

Challenges in assessing and assuring the quality of new data sources for population and housing census 2025

Dhafer Al-Shawawreh/Department of Statistics Jordan/Dhafer.AlShawawreh@dos.gov.jo

Abstract

The General Population and Housing Census 2025 is the most important statistical product that serves the public and private sectors in precisely knowing the current situation and planning for the future. The General Population and Housing Census 2025 is designed to be a hybrid census which inputs consist of administrative records and data collected from households and establishments.

Using administrative records as input will provide many benefits, such as reducing time, effort, and cost, more accuracy, and reliability, but at the same time it presents many challenges for the quality check of the census, which makes it necessary to ensure the quality of the administrative record before using it as input in the population census.

The stages of the quality check of the administrative record are divided into four stages: The source stage (understanding and evaluating the source), The data stage (receiving the actual data and evaluating its quality), The processing stage (processing administrative data and using it in the census), and The output stage (assessing the quality of the census that uses administrative data). Both of the source stage and the data stage represent the quality of administrative record inputs

Before accepting them as input to the census.

we applied the Statistical Code of Practice for the European (ENP-south) to evaluate the quality of each stage separately for the new administrative record. Many challenges appeared in evaluating the source stage, the data stage.

In this paper, we will discuss the general plan for examining the quality of the record as a good input for the census then revealing the problems and challenges that faced the quality team, measuring the problem, and trying to solve it in the processing stage to accept it or merge it with a new record from another source so that it has a value commensurate with the conditions of the quality of the census inputs, or to accept a part of it as inputs for the census or for the purposes of confirming the quality of the census or rejecting the record. Each stage applied part of the Statistical Code of Practice for the European (ENP-south) assessment of the quality of inputs to the administrative register.

Keywords

administrative record, quality stages, census, accept new data, examining the quality

1. Introduction

The population census is considered as the most important statistical product, as it serves all statistical studies in different sectors, policy-making, planning & development, decision-making, and provides data and characteristics for all inhabitants, households, and houses.

Data quality as defined by Eurostat is a set of attributes and characteristics that exist in the statistical product or service to meet the user's need and user's satisfaction.

By applying the principles of Statistical Practices' Code and the guidelines for assessing the quality of administrative resources to be used in censuses produced by the United Nations Economic Commission for Europe, and by applying the principles of comprehensive quality management in understanding and managing the quality of sources, processes, and products, and applying them to all stages of the general population and housing census; thus, the quality of data in statistical work is achieved by achieving The following important elements (relevance to reality, confidentiality, accurate time and timeliness, easy accessibility, comparability for consistency and completeness).

In order to complete a hybrid population and housing census, which relies partly on administrative records in collecting data and ensuring quality, it is necessary to adhere to the Guide of Quality Assurance Standards and Procedures for Evaluating Administrative Records Data, which assist in establishing an integrated system of procedures to control the quality of administrative records data for the Department of Statistics with government entities.

The aim of this paper is to examine the administrative record at the source stage (understanding and evaluating the source), the data stage (receiving the actual data and evaluating its quality), and measuring it as good valid data to be part of the census data, and identifying the problems and challenges that faced the quality team and trying to solve them; by accepting it or merging it with a new record from another source that has a value that is consistent with the quality conditions of the census inputs, or accepting part of it as inputs to the census, or for the purposes of confirming the quality of the census or rejecting the record.

2. Methodology

This paper relies on applying the dimensions of administrative data quality that identified by the guidelines for assessing the quality of administrative sources for use in censuses issued by the United Nations Economic Commission for Europe to the source stage and the data stage, measuring the extent of their effectiveness in accepting and integrating administrative sources into the 2025 Population and Housing Census, and noting the

challenges facing the quality team during the application of the quality of administrative records to evaluate the quality of the record at the source stage and the data stage.

3. Study Questions

The study question focuses on the challenges facing the quality team when applying the dimensions of administrative data quality to the source stage and the data stage of the administrative record, which is stemmed from the guidelines for assessing the quality of administrative sources for use in censuses issued by the United Nations Economic Commission for Europe.

4. Population Census and Quality Procedures

In 2015, the Department of Statistics implemented the E-population census and relied on tablet devices to collect data at that time.

The role of quality was parallel to the census from the beginning of the planning stage to the end of the evaluation stage according to GSPBM procedures. In the current hybrid census, the quality team also performs a parallel procedure according to the GSPBM model, with the presence of a new and additional challenge which is the quality of administrative records.

4.1. Census Methodology

The Department of Statistics works to collect administrative records from strategic partners, and performs complex operations in all stages of the administrative record to be accepted and approved as part of the census data. Then, the census teams work through Field inventory by linking the address record to the national number of the head of the household or the national number of the establishment that is distinguished to the administrative record so that the data is transferred to the tablet device of the numerator before the enumeration stage, and then the numerator collects the remaining data from households to reduce the burden on the respondent and the numerator in the population enumeration stage. The administrative data will be used to confirm the quality of operations and ensure the quality of the enumeration stage. DOS uses the administrative data to reach preliminary estimates and to serve as a roadmap during the census operations.

4.2. Types of administrative records used in the 2025 Census:

- Records of the Civil Status and Passports Department
- Social Security corporation records
- Records of the Companies Control Department
- Amman Municipality records
- Records of the Ministry of Local Administration
- Ministry of Education records
- Records of the Ministry of Higher Education
- Records of the Ministry of Industry and Trade
- Health insurance records

4.3. Quality Assurance Standards

Includes the following main standards:

The first main standard: Managing institutional arrangements that support quality: includes the following sub-standards:

- ❖ Organizational structure and institutionalization of procedures.
- ❖ Providing resources and enhancing statistical capabilities and resources.

The second main standard: Managing the quality of administrative records data:

- ❖ Clarity and soundness of methodologies.
- ❖ Suitability
- ❖ Accuracy, reliability and comprehensiveness
- ❖ Consistency and comparability
- ❖ Availability and accessibility
- ❖ Modernity and temporality

The third main standard: Self-evaluation for the statistical quality of administrative records data:

Self-assessment of the statistical quality of DoS's registered data allows monitoring the extent of compliance in applying quality standards and procedures, identifying the strengths and opportunities for improvement, and developing a plan to enhance the quality level of administrative records data, and it is preferable to carry out a self-evaluation process for quality on a regular and organized basis.

4.4. Quality team documents and indicators for the three main standards

Quality team reviews the administrative records directorate and verifies documents and indicators for the three main standards as follows:

Documents and indicators of the first main standard: Managing institutional arrangements that support quality

- ❖ A document of the tasks and duties of the Records Directorate concerned with statistics and records data management.
- ❖ A document of providing the department with the required data.
- ❖ A document that includes a list of data available with the data provider.
- ❖ A document that includes a list of data that can be provided to the department of statistics.
- ❖ Data confidentiality classification document from the data provider.
- ❖ A document of the number of employees working in the records concerned with data management.
- ❖ Document the number of employees who received training in the field of data records management.
- ❖ A document of the average number of training hours in courses or programs related to capacity building in the field of data records management.
- ❖ Document on computer programs and electronic systems for managing, exchanging and linking administrative records data.

Documents and indicators of the second main standard: Managing the quality of administrative records data

- ❖ A document containing a methodology for managing records data.

- ❖ The percentage of descriptive information received from the administrative databases by the entity providing the administrative records.
- ❖ A list of data users and the nature of their use of record data from the entity providing administrative records.
- ❖ Level/percentage of satisfaction with the data and statistics produced by the entity.
- ❖ A methodological document and procedures for auditing and examining the entity's records.
- ❖ The percentage of errors in the data set by type and cases.
- ❖ Reports of the results of auditing the quality of data log, including errors and treatment methods.
- ❖ Document for procedures for exchanging data log.
- ❖ The percentage of achieving electronic connectivity.
- ❖ An agreed timetable showing the dates the data will be available.
- ❖ The percentage of data provided to the department from the list of data required by the entity.
- ❖ The percentage of data that is provided according to the schedule.

Documents and indicators of the third main standard: Self-assessment of the statistical quality of administrative records data

- ❖ Prepare a checklist of all the procedures, documents, and indicators mentioned above, with the possibility of adding any other procedures that were not mentioned in the guide, provided that they are appropriate to the standards contained in the guide.
- ❖ Determine the level of application of all the mentioned procedures, with an explanation of the reason for not implementing any of them.
- ❖ Preparing a report explaining the most prominent quality procedures applied and improvement proposals to improve the quality of the department's registered data.
- ❖ Preparing a time plan to implement the quality procedures approved in the department and following up on the implementation of this plan.
- ❖ Repeat the quality assessment process and measure its indicators periodically in order to follow up on the progress made in improving and developing the quality of registry data.

4.5. Quality Stages of Administrative Record

The quality of the administrative record has been divided into several stages to facilitate and organize dealing with the administrative record in a logical, easy and clear manner.

- **Source Stage**

in this stage, the quality team must define the main dimensions of this stage and determine the evaluation process and the tools and indicators associated with it to evaluate the quality of the administrative data source. Based on the evaluation of this stage, the administrative source is accepted or rejected by the census, and the dimensions of source quality are the following:

1. **Appropriateness and accuracy.** Appropriateness refers to the extent to which the administrative data source meets the needs of the census. Accuracy refers to the degree to which the data correctly describes the phenomenon it was designed to measure, and determines the level of coverage errors to know the degree of compatibility of the target units with the census units. Here, it is necessary to know what the unit is in the source, laws and regulations that determine the data unit in the administrative source, making a comparison between the unit in the administrative source and the needs of the census, and knowing the tests that the data providers performs to ensure the validity of these units and whether they cover the target units. Thus, it is necessary to know the missing coverage and excess coverage and know the other administrative sources to cover the deficiency, and to determine the level of measurement errors to identify the availability of the characteristics and variables targeted for the census and whether the concepts, definitions and classifications of variables match the concepts and needs of the census, knowing the errors of delay in filling out the variables in the administrative source, such as delaying the registration of births and deaths in the civil registry, and knowing the quality of collection of variables so that the accuracy of the administrative source variables depends on incentives. The laws, goals, importance of the variable, and the absence of any biases also depend on the review and correction tools used by the administrative source.

2. **Institutional environment:** It is described and calculated by calculating the index of the strength of the relationship with the source of administrative data, calculating the

index of previous relationships and experiences with the source of administrative data, calculating the index of official agreements, and calculating the index of the status of the data supplier and its commitment to quality standards. When applying this through our work, the strength of the relationship with the data source was affected by a change in administration or policy, or by a change in the data structure of the administrative source.

3. Timing: The timing must be available at the appropriate time for the census, and this is verified through the indicator of the time difference between the end of the reference period for administrative data and the date of providing the Department of Statistics, the indicator of the number of times administrative data is made available for the census, knowledge of the requirements for making data available, and the extent of its temporal impact on data delivery, and an indicator Processing time within the Department of Statistics after receiving the data.

4. Consistency and comparability: to achieve this characteristic, there must be the possibility of integrating the administrative source with other sources and with census data. This requires verifying the compatibility of definitions and data characteristics, verifying the possibility of comparison across time, field, and geographic regions, and verifying the indicator of the possibility of linking between records through the presence of a unique identifier in all of them. Records to be linked to the census.

5. Accessibility and interpretation: It is a study of the degree of ease of access to administrative data and its interpretation, and it is calculated through the index of calculating the restrictions imposed on accessing and using the data at the source, calculating the index of public opinion acceptance of the use of administrative data, calculating the index of ease of transferring data, and calculating the index of the degree of clarity of metadata.

- **Data Stage**

In this stage, the quality of administrative data is evaluated based on the outputs of the source stage and the extent of conformity with census requirements. The result of evaluating the quality of the data stage is either re-requesting data from the source according to certain characteristics, linking it to other sources, or recommending the treatments that will be carried out in the subsequent stage (the processing stage) or dimensions. Data quality at this stage is as follows:

1. Alignment and validation of data: in this dimension, data transfer processes are verified, the data provided is verified against the metadata, the reference period is verified, and then data formatting and recoding operations are carried out, such as encoding missing values, removing duplicate records, and performing many data filtering operations.
2. Accuracy and reliability: In this dimension, representation errors are checked, over-coverage and under-coverage are checked, to ensure the percentage of coverage of statistical units, the number of statistical units in the register, and the percentage of duplicate statistical units. Measurement errors are also studied, and the percentage of illogical or missing values is checked and verified. The reasons that led to this include transmission errors, failure to update data, or changes in the data structure at the supplier.
3. Timing and adherence to deadlines: The Department of Statistics tries to keep the difference between the reference date indicated by the data and the date of submitting the data to the Department of Statistics at its minimum to meet the needs of the census in the optimal way.
4. Link-ability: The quality of the administrative record requires the ability to link it to other records at the record level through a shared unique identifier and knowing the percentage of records that have a unique identifier.

4.6. Method of Index Calculating

Each dimension of quality consists of one or more indicators that can be measured through the data that is made available for DOS to reflect the level of achievement of the quality measure, each indicator takes three values: it takes the value (0) if it is not verified or achieved slightly, it takes the value (0.5) if the indicator is partially achieved, and it takes the value (1) if it is fully verified and verified.

Then, the value is multiplied by the weight of the indicator, and if the indicator does not apply, the weight and degree of the indicator are distributed among the rest of the indicators. In the end, the values of all indicators are collected to reach a result, and based on this result, a decision is made in the administrative record.

Throughout the application, it was difficult to apply indicators and weights for quality dimensions to all administrative sources in an equal manner, because each administrative source has its own specificity, importance, controls, and uses that are different from the other.

4.7. Objectives of the quality of administrative records

The quality team aims to ensure the continuity and continuity of the flow of registered data, the adequacy, comprehensiveness and up-to-datedness of the registered data for preparing official statistics, the integrity of the data collection systems and their suitability for preparing the required statistics, the consistency of the registered data with statistical definitions and classifications in accordance with international guidelines, and the existence of mechanisms for auditing and processing the registered data that are consistent with practices. Sound, internationally recognized, periodic evaluation and continuous review of administrative records for statistical uses.

5. Implementing the quality team

The quality team developed a final assessment of the input phase of the administrative record. This is done by collecting evaluations of quality dimensions indicators that were completed in the source phase and the data phase to reach a final decision regarding the administrative record and its suitability for census uses. We sometimes face difficulty in convincing the administration if the record is not accepted as an input for the census. Especially since there is a painstaking and expensive effort in the agreements, processes and study of the administrative record. We also faced challenges in convincing the administration to provide cadres, budget and training to address the lack of data, audit and process it and deal with tools for linking with other administrative records. We also faced difficulty and delay in measuring link ability because it requires an administrative source. Another, other agreements, and a long series of works are usually not available at the same time to the first administrative source, and from the reality of application we found that the incompleteness of the descriptive data from the administrative source may affect all dimensions of quality because the descriptive data is what shows the administrative record and its variables and all the restrictions and the importance of the variables and their mechanism. Collecting and processing it, and the mechanism and time of updating the record data and controlling its quality at the source.

6. Challenges

When applying the guidelines for evaluating administrative sources for use in censuses issued by the United Nations Economic Commission for Europe, the quality team faces many challenges, including:

- 1- There are no fixed and unified indicators and unified weights for quality dimensions that are appropriate for all administrative sources. Therefore, you need experience and training in dealing with the administrative source.
- 2- Commitment to quality at the source stage is affected by changing officials, policies, and data structure
- 3- It takes a long time to obtain, understand and evaluate administrative data sources, especially when there is a need for a work program to develop records for use in the census.
- 4- Administrations accept that elements, definitions, concepts, time reference periods, coverage and measurement errors within an administrative source may not be consistent with the objectives of the census.
- 5- Incomplete data and linking processes to other administrative records.
- 6- A lot of time and effort is wasted and sometimes fails if administrative resources are not available at the same time, which can be linked to build a good administrative record.

7. Results and recommendations:

The paper makes the following recommendations for consideration in preparing the next updated version of the Guidelines to resolve the challenges faced by the Quality Team during the implementation of the UNECE Guidelines for Evaluating Administrative Sources for Use in Censuses

- 1- Reformulating the evaluation of indicators and weights of quality dimensions in a more precise manner, Clarifying the qualifications and training courses, and the preparation and role of workers on the quality of the administrative record.
- 2- Develop a clear model that contains the basic conditions for any agreement to overcome the challenges of changing officials, policies, and structuring data at the source.
- 3- For the purposes of planning, organizing, and optimal use of resources, a time-related process flow chart may be required.
- 4- Clarifying the risks of each dimension of quality if it is not achieved ideally to include it in the quality report for management so that the report is based on high reliability.
- 5- Include the names and sources of ready-made software that helps complete data and link administrative records.
- 6- Clarifying the work on the administrative sources that make up the administrative record in parallel to reduce time and effort.

Conclusion

In this paper, an introduction is provided on the challenges of evaluating and ensuring the quality of new administrative data sources for the population and housing census using the methodology of the Guidelines for Assessing the Quality of Administrative Sources for Use in Censuses issued by the United Nations Economic Commission for Europe.

The question of the study was: What are the challenges facing this methodology when applied in practice on the Jordanian General Population and Housing Census 2025. The paper showed the census methodology and the administrative records used in the census. The paper included quality standards, documents and indicators that the quality team in the DOS are working on. The quality stages of the administrative record have been clarified and the quality dimensions of the source stage and the data stage have been studied.

The mechanism for calculating quality indicators, the goals of the quality of administrative records, and the mechanism for applying the quality indicators on the administrative record to come up with a decision on the quality of the administrative record and the challenges that the team faced while measuring the quality dimensions that were explained. This paper revealed several recommendations to facilitate the work of the quality methodology used in accepting the administrative record as an input for the general census of population and housing 2025.

References

United Nations Geneva, 2021. Guidelines for Assessing the Quality of Administrative Sources for Use in Censuses. https://unece.org/sites/default/files/2021-10/ECECESSTAT20214_WEB.pdf

UK Statistics Authority, 2013 • Quality Assurance and Audit Arrangements for Administrative Data. https://osr.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-ados_tcm97-43970.pdf

STATISTICS CODE OF PRACTICE FOR THE EUROPEAN NEIGHBOURHOOD SOUTH COUNTRIES.

<https://ec.europa.eu/eurostat/documents/4472553/6753624/Statistics+Code+of+Practice+for+the+European+Neighbourhood+South+countries+%28Based+on+the+ESCoP%29/bd1d86c0-445d-4aca-93e9-990ff6294a93>

Cerroni, F., G. Di Bella and L. Galiè (2014). Evaluating administrative data quality as input of the statistical production process.

https://unece.org/sites/default/files/2021-10/ECECESSTAT20214_WEB.pdf

Daas P., S. Ossen, R. Vis-Visschers and J. Arends- Tóth (2009). Checklist for the Quality evaluation of Administrative Data Sources. The Hague: Statistics Netherlands.

<https://ec.europa.eu/eurostat/documents/64157/4374310/45-Checklist-quality-evaluation-administrative-data-sources-2009.pdf/24ffb3dd-5509-4f7e-9683-4477be82ee60>

Eurostat BLUE-ETS (2011). List of quality groups and indicators identified for administrative data sources.

https://mtennekes.github.io/downloads/publications/BLUE-ETS_WP4_Del1.pdf