United Nations ECE/CES/GE.41/2024/16

Economic and Social Council

Distr.: General
2 August 2024

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Twenty-sixth Meeting
Geneva, 2–4 October 2024
Item 2 (b) of the provisional agenda
Revising the Conference of European Statisticians Recommendations for Population and Housing Censuses for the 2030 round: Quality Assessment and Quality Management

Developing the Recommendations on Quality Assessment and Quality Management: part 2 (Annex on Quality management programme implementation)

Note by the Conference of European Statisticians Task Force on Quality Assessment and Quality Management*,**

Summary

This document contains the draft text for the Annex of the Conference of European Statisticians (CES) Recommendations for the 2030 round of population and housing censuses, entitled Quality management programme implementation. It is intended to be read in conjunction with document ECE/CES/GE.41/2024/15 which introduces the topic of quality management and quality assessment as a whole, gives a summary of the changes introduced in comparison to the Recommendations for the previous, 2020 round, and presents the draft text of the chapter on Quality management.

* The Conference of European Statisticians Task Force on Quality Assessment and Quality Management currently consists of the following members: Donatella Zindato (Italy, chair), Gladys Massé (Argentina), Daniel Scheuregger (Eurostat), Saskia Fuchs (Germany), Martin Streng (Germany), Carole Schmitz (Mexico), Mauricio Rodríguez Abreu (Mexico), Jacco Daalmans (Netherlands), Suzana Stanojkovic (Serbia), Yu Han (United Arab Emirates (UAE)), Nasser Mohamed Dayan (UAE), Cal Ghee (United Kingdom), Deborah Stempowski (United States of America) and Haoyi Chen (United Nations Statistics Division).

NOTE: The designations employed in this document do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

** This document was submitted to the conference services for processing after the deadline for technical reasons beyond the control of the submitting office.

The main purpose of the two documents is to elicit comments and suggestions from national census experts on the proposed draft text, to ensure that it reflects the needs and priorities of national statistical offices and the latest methodological developments.

I. Introduction

1. Every ten years the Conference of European Statisticians (CES) issues Recommendations to guide countries in conducting their population and housing censuses. The Recommendations are developed by expert task forces overseen by the CES Steering Group on Population and Housing Censuses.

2. Section II of this document presents the draft Annex on Quality management programme implementation for the CES Recommendations for the 2030 round of population and housing censuses. This is intended to be read in conjunction with document ECE/CES/GE.41/2024/15, which contains the draft chapter on Quality management as well as a summary of the changes introduced in all the content on quality management, in comparison to the Recommendations for the previous, 2020 round.

3. The main purpose of the two documents is to elicit comments and suggestions from national census experts on the proposed draft text, to ensure that it reflects the needs and priorities of national statistical offices and the latest methodological developments.

II. Draft text for the Annex on Quality management programme implementation

A. Introduction

4. As described in the chapter on Quality management, quality must be managed in an integrated fashion within the broader context of undertaking the entire census programme. Census management will require input and support from all functional areas and it is within this context that trade-offs necessary to ensure an appropriate balance between quality and concerns of cost, response burden and other factors will be made. There needs to be adequate staff with people able to speak with expertise and authority while being sensitive to the need to weigh competing pressures regarding dimensions of quality and other factors to reach a consensus. Those responsible for each aspect of census work must be equipped with appropriate expertise. Each of them will need to develop and implement strategies addressing many aspects of quality. In doing so they must be sensitive not only to their own quality requirements but also to their interactions with quality requirements of others. Strategies to facilitate the necessary information sharing and joint consideration of cross-cutting quality issues are vital.

5. Quality requirements need to receive appropriate attention during design, implementation and assessment. Subject matter experts will bring knowledge of content, client needs, relevance and coherence. Statistical methodologists bring their expertise on statistical methods and data quality trade-offs, especially with respect to accuracy, timeliness and cost. Operations experts bring experience in operational methods, and concerns for practicality, efficiency, field staff, respondents and operational quality control. The systems experts bring knowledge of technology standards and tools that will help facilitate achievement of quality, particularly in the timeliness and accuracy dimensions. In collaboration with subject matter experts, dissemination experts will bring a focus to accessibility and interpretability.

6. This annex provides further guidance about implementing a quality management programme, building on the ideas that were introduced in the chapter on Quality management. Firstly, the six dimensions of quality are taken in turn, with a description of how the five components of the quality management framework might apply to each. Later sections then provide further detail on:

(a) Operational quality control;

(b) Questionnaire design;

(c) Management of coverage error;

(d) Systems development;

(e) Census evaluation;

(f) Measuring record-linkage quality; and

(g) Measuring editing and imputation quality in register-based censuses.

7. As a reminder, the six dimensions of quality are:

(a) Relevance;

(b) Accuracy;

(c) Timeliness;

(d) Accessibility;

(e) Interpretability; and

(f) Coherence.

and the five components of the quality management framework are:

(a) Setting quality targets;

(b) Quality design;

(c) Operational quality control;

(d) Quality assurance and improvement; and

(e) Quality evaluation and reporting.

B. Managing relevance

8. As noted in the chapter on Quality management, the programmes and outputs of a National Statistical Office (NSO) must reflect the country's most important information needs. Relevance for the census must therefore be managed within this broader context. At the stage of *setting quality targets* it is necessary to discuss how much change in the questionnaire (or the information to be taken from registers) will be contemplated. In cases of severe budget restrictions, some countries have agreed to a policy of minimal change or no change to minimize testing requirements and quality risks. Clearly, this impacts the relevance of the final statistics, but having such discussions at the outset, with stakeholders, is essential.

9. At the stage of *quality design*, relevance is managed through processes to assess the relevance of previous census content and to identify new or emerging information gaps that may be appropriately filled via the census. Major processes to achieve this can be described as client and stakeholder feedback mechanisms and programme review and data analysis. Information from these processes can then be used to ensure the relevance of census content and outputs.

10. Important feedback mechanisms might include: consultations with key government departments and agencies; advice from professional advisory committees in major subject matter areas; user feedback and market research; ad hoc consultations with interested groups; and liaison with statistical offices from other countries.

11. While the primary purpose of data analysis is to advance understanding of phenomena, it also provides feedback on the adequacy and completeness of the data used in the analysis. By identifying questions that the census data cannot answer it can pinpoint gaps and weaknesses. This must be taken in the context of the analytic potential of other data holdings of the NSO. There is a reduced focus on relevance during *Operational quality control* and *Quality assurance and improvement*, but the emphasis increases again for *Quality evaluation and reporting*, when the published outputs can be reviewed to consider how well they met the originally stated information needs.

C. Managing accuracy

12. Management of accuracy requires attention during all five steps in the quality management framework. Firstly, when *setting quality targets*, as described in the chapter on Quality management, accuracy targets should be set as these will fundamentally affect the census costs and design.

13. During *quality design*, parameters and decisions will have a direct impact on accuracy. This includes the design of later components of the quality management framework. The accuracy achieved will depend on the explicit methods put in place for operational quality control and quality assurance and improvement. If these processes are not built in from the outset, including the required data collection processes and feedback loops, it will be much more challenging for them to be implemented effectively.

14. A number of key aspects of design should be considered in every census to ensure that accuracy concerns are given appropriate attention:

(a) Explicit consideration of overall trade-offs between accuracy, cost, timeliness and respondent burden during the design phase;

(b) Adequate justification for each question asked and appropriate pre-testing of questions and questionnaires in each mode of collection, while also ensuring that the set of questions is sufficient to meet requirements;

(c) Assessment of the coverage of the target population. This relates to the adequacy of the geographic infrastructure upon which collection and dissemination geography will be based. It may also relate to the adequacy of address lists to be used in areas where mail out of census questionnaires takes place;

(d) Proper consideration of sampling and estimation options. For example, sampling could be used at the collection stage through the use of short and long form questionnaires in order to reduce respondent burden and collection costs. Alternatively, sampling could be introduced after collection, by processing only a sample of records, at least for a subset of characteristics, in order to produce more timely results or to control processing costs. In either case, careful consideration should be given to the size and design of the sample and to the weighting and other estimation procedures needed;

(e) Adequate measures in place for facilitating and encouraging accurate response, following up non-response and dealing with missing data;

(f) Proper consideration of the need for operational quality control;

(g) Appropriate quality assurance for the final statistics.

15. While individual programme managers may have considerable flexibility in implementing specific practices and methods, this should be done in an integrated fashion within the overall management of census quality.

16. A good design will always contain protection against implementation errors through, for example: adequate selection and training of staff; suitable supervisory structures; carefully written and tested procedures and systems; and operational quality control procedures.

17. Mechanisms for *operational quality control* should be built into all processes as part of the design. Information is needed to monitor and correct problems arising during

implementation. This requires a timely information system that provides managers with the information they need to adjust or correct problems while work is in progress. There is an overlap here with *quality assurance and improvement* and *quality evaluation and reporting* as much of the information collected during operational quality control is also needed to assess whether the design was carried out as planned, and to identify problem areas and lessons learned from operations in order to aid design for future censuses.

18. Some examples of activities that could be undertaken to manage and monitor accuracy during implementation and operations are:

(a) Regular reporting and analysis of response rates and completion rates during collection;

(b) Monitoring non-response follow-up rates;

(c) Monitoring interviewer feedback;

(d) Monitoring coverage checks and controls;

(e) Monitoring of edit failure rates and the progress of corrective actions;

(f) Monitoring of results of quality control procedures during collection and processing;

(g) Monitoring of expenditures against progress; and

(h) Development, implementation and monitoring of contingency plans.

19. Where applicable, the activities outlined in paragraph 18 above should be at different geographic levels or aggregations that are useful for each level of management, including those suitable for supervising and correcting the actions of groups or individuals involved.

20. Accuracy is multidimensional. Indicators may touch on many aspects of census collection, processing and estimation. Primary areas of assessment include the following:

a) Assessment of coverage error, both under-coverage and over-coverage. In most countries this is done via a post-enumeration coverage survey and using dual system estimation methods. Comparisons with official population estimates, typically projections from the previous census, are often also used as an assessment tool;

b) Non-response rates and imputation rates (including imputation rate of total non-response);

c) Data capture error rates, coding error rates;

d) For register-based/combined censuses, linkage error rates (i.e. errors in the record-linkage process: e.g. false positives/negatives);

e) For register-based/combined censuses, editing rate i.e. the amount of corrections needed to resolve inconsistencies between sources at micro-data level;

f) Measures of sampling error, where applicable; and

g) Any other serious accuracy or consistency problems with the results. This relates closely to coherence and allows for the possibility that problems were experienced with a particular aspect of the census resulting in a need for caution in using results.

21. Further advice on management of accuracy through management of coverage error and through operational quality control is provided later in this annex.

D. Managing timeliness

22. Planned timeliness is a decision to be made when *setting quality standards*, refined if necessary during *quality design*, if resources or practicalities indicate that the ideal timeframes are not achievable. There are often important trade-offs to be made with accuracy and relevance. More timely information may be more relevant but less accurate. So, although timeliness is important it is not an unconditional objective. Many of the factors described under accuracy apply equally here. Timeliness is also directly affected by fundamental time

requirements to collect and process census data giving an adequate allowance for *operational quality control* and *quality assurance and improvement*. It might be tempting to aim for challenging timeframes for outputs at the early stages in census preparations, but these should be tempered by the experience gained from any *quality evaluation* of previous census operations.

23. Major information releases should have publication dates announced well in advance. This helps users plan and provides internal discipline in working within set time frames.

24. For customized/commissioned information retrieval services, the appropriate timeliness measure is the elapsed time between the receipt of a clear request and the delivery of the information product to the client. Service standards should be in place for such services and announced beforehand.

E. Managing accessibility

25. Census information must be readily accessible to users. Statistical information that users are unaware of, cannot locate, are unable to access, or cannot afford to purchase is of no value to them. In most statistical offices, corporate-wide dissemination policies and delivery systems will determine most aspects of accessibility. Decisions about output dissemination methods and policies are often made late in the census process, as the focus is often on the challenges of data collection and processing. This can lead to later time and resource pressures, to the detriment of accessibility. Explicitly setting aims and polices when *setting quality targets* for the other dimensions of quality can help reduce this impact, as it enables costs and development timescales to be better estimated during *quality design*.

26. In determining information product definition and design, managers should take careful account of client demands. Market research and client liaison will help determine these. The proposed aims and output designs, having been defined, can then be discussed with users, with appropriate modifications, in a controlled way. Whilst not strictly *operational quality control* or *quality assurance and improvement* as defined in the chapter on Quality management, there are nonetheless parallels to these components in such discussions.

27. In today's world the Internet is the primary dissemination vehicle. It should include not only the data released but also information about the data (metadata) such as data quality statements and descriptions of the concepts, definitions and methods used. Use should also be made of appropriate links to the NSO's corporate dissemination vehicles.

28. Finally, as part of *quality evaluation*, client feedback should be monitored on the content of the output products and on the mode of dissemination with a view to future improvements.

29. The information needs of the analytic community present some particular requirements. Analysts often need access to microdata records to facilitate specific analyses. This presents special challenges in order to continue to respect the requirements for the maintenance of the confidentiality of census data. A number of means could be used to address these needs. Public use microdata files, typically a sample of census records that have been pre-screened to protect confidentiality can be valuable for analysts. Custom retrieval services where specific analyses, designed by external analysts, can be conducted by staff of the statistical office may meet the needs of some analysts.

F. Managing interpretability

30. Managing interpretability is primarily concerned with providing metadata. Information needed by users to understand census information falls under three broad headings: the concepts, definitions and classifications that underlie the data; the methods used to collect and process the data; and measures of data quality. The first of these also relates to coherence.

31. A further aid to users is the interpretation of census information as it is released. Commentary on the primary messages that the data contains can assist users in initial understanding of the information. As with accessibility, interpretability can be addressed in *all five components* of the quality management framework.

32. As for register-based/combined censuses, the complexity of administrative data may impact the accessibility and clarity output quality dimension from a data user's perspective. That is, users of the census data may find it difficult to understand the use of administrative data in the census and the impact this use has on the quality of the census outputs. Countries are encouraged to provide users with the relevant information on the use of administrative data and on the impact this may have on the quality of census outputs, especially when transitioning from a field-based to a register-based census.

G. Managing coherence

33. Coherence is multidimensional. Objectives for coherence of census data include:

(a) Coherence of census data within itself;

(b) Coherence with data and information from prior censuses;

(c) Coherence with other statistical information available from the statistical office on the same or related phenomena; and

(d) Coherence with information from censuses of other countries.

34. Aims for coherence should be set when *setting quality standards* as these will drive decisions during *quality design*. For example, there will be trade-offs to be made about the degree with which to standardise across programmes within NSOs and, for international standards, between countries. Subsequent decisions during quality design will need to be made about the development and use of standard frameworks, concepts, variables, classifications and nomenclature for all subject matters that are measured.

35. The census must ensure that the process of measurement does not introduce inconsistency between its data and that from other sources. Managers of other statistical programmes are, of course, equally responsible for this aspect of coherence.

36. There is usually less of an emphasis on coherence during *operational quality control*, but the emphasis increases again during *quality assurance and improvement* when the emerging census results can be compared with other available sources (whether published statistics or administrative data, for example). This can highlight differences in interpretation of definitions between statistical outputs, or, indeed, errors in either the census or other surveys. Although this is more correctly concerned with managing accuracy, there is an overlap with managing coherence when issues of definition and their interpretation arise.

37. After publication of census results, analysis of those data that focuses on the comparison and integration of information between the census and other sources will give insights to inform *quality evaluation and reporting* and the degree to which quality has been achieved in coherence. The census data should be analysed for domains and aggregations, both large and small that are considered important. Such analysis should consider totals, distributions, relations between variables or sets of variables, relations between domains, growth rates, etc. as appropriate. Comparisons should be made to data from prior censuses and to comparable survey data.

38. While other sources certainly offer views on the quality of the census, there are limitations to the information they provide and the comparisons that can be made according to the sources. Therefore, data users should be guided in exercising caution in how they interpret differences.

39. NSOs should explain which population comparisons are possible, and provide guidance on how to interpret differences between demographic benchmarks and census results. E.g. results from the Labour Force Survey data are generally used for making comparisons about the population labour market (economic activity) status. In this case, users should be helped in interpreting differences due to data collection and questions design.

40. As for register-based/combined censuses, the administrative sources may be subject to changes over time and inconsistencies in the way the data are collected across segments of the population. It might also happen that a new data source (an administrative register or any other source) is chosen for a census variable (e.g. in The Netherlands the educational attainment register (integrated source from several registers) has replaced the Labour Force Survey as a source for educational attainment. The impact of these changes on census results should be carefully evaluated and explained to census users.

H. Operational quality control

1. Census activities requiring operational quality control

41. A number of census processes involve massive operations, either manual or automated. Examples of such operations include dwelling listing operations, preparation of maps, printing of census materials, enumeration procedures, data capture and editing and coding (both manual and automated). Specific operational quality control procedures are particularly relevant and important for each of these.

42. Dwelling listing operations are commonly conducted by enumerators prior to, or as, questionnaires are dropped off at dwellings. It is particularly important at this stage to minimize both under-coverage and over-coverage of dwellings. To that end, enumerators' procedures should include quality checks to ensure the quality of their work. As well, supervisors should have planned spot checks as listing work starts, and planned quality control procedures to be applied as work is completed.

43. When census questionnaires are delivered, it is usually done on the basis of a list of addresses extracted from an address register. Address register maintenance itself will involve several steps of quality management. Nonetheless, prior to its use, the address list should be validated to confirm that each dwelling exists and is included with correct address and geo-coding information, and that no non-dwellings are included. Allowance should be made for dwellings under construction that may be completed prior to the census. If not done via administrative sources, including the use of postal code information and addresses, this validation requires a large operation in the field and is subject to errors. Since this work must be parcelled out to individual employees in batches, acceptance sampling quality control procedures will be appropriate. Again, spot-checking and close communications with supervisors will be important quality assurance steps.

44. Enumeration, whether by interviewing or by collecting completed questionnaires from the dwellings on the list, is similar. Usually one enumerator is responsible for all work in an enumeration area and will be required to implement a number of quality checks on their own work. Further acceptance sampling procedures, implemented by supervisors, will ensure the quality of various aspects of the enumerators' work.

45. Data processing is one of the crucial steps by which raw census data are converted into a complete edited and coded master file useable for tabulations. In some of these processes the data are being transformed (for example data capture, coding) while in others the data are being corrected (for example edit and imputation). New errors can occur in any of these operations.

46. Countries may find it useful to use applications to provide tools that help ensure data quality, such as:

(a) A *collection application* that can provide consistency and completeness checks, automatic skips and alert messages, and GPS coordinates of the household dwelling to check the enumerator's presence;

(b) A *control application* for the enumeration supervisor that provides the results above as well as other summary information about each enumerator's cases, including information that signals outliers or other unexpected results;

(c) A *data transfer application* at central level, with all the necessary security features (data encryption, data bandwidth fluidity) to avoid data loss in the field;

(d) A *web application/dashboard* for monitoring data quality via real-time indicators for management teams in the field and at central level.

2. Operational quality control methods

47. Clearly a census operational quality control regime comprises a wide variety of mechanisms and processes acting at various levels throughout the census programme. An important technique applicable in many census operations is statistical quality control. It primarily addresses accuracy, although depending on the operation it may also address other elements of quality. What follows is some brief basic information on quality control. For a complete explanation of these methods, the reader should refer to a standard text or reference such as Duncan (1986), Hald (1981) or Schilling (1982).[1]

48. The success of any operational quality control programme depends on laying down quality standards or requirements; determining appropriate verification techniques; measuring quality; and providing for timely feedback from the results of the programme so that effective corrective action may be taken.

49. Sample verification, complete (or 100 per cent) verification, or spot checks are the usual quality control techniques adopted in censuses.

50. Verification can be dependent or independent. In dependent verification, a verifier assesses the work of a census worker by examining that work. However, the verifier may be influenced by the results obtained in the initial operation. In independent verification a job is verified independently by a verifier without reference to the original work. The original results and those of the verifier are compared; if the results agree then the work is considered correct; if not, a third, often expert, verifier may resolve the difference.

51. For each operation, the quality control program should determine which approach is appropriate and will achieve the goals of the program while also staying within the budget[2].

I. Questionnaire design

52. The design of the census questionnaire(s) takes into account the statistical requirements of the data users, administrative requirements of the census, the requirements for data processing, as well as the characteristics of the population. Because censuses often involve multiple collection methods, testing must be performed to ensure that questionnaires will work properly for all applicable methods. The questionnaire should include elements aimed at ensuring accurate coverage of the population (for example who to include, who not to include, where to be enumerated). Qualitative testing is required to check these issues and should cover an adequate variety of situations encountered in the population. In terms of content, quality management approaches for a census are similar to those for a sample-based

[1] Duncan, A.J. 1986. Quality Control and Industrial Statistics. Fifth edition. R.D. Irwin Inc., Illinois. Hald, A. 1981. Statistical Theory of Sampling Inspection by Attributes. Academic Press, New York. Schilling, for example 1982. Acceptance Sampling in Quality Control. Marcel Dekker, New York.

[2] Complete verification theoretically assures a complete check of the work in an operation. However, verifying all items can be time consuming and very costly. In many operations, complete verification is only used as the operation is starting up. Once it is shown that the quality is meeting the required standard, sample verification procedures may be implemented. Sample verification reduces the cost and can yield results almost as reliable as 100 per cent verification, if the sample is selected using probability sampling. Acceptance sampling is a quality control technique that establishes a sample design and decision rules to determine which batches are acceptable or unacceptable, and is usually used in jobs like manual editing, coding, and key entry data capture where work is assembled in lots or batches. Each batch is either accepted or rejected on the basis of the verification of a sample chosen from the batch based on probability methods. The sampling plan is designed so as to provide an outgoing error rate below a certain value, called the average outgoing quality limit. This method is applicable to processes which are fairly predictable in terms of their outputs and which consistently produce output that meets the quality standard — the process is then ‘in control’. Census operations where this may be applicable include: the printing of forms; automated data capture via intelligent character recognition (ICR) or optical mark recognition (OMR); and the scanning of forms for ICR/OMR.

survey. Qualitative tests and cognitive interviews should be planned to ensure that questions are clear and properly understood not only by the general population but also by any special groups to whom certain questions are targeted or for whom there are particular issues of concern (for example, the elderly, persons living alone, or those with language difficulties).

53. Web-based questionnaires can provide options not available to their printed counterparts. These options can ensure greater quality in terms of question response and coverage. Checks on such serve as opportunities for detecting inconsistencies and presenting them to respondents for correction or confirmation.

54. The design of electronic questionnaires for data collection via CAPI, CAWI and CATI methods require additional considerations to make the data entry process intuitive for the enumerator or respondent. Some essential functional features that should be used in the design of the electronic questionnaires include:

(a) *Questionnaire navigation* should allow enumerators/respondents to move relatively freely through the questionnaire in order to enter responses in the most effective way, giving the ability to pause and resume at the last answered question with a "save and continue later" functionality. On the other hand, the design should impose some restrictions on navigation, for example, by preventing enumerators/respondents from entering certain questions without having first obtained responses from other, earlier, questions;

(b) *Skipping/automated routing* is one of the most important error reducing features in electronic questionnaires. It obviates responding to questions that should be skipped. It also avoids the converse – skipping questions that should be asked, thus minimizing the need to impute for missing responses. Basic skips allow the response to a particular question to determine whether or not the next question is relevant, while complex skips are those that either use responses from several previous questions to determine whether the next question is relevant;

(c) *Pre-coding* allows relevant questions to be answered from pre-coded drop-down menus. In some cases, drop-down menus could be altered dynamically, depending on previous responses, so that the interviewer is never presented with an impossible response code;

(d) *Validation:* Real-time data validation checks can correct inadmissible or inconsistent responses that could be the result of either interviewer or respondent error, thus reducing the amount of post-enumeration data edits.

55. The testing programme should ensure that these features are thoroughly tested prior to questionnaire finalization.

56. All of these factors should be tested on a small scale (qualitative testing) and then on a large one with a significant number of respondents. A large-scale test can detect a variety of potential issues that qualitative testing cannot. Such tests also make it possible to compare different design and format possibilities via split sample designs. The large-scale test also facilitates assessing how well the questionnaire fits into other census operations (for example collection, data input, coding).

57. The design and presentation of a web-based questionnaire to the respondent will differ from the paper version. Special care must be taken to minimize any potential mode effects arising from differences between the paper and electronic versions of the questionnaire, and between the different modes of compilation of electronic questionnaires (e.g. between CAWI i.e. self-enumeration and CAPI where an interview will be conducted). Hence, this should be an important topic to be considered in the testing programme for the questionnaire.

J. Measuring record-linkage quality

58. One key aspect in the usage of administrative sources for censuses is the linkage process of various data sources. Often, determining the quality of a dataset will require its linkage to another dataset for comparison. Also, if the NSO relies on more than one source of administrative data for its census, it is necessary to be able to link data from the different

sources at the unit/record level. The degree of success of such linkage will affect both the accuracy and the relevance of the input data.

59. A common unique identifier reduces the effort required to link the data by making it easier to evaluate the completeness and accuracy of matching. In the absence of such an identifier, it is more difficult to link data reliably. In this case, record linkage using multiple variables that are common to the units in each data source (typically, name, date of birth, sex, and address) may be possible. In this case, the NSO needs to be assured that such 'matching' variables are of sufficient quality in all sources. Such a process may require deterministic and probabilistic record-linkage strategies, as it is the case in the United Kingdom's census of England and Wales 2021.

60. This process may take place inside the NSO as part of the process stage, or outside the NSO, e.g. within the population register. In the case of record-linkage taking place outside the NSO, a thorough documentation of the processes by the data-providing agency is crucial for quality assessment. Even though the existence and usage of universal identifiers is a strong indicator of quality, the quality of record-linkage performed by the third-party agency should be assessed in the source and data stage through the use of metadata and information provided by the data-providing agency. In addition, a cross-validation of the data provided is possible in the processing stage, if the data structure allows it. If record-linkage is carried out inside the NSO as part of the census, the NSO has much more direct control over the process. The quality of record-linkage can be assessed by statistical analysis, clerical reviews, or additional surveys. These allow for the calculation of quality measures such as false positives/negatives and recall value (see Quality measures for record-linkage). However, regardless of the stage, a detailed and transparent measurement and reporting of record-linkage processes' quality is important in order to ensure the quality of the final product, a detailed and transparent documentation is vital to quality assessment and potential future improvements.

Quality measures for record-linkage

In the case of the register-based population census test in Germany, a 12 percent sample of the German population was used to compare the results of the original 2022 census survey with that of a probabilistic record-linkage model.

The census test is carried out under a legal framework, the "Registerzensuserprobungsgesetz" (RegZensErpG). In this process, the record-linkage aims at matching the population register to records from ten administrative sources using a probabilistic record-linkage strategy. The use of survey data from the 2022 census as external reference data allows for the calculation of quality measures for each administrative source and for the census database as a whole. The record-linkage process is complemented by clerical reviews of potential matches and statistical analysis in order to optimize and validate the record-linkage model.

The confusion matrix is the baseline for the calculation of quality measures in record-linkage using external reference data.

The basic classification problem in record-linkage is the match of records present in different sources and referred to the same person. In this perspective, the record linkage process aims at matching units already present into the census dataset (e.g. in the population register) with units in different administrative sources, and at associating them with the same code or id.

In a simplified form, there exists a classification decision for every unit/record representing a person in the census database. Either a unit has a match with another record in a different source, or not. The reference data, e.g. a survey or another administrative source, serves as the external benchmark. A unit can exist in the reference data, or not. In that, through the cross-tabulation between the variable existence/non-existence of the unit in the reference data and the variable existence (match)/non-existence (non-match) of the classification model, it is possible to categorize the data under four cases. From these four

cases, the quality measures precision, recall, and accuracy can be derived (see table 2).

Table 1
Confusion matrix in record-linkage

		Classification Model	
		Match (Predicted positive)	Non-Match (Predicted negative)
Actual population/ Reference data	Existence (Positive)	True Positives (TP)	False Negatives (FN)
	Non-Existence (Negative)	False Positives (FP)	True Negatives (TN)

Table 2
Quality measures for record-linkage

Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
Accuracy	$\frac{TP + TN}{TP + TN + FN + FP}$

K. Measuring imputation quality in register-based censuses

61. The need for data editing and imputation does not decrease when administrative sources are used, as the existence of multiple data sources can be in itself a source of inconsistencies (when several register sources are used simultaneously to define, for each statistical unit, the value of the relevant variable, it becomes necessary to include rules for prioritization between different sources in the event of contradictory data). Under-coverage of data sources, data linkage problems and the use of sample data can lead to missing values that might require specific solutions. Improper imputation methods can lead to erroneous results. Therefore, an increasing need arises to assess the quality of data editing and imputation procedures.

62. An important aspect of quality assurance for register-based censuses is imputation of missing values. This process is especially relevant for census variables that are taken from a non-integral source (i.e. a source that does not cover the entire population). In that case, mathematical methods are often used to make inferences for the population, by using deterministic and probabilistic imputation and weighting.

63. For a sound interpretation of results, it is important to disseminate imputation rates for each cross-tabulation in the publication. Results for certain cross-tabulations that are insufficiently covered by any data source might not be published at all. For instance, Statistics Netherlands uses Labour Force Surveys as a source for the census variable Occupation. Any cell value that is covered by less than 25 observations is not published, due to a lack of confidence.

64. If imputation or weighting is applied, the accuracy of these methods should be evaluated. Several methods are available. The distribution of a variable with and without imputations can be compared, comparisons can be made with other sources and mathematical methods can be applied. The latter includes: 1) analytical expressions for estimating confidence intervals, 2) resampling techniques like Bootstrapping and Jacknife and 3) cross-validation methods that apply estimation methods to observed records to assess estimation error.

65. The assessment of imputation quality is not limited to the accuracy of individual imputations, but especially for a census it is also relevant to consider implications for all possible cross-tabulations, e.g. the number of 53 year-old, females with a certain educational level. In this respect, Chambers (2006) defines four criteria to evaluate imputation accuracy: 1) predictive accuracy 2) distributional accuracy; 3) estimation accuracy and 4) imputation plausibility[3].

Use of imputation for missing values in a register: The imputation of educational attainment in the Dutch register-based census

In the Dutch register-based census, the variable 'educational attainment' is derived from the so-called Educational Attainment File (EAF).

The coverage of this data sources is selective as the data cover approximately two thirds of the population; while younger persons are integrally observed, older persons are covered on a sample basis. All missing educational levels are estimated at micro level. This is done by a so-called multinomial logistic regression model, which takes the selectivity of data collection into account. The model has been exclusively designed for the census. This means that the imputed educational levels are not used for any other purpose.

A first step in the validation consists of a logical check of the estimated model. The estimated regression coefficients should not be extremely small or high, as this might point out to technical difficulties in estimating the model. Moreover, the estimated regression coefficient should conform to logical expectations, e.g. a positive correlation between income and educational level.

The final validation consists of a thorough comparison of the results with other educational statistics. Comparisons are not only made for the Dutch population as a whole, but also for detailed subpopulations, e.g. the distribution of the educational levels by geographic area, sex and age. Any significant deviations from other publications should be explainable.[4]

L. Management of coverage error

66. Coverage is a critical element of accuracy. It has a direct influence on the quality of population counts and an indirect impact on the quality of all other data produced by the census. Thus the coverage concerns should be taken into consideration in the design and implementation of most census activities. Enumeration area boundaries should be carefully defined and mapped to ensure that no area is omitted or included twice. Instructions and training on dwelling coverage for staff engaged in dwelling listing and enumeration should be clear, explicit and easy to understand. The target population must be well defined and related instructions and questions for both interviewers and respondents need to be carefully developed and thoroughly tested.

67. Clarity and simplicity of instructions concerning place of residence for enumeration is vital to help ensure people are enumerated exactly once and at the correct location. This is particularly important in minimizing over-coverage. Questionnaires should include guidance or questions to assist with situations where it may be unclear whether certain persons should be included or not. Special procedures should be developed for difficult to enumerate population groups (for example remote areas, collectives or group quarters, persons with literacy or language difficulties). Processing procedures should be developed with a view to

[3] For further details, see Ray Chambers (2006): "Evaluation Criteria for Editing and Imputation in EUREDIT". In: Statistical Data Editing. Vol. 3. United Nations Statistical Commission, 2006, pp. 17–27, https://unece.org/DAM/stats/publications/editing/SDE3.pdf.

[4] For further detail see: Statistics Netherlands (2017), Mass imputation for Census estimation, https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2017/Meeting-Geneva-Oct/GE_41_2017_3_ENG.pdf and Jacco Daalmans (2017), Mass imputation for Census estimation, Statistics Netherlands Discussion paper 2017/4, https://www.cbs.nl/-/media/_pdf/2017/11/mass-imputation-census-2017.pdf.

minimizing the risk of erroneously cancelling, losing or artificially creating households. A well-crafted publicity campaign can play an important role in promoting census awareness and response, thus helping minimizing coverage error.

68. All of these steps, along with appropriate training, supervisory checks and quality control during operations will help minimize coverage error. Nonetheless some coverage error is unavoidable. Hence it is important to measure, analyse and report on coverage error. This is best done via an independent post-enumeration survey of a sample of census areas or via a Reverse Record Check methodology. Results of coverage studies provide an important evaluation of the current census and can also provide valuable guidance for the next census. Results in conjunction with the census counts themselves are a critical input for population estimation programmes. Analysis of census results vis-à-vis demographic projections of the population from the previous census can also be informative.

69. In the case of a register-based census (or of the register-based component of a combined census), the issue of over-coverage of registers is frequently considered. If, for example, people have not officially declared their emigration, their records may be kept unchanged in administrative registers, and consequently be the cause of over-coverage in the census. In Slovenia, a special survey with a focus on plausible unregistered emigration was conducted in 2016[5]. On the other hand, when the unregistered population is deemed significant, the under-coverage of registers is also assessed. In the Netherlands, for the 2010 census round, the difference between the registered population and the usual resident population was measured by estimating the under-coverage of the population register through a capture-recapture methodology (applied to the Population Register linked to the Employment Register and the Crime Suspects Register)[6].

70. Another possible means of measuring coverage errors in register-based censuses is to create a residency index on the basis of the records held in multiple registers in order to determine a so-called ‘signs of life’ score. The approach is to define for all possible residents the ‘sign of life’ as a binary score (with a value 0 or 1) for each record in each register. Using these signs of life as explanatory variables, it is possible to build a model to estimate the size of under- and over-coverage.

M. Systems development

71. Systems development is a cross-cutting topic which can have a major impact on quality. In particular, the related dimensions of quality are accuracy, timeliness and accessibility. A modern census makes use of numerous automated systems to operate, manage and control a range of activities from the field staff payroll to data capture, edit and imputation, coding, dissemination and others. This pervasive influence makes it very important that an integrated view be taken in the design of the overall architecture as well as the individual design and implementation of systems.

72. A standard methodology for systems development should be implemented and should include steps such as:

(a) Overall system architecture design;

(b) Design and analysis of individual systems;

(c) Programming or building of systems;

(d) Functional testing of components and then of systems;

(e) Testing of interfaces between systems;

(f) Volume testing and user acceptance testing;

[5] The sample frame consisted of persons for whom labour force status had to be imputed in the 2015 Census (15,500 persons); the second sample group consisted of persons marked in the administrative data (CPR) as non-residents, but where data on their labour force status in Slovenia could be found in at least one out of the nine sources (2,700 persons).

[6] For details see: https://webmail.istat.it/service/home/~/?auth=co&loc=it&id=496989&part=2.

(g) System delivery and implementation; and

(h) Evaluation.

73. This should be done within a configuration management approach to:

(a) Manage change;

(b) Accommodate the re-use of standards and best practices;

(c) Ensure that all requirements remain clear and valid;

(d) Communicate each of these to developers and users promptly and precisely; and

(e) Ensure that results conform to requirements.

74. Specifications should be well written and carefully analysed to produce functional requirements. A standardized approach for change management is required. Ensuring the interoperability of different systems that can communicate with each other is particularly important. At each stage, the performance should be evaluated, and outputs should be checked to conform to requirements. Many of the systems developed for a census will be used by numerous key entry, coding, editing and other clerical staff. Consequently it is very important that user interfaces be carefully designed and thoroughly tested. More generally, a well-developed standardized testing strategy should be applied throughout in an integrated fashion.

75. The transition from paper to digital data collection introduces a new operation: that of developing a data capture system. Specific quality concerns related to the development of Computer Assisted Personal Interviewing (CAPI) applications and Computer Assisted Web Interviewing (CAWI) applications have been discussed in the section on Questionnaire design.

76. Consideration should be given to the development of an operational control system to handle tasks that are not directly related to the interview. This will include logging in to the system, the ability to assign and receive assignments, complete and review questionnaires, track progress, and the ability to synchronize data. Once an enumerator is logged into the operational control system and has selected an assignment the questionnaire can be prefilled with data such as the enumerator and geographical codes which will reduce human error. Additionally, it can be helpful to prefill paradata such as device id, interview start time, and software versions.

N. Methods of census evaluation

77. Given the importance of the accuracy of census statistics, and the budgets involved, it is common for census programmes to undertake extensive evaluation of the census process. Census coverage and content evaluation are a particular focus of such exercises, but they often also cover wider issues such as the effectiveness of publicity campaigns, project management or contract management.

1. Coverage evaluation

78. As noted in the chapter on Quality management, in some countries the evaluation of census coverage forms an integral part of the census quality assurance and improvement component of the quality management framework, and feeds directly into the published census results. In other countries, it is a separate exercise, with the results used to adjust the published census results when producing subsequent population estimates.

79. Both gross and net error must be taken into account in developing the overall evaluation plan. Gross coverage error in a census is defined as the total of all persons omitted, duplicated, or erroneously enumerated. Net coverage error takes into account the underestimates due to omissions and the overestimates due to duplications and erroneous inclusions. When omissions exceed the sum of duplications and erroneous inclusions, a net

undercount is said to exist; otherwise, a net over-count results. Similarly, both gross and net content errors have to be considered in the evaluation design.

80. The choice of coverage evaluation methods to be used depends upon the evaluation objectives, and the method of undertaking the census (that is, whether based on field work or population registers). Common methods (that can be applied to both field and register based censuses) include:

(a) Post-enumeration surveys to measure the accuracy of the census (by independently surveying a sample of the population, the post-enumeration survey estimates the proportion of people and housing units potentially missed or counted erroneously in the census, usually within a Dual System Estimation model or so-called capture-recapture method, with the PES being the second 'capture' while the census itself is the first 'capture');

(b) Comparisons of results with other data sources including previous censuses, current household surveys, and/or administrative records; and

(c) Demographic analysis; and

(d) Ethnographic and social network methods to study the effects of mobility on census coverage or to measure census coverage of specific sub-populations.

81. For register-based censuses, it is also possible to compare data from a past traditional census with register data from the same time. Data linkage at the individual level can enable estimation of both under and over coverage and longitudinal databases would enable such estimates to be carried forward.

2. Content evaluation

82. Common methods for content evaluation include:

(a) Post-census surveys designed to measure error in specific census questions (see a case-study below);

(b) Record-checking, in which individual census records are matched against alternative sources and specific data items are checked for accuracy;

(c) Quality control techniques such as internal consistency checks;

(d) Surveys to determine customer satisfaction with data collection instruments or questionnaire assistance; and

(e) Focus group interviews to learn how or why respondents behave in a certain way.

The use of a post-enumeration survey to assess content error (UK Census Quality Survey)

The Census Quality Survey (CQS) used in the UK is a voluntary post-enumeration re-interview survey which asks each of the census questions again. It is a much smaller operation than census collection, using trained interviewers.

The UK ran a CQS after their 2021 Census. Invitation letters were sent to 110,000 households asking them to contact headquarters to arrange an interview. The achieved number of interviews was 8,700 households containing 16,000 residents: a household response rate of 7.9%, sufficient for analysis.

The achieved sample differed demographically from the known population, and so weighting was used to produce more representative results. For individuals, this was based on age, sex, ethnic group, whether the respondent lived in England or Wales, and whether the individual's response was given by proxy. For household-level questions, weighting was based on the demographics of residents in the household, plus the household's accommodation type and whether the household responded to the census online or on paper.

An agreement rate and associated confidence interval were calculated for each question.

Where people gave different answers, this might have been a mistake in completing the census or a mistake answering the survey. This means the CQS does not attempt to give an exact measure of 'respondent error' in the census, but does indicate which questions might be more subject to that error, and the possible scale of such error.

Trained interviewers carried out the CQS, and took extra measures to help people interpret the questions correctly: explaining categories and definitions more clearly than they may have noticed when they filled in their own census form.

The CQS results were not used to amend the final data, but were published to help users understand likely quality issues.

The highest agreement rate overall in 2021 was for sex, at 99.3%. The lowest was for national identity, at 59.2%. Agreement rates tend to be lower for questions that are subjective or harder to evaluate: where the answer changes over time, and where there are more, or finer-grained, options[7].

Examples of a household and an individual question:

Landlord

The household section of the Census 2021 questionnaire asked people who did not own their own home who their landlord was, choosing from one of six options.

The weighted agreement rate for this question was 88.0%.

The most common disagreement was where respondents stated that they rent from a "council/local authority" on one survey, but on the other survey they say they rent from a "housing association, housing co-operative, charitable trust, registered social landlord".

The disagreements may simply reflect uncertainty of the respondents, as social housing could be allocated by applying to the local council, and this may have been clarified by the interviewer.

The issue of respondent error on this question in the 2021 Census is noted in the Housing quality information for Census 2021 methodology[8].

National Identity

Respondents were asked whether they identified as British, English, Welsh, Scottish, Northern Irish, or any other national identity. Respondents could select as many options as they wanted.

The agreement rate for this question was 59.2%, the lowest of any question on the survey. This was driven almost entirely by respondents disagreeing on whether they identify as British, English, or both British and English. Three-quarters of disagreements in the CQS data are among and between these three categories.

Agreement rates vastly improve to 93.1% when generalising to just three groups – "any individual or combination of UK national identities", "any UK plus 'other'" and "(just) other". In this case, most disagreement occurs when

[7] See https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/censusqualitysurveyforcensus2021inenglandandwales#overview.

[8] See https://www.ons.gov.uk/peoplepopulationandcommunity/housing/methodologies/housingqualityinformationforcensus2021.

people identified with just UK national identities on the census, but UK plus another nationality on the CQS. This may be the result of CQS interviewers clarifying the question and probing for further categories of response.

3. Designing an evaluation programme

83. The following basic recommendations can be applied to any evaluation programme:

(a) Begin planning the evaluation programme early in the census cycle. Early planning and design of a structured evaluation programme allows appropriate consideration and accommodation of evaluation and experiment needs during the census design;

(b) Decide the high-level scope and focus of research programmes before developing research proposals. Define general selection guidelines or criteria, select research topics, and identify high-level research questions before designing the evaluations and experiments. Identify areas to meet the needs of external data users and internal census planners and set evaluation priorities accordingly;

(c) Develop study plans for each evaluation and experiment. These project-level plans then become the designated baseline documentation for achieving programme research goals;

(d) Develop a standardized Change Control Plan, which describes a protocol to initiate a change process. Recommendations for change (including the reasons for the change and critical implications) are submitted to a Change Control Board which then assesses implications of the change and approves or disapproves it;

(e) Develop a milestone schedule for planning, designing, and implementing the research programme. Include in the milestone schedule dates for issuing results of the operational assessments, evaluations, and experiments. Changes to the schedule should also go through the Change Control process;

(f) Anticipate delays or the need to cancel some planned evaluations. During a census, staff may become overburdened with either too much evaluation work or too much of a combination of evaluation and production work. Attrition of project managers is often inevitable and can also be a reason to delay or cancel evaluations;

(g) Explore ways to incorporate real-time evaluations during the course of the census.

III. Conclusion

84. The draft annex on Quality management programme implementation for the 2030 round of population and housing censuses is presented for comments and discussion.

85. This proposal should be read in conjunction with document ECE/CES/GE.41/2024/15 containing the draft chapter on Quality management.