uses of the statistics. The dialogue with data suppliers, on the other hand should address how their outputs affect the accuracy of the final outputs of the statistical production process. Users and data suppliers are grouped in the same criterion even though communication with them can be quite different.

Figure 4 shows how the six criteria relate to the cyclical procedure for evaluating the quality in statistics (Section 2.3).

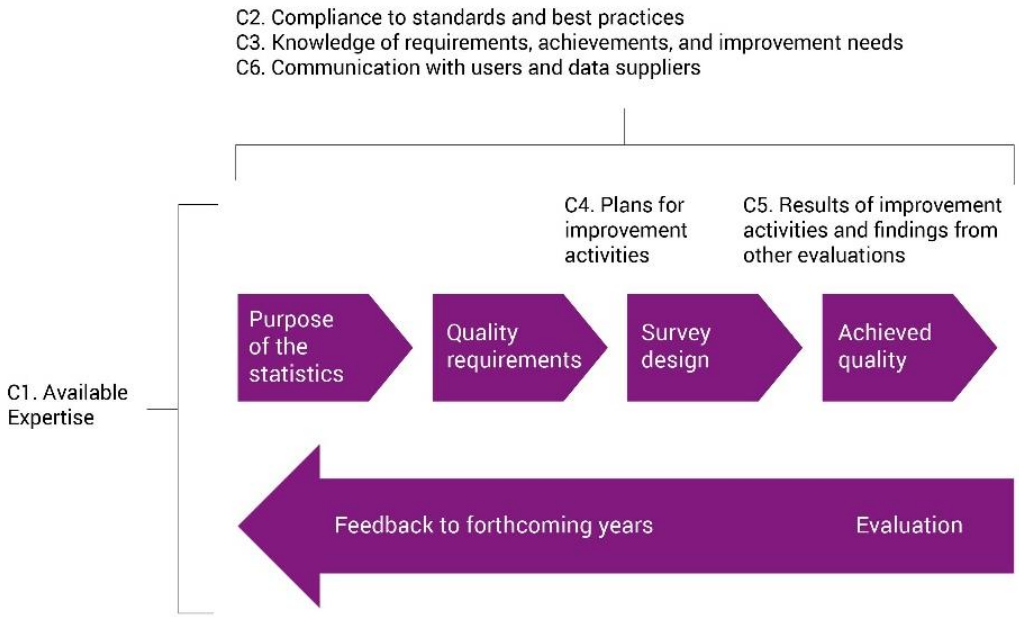

Figure 4. Relationship between evaluation criteria and the cyclical procedure.

Available expertise (C1) is an essential factor to work with improvements to accuracy and with the cyclical procedure in general. *Compliance to standards and best practices* (C2), *Knowledge of requirements, achievements, and improvement needs* (C3), as well as *Communication with users and data suppliers* (C6) apply throughout the whole cyclical procedure. *Plans for improvement activities* (C4) can arise with quality requirements, recommendations or findings from evaluations and can affect the survey design and/or the production process. *Results of improvement activities and findings from other evaluations* (C5) should be evident in the achieved quality and give input to the evaluation.

3.3. Ratings and average scores

Ratings are assigned for each accuracy component per criterion combination aided by evaluation requirements in the so-called checklists. Computations are made to show average scores and the results are shown in a tabular presentation.

3.3.1. The rating process and checklists

A two-step rating process is used. First, an assessment is made of the importance of each source of uncertainty relative to Overall accuracy accordingly to low, medium, and high. This is followed by the assignment of a rating for each accuracy component per criterion combination. This process is aided by three sets of checklists – one for each of the subcomponents of *Accuracy* as shown in the table below along with their applicability to different types of products.

Table 1. Subcomponents of Accuracy and applicability to products

Subcomponent of Accuracy	Brief description of subcomponent		Applicability to products
Overall accuracy	The subcomponent of Accuracy that is of primary interest for many users and evaluated in its own right.		All products
Source of uncertainty	This subcomponent includes the following sub-subcomponents:		All products for each applicable source of uncertainty (potentially 6 sources for statistics and 5 for registers)
	Statistical products:	Statistical registers:	
	-Sampling -Frame coverage -Measurement -Non-response -Data processing -Model assumptions	-Coverage -Measurement -Non-response -Data processing -Model assumptions	
Preliminary statistics (register) compared to final statistics (register)	The subcomponent refers to the accuracy of the preliminary statistics (register), measured here by comparing to the final statistics (register). The checklists refer simply to *revisions*.		Those products that disseminate preliminary statistics (produce one or more preliminary registers)

The product area completes one set of checklists for each applicable sub and sub-sub component of Accuracy rendering a maximum of eight separate checklists.

Common to all three sets of checklists is that they are structured according to the six evaluation criteria (Section 3.2). For each criterion, there is a list of ten levels which correspond to numerical ratings. The ten numerical ratings form five pairs of qualitative ratings: *weak* (1, 2), *fair* (3, 4), *good* (5, 6), *very good* (7, 8) and *excellent* (9, 10). Higher levels will often require more items or more details compared to lower levels. Compliance to a higher level in the list assumes compliance with the previous lower levels or, depending on wording, that the product has made progress compared to lower levels.

The product area indicates for each criterion the highest level of the product's full compliance and provides comments to justify that level. By doing so, the product signifies that it complies with or has made progress compared to lower levels.

Each set of checklists has been adapted to its particular subcomponent of Accuracy, but the thinking is basically the same regarding the level of work that is expected. See the table below for an example regarding the criterion C3. *Knowledge of requirements, achievements, and improvement needs*.

Table 2. Comparison of requests for criterion C3 between two sets of checklists

C3. Knowledge of requirements, achievements, and improvement needs, rating level 5	
Checklist for Overall accuracy	**Checklist for Source of uncertainty**
Key accuracy requirements are considered in the product design; some accuracy-related targets are set. Some comparisons are made between the design and the achievements.	When key accuracy requirements are considered in the product design, this source of uncertainty is included if it is clearly influential. Some comparisons are made between the design and the achievements.

For more details regarding the checklists, see appendices 3, 4 and 5.

3.3.2. The computation of average scores

Table 3 shows a tabulation of ratings for a typical product.

Table 3. Tabulation of ratings

Row	Sub and sub-sub-components of Accuracy	Average Score	C1. Available Expertise	C2. Compliance with standards & best practices	C3. Knowledge of requirements, achievements, and improvement needs	C4. Plans for improvement activities	C5. Results of improvement activities and findings from other evaluations	C6. Communication with users and data suppliers	Importance to Overall accuracy (single sources of uncertainty)
1	Overall accuracy	35	5	3	3	4	3	3	
2	Sources of uncertainty:	35							
3	-Sampling	32	5	3	3	5	1	2	M
4	-Frame Coverage	28	5	3	3	3	1	2	M
5	-Measurement	40	7	6	3	5	1	2	M
6	-Non-response	42	7	5	5	4	1	3	L
7	-Data processing	43	6	4	3	6	5	2	L
8	-Model assumptions	32	4	4	3	6	1	1	H
9	Preliminary statistics compared with final statistics	35	6	3	4	5	1	2	

The average score for each of the accuracy components is in principle the row average of the numerical ratings for each of the accuracy components. These averages are scaled and range from zero to 100. Currently, the evaluation criteria are weighted equally in the computation of average scores.

The assessments of low, medium or high regarding the importance of each source of uncertainty to Overall Accuracy, are seen in the right-hand column. The categorisations of L, M, and H correspond in turn to weights of 1, 2 and 3, respectively. These weights affect the compilation of the average score for Sources of uncertainty (row 2).

The computation of the average score for the each source of uncertainty (rows 3-8) is shown below (Formula 1).

Let r_i denote the rating for the ith criterion, i = 1,.., 6, for single sources of uncertainty i.e. *Sampling, Frame coverage, Measurement, Non-response, Data processing or Model assumptions*. The average score for this single source of uncertainty is then denoted by S_{SSU} ,

$$AS_{SSU} = \frac{10}{6}\sum_{i=1}^{6} r_i \qquad (1)$$

This is essentially the average rating for each source of uncertainty expressed as a percentage. The average scores for Overall Accuracy (row 1) and Preliminary statistics compared to final statistics (row 9) are computed in a similar way.

The average score for *Sources of uncertainty* (row 2) is a weighted average based on the single sources of uncertainty applicable to the product. The computation of the average score for, *Sources of uncertainty,* is shown below (Formula 2).

$$AS_{SoU} = \sum_{AS_{SSU}} \frac{AS_{SSU} \times (weight)}{weight\ sum} \qquad (2)$$

The *weight* (1, 2, or 3) corresponds to the assessment of how important the source of uncertainty is to Overall accuracy (low, medium or high) – and *weight sum* is the sum of these weights over the product's applicable sources of uncertainty.

Table 4 portrays the ratings given in table 3 as so-called Harvey ball symbols which gives a good overview of the results and facilitates interpretation.

Table 4. Tabulation of the ratings in table 3, using Harvey ball symbols

Sub and sub-sub-components of Accuracy	Average Score Previous Round	Average Score Current Round	C1. Available Expertise	C2. Compliance with standards & best practices	C3. Knowledge of requirements, achievements, and improvement needs	C4. Plans for improvement activities	C5. Results of improvement activities and findings from other evaluations	C6. Communication with users and data suppliers	Importance to Overall accuracy (single sources of uncertainty)
Overall accuracy	-	35	○	◔	◔	◔	◔	◔	
Sources of uncertainty:	-	35							
-Sampling		32	○	◔	◔	○	●	●	M
-Frame Coverage	-	28	○	◔	◔	◔	●	●	M
-Measurement	-	40	◒	○	◔	○	●	●	M
-Non-response	-	42	◒	○	○	◔	●	◔	L
-Data processing	-	43	○	◔	◔	○	○	●	L
-Model assumptions	-	32	◔	◔	◔	○	●	●	H
Preliminary statistics compared with final statistics	-	35	○	◔	◔	○	●	●	

Ratings					Importance to Overall accuracy			
●	◔	○	◒	◎	Not applicable (N/A)	Low (L)	Medium (M)	High (H)
Weak	Fair	Good	Very good	Excellent	Weights			
1,2	3,4	5,6	7,8	9,10	0	1	2	3

The average score for *Overall accuracy* and the average weighted score for *Sources of uncertainty* are allowed to differ. There are interdependencies between all the subcomponents of Accuracy – *Overall accuracy*, *Sources of uncertainty* and *Preliminary statistics compared to final statistics* – which should be acknowledged and discussed between the product staff and the reviewers.

4. The review process

The previous chapter gave the theoretical underpinnings and technical justifications for ASPIRE. This chapter describes the implementation regarding the general timeline for the process and subsequently regarding the actual review process. This will differ depending upon whether it is a new product to the review process or not. For new products, time must be devoted to learning about the product, understanding which accuracy components are applicable, the importance of single sources of uncertainty to *Overall accuracy*, and establishing the baseline ratings for each of the six evaluation criteria per accuracy component. For previously reviewed products, the review focuses more on changes since the previous round.

4.1. Timeline

The evaluations at Statistics Sweden are normally scheduled in May so that the results including recommendations can be considered within the context of the agency's general planning and follow-up cycle for the agency's operational plan and for the portfolio of development projects.

Table 5. The timeline for the ASPIRE process

Time in relation to review week	Activity
5 months prior	Preparation starts with reviewers and product staff
16 weeks prior	Internal meeting at Statistics Sweden with unit heads for selected products
12 weeks prior	Checklist completion is underway with the products
8 weeks prior	Completed checklists and other relevant documentation are sent to reviewers
Review week with interviews of products (normally in May)	
4 weeks after	Report delivered to Statistics Sweden (normally by mid-June)

4.2. New products

The process begins several weeks before the review week (Table 3). Each external reviewer receives a collection of materials that provides background and relevant information about the product to be evaluated. At a minimum, this will include the quality report and complementary documentation on the design and production of the statistics, the completed checklists, and any relevant quality studies or other evaluations. The comments in the checklist should generally be consistent with the information in the other materials provided.

The quality report describes the purpose of the statistics, the primary uses and users of the product, the target characteristics and what is

known about the quality in the statistics. There is also a brief description of the design, descriptions of previous or current quality improvement or evaluation studies, links to questionnaires, training manuals, and references to reports or recent studies related to quality. The quality report is mainly directed towards the key users of statistics who are interested in more brief and non-technical descriptions regarding design and processing.

The reviewers read the materials prior to the product evaluations. So that they gain a good understanding of the main purpose of the statistics. During the interview, the reviewers lead discussions on the prioritised uses of the statistics and the user quality (accuracy) requirements to be fulfilled by the survey design and the production process. The reviewers are prepared to ask questions for clarification on the quality, design and production process.

The reviewers divide among them the main responsibility for each product to be reviewed. This will entail leading the discussions before, during and after a meeting with the product, formulating recommendations for improvement activities and the reporting of the product's results.

Before the interview, the reviewers collaborate to agree upon the main quality issues for the products and where there is need for clarification. For example, the self-rating in the checklists may not give sufficient justification for a certain level or there could be apparent inconsistencies in the product's documentation.

4.2.1. The interview

The product's ratings are, for the most part, determined at this meeting and recommendations for improvements are discussed. All the reviewers participate in the interview and the reviewer responsible for the product leads the discussion. The product is represented by the product manager, production manager (from data collection), the responsible methodologist and the unit head. Sometimes an additional expert participates who works with the product or group of products. Largely, it is the key members from the Production management team for the product who are represented at this meeting. In addition, a Statistics Sweden facilitator participates to assist the discussions in different ways, take notes and tabulate the ratings.

The interview starts with a brief introduction by the product starting with the purpose of the statistics, the key users and uses, the production process and the key quality issues and activities associated with these. The next point is to step through the quality report together to clear up any ambiguities that the reviewers have identified and to make sure there is clarity on the importance of individual sources of uncertainty to Overall accuracy. The discussion then continues with a review of the completed checklists starting with Overall accuracy, which is funda-

mental, and proceeding in the order given by tables 3 and 4. The reviewers will focus on the most important areas, for example, regarding those sources of uncertainty of high importance to Overall accuracy where reviewers may question the self-rating and the justifying comment. After the discussion, it is up to the reviewers to decide if the rating should be adjusted up or down. The reviewers provide justification for any changes.

Following the discussion on the ratings, or intertwined with this discussion, possible recommendations for improvement activities are considered. It is important that the product staff understand the recommendations and agree, at least in principle, that they should be prioritised.

Finally, and throughout the whole process, it is important to clarify to staff that *product* ratings should not be interpreted as *personnel performance* ratings. Indeed, many aspects of quality are beyond the control of the product staff and are really higher-level or *system* issues. For example, staff may be aware of a potential cause of uncertainty but lack the resources to deal with it or there may be deficiencies in a data source over which they have minimal control. External reviewers should emphasize that what is being evaluated is the product quality, not personnel performance.

4.2.2. Post-interview activities

The review team jointly discusses the output from each interview meeting as well as any concerns that may have surfaced during the meeting. The reviewers further refine and finalize the ratings, resolving any discrepancies or inconsistencies in the ratings. In particular, the reviewers should examine a product's ratings to identify any apparent inconsistencies across the accuracy components or across the criteria within the product. Any apparent rating inconsistencies or inequities across products should also be identified and resolved.

For this review process, the ratings and their written justifications are shared with the product teams who are asked to correct any inaccuracies or misleading information. Product teams are encouraged to appeal ratings they believe are not well justified. Once the appeal process is completed, the ratings and justification narratives are finalised.

There should also be discussion of possible cross-cutting recommendations that are noted for several products. For example, in the past, recommendations have been made on the treatment of non-response.

The review process is completed when the interviews have been conducted and the ratings have been reviewed and vetted by the reviewers and the product teams. At that point, the review team will begin the process of writing a report in which the ratings, their justifications and recommendations for improvement constitute a major part.

4.3. Previously reviewed products

For previously reviewed products, the preparations for the product team involves updating their checklists with a focus on changes that affect accuracy. Reviewers will also still read the background materials for the product, note changes and understand their implications for the rating process. Changes may include staff changes, meetings with users or data suppliers, studies or experiments that have been newly launched as well as progress on experiments, plans for new studies or investigations, process changes, reorganisations as well as external conditions. Any progress made on addressing the recommendations made in the prior round's report should also be discussed at the interview.

Following the discussions, the ratings that have been affected by changes during the review period are assessed and adjusted as appropriate. In addition, any ratings of *Very Good* and *Excellent* should be reconsidered to determine whether they can be sustained or should default to the next lowest rating if they are not subject to an increase. In order for a product to maintain a rating of *Very Good* or *Excellent*, evidence must be presented that current activities and performance still warrants the high rating. This is particularly important for products whose ratings have been consistently high for several rounds.

The next order of business is to discuss what progress can be made in the coming year to improve some of the lowest ratings, particularly for those components associated with a high (H) or medium (M) influence to *Overall accuracy*. The recommendations for improvement are reviewed and updated as applicable.

4.4. Corrections compared to previous rounds

The process of assigning firstly, the levels of importance to *Overall accuracy* (high, medium or low) for single sources of uncertainty and secondly, the ratings (from 1 to 10) to each accuracy component/criterion combination can be tricky. It is dependent on the product's complexity, the knowledge of the reviewers about the product and its processes, and the accuracy and/or level of detail of the information supplied by the product team. Occasionally, misunderstandings can occur with that only become known in some subsequent review round as more information is obtained. For example, it may be recognised that a single source of uncertainty has high rather than medium influence on *Overall accuracy*. A correction like this will affect the scores on different levels. In the case of a correction like this, practice has been to also correct the prior round's levels so that the change in score will not be exaggerated in the current round.

5. The Report

Nine reports (winter 2021) have been delivered to Statistics Sweden which are publically available upon request.

5.1. Purpose of the final report

Although the primary audience for the overall results is the Swedish government (see Section 5.4), the main recipient of the final report is the product areas and Statistics Sweden's top management. The final report serves at least four purposes.

- It provides an assessment of the present status of accuracy for each selected product.
- For previously reviewed products, the review focuses on and documents improvements made to accuracy since the prior review.
- It flags areas for products whose scores have decreased and offers guidance in the form of recommendations for improvements and, where possible, setting out priorities.
- It provides general recommendations to Statistics Sweden on cross-cutting issues that the external reviewers have noted during the deliberations.

The report is considered as mandatory reading for top management and the production teams for selected products in the review.

It may be difficult for a product to address all the recommendations within the review cycle, presently, of two years. Some of the recommendations may also require extra funding and additional expertise in order to implement them. However, it should be encouraged that products selected for the reviews, are prepared to plan for improvement activities within their existing resources, as much as possible. Some recommendations may take several years to be fully addressed and should be considered as continuous improvement activities. Careful consideration together with the product area of the cost, feasibility, and impact on accuracy of the recommendations in the report will likely increase the chance of these being implemented.

Statistics Sweden documents its standpoint on each recommendation and the progress made during the review period which is made available to the reviewers in advance of the next review round. This documentation and feedback has proven to be quite useful to the review team because it has helped them to improve the relevance of the recommendations and enabled them to be more sensitized to organisational constraints and objectives.

5.2. Report contents

The report begins with an *Executive Summary*, which summarises the key findings and recommendations. The *Introduction* highlights the changes to methodology since the prior round.

The next section, *Product Reviews*, contains the detailed ratings for all products that were evaluated. This section begins with *General Observations*, which is a summary of the major findings and highlights any notable changes in overall scores since the previous round. *Product Ratings and Recommendation*s, consists of one subsection per product according to the following:

- *Context and Discussion* discusses the purpose of the statistics and their main users and uses. It highlights the most important sub and sub-subcomponents of *Accuracy* and provides an overview of the present situation with these.
- *Progress towards prior recommendations* and *Other accomplishments* are listed for previously reviewed products.
- *Key recommendations for the coming two years* provides the highest priority recommendations until the next round.
- *Other areas for consideration* are listed, recommended but with not as high priority.
- A tabulation of the product's ratings is presented indicating improvements and deteriorations in ratings for previously reviewed products.
- Lastly comes *Crosscutting Issues and Recommendations*. In round 9 six recommendations were highlighted (Linacre and Penneck 2020)

Traditionally, the responsibility to draft the report's different sections are divided up among reviewers. A Statistics Sweden facilitator who also formats the document then assembles the sections. After a couple of rounds of comments of the draft report it is finalised and subsequently submitted to Statistics Sweden, approximately four weeks after the interviews.

Table of Contents

Figure 5. Outline of the Round 9 report

5.3. Meetings with management

It is custom to commence each review round with an introductory meeting with the review team and of top management including the Director General, the Deputy Director General, the Director of Quality Management, and other appropriate directors of different departments. This introductory meeting provides an important context for the coming review round, the purpose being to:

- review the progress made on the previous high priority recommendations,
- review current priorities at the agency,
- discuss organisational and staff changes that may bear on the current review, and
- vent possible quality concerns that should be considered with the current review.

It has also been suitable to end the round with a final debriefing with the Director General to share preliminary results. This is an opportunity for the reviewers to speak more informally or candidly with the Director General about major findings, concerns, crosscutting issues, recommendations, particularly as they pertain to accuracy in the statistics.

5.4. Annual reporting to the Swedish government

Since 2011, Statistics Sweden has presented a summary of the results of the ASPIRE evaluations in its annual report to the government with the exception of 2018 (see Section 1.2 and Appendix 1). The main issues are summarised as well as the main recommendations. In the annual report for 2019, a commented summary was provided for each product's average score for Overall accuracy as well as the weighted score for Sources of uncertainty (Statistics Sweden 2020c).

The government offices, as well as the Swedish National Audit Office, annually follow the results and developments from the ASPIRE reviews and appreciate the assessments in quantitative terms of accuracy over time.

6. Concluding remarks

In this manual, the basics of ASPIRE and the review process have been described in a few brief chapters. While there is no substitute for reviewers to hands-on experience, this manual is intended to answer important questions that new external reviewers may have before they start. ASPIRE has evolved into its current state since 2011 and it will continue to evolve and adapt to Statistics Sweden's needs. As it does, it will be revised to document and keep pace with these changes.

By way of conclusions, some of the strengths and weaknesses as well as benefits are considered. Hopefully, the strengths can be maintained or enhanced and the weaknesses handled adequately or even reduced with further development of the approach.

6.1. Strengths and weaknesses

The ASPIRE approach makes use of external reviewers which conceivably adds value to the process and the results, compared to self-assessments and internal reviewers. An advantage is that external reviewers are not limited by internal principles or culture. This aspect leads to reviews that are more objective yielding more credibility to the results. Another advantage is that external reviewers provide different perspectives regarding various questions and in formulating recommendations for improvements. The feature of doing reviews of several important products in a single round also allows the external reviewers to reflect on and give recommendations to the agency on cross-cutting issues.

The evaluation criteria comprise good practice reflecting the general Plan-Do-Check-Act (PDCA) cycle recommended for continuous improvements, also further manifested in Statistics Sweden's cyclical procedure for quality evaluations. This is particularly evident with the latest changes made to the criteria. The evaluations requirements also stimulate the challenging task that a statistical agency has of measuring sources of uncertainties as well as the assessment (whether subjective or objective) of overall accuracy. This gives greater knowledge to the producer of statistics, which can in turn give benefits by way of improvements that increase the fitness for purpose of the statistics.

One potential weakness of the approach is that it can be subjective in that it relies on the knowledge, skill, impartiality of the reviewers, and the ability to familiarize themselves with the European quality framework in general and more specifically, the Swedish quality framework. The subjectivity aspect is one that becomes particularly evident when the review team is renewed. However, it would not be desirable to remove all the subjectivity from the process because that would essentially automate the review process and remove many

beneficial human elements from it. In addition, a purely objective process may not fully incorporate the expertise of the external reviewers nor allow for more complex judgments to be applied to the process. It is important, however, that any subjectivity in the ratings and scores does not lead to inequities and inconsistencies across reviews. Some safeguards are in place to prevent these potentially adverse effects including the checklists with evaluation requirements, the rating correction process, the appeal process and the participation of at least two reviewers.

The dependency on external reviewers can also involve some vulnerability for the statistical agency who is employing them. It has proved to be wise to engage a team of three or more experts to allow for possible exits when the need arises. Experts for the review team should also be prepared to allocate time corresponding to two weeks for the whole review process, something that hasn't always been compatible with work situations for some who are invited to join the review team.

A difficulty that Statistics Sweden has encountered over the years with the process is that it has not always been clear for staff on how to relate to the reviewers' recommendations which are neither entirely mandatory, nor should they be taken lightly. Statistics Sweden's view has been that the agency has expectations on important products that they engage actively in improvements such that statistics be fit for their purpose. It is therefore important that product staff engage with reviewers in the articulation of the recommendations to agree and take ownership of recommended measures.

Over the years, a challenge and possible limitation to the reviews has been that they are conducted in English, due to the employment of English-speaking reviewers. English language skills vary across the organisation and in the case where conversing in English is a challenge, it is possible that all appropriate information and documentation has not been communicated to the reviewers in an adequate way which is unfortunate. It is a crucial aspect of the process that the reviewers can access the appropriate documentation, in English, and that outstanding issues can be sufficiently clarified. However, in many cases the product staff and the Statistics Sweden facilitator have been able to give support to each other in this regard. For other staff the challenge to have to communicate in English has been mostly stimulating and not a problem.

6.2. Selected benefits

ASPIRE has provided a number of important benefits to Statistics Sweden over its nine rounds. Most products have shown slow but steady improvements over the average total scores have increased each year, although the relative increases have been increasingly smaller for most products. One explanation 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 in early

rounds. The achievement of further improvements has necessarily required greater commitments of resources and personnel and more innovative methods.

However, for round 8 (2019) lower average scores were noted for Overall accuracy compared to the total average scores with the previous approach. This was due to the adaption made in 2018 to the Swedish quality framework including the shift in perspective to fitness for purpose as well as the introduction of the new review team. In addition, the latest report explains that the three new product areas all had lower scores compared to the products reviewed in earlier rounds. This was partly due to the tightened evaluation requirements for achieving higher ratings, i.e. the clarifications made of the checklists prior to round 9 (2020). This will probably also affect the other products whose review is coming up in 2021.

Some examples of specific instances where recommendations have been followed and which have led to tangible quality improvements are:

- Use of *before-and-after* analysis to evaluate the impact of the editing,
- Process-mapping to identify problem areas,
- Increased use of modelling to identify uncertainties in registers,
- Use of sensitivity analysis to examine the impact of various sources of uncertainty in GDP,
- A raised awareness of systematic deviation due to non-response uncertainty,
- A handbook on methods to quantify measurement errors.

Another notable benefit with the current approach and in particular with the current review team is that the products selected for review are being pressed to work more concretely with the fitness for purpose perspective. This includes promoting the dialogue with users to increase staff's understanding of user needs and requirements for the statistics (or register) and adapting their statistics (or register) accordingly. Another current benefit is the emphasis that the review team is making on what they call *statistical leadership* which entails staff taking a more expert and leadership role in communicating quality to users. Another more cultural aspect of statistical leadership is to ensure that all expertise working with a product is taking active responsibility for the whole, including the production process and the outputs. (Linacre and Penneck 2020)

7. References

Biemer, Paul, and Dennis Trewin. 2012. *Development of Quality Indicators at Statistics Sweden*. Internal Statistics Sweden report.

Biemer, Paul P, Dennis Trewin. Heather Bergdahl, and Lilli Japec. 2014. "A System for Managing the Quality of Official Statistics". *Journal of Official Statistics,* 30, no. 3: 381-415.

Biemer, Paul P, Dennis Trewin, Heather Bergdahl, and Yingfu Xie. 2017 "ASPIRE: An approach for Evaluating and Reducing the Total Error in Statistical Products with Application to Registers and the National Accounts," in *Total survey error in practice*. edited by. Biemer, Paul P., Edith de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker, Brady T. West. 359-85. Hoboken, NJ: John Wiley & Sons.

Committee of Sponsoring Organizations of the Treadway Commission, COSO. 2013 "COSO Internal Control —Integrated Framework Principles". https://www.coso.org/Pages/ic.aspx Accessed January 29, 2021)

European Commission. 2017. "European Statistics Code of Practice – revised edition 2017". Eurostat. https://ec.europa.eu/eurostat/web/products-catalogues/-/KS-02-18-142 (Accessed January 29, 2021)

European Union. 2009. "Regulation (EC) No 223/2009 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 March 2009 on European statistics**".** EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32009R0223 (Accessed January 12, 2021)

European Union. 2020. "European Statistical System handbook for quality and metadata reports". Eurostat. ttps://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-19-006 (europa.eu) (Accessed January 12, 2021)

Linacre, Susan, and Stephen Penneck. 2020. *A Ninth Application of ASPIRE for Statistics Sweden*. Report to Statistics Sweden. https://www.scb.se/contentassets/8c5cc234fbae4e0ba731b1f9c8ea2042/a-ninth-application-of-aspire-for-statistics-sweden.pdf (Accessed February 3, 2021)

Regeringskansliet. 2016. "Förordning (2016:822) med instruktion för Statistiska centralbyrån", Regeringskansliets rättsdatabaser. http://rkrattsbaser.gov.se/sfst?bet=2016:822 (Accessed January 12, 2021)

Statistics Sweden. 2002. ”Författningsnummer: SCB-FS 2002:16 Statistiska centralbyråns föreskrifter och allmänna råd för offentliggörande m.m. av officiell statistik” partially amended by ”SCB-FS 2016:27” and ”SCB-FS 2020:19”. SCB. https://www.scb.se/om-scb/scbs-verksamhet/regelverk-och-policyer/foreskrifter/sos/scb-fs-200216/ , https://www.scb.se/om-scb/scbs-verksamhet/regelverk-och-policyer/foreskrifter/sos/scb-fs-201627/ and https://www.scb.se/om-scb/scbs-verksamhet/regelverk-och-policyer/foreskrifter/sos/scb-fs-2020-19/ (Accessed January 12, 2021)

Statistics Sweden. 2016. “Constitutional number: SCB-FS 2016:17 Statistics Sweden’s regulations on quality in official statistics”. SCB. https://www.scb.se/en/About-us/main-activity/regulations-and-policies/1.regulations-and-general-guidelines---new/sos/scb-fs-201617/ Accessed January 12, 2021)

Statistics Sweden. 2017. ”Författningsnummer: SCB-FS 2017:8 Statistiska centralbyråns föreskrifter om utvärdering av den officiella statistikens kvalitet” partially amended by ”Författningsnummer: SCB-FS 2018:17 Statistiska centralbyråns föreskrifter om ändring i Statistiska centralbyråns föreskrifter (SCB-FS 2017:8) om utvärdering av den officiella statistikens kvalitet” SCB. https://www.scb.se/om-scb/scbs-verksamhet/regelverk-och-policyer/foreskrifter/sos/scb-fs-20178/ and https://www.scb.se/om-scb/scbs-verksamhet/regelverk-och-policyer/foreskrifter/sos/scb-fs-201817/ Accessed January 12, 2021)

Statistics Sweden. 2018. “A Handbook on Evaluation of Quality of Official Statistics of Sweden”. SCB. https://www.scb.se/publication/35895 (Accessed January 12, 2021)

Statistics Sweden. 2018. “Revision policy”. SCB. https://www.scb.se/contentassets/99b579ae03d245c1a5700c12947746cf/revideringspolicy_2018_eng.pdf (Accessed January 12, 2021)

Statistics Sweden. 2018. ”Riktlinjer för vad som är officiell statistik”. SCB. Accessed September 7, 2020. https://www.scb.se/contentassets/0cd92207266d40eb8829244d51d90b94/riktlinjer-for-vad-som-ar-officiell-statistik.pdf (Accessed January 12, 2021)

Statistics Sweden. 2019. “Official Statistics of Sweden - Annual Report 2018”. SCB. https://www.scb.se/en/finding-statistics/statistics-by-subject-area/other/other/other-publications-non-statistical/pong/publications/official-statistics-of-sweden--annual-report-2019/ (Accessed January 12, 2021)

Statistics Sweden. 2019. “Production and Quality of the Statistical Register”. SCB. Internal document.

Statistics Sweden. 2020. “A Handbook on Quality for Official Statistics of Sweden”. SCB. https://www.scb.se/hitta-statistik/statistik-efter-amne/ovrigt/ovrigt/ovriga-publikationer-ej-statistik/pong/publikationer/kvalitet-for-den-officiella-statistiken---en-handbok/ (Accessed January 12, 2021)

Statistics Sweden. 2020. “Quality Policy”. SCB. https://www.scb.se/contentassets/99b579ae03d245c1a5700c12947746cf/quality-policy.pdf. (Accessed January 12, 2021)

Statistics Sweden. 2020. ”Årsredovisning 2019”. SCB. https://www.scb.se/contentassets/e6e3bb34eaa84f7b9e2244283c4cc8b0/arsredovisning-2019.pdf (Accessed January 12, 2021)

Statistics Sweden. n.d. “Statistics Sweden’s Operations”. SCB. https://www.scb.se/en/About-us/main-activity/. (Accessed January 12, 2021).

Sveriges Riksdag. 2019. ”Lag (2001:99) om den officiella statistiken”. Sveriges Riksdag. http://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/lag-200199-om-den-officiella-statistiken_sfs-2001-99 (Accessed January 12, 2021)

Sveriges Riksdag. 2020. ”Förordning (2001:100) om den officiella statistiken”. Sveriges Riksdag. http://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/forordning-2001100-om-den-officiella_sfs-2001-100 (Accessed January 12, 2021)

UNECE. 2013. “Generic Statistical Business Process Model version 5”. High-Level Group for the Modernisation of Official Statistics. Last modified April 1, 2019. https://statswiki.unece.org/display/GSBPM/GSBPM+v5.0 (Accessed January 12, 2021)

Appendix 1. Year-by-year changes 2011-2020

Changes to the approach and review team

Table 6. Changes to the ASPIRE approach and review team, 2011-2020

Year: round	Changes to ASPIRE	Reviewers
2011:1	Baseline measurements are performed based on 5 evaluation criteria with guidelines to aid the rating process of 8 potential error sources for statistical products and 6 potential error sources for registers.	-Paul Biemer, -Dennis Trewin
2012:2	ULF/SILC is added to the set of products. GDP is split in two one for quarterly and one for annual accounts. Year-to-year change figures are available. Checklists are developed to support the rating process.	-Paul Biemer, -Dennis Trewin
2013:3		-Paul Biemer, -Dennis Trewin
2014:4	A separate set of error sources is developed for the GDP products. (Biemer et al 2017)	-Paul Biemer, -Dennis Trewin
2015:5	GDP-A is dropped and PPI comes as a new product. Two new reviewers are introduced to ASPIRE. A sixth criterion is introduced to give more emphasis to the effective-ness of achievements.	-Paul Biemer, -Dennis Trewin, -Dan Kasprzyk -Jesper Hansson
2016:6	Interviews are carried out with 2 pairs of reviewers who meet before and after to collaborate on the ratings.	-Paul Biemer, -Dennis Trewin, -Dan Kasprzyk -Jesper Hansson
2017:7	The LCS/SILC is dropped due to a redesign. Three reviewers remain in the team.	-Paul Biemer, -Dennis Trewin, -Dan Kasprzyk
2018	Transition year after an internal to the agency evaluation with training of new review team by Biemer and Trewin. Adaptions are made to the Swedish quality framework for Accuracy including sources of uncertain-ty and the fitness for purpose perspective (see below for more details). Work is commenced on an ASPIRE training manual.	-Paul Biemer, -Dennis Trewin, -Stephen Penneck, -Susan Linacre, -Johanna Laiho-Kauranne
2019:8	FTG, RS, SBS, and TPR are dropped to provide room for 3 new products. The adaptions are implemented in the checklists making comparisons of ratings and scores diffi-cult with earlier years. A new ASPIRE team is in place guided by Trewin. Reviews per product are conducted every 2 years.	-Dennis Trewin, -Stephen Penneck, -Susan Linacre, -Johanna Laiho-Kauranne
2020:9	More adaptions are needed to the checklists to better reflect the fit-ness for purpose perspective (guidelines are now dropped). Adaptions are implemented with 3 new products. Two reviewers remain in the team Reviews are done largely by written questions and answers between reviewers and products due to Covid 19. The ASPIRE training manual (version 1) is completed.	-Stephen Penneck, -Susan Linacre

Changes in the product mix

Table 7. Reviewed products per round and year, 2011-2020

	Rounds Reviewed								
	1	2	3	4	5	6	7	8	9
Products	2011	2012	2013	2014	2015	2016	2017	2019	2020
Foreign Trade of Goods Survey (FTG)	x	x	x	x	x	x	x		
Labour Force Survey (LFS)	x	x	x	x	x	x	x	x	
Annual Municipal Accounts (RS)	x	x	x	x	x	x	x		
Living Conditions Survey (LCS/SILC)		x	x	x	x	x	x		
Structural Business Statistics (SBS)	x	x	x	x	x	x	x		
Consumer Price Index (CPI)	x	x	x	x	x	x	x	x	
Producer and Import Price Index (PPI)					x	x	x	x	
Business Register (SBR)	x	x	x	x	x	x	x	x	
Total Population Register (TPR)	x	x	x	x	x	x	x		
Gross Domestic Product, quarterly (GDP-Q)	x	x	x	x	x	x	x	x	
Gross Domestic Product, annual (GDP-A)		x	x	x					
Production Value Index (PVI)									x
Quarterly Air Emission Accounts (QEA)									x
Building permit statistics (BPS)									x

More on the significant changes in 2018-2019

Following round 7 in 2017, an evaluation of the ASPIRE approach was conducted internally at Statistics Sweden. Based on this evaluation, Statistics Sweden took steps in 2018 to make five changes:

1. Adaptions to the legally-binding quality framework of the SOS.
2. Five products continue with the reviews and five products are phased out to make room for new products.
3. Products are reviewed every two years to enable more time for improvement activities.
4. The renewal of the review team to reduce personal dependency.
5. The development of the present training manual to facilitate the training of new reviewers.

Adaptation 1 above affected the approach such that comparisons to previous years are not meaningful given the following discontinuities:

1. The fitness for purpose perspective replace the previous implicit objective of excellence.
2. The accuracy components (see Section 2.4.3) replace the previous eight error sources.
3. Sources of uncertainty are assessed for importance to Overall accuracy, not risk for error.

Following an evaluation of round 8 with the new external review team the need for further clarification of the checklist was necessary to make the rating process more comprehensible, transparent and reliable both for the experts and for the product teams.

Clarifications were therefore made in collaboration between Statistics Sweden and the review team. Briefly, this involved:

1. renaming three of the six criteria to better reflect the cyclical process of improving quality in relation to quality requirements such that the statistics be fit for purpose,
2. breaking down the requirements in the previous checklists into the same number of levels as there are ratings i.e. 10 levels, and
3. specifying in more detail what is required for each of the levels.

The above changes seemed effective in improving the rating process as needed. See appendices 3, 4 and 5 for the current versions of the checklists.

Appendix 2. Description of the evaluation criteria

C1. Available expertise (members of the production team or other internal expertise)

Available expertise refers to the existence of appropriate skills with the production team or other internal expertise and that these skills are available in sufficient quantity. Available expertise is an important aspect so that the production team can

- apply standards and best practices (C2),
- pursue knowledge of accuracy requirements, achievements and improvement needs (C3),
- plan and implement improvement activities (C4),
- observe results of improvement activities and findings from other evaluations (C5), and
- communicate in a suitable way with users and data suppliers (C6).

Besides subject matter, methodological, and data collection competencies, other skills are needed to improve accuracy from a fitness for purpose perspective or attain efficiency gains, such as a good understanding of the thinking behind Statistics Sweden's cyclical procedure (see Figure 1), communication skills, and the ability to formulate improvement strategies. It is important for the production team to bear in mind that additional internal expertise may need to be utilised from time to time. Note that this criterion stresses the actual *availability* of the expertise to work with the product, in the sense that they can engage in the work with a particular issue, when and to the extent needed, i.e. that they are resourced.

C2. Compliance with standards and best practices

Statistics shall be developed, produced and disseminated on the basis of uniform standards and harmonised methods according to EU's statistics regulation and the Swedish Official Statistics Act. This is a basic condition for accuracy in statistics, for coherence between statistics, and for an effective use of resources.

Standards, whether internal or external, are regarded as obligatory. Internal standards are generally provided in Statistics Sweden's Process Support System. National standards are communicated in legislation and in guidelines for official statistics. International standards are communicated in EU-regulations or by other international organisations.

Best practice is a procedure that has been shown by research and experience to be a very good, or best, procedure. This is a judgment given

the current level of knowledge. Best practice is often recognised in some way or recommended in the statistical community. International organisations frequently provide support. In the context of ASPIRE, best practices can often be found by investigating how other countries perform tasks or apply methods.

The production team should therefore be aware of and comply with applicable standards regarding different accuracy components. The production team should also be aware of best practices as they relate to their product and comply with these. For higher levels, the production team should also be aware of evolving standards and best practices. Also, suitable expertise is contributing to these, when appropriate. In many cases, best practice will involve a higher ambition than current standards. In some cases, best practice will give guidance where standards do not exist.

C3. Knowledge of requirements, achievements, and improvement needs

The production team that oversees a production process should be knowledgeable of the fitness for purpose and the key uses of their product's statistics. From this perspective, they should also be knowledgeable of

- the accuracy requirements for their product,
- the accuracy attained in the design,
- the accuracy achieved in the production process, and
- improvement needs that should be addressed in future production rounds.

These elements reflect the steps in Statistics Sweden's cyclical procedure. The knowledge that the production team acquires should reasonably accumulate over time as the team members draw from studies, earlier experience and findings from previous evaluations.

C4. Plans for improvement activities

Plans for improvement activities are made from the perspective that statistics should be fit for their purpose. There may be a wide range of plans that can include changes in the design or production process, as well as the conducting of experiments or launching of studies. The objective for such plans is generally to acquire better knowledge, to improve accuracy (i.e. reduce uncertainties) or to attain a more effective use of resources (increased efficiency).

Criterion 4 draws attention to the fact that improvement projects require planning before implementation and before they can yield effective results according to their objectives. The plans act upon recommendations for improving accuracy and from the findings from previous

evaluations, the latter of which may stem from process data, assessments, improved knowledge including external factors and user input, or the measured effects of recently implemented activities. The activities should be in line with the current knowledge (see C3). An important element in the planning process is to devise follow-up measures to evaluate the effects of improvement activities (see C5). For example, it might be necessary to set up *before-and-after* measures regarding implementation, or to measure a controlled experiment. Generally, plans will change over time as priorities change. Current priorities may not include every source of uncertainty – the emphasis is rather on the most influential sources, also where clear improvements are expected.

C5. Results of improvement activities and findings from other evaluations

Criterion 5 has two aspects. It follows up the results and the effectiveness of any improvement activities referred to in C4, if there has been any implementation since the previous ASPIRE round. Also, it always acknowledges findings from the most recent regular evaluations of the statistical production process (according to the cyclical procedure) and the accuracy achieved. For example, there may be both results from an experiment on contact strategies (an improvement activity) and findings from process data concerning editing (a finding from a regular evaluation).

In general, it is advisable that appropriate measures, which can capture direct and indirect information about accuracy and efficiency, be incorporated in the production process and followed up in future evaluations. In following the cyclical procedure, the production team will first compare the most recent evaluation findings to the design and possibly with previous production rounds, and then draw suitable conclusions to feed back into future rounds.

C6. Communication with users and data suppliers

Users should be made aware of the uncertainties in the statistics they are using to help them take due consideration in their uses and to suitably interpret the results. Users are normally interested in the overall accuracy of the statistics but it can also be beneficial to involve the prioritised or key users in discussions regarding single sources of uncertainty that influence overall accuracy the most or if a single source of uncertainty is instable. Users should also be involved in discussing any revision practices of statistics and other forward-looking needs. To attain higher levels, it should be evident that the production team and users hold discussions based on a mutual understanding of the key uses of the statistics and the important accuracy issues associated with these.

Suppliers of micro data or statistics, which we refer to as data suppliers, are defined here as the internal or external body who supplies input to the department at Statistics Sweden responsible for the statistical product or register. In the case of direct data collection, the data supplier is

usually the data collection department. For registers, the data supplier is commonly another government agency such as the Swedish Tax Agency. Communication with data suppliers will not necessarily be applicable to all accuracy components and will vary depending on the product. Data suppliers should be aware of and involved in a dialogue on how the quality, and particularly accuracy, of their outputs affects the final outputs of the statistical production process. The production team including data suppliers should strive together to improve the inputs.

Communication with users and data suppliers can be quite different. They remain grouped in the same criterion but the requirements for users and data suppliers are stated separately, on each level.

Appendix 3. Checklists for Overall accuracy

For each criterion (C1-C6), highlight in bold text the statement in the left-hand column that corresponds to the highest level to which the product complies. Then provide comments in the right-hand column to justify that level.

C1. Available Expertise (members of the production team or other internal expertise)		**Comments**
1.	There is no capable expertise available regarding methods and techniques to even crudely assess overall accuracy. *(Weak)*	
2.	The available expertise is insufficient when it comes to understanding the accuracy requirements corresponding to key uses and the development of a basic assessment of overall accuracy. *(Weak)*	
3.	There is basic expertise available to understand the accuracy requirements corresponding to key uses and to study overall accuracy, regarding the most important contributing factors. *(Fair)*	
4.	There is basic expertise available to discuss accuracy requirements and assessments of achieved accuracy with the production team and key users. *(Fair)*	
5.	There is a good level of expertise available to study overall accuracy and to communicate with the production team and discuss with users. There is good understanding of the cyclical procedure. *(Good)*	
6.	There is a good level of expertise available to study overall accuracy and to communicate with the production team and discuss with key users; there is also some capability of contributing improvement ideas for overall accuracy. *(Good)*	
7.	There is a very good level of expertise available to study overall accuracy and discuss the results with key users, taking their uses of the statistics into account. There is a good working relationship within the production team regarding the cyclical procedure. *(Very good)*	
8.	There is a very good level of expertise available to derive measures of overall accuracy, including some objective measures. There is some ability to extend standard theory to develop new techniques and measures. There is expertise to develop improvement strategies including measures of effects. *(Very good)*	
9.	There is expertise available to have an in depth understanding of accuracy requirements, to develop objective measures of overall accuracy and to formulate prioritised improvement strategies. The expertise has international contacts and cooperation. *(Excellent)*	
10.	The expertise extends themselves such that they have the ability to innovate and develop new approaches, best practices, and new standards. There is a clear strategy in place to ensure this level of expertise into the future. *(Excellent)*	

C2. Compliance with standards and best practices		Comments
1.	The production team is mainly unaware of standards that relate to the product and overall accuracy. *(Weak)*	
2.	The production team has some, but limited, knowledge of standards that relate to the product and overall accuracy. *(Weak)*	
3.	The production team is generally aware of standards (internal and external regulations, policies etc.) that relate to the product and overall accuracy. There is some evidence of compliance with these. *(Fair)*	
4.	The production team is generally aware of standards that apply to the product and there is evidence of compliance with these. There are plans in place to address significant areas that do not satisfy standards. *(Fair)*	
5.	The production team has good knowledge of relevant standards in relation to overall accuracy and the product. There is knowledge of relevant internal best practices. There is evidence of application of standards and best practices, at least where significant. *(Good)*	
6.	The production team has good knowledge of standards in relation to overall accuracy and the product. There is knowledge of relevant best practices, for example from similar products in other countries. There is evidence of application of both standards and best practices. *(Good)*	
7.	The production team has very good knowledge of standards and best practices related to overall accuracy and the product. The knowledge of best practices is regularly updated through contacts with other expertise. Relevant standards and best practices are applied where significant. *(Very good)*	
8.	The production team has very good knowledge of standards and best practices related to overall accuracy and the product. The knowledge of best practices is kept up-to-date through close contacts with other expertise. Relevant standards and best practices are applied wherever appropriate. *(Very good)*	
9.	The production team conducts cooperative activities with other expertise to ensure that compliance with standards and best practices related to overall accuracy is maintained, also to participate in the development of best practices related to overall accuracy and the product. *(Excellent)*	
10.	The production team is sought out by other national and international experts to participate in the development of standards and best practices related to overall accuracy and the product. *(Excellent)*	

C3. Knowledge of requirements, achievements, and improvement needs		Comments
1.	The knowledge of key users and uses is vague, as is the knowledge about accuracy requirements and accuracy achieved. *(Weak)*	
2.	The knowledge of key users and uses is limited. The knowledge about accuracy requirements and accuracy achieved is vague. *(Weak)*	
3.	Some work has been done to quantify accuracy requirements with regard to key uses. There are crude assessments of the achieved overall accuracy. *(Fair)*	
4.	Accuracy requirements have been quantified with regard to key uses. There are assessments of the achieved overall accuracy. *(Fair)*	
5.	Key accuracy requirements are considered in the product design; some accuracy-related targets are set. Some comparisons are made between the design and the achievements. *(Good)*	
6.	There are some quantitative measures of achieved accuracy, available at least on an *ad hoc*-basis. The sources of uncertainty which influence accuracy the most are identified. A working list exists of possible activities in order to improve knowledge or accuracy. *(Good)*	
7.	There are assessments of the overall accuracy, based on both subjective and objective intervals which are available at least on an *ad hoc*-basis. The measures take into account the influential sources of uncertainty. The product design has accuracy requirements as a starting point. Comparisons with achieved accuracy are utilised to formulate suggestions that will improve knowledge or accuracy. *(Very good)*	
8.	There are objective measures of overall accuracy that are updated regularly. These measures are utilised in the ongoing work with the product design, which has accuracy requirements as a starting point. The working list with potential activities and rough priorities is regularly updated. *(Very good)*	
9.	The product design is based on the accuracy requirements and on previous experience of achieved accuracy. The need for more knowledge is considered and prioritised with regard to achieved accuracy, the effective use of resources, and recent information that has become available. Possible future factors are considered. *(Excellent)*	
10.	The product design is based on the accuracy requirements and on previous experience of achieved accuracy. The need for more knowledge is prioritised with regard to achieved accuracy, the effective use of resources, and recent information that has become available. Regular contact is maintained internally with similar products and with other countries in order to improve knowledge and suggestions for improvements. Possible future factors are considered. *(Excellent)*	

In many cases, improvement activities relate to one or more sources of uncertainty rather than to overall accuracy. Double counting should be avoided. Here improvement activities that explicitly refer to overall accuracy are considered.

C4. Plans for improvement activities		**Comments**
1.	There are no recent or current efforts to launch improvement activities with respect to overall accuracy, although there is a clear need. *(Weak)*	
2.	There are no recent or current efforts to launch improvement activities, although areas of shortfall with respect to overall accuracy have been identified. *(Weak)*	
3.	Some activities to address particular shortfalls with respect to overall accuracy are planned, and they are about to start. *(Fair)*	
4.	Some activities to address particular shortfalls with respect to overall accuracy are planned, and a few have started. *(Fair)*	
5.	The important areas of shortfall with respect to overall accuracy have been identified, and a management approved plan to address the key areas exists. *(Good)*	
6.	Resources have been allocated for at least one prioritised area, and an overall plan exists for the other addressed key areas. *(Good)*	
7.	There is a plan in line with all identified priorities. The plan has management approval and allocated resources at least in the short run. It includes some measures to evaluate the effectiveness of the improvement activities. *(Very good)*	
8.	Well-reasoned and prioritised plans are developed on the basis of the regular evaluations. These plans are resourced, and they include the evaluation of the effectiveness of the improvement activities. *(Very good)*	
9.	The prioritised improvement activities not only are resourced and include evaluation of effectiveness, but they are also coordinated within the product and, where possible, with similar products in order to enhance improvements. Unintended consequences are avoided. *(Excellent)*	
10.	The prioritised improvement activities not only are resourced and include evaluation of effectiveness, but they are also coordinated within the product and with similar products in order to enhance improvements. Unintended consequences are avoided. *(Excellent)*	

Here improvement activities and findings from other evaluations that explicitly refer to overall accuracy are considered.

C5. Results of improvement activities and findings from other evaluations		Comments
1.	There is no study of effectiveness from improvement activities. Evaluations hardly consider overall accuracy. *(Weak)*	
2.	The study of effectiveness from improvement activities is not very conclusive. There is no regular work to evaluate the production process with respect to overall accuracy. *(Weak)*	
3.	Improvement activities seem to have improved achieved accuracy, but the results are just indicators. There are assessments of overall accuracy, possibly crude, that are regularly followed up. Unexpected results are noted. *(Fair)*	
4.	There are indications that improvement activities have improved overall accuracy when it comes to key uses, although the indicators may be qualitative. The possibly crude assessments of overall accuracy are regularly followed up, and any unexpected results are considered. *(Fair)*	
5.	The results of the improvement activities have been analysed and objectively shown to move the achieved overall accuracy towards the accuracy requirements for many, but possibly not all, key uses. The regular evaluations of overall accuracy include accuracy requirements and the most influential sources of uncertainty. *(Good)*	
6.	The results of the improvement activities have been analysed and objectively shown to clearly move the achieved overall accuracy towards the accuracy requirements for many, but possibly not all, key uses. The regular evaluations of overall accuracy include accuracy requirements and the most influential sources of uncertainty. Results are documented, especially any unexpected findings. *(Good)*	
7.	Improvement activities to improve overall accuracy have, where warranted, been undertaken, evaluated, and shown to provide an improved level of accuracy. The implemented regular evaluation measures provide a better understanding of the causes of uncertainty, which will be useful for coming production rounds. The measures cover some important areas of improvement. *(Very good)*	
8.	Improvement activities to improve overall accuracy, where warranted, have been undertaken, evaluated, and shown to provide a clearly improved level of accuracy. There are implemented regular quantitative evaluation measures that provide a better understanding of the causes of uncertainty, which will be useful for coming production rounds. The measures cover the essential areas of improvement. Results are documented and communicated. *(Very good)*	
9.	The evaluation measures for improvement activities show clear improvements in overall accuracy or in the effective use of resources. The achieved accuracy is in line with the key accuracy requirements. The implemented regular evaluation measures are conclusive and provide a better understanding of the causes of uncertainty. They cover the essential areas of improvement. Results are documented and communicated. *(Excellent)*	
10.	There are regular comparisons between current accuracy requirements, design, and achieved accuracy. The achieved accuracy is in line with the key accuracy requirements. Deviations from expectations are analysed. Results are documented and communicated. There is cooperation with similar products to improve the evaluations, including measures and priorities. *(Excellent)*	

C6. Communication with users and data suppliers		Comments
1.	There is some basic communication with users, but the communication on overall accuracy is limited even with key users. The communication with data suppliers (micro data or statistics) is limited, for instance, it deals essentially with the time schedule. *(Weak)*	
2.	There is some basic information on overall accuracy, but it is crude and not user-friendly. The communication is limited even with key users. The communication with data suppliers mainly stresses the time schedule and the processing of input data (statistics). *(Weak)*	
3.	Key uses and accuracy requirements have been discussed with the main users (national and international). Users are informed about achieved overall accuracy, but the assessments and the descriptions are crude. The communication with data suppliers includes factors of importance in their input for overall accuracy. *(Fair)*	
4.	Key uses and accuracy requirements have been discussed with the main users. The information on the assessments of achieved overall accuracy is related to key uses. The communication with data suppliers emphasises factors of importance in their input with regard to overall accuracy. *(Fair)*	
5.	Communications with key users include main uses, requirements for overall accuracy, and the (possibly crude) assessment of the achieved overall accuracy. The information includes influential uncertainty sources. The communication with data suppliers discusses factors of importance in their input for overall accuracy. *(Good)*	
6.	Communications with key users also include priorities. Communication with data suppliers discusses factors of importance in their input for overall accuracy and sets priorities for their work. *(Good)*	
7.	Communications with users of statistics (registers) share information on measures for the overall accuracy in order to enable interpretation and adequate use of the statistics (registers). Possible improvements and priorities are discussed with key users. Communication with data suppliers provides information of the effects of their outputs on accuracy, and they are somewhat involved in the cyclical procedure with possible improvements. *(Very good)*	
8.	Communications with users of statistics (registers) share information on relevant measures for the overall accuracy in order to enable interpretation and adequate use of the statistics (registers). Possible improvements and priorities are discussed with key users, including conflicting needs. Communication with data suppliers provides information of the effects of their outputs on accuracy, and they are involved in the cyclical procedure with possible improvements. *(Very good)*	
9.	There is regular discussion with users regarding the product and on accuracy requirements. Is-sues on accuracy for key uses are discussed. Communication with data suppliers is sufficiently detailed regarding the effects of their outputs on accuracy. Data suppliers are involved in the process of making possible improvements. *(Excellent)*	
10.	The consultation with users also includes prioritisation of resources. Communication with data suppliers is detailed regarding effects of their outputs on accuracy. Da-ta suppliers are involved in the process of making possible improvements and, where relevant, in prioritisation of resources. *(Excellent)*	

Appendix 4. Checklists for Sources of uncertainty

For each criterion (C1-C6), highlight **in bold text** *the statement in the left-hand column that corresponds to the highest level to which the product complies. Then provide comments in the right-hand column to justify that level.*

C1. Available Expertise (members of the production team or other internal expertise)		**Comments**
1.	There is no capable expertise available regarding methods and techniques to even crudely assess the effects of this source of uncertainty. *(Weak)*	
2.	The available expertise is insufficient when it comes to methods and techniques to assess the effects of this source of uncertainty. *(Weak)*	
3.	There is basic expertise available to study this source of uncertainty and its importance for overall accuracy. *(Fair)*	
4.	There is basic expertise available to assess the effects of this source of uncertainty, also to discuss it with the production team and key users. *(Fair)*	
5.	There is a good level of expertise available to study this source of uncertainty, also to communicate with the production team and discuss with key users. There is good understanding of the cyclical procedure. *(Good)*	
6.	There is a good level of expertise available to study this source of uncertainty, also to communicate with the production team and discuss with users; there is also some capability of contributing improvement ideas. *(Good)*	
7.	There is a very good level of expertise available to study this source of uncertainty and to discuss the results with users. There is a good working relationship within the production team regarding the cyclical procedure. *(Very good)*	
8.	There is a very good level of expertise available to derive measures of effects of this source of uncertainty, including some objective measures. There is ability to extend standard theory to develop new techniques and measures. There is expertise to develop improvement strategies including measures of effects. *(Very good)*	
9.	There is expertise available to have an in depth understanding of accuracy requirements, to develop objective measures of effects of this source of uncertainty, and to formulate prioritised improvement strategies. The expertise has international contacts and cooperation. *(Excellent)*	
10.	The expertise extends themselves such that they have the ability to innovate and develop new approaches, best practices, and new standards. There is a clear strategy in place to ensure this level of expertise into the future. *(Excellent)*	

C2. Compliance with standards and best practices		Comments
1.	The production team is mainly unaware of standards that relate to the product and this source of uncertainty. *(Weak)*	
2.	The production team has some, but limited knowledge of standards that relate to the product and this source of uncertainty. *(Weak)*	
3.	The production team is generally aware of standards (internal and external regulations, policies etc.) that relate to the product and this source of uncertainty. There is some evidence of compliance with these. *(Fair)*	
4.	The production team is generally aware of standards that apply to the product and there is evidence of compliance with these. There are plans in place to address significant areas that do not satisfy standards. *(Fair)*	
5.	The production team has good knowledge of relevant standards in relation to this source of uncertainty and the product. There is knowledge of relevant internal best practices. There is evidence of application of standards and best practices, at least where significant. *(Good)*	
6.	The production team has good knowledge of standards in relation to this source of uncertainty and the product. There is knowledge of relevant best practices, for example from similar products in other countries. There is evidence of application of both standards and best practices. *(Good)*	
7.	The production team has very good knowledge of standards and best practices related to this source of uncertainty and the product. The knowledge of best practices is regularly updated through contacts with other expertise. Relevant standards and best practices are applied where significant. *(Very good)*	
8.	The production team has very good knowledge of standards and best practices related to this source of uncertainty and the product. The knowledge of best practices is kept up-to-date through close contacts with other expertise. Relevant standards and best practices are applied wherever appropriate. *(Very good)*	
9.	The production team conducts cooperative activities with other expertise to ensure that compliance with standards and best practices related to this source of uncertainty is maintained, also to participate in the development of best practices related to this source of uncertainty and the product. *(Excellent)*	
10.	The production team is sought out by other national and international experts to participate in the development of standards and best practices related to this source of uncertainty and the product. *(Excellent)*	

C3. Knowledge of requirements, achievements, and improvement needs		Comments
1.	The knowledge of key users and uses is vague, as is the knowledge about this source of uncertainty, its importance, and its effects. *(Weak)*	
2.	The knowledge of key users and uses is limited. The knowledge about this source of uncertainty, its importance, and its effects is vague. *(Weak)*	
3.	Some work has been done for this source of uncertainty to quantify its importance and effects. *(Fair)*	
4.	The importance of this source of uncertainty has been assessed and related to overall accuracy requirements with regard to key uses. *(Fair)*	
5.	When key accuracy requirements are considered in the product design, this source of uncertainty is included if it is clearly influential. Some comparisons are made between the design and the achievements. *(Good)*	
6.	There are some quantitative measures of uncertainty from this source, available at least on an ad hoc-basis and especially if it is influential. A working list exists of possible activities in order to improve knowledge or accuracy. *(Good)*	
7.	There are assessments for this source of uncertainty, for example, in some form of interval, at least if the source is influential. The product design includes this information in its starting point. *(Very good)*	
8.	There are objective measures of uncertainty from this source that are updated regularly. These measures are utilised in the ongoing work with the product design, which has accuracy requirements as a starting point. The working list with potential activities and rough priorities is regularly updated regarding this source of uncertainty, especially if it is influential. *(Very good)*	
9.	The product design regards this source of uncertainty. The need for more knowledge is considered and prioritised with regard to achieved accuracy, the effective use of resources, and recent information that has become available. Possible future factors are considered. *(Excellent)*	
10.	The product design is based on the accuracy requirements and on previous experience of achieved accuracy. The need for more knowledge is prioritised with regard to achieved accuracy, the effective use of resources, and recent information that has become available. Regular contact is maintained internally with similar products and with other countries in order to improve knowledge and suggestions for improvements. Possible future factors are considered. *(Excellent)*	

C4. Plans for improvement activities		Comments
1.	There are no recent or current efforts to launch improvement activities with respect to this source of uncertainty, although there is a clear need. *(Weak)*	
2.	There are no recent or current efforts to launch improvement activities, although particular areas of shortfall with respect to this source of uncertainty have been identified. *(Weak)*	
3.	Some activities to address particular shortfalls with respect to this source of uncertainty are planned, and they are about to start. *(Fair)*	
4.	Some activities to address particular shortfalls with respect to this source of uncertainty are planned, and a few have started. *(Fair)*	
5.	The important areas of shortfall with respect to this source of uncertainty have been identified, and a management approved plan to address the key areas exists. *(Good)*	
6.	Resources have been allocated for at least one prioritised area, and an overall plan exists for the other addressed key areas. *(Good)*	
7.	There is a plan in line with all identified priorities. The plan has management approval and allocated resources at least in the short run. It includes some measures to evaluate the effectiveness of the improvement activities. *(Very good)*	
8.	Well-reasoned and prioritised plans are developed on the basis of the regular evaluations. These plans are resourced, and they include the evaluation of the effectiveness of the planned improvement activities. *(Very good)*	
9.	The prioritised improvement activities not only are resourced and include evaluation of effectiveness, but they are also coordinated within the product and, where possible, with similar products in order to enhance improvements. Unintended consequences are avoided. *(Excellent)*	
10.	The prioritised improvement activities not only are resourced and include evaluation of effectiveness, but they are also coordinated within the product and with similar products in order to enhance improvements. Unintended consequences are avoided. *(Excellent)*	

C5. Results of improvement activities and findings from other evaluations		Comments
1.	There is no study of effectiveness from improvement activities. Work to evaluate the production process hardly considers this source of uncertainty. *(Weak)*	
2.	The study of effectiveness from improvement activities is not very conclusive. There is no regular work to evaluate the production process with respect to this source of uncertainty. *(Weak)*	
3.	Improvement activities seem to have improved achieved accuracy, but the results are just indicators. There are assessments of this source of uncertainty, possibly crude, that are regularly followed up. Unexpected results are noted. *(Fair)*	
4.	There are indications that improvement activities have reduced the uncertainty from this source when it comes to key uses. The regular evaluations include this source of uncertainty if influential. *(Fair)*	
5.	The results of the improvement activities have been analysed and objectively shown to move the uncertainty from this source towards the accuracy requirements for many, but possibly not all, key uses. The regular evaluations include appropriate measures for this source of uncertainty (if influential). *(Good)*	
6.	The results of the improvement activities have been analysed and objectively shown to clearly move the uncertainty from this source towards the accuracy requirements for many, but possibly not all, key uses. The regular evaluations include appropriate measures of uncertainty from this source. Results are documented, especially any unexpected findings. *(Good)*	
7.	Improvement activities with respect to this source of uncertainty have, where warranted, been undertaken, evaluated, and shown to provide an improved level of accuracy. The implemented regular evaluation measures provide a better understanding of this source of uncertainty, which will be useful for coming production rounds. The measures cover some important areas of improvement. *(Very good)*	
8.	Improvement activities with respect to this source of uncertainty have, where warranted, been undertaken, evaluated, and shown to provide a clearly improved level of accuracy. There are implemented regular quantitative evaluation measures that provide a better understanding of the causes of this source of uncertainty, which will be useful for coming production rounds. The measures cover the essential areas of improvement. Results are documented and communicated. *(Very good)*	
9.	The evaluation measures for improvement activities show clear improvements in overall accuracy or in the effective use of resources when it comes to this source of uncertainty. The implemented regular evaluation measures are conclusive and provide a better understanding of the causes of uncertainty from this source. The measures cover the essential areas of improvement. Results are documented and communicated. *(Excellent)*	
10.	This source of uncertainty is included in the regular comparisons between current accuracy requirements, design, and achieved accuracy. The achieved accuracy is in line with the key requirements. Deviations from expectations are analysed. Results are documented and communicated. There is cooperation with similar products to improve the evaluations, including measures and priorities. (*Excellent)*	

C6. Communication with users and data suppliers		Comments
1.	There is some basic communication with users but the communication on this source of uncertainty is limited even with key users. The communication with data suppliers (micro data or statistics) is limited, for instance, it deals essentially with the time schedule. *(Weak)*	
2.	There is some basic information on this source of uncertainty, but it is crude and not user-friendly. The communication is limited even with key users. The communication with data suppliers mainly stresses the time schedule and the processing of input data (statistics). *(Weak)*	
3.	Key uses and accuracy requirements have been discussed with the main users (national and international). Users are informed about achieved accuracy when it comes to this source of uncertainty, but the assessments and the descriptions are somewhat crude. The communication with data suppliers includes factors of importance in their input with regard to this source of uncertainty, if influential. *(Fair)*	
4.	Key uses and accuracy requirements have been discussed with the main users. The information on the assessments of achieved accuracy when it comes to this source of uncertainty is related to key uses. The communication with data suppliers emphasises factors of importance in their input with regard to this source of uncertainty, if influential. *(Fair)*	
5.	Communications with key users include this source of uncertainty and its effects, if influential. The communication with data suppliers discusses factors of importance in their input for this source of uncertainty with regard to overall accuracy. *(Good)*	
6.	Communications with key users also include priorities when it comes to this source of uncertain-ty. Communication with data suppliers discusses this source of uncertainty and its importance for overall accuracy and sets priorities for their work. *(Good)*	
7.	Communications with users of statistics (registers) share information on measures for this source of uncertainty in order to enable interpretation and adequate use of the statistics (registers). Possible improvements and priorities are discussed with key users. Communication with data suppliers provides information of the effects of their outputs on un-certainty from this source, and they are somewhat involved in the cyclical procedure with possible improvements. *(Very good)*	
8.	Communications with users of statistics (registers) share information on relevant measures for this source of uncertainty in order to enable interpretation and adequate use of the statistics (registers). Possible improvements and priorities are discussed with key users, including conflicting needs. Communication with data suppliers provides information of the effects of their outputs on un-certainty from this source, and they are involved in the cyclical procedure with possible improvements. *(Very good)*	
9.	There is regular discussion with users regarding the product and issues concerning this source of uncertainty for key uses. Communication with data suppliers is sufficiently detailed regarding the effects of their outputs on uncertainty from this source. Data suppliers are involved in the process of making possible improvements. *(Excellent)*	
10.	The consultation with users also includes prioritisation of resources. Communication with data suppliers is detailed regarding effects of their outputs on this source of uncertainty. Data suppliers are involved in the process of making possible improvements and in prioritisation of resources. *(Excellent)*	

Appendix 5. Checklists for Preliminary statistics compared with final statistics

For each criterion (C1-C6), highlight **in bold text** *the statement in the left-hand column that corresponds to the highest level to which the product complies. Then provide comments in the right-hand column to justify that level.*

C1. Available Expertise (members of the production team or other internal expertise)		Comments
1.	There is no capable expertise available regarding methods and techniques required to study revisions: size and direction. *(Weak)*	
2.	The available expertise is insufficient when it comes to methods and techniques required to study revisions: size and direction. *(Weak)*	
3.	There is basic expertise available to understand requirements for the size and direction of revisions. *(Fair)*	
4.	There is basic expertise available to analyse revisions, also to discuss requirements and achievements with the production team and key users. *(Fair)*	
5.	There is a good level of expertise available to analyse revisions and to communicate with the production team and discuss with users. There is good understanding of the cyclical procedure. *(Good)*	
6.	There is a good level of expertise to analyse revisions and to communicate with the production team and discuss with key users; there is also some capability of contributing improvement ideas for the size and direction of revisions. *(Good)*	
7.	There is a very good level of expertise available to analyse revisions and to discuss the results with key users. There is a good working relationship within the production team regarding the cyclical procedure. *(Very good)*	
8.	There is a very good level of expertise available to analyse revisions, including causes for them. There is some ability to extend standard theory to develop new techniques for preliminary statistics (registers). There is expertise to develop improvement strategies including measures of effects. *(Very good)*	
9.	There is expertise available to have an in depth understanding of revisions and their causes, also to formulate prioritised improvement strategies. The expertise has international contacts and cooperation. *(Excellent)*	
10.	The expertise extends themselves such that they have the ability to innovate and develop new approaches, best practices, and new standards. There is a clear strategy in place to ensure this level of expertise into the future. *(Excellent)*	

C2. Compliance with standards and best practices		Comments
1.	The production team is mainly unaware of standards that relate to the product and preliminary statistics (register). *(Weak)*	
2.	The production team has some, but limited, knowledge of standards that relate to the product and preliminary statistics (register). *(Weak)*	
3.	The production team is generally aware of standards (internal and external regulations, policies etc.) that relate to the product and preliminary statistics (register). There is some evidence of compliance with these. *(Fair)*	
4.	The production team is generally aware of standards that apply to the product – especially preliminary statistics (register) – and there is evidence of compliance with these. There are plans in place to address significant areas that do not satisfy standards. *(Fair)*	
5.	The production team has good knowledge of relevant standards in relation to preliminary statistics (register) and the product. There is knowledge of relevant internal best practices. There is evidence of application of standards and best practices, at least where significant. *(Good)*	
6.	The production team has good knowledge of standards in relation to the product and preliminary statistics (register). There is knowledge of relevant best practices, for example from similar products in other countries. There is evidence of application of both standards and best practices. *(Good)*	
7.	The production team has very good knowledge of standards and best practices related to preliminary statistics (register) and the product. The knowledge of best practices is regularly updated through contacts with other expertise. Relevant standards and best practices are applied where significant. *(Very good)*	
8.	The production team has very good knowledge of standards and best practices related to preliminary statistics (register) and the product. The knowledge of best practices is kept up-to-date through close contacts with other expertise. Relevant standards and best practices are applied wherever appropriate. *(Very good)*	
9.	The production team conducts cooperative activities with other expertise to ensure that compliance with standards and best practices related to preliminary statistics (register) is maintained, also to participate in the development of best practices related to preliminary statistics (register) and the product. *(Excellent).*	
10.	The production team is sought out by other national and international experts to participate in the development of standards and best practices related to preliminary statistics (register) and the product. *(Excellent)*	

C3. Knowledge of requirements, achievements, and improvement needs		Comments
1.	The knowledge of key users and uses is vague, as is the knowledge for the preliminary statistics (register) about requirements and achievements for the size and direction of revisions. *(Weak)*	
2.	The knowledge of key users and uses is limited. The knowledge about the preliminary statistics (register) is vague when it comes to requirements and achievements for the size and direction of revisions. *(Weak)*	
3.	Some work has been done to quantify requirements on size and direction of revisions with regard to key uses. There are some measures of revision size and direction. *(Fair)*	
4.	Requirements on size and direction of revisions have been quantified with regard to key uses. There are measures of the size and direction of revisions. *(Fair)*	
5.	Key requirements for revisions (size and direction) are considered in the product design; some targets are set. Comparisons are made between the design and the achievements. *(Good)*	
6.	There are regular measures of the size and direction of revisions. The sources of uncertainty which influence revisions the most are identified. A working list exists of possible activities in order to improve knowledge or revisions. *(Good)*	
7.	There are analyses of revisions (size and direction). The product design includes requirements for revisions in its starting point. Comparisons with achieved revisions are utilised to formulate suggestions that will improve knowledge or accuracy of the preliminary statistics (register). *(Very good)*	
8.	The analyses of revisions are updated regularly. These measures are utilised in the ongoing work with the product design, which has accuracy requirements as a starting point. The working list with potential activities and rough priorities is regularly updated. *(Very good)*	
9.	The product design is based on accuracy requirements and on previous experience of achievements. The need for more knowledge is considered and prioritised with regard to achieved revisions, the effective use of resources, and recent information that has become available. Possible future factors are considered. *(Excellent)*	
10.	The product design is based on the accuracy requirements and on previous experience of achieved accuracy of the preliminary statistics (register). The need for more knowledge is prioritised with regard to achieved accuracy, the effective use of resources, and recent information that has become available. Regular contact is maintained internally with similar products and with other countries in order to improve knowledge and suggestions for improvements. Possible future factors are considered. *(Excellent)*	

C4. Plans for improvement activities		Comments
1.	There are no recent or current efforts to launch improvement activities with respect to revisions (size and direction), although there is a clear need. *(Weak)*	
2.	There are no recent or current efforts to launch improvement activities, although areas of shortfall with respect to revisions have been identified. *(Weak)*	
3.	Some activities to address particular shortfalls with respect to revisions are planned, and they are about to start. *(Fair)*	
4.	Some activities to address particular shortfalls with respect to revisions are planned, and a few have started. *(Fair)*	
5.	The important areas of shortfall with respect to revisions have been identified, and a management approved plan to address the key areas exists. *(Good)*	
6.	Resources have been allocated for at least one prioritised area, and an overall plan exists for the other addressed key areas. *(Good)*	
7.	There is a plan in line with all identified priorities. The plan has management approval and allocated resources at least in the short run. It includes some measures to evaluate the effectiveness of the improvement activities. *(Very good)*	
8.	Well-reasoned and prioritised plans are developed on the basis of the regular evaluations. These plans are resourced, and they include the evaluation of the effectiveness of the planned improvement activities. *(Very good)*	
9.	The prioritised improvement activities not only are resourced and include evaluation of effectiveness, but they are also coordinated within the product and, where possible, with similar products in order to enhance improvements. Unintended consequences are avoided. *(Excellent)*	
10.	The prioritised improvement activities not only are resourced and include evaluation of effectiveness, but they are also coordinated within the product and with similar products in order to enhance improvements. Unintended consequences are avoided. *(Excellent)*	

Here improvement activities and findings from other evaluations that explicitly refer to preliminary statistics (register) compared to final statistics (register) are considered.

C5. Results of improvement activities and findings from other evaluations		Comments
1.	There is no study of effectiveness from improvement activities. Work to evaluate the production process hardly considers revisions (size and direction). *(Weak)*	
2.	The study of effectiveness from improvement activities is not very conclusive. There is no regular work to evaluate the production process with respect to revisions. *(Weak)*	
3.	Improvement activities seem to have improved revisions, but the results are just indications. There are revision measures that are regularly followed up. Unexpected results are noted. *(Fair)*	
4.	There are indications that improvement activities have improved revisions when it comes to key uses. Revision measures are regularly followed up, and any unexpected results are considered. *(Fair)*	
5.	The results of the improvement activities have been analysed and objectively shown to move the achieved revisions towards requirements for many, but possibly not all, key uses. The regular evaluations of revisions include accuracy requirements and analyses. *(Good)*	
6.	The results of the improvement activities have been analysed and objectively shown to clearly move the achieved revisions towards requirements for many, but possibly not all, key uses. The regular evaluations of revisions include requirements and analyses, for example with respect to causes. Results are documented, especially unexpected findings. *(Good)*	
7.	Activities to improve revisions have, where warranted, been undertaken, evaluated, and shown to provide improvements for revisions. The implemented regular evaluation measures provide a better understanding of the causes of revisions, which will be useful for coming production rounds. The measures cover some important areas of improvement. *(Very good)*	
8.	Activities to improve revisions where warranted, have been undertaken, evaluated, and shown to provide clear improvements for revisions. There are implemented regular quantitative evaluation measures that provide a better understanding of the causes of revisions, which will be useful for coming production rounds. The measures cover the essential areas of improvement. Results are documented and communicated. *(Very good)*	
9.	The evaluation measures for improvement activities show clear improvements for revisions or in the effective use of resources. The achievements are in line with the key accuracy requirements. The implemented regular evaluation measures are conclusive and provide a better understanding of the causes of revisions. The measures cover the essential areas of improvement. Results are documented and communicated. *(Excellent)*	
10.	There are regular comparisons between current requirements for revisions, design, and achievements. The achievements are in line with the key requirements. Deviations from expectations are analysed. Results are documented and communicated. There is cooperation with similar products to improve the evaluations, including measures and priorities. *(Excellent)*	

C6. Communication with users and data suppliers		Comments
1.	There is some basic communication with users but the communication on revisions is limited even with key users. The communication with data suppliers (micro data or statistics) is limited, for instance, it deals essentially with the time schedule. *(Weak)*	
2.	There is some basic information on revisions, but it is crude and not user-friendly. The communication is limited even with key users. The communication with data suppliers mainly stresses the time schedule and the processing of input data (statistics). *(Weak)*	
3.	Key uses and requirements on revisions have been discussed with the main users (national and international). Users are informed about achieved revisions, but the assessments and the descriptions are somewhat crude. The communication with data suppliers includes factors of importance in their input for preliminary statistics (register). *(Fair)*	
4.	Key uses and requirements on revisions have been discussed with the main users. The information on achieved revisions is related to key uses. The communication with data suppliers emphasises factors of importance in their input with regard to preliminary statistics (register). *(Fair)*	
5.	Communications with key users include main uses, requirements for revisions, and the revision measures. The communication with data suppliers discusses factors of importance in their input with regard to revisions. *(Good)*	
6.	Communications with key users also include priorities. Communication with data suppliers discusses factors of importance in their input with regard to revisions and sets priorities for their work. *(Good)*	
7.	Communications with users of statistics (registers) share information on revisions in order to enable interpretation and adequate use of the preliminary statistics (registers). Possible improvements and priorities are discussed with key users. Communication with data suppliers provides information of the effects of their outputs on revisions, and they are somewhat involved in the cyclical procedure with possible improvements. *(Very good)*	
8.	Communications with users of statistics (registers) share information on relevant measures for revisions in order to enable interpretation and adequate use of the preliminary statistics (registers). Possible improvements and priorities are discussed with key users, including conflicting needs. Communication with data suppliers provides information of the effects of their outputs on revisions, and they are involved in the cyclical procedure with possible improvements. *(Very good)*	
9.	There is regular discussion with users regarding the product and on revision requirements. Is-sues on accuracy of preliminary statistics (registers) for key uses are discussed. Communication with data suppliers is sufficiently detailed regarding the effects of their outputs on revisions. Data suppliers are involved in the process of making possible improvements. *(Excellent)*	
10.	The consultation with users also includes prioritisation of resources. Communication with data suppliers is detailed regarding effects of their outputs on revisions. Data suppliers are involved in the process of making possible improvements and in prioritisation of resources. *(Excellent)*	

List of figures

List of tables

Statistics Sweden provides society with statistics for decision-making, debate and research on behalf of the Government, government agencies, researchers and industry. These statistics contribute to fact-based public discourse and informed decisions.