



**United  
Nations**

Department of  
Economic and  
Social Affairs

# Quality assessment of administrative data sources

**Meryem Demirci**  
UN Statistics Division



**DESA** | Statistics Division



## Quality assessment

- ❑ **An integral part of the census operation** regardless of types of census methodologies
- ❑ **Overarching process** covering all phases of censuses-quality of one phase has an impact on the quality of next phase
  - ✓ Quality assessment process is designed differently for censuses which are conducted with administrative data sources compared to the traditional census





## What we will learn

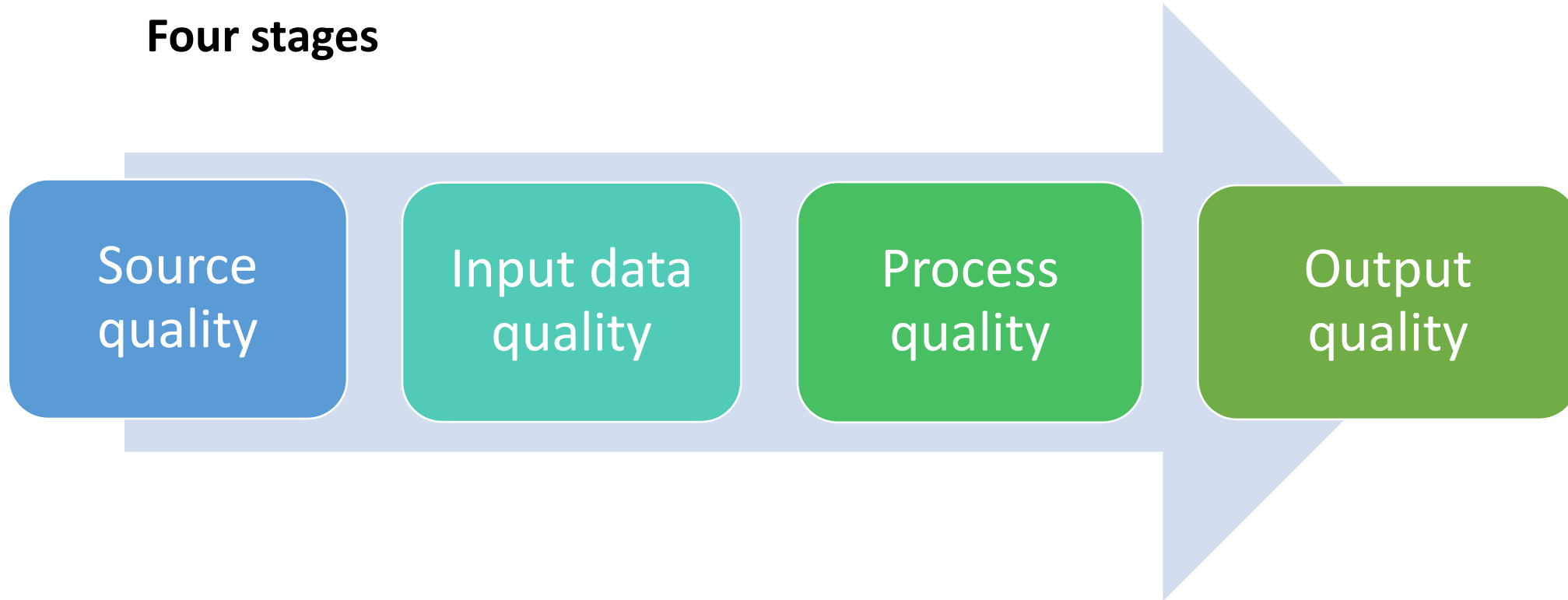
- How to design quality assessment process ?
- ✓ **Stages** of the quality assessment for administrative data sources
- What to check in each stage ?
- ✓ **Dimensions** (components) of the quality assessment
- How to measure quality
- ✓ **Indicators** for measuring quality





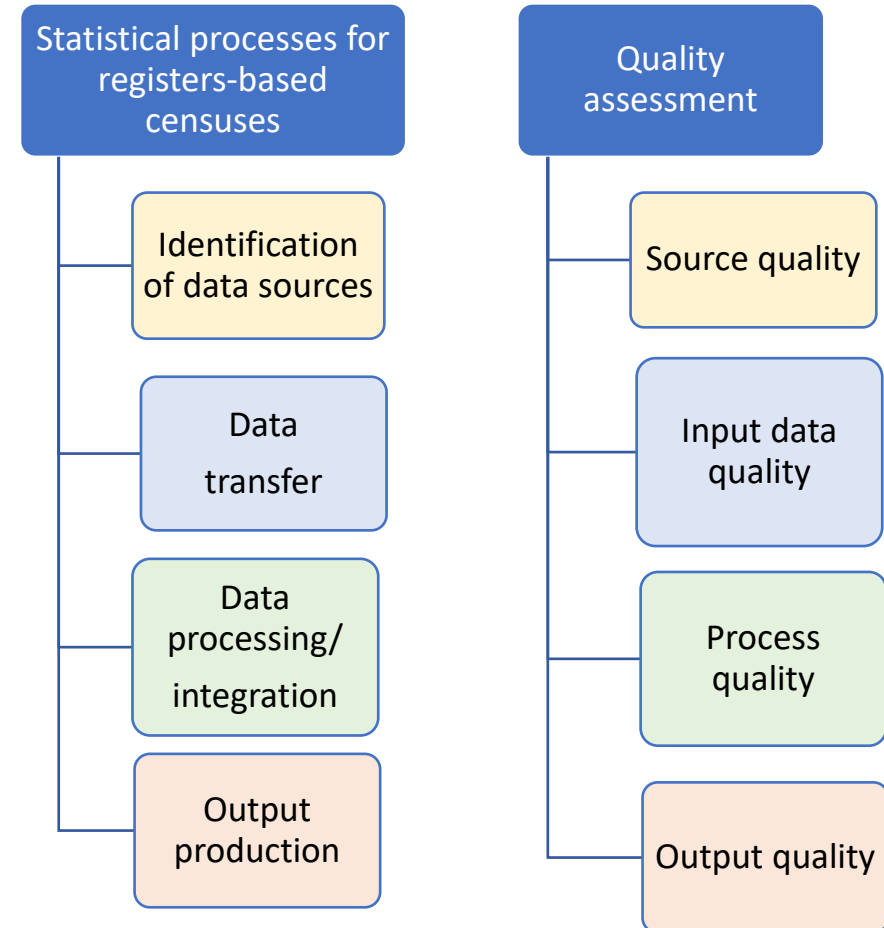
## Stages of quality assessment

### Four stages



## Stages of quality assessment for administrative data sources

- ❑ Stages of the quality assessment of administrative registers and the census data derived from them are broadly corresponding to the stages of the statistical processes of the census
- ❑ Designing the quality assessment process through these four stages will help ensure that census estimates are based on the most appropriate sources and methods





## Quality assessment dimensions



**Relevance** - the degree to which statistical outputs meet current and potential user needs, in terms of data availability, concepts and definitions

**Accuracy and reliability**- the degree to which the information correctly describes the phenomena, such as usual resident population

**Timeliness** - the delay between the date to which the data refer (census day) and the date on which the information becomes available

**Coherence and comparability**- the degree to which data that are derived from different sources or methods, are similar, – the degree to which data are comparable over time - the degree of consistency across data sources and time

**Accessibility and interpretability**- the ease with which users are able to access census data – availability of metadata describing about sources, methods and definitions



## Stages of quality assessment

Source  
quality

Input data  
quality

Process  
quality

Output  
quality

Quality assessment of  
**administrative data  
sources**

*Metadata-based  
assessment*

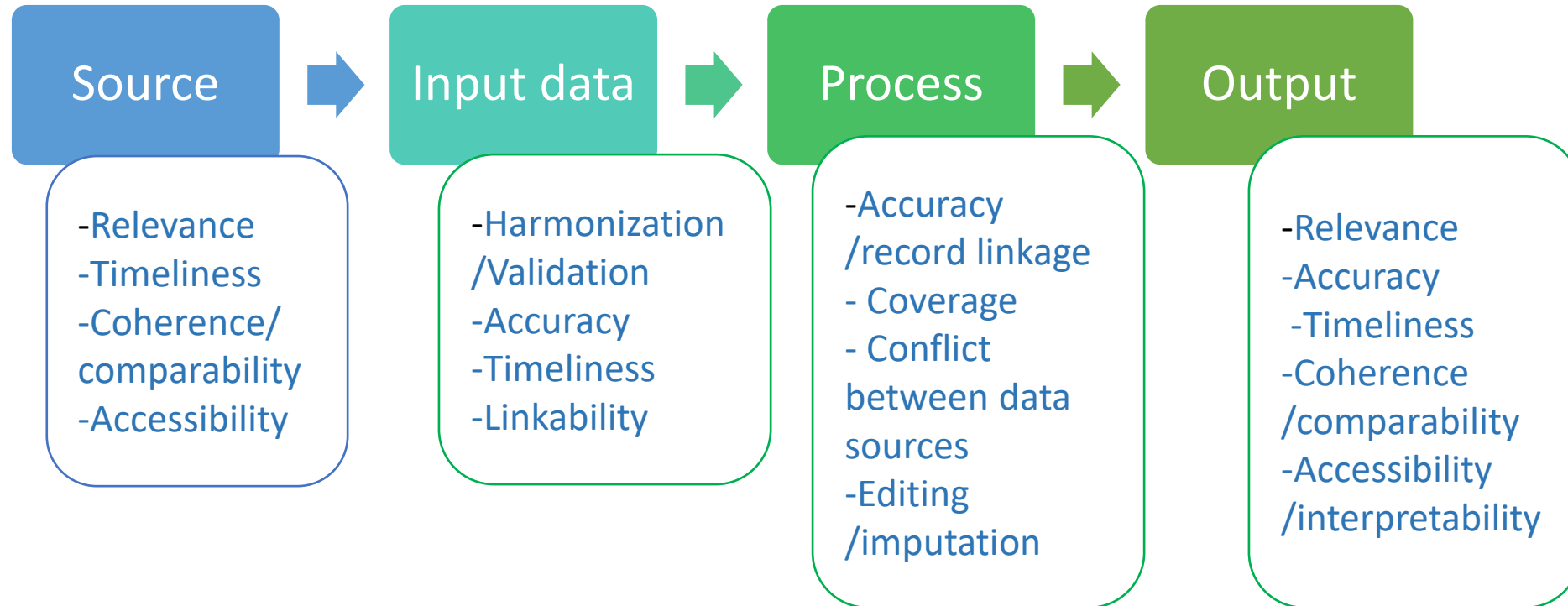
-Quality assessment  
of **raw admin data**,  
as it is supplied to  
NSO by the  
administrative  
authorities

-Assessment of  
changes in the quality  
of data which **results  
from data integration  
and processing of the  
admin data**

- Overall quality  
assessment of the  
statistical results as  
disseminated to users



## Quality dimensions of each stage







## Source quality (1)



Assessment of **Representation errors and Measurement errors**

- **Representation error**

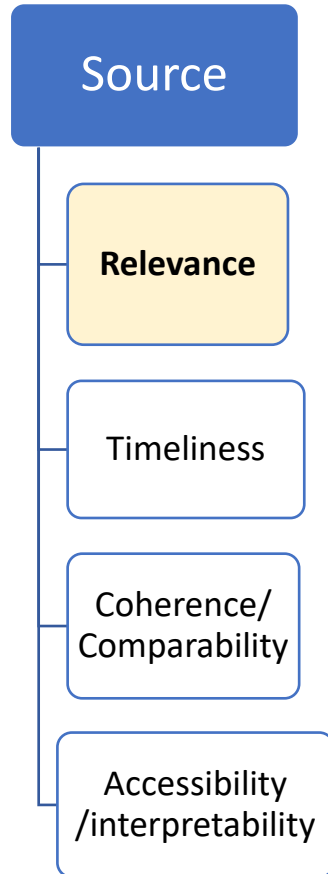
- Alignment of the units in the register with the census target units (persons, households, housing units)
- Information about **what laws and regulations** define who will be included/excluded in administrative data sources
- Information about **what methods/procedures** are used to include/update/exclude units

Assessment indicator

- \* does the coverage of the population register meet the needs of the census?
- \* the evidence of under and/or over-coverage –make an assessment for all population groups that should be included in the population register



## Source quality dimensions



- **Measurement error**

- Alignment of concepts and definitions of variables in the registers with the concepts and definitions of the census topics

- \*whether or not the register includes the variables needed for the census?

- \*whether or not the administrative concepts, definitions and classifications for such variables are consistent with those adopted in the census?

- \*in case of inconsistency, whether a transformation of the variables is possible to satisfy the requirements of the census?

- \*if not possible, whether or not it would provide similar information?



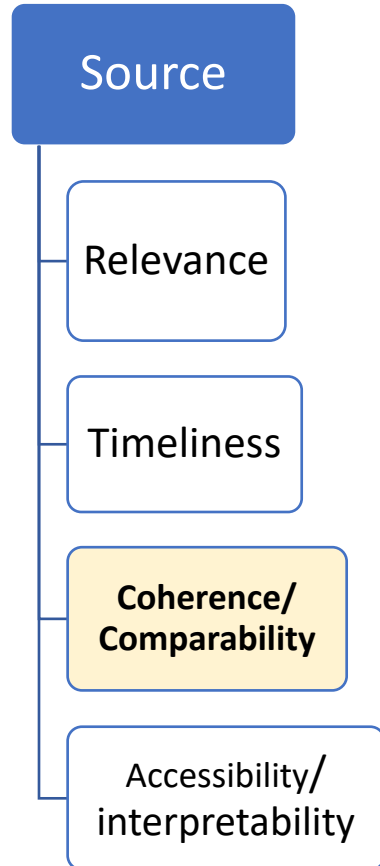
## Source quality- Timeliness



- The difference between *the reference date to which the data refer and the date on which they are supplied to the NSO*- the longer the delay the less relevant
- Some examples of information that can be used to assess timeliness
  - \* what is the time lag between date of occurrence and date of registration?
  - \* What is the time lag between date of registration and date on which the data are supplied to the NSO?
  - \* whether or not the register has been completely updated when provided to the NSO?
  - \* how frequently the data can be supplied to the NSO for updates or new persons or dwellings?



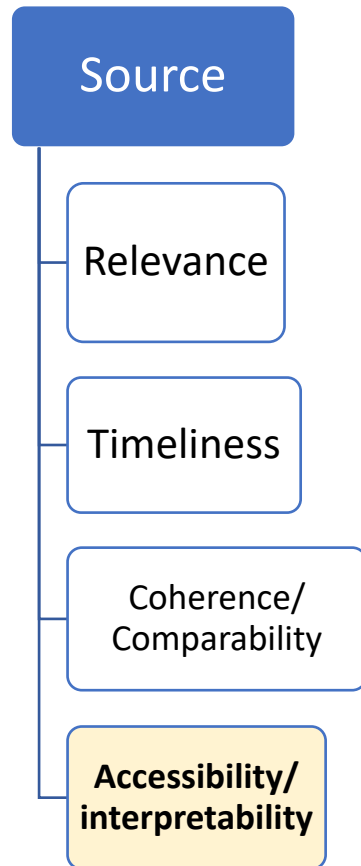
## Source quality- Coherence and comparability



- assess the degree to which an administrative source can be **successfully linked and combined with other data sources** for use in the census
- Some examples of information that can be used for this assessment :
  - \* Does the source include a unique identifier (such as PIN) that is common with the unique key required for the census linkage?
  - \* If so, is the identifier available for all of the relevant individuals/ addresses in the source or only for special population groups or geographic areas?
  - \* Does the source include a unique combination of variables (such as name, date of birth and address), which could be used for the census linkage?



## Source quality- Accessibility and interpretability



- It is important to **identify any restrictions that may impact on the NSO's ability to acquire and use** an administrative source, such as existing data protection restrictions

*\*What is the level of public acceptability?*

Whether or not an NSO decides to access a particular data source for use in the census may also depend on public acceptance

*\*How easy is it to transfer data?*

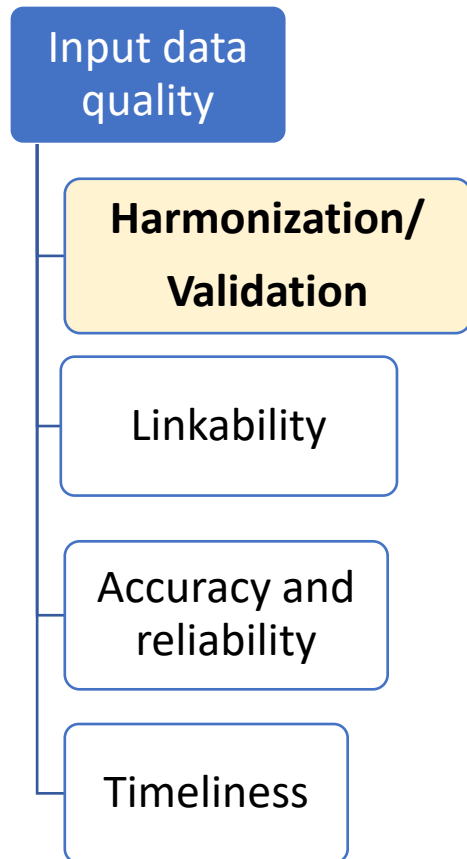
The data supplier might adopt very different data models, formats, schemas, software and hardware to that with which the NSO is familiar

*\* Is there clear and comprehensive metadata?*

An assessment of interpretability relates to the existence and availability of comprehensive and clear metadata and documentation about the administrative source



## Input quality- Harmonization and Validation



- ❑ It is crucial for the NSO to ensure that the transmitted data files are in the required 'readable' format; the databases are structured in a way which can be ingested and read by the NSO's systems

Some indicators can be used to assess the validity including:

- \* Whether or not the variables supplied are correctly named and formatted (e.g., numerical, categorical, text information, etc.),
- \* Whether or not the correct reference period has been supplied
- \* Whether or not the variables match the expected pre-defined content, established through the metadata collected at the Source Stage



## Input quality- Linkability

Input data  
quality

Harmonization/  
Validation

**Linkability**

Accuracy and  
reliability

Timeliness

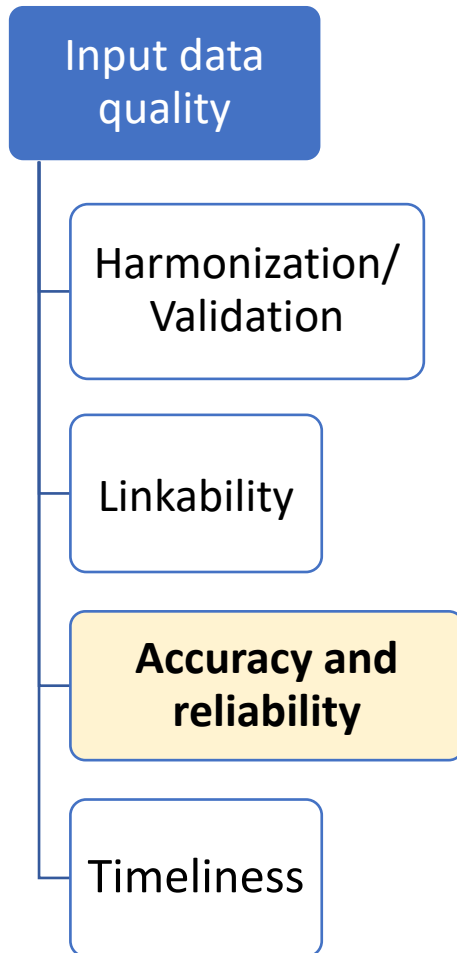
- ❑ Assessment of the variables in each administrative data source used in the linkage – informing the design of a successful linkage in the process phase

Some indicators can be used to assess the linkability including:

- \* Percent of unique values or a combination of variables to be used in linkage
  - ex. Percentage of unique Personal Identification Number or combination of age, date of birth and address
- \* Measurement errors within linkage variables
  - Percentage of missing values, implausible values, etc.
- \* Prevalence of biased distribution
  - is there a significantly higher proportion of out of range or missing values for a key linkage variable(s) within certain geographies



## Input quality- Accuracy and reliability

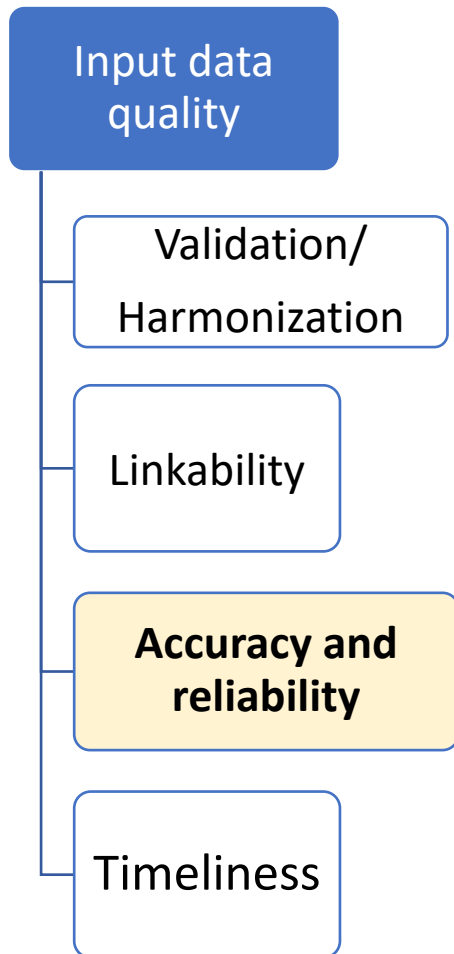


- In assessing the accuracy of the input data, NSOs should distinguish between
  - **‘representative errors’** (those relating to the coverage of target population) and
  - **‘measurement errors’** (those relating to the particular variable being considered).
- Basic indicators to assess **representation errors** include:
  - the total number of units (persons/housing units) received (for comparison against expected count);
  - the percentage of duplicate units
- A key indicator in assessing **under-coverage** would be:
  - the percentage of units in the reference source (traditional census or a complete base register) that are missing in the supplied (administrative) source.while **over-coverage** can be assessed by:
  - the percentage of units in the (supplied) source not belonging to the target resident population of the NSO





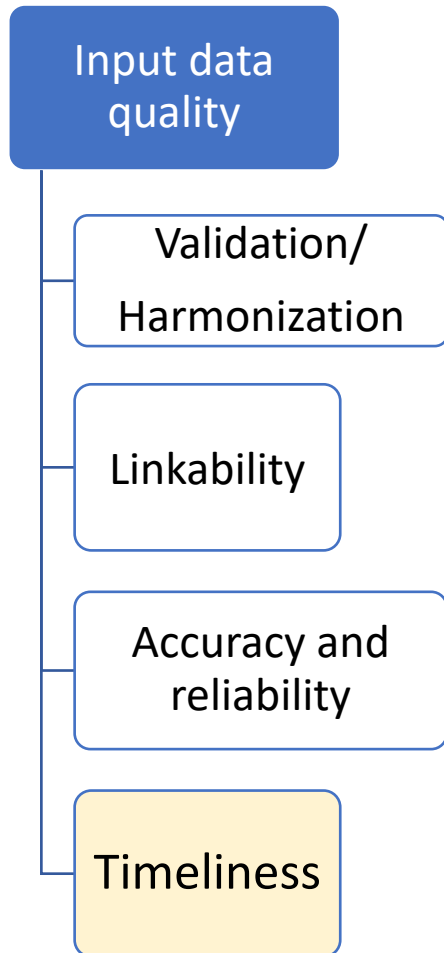
## Input quality- Accuracy and reliability



- Assessment of **measurement errors**
- Basic indicators to measure the completeness of the characteristic variables supplied within administrative datasets at the aggregate level (such as age, sex, ethnicity, etc) include following:
  - number and percentage of missing values within key variables (such as date of birth and sex);
  - number and percentage of out-of-range values within key variables (for example a recorded age of 120 years);
  - number and percentage of implausible values (based on, for example, cross-tabulations of different variables);
  - prevalence of unexpected frequencies, patterns or outliers, based on frequency/distributional analysis of key variables



## Input quality- Timeliness



Measures of **timeliness** can be determined relatively easily by comparing

- the reference date,
- the specified delivery date, and
- the actual delivery date of the data

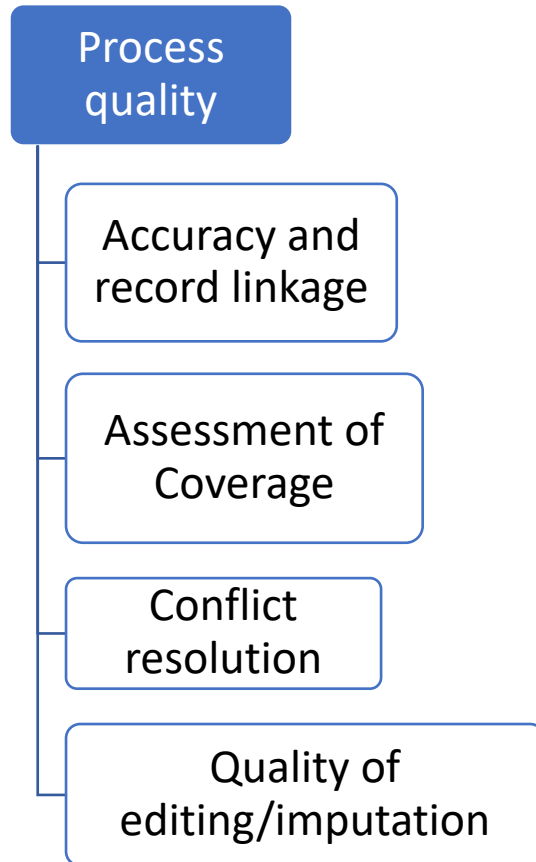
Two aspects of timeliness can be assessed by:

- **the difference between date of occurrence and registration**
  - the date of capturing any change to the data in the source by the register holder and the date on which the change actually occurred in the population;
- **the difference between** the date of receipt of the data by the NSO and the date of the reference period to which the data refers,

the longer the delay, the less relevant



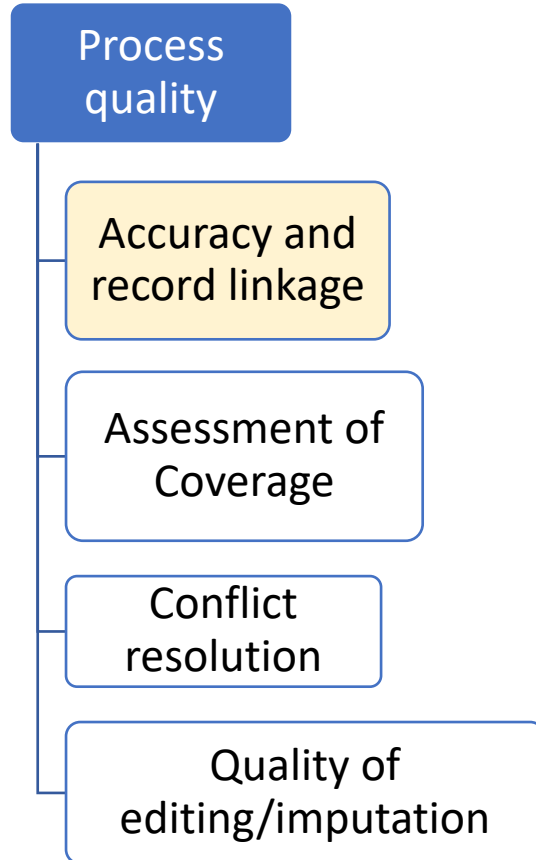
## Process quality



- As data held in an administrative source are not collected for statistical purposes, they must be transformed by the NSO in some way for use in the census.
  - Linkage of data through a common identifier
  - Constructing/updating a statistical population register
  - Data processing
    - Dealing with duplications
    - Conflict resolution
    - Updating and **Signs of life** method for improving the quality of coverage of statistical population register
    - Editing and imputation
    - Validation of census outputs



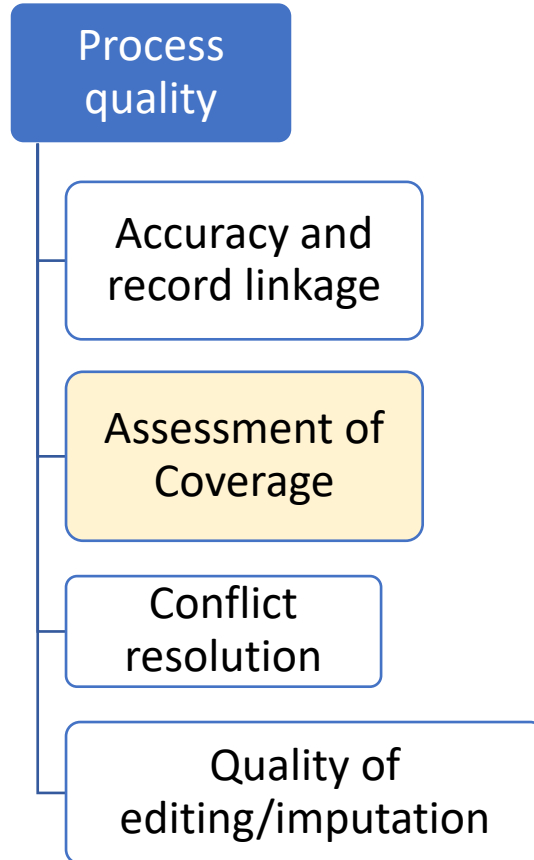
## Process quality-Accuracy and record linkage



- Common methods for assessing linkage quality are:
  - determining the proportion of records that are not, or cannot be, linked
  - the number and percentage of duplicate linkage key
  - absence of linkage keys, or the number of missing or implausible values
  - Comparison of the distributions of characteristics of linked and unlinked records, by for example such variables as age and sex and by region and population sub-groups
    - Differences in characteristics suggest that some bias is introduced by linkage error,



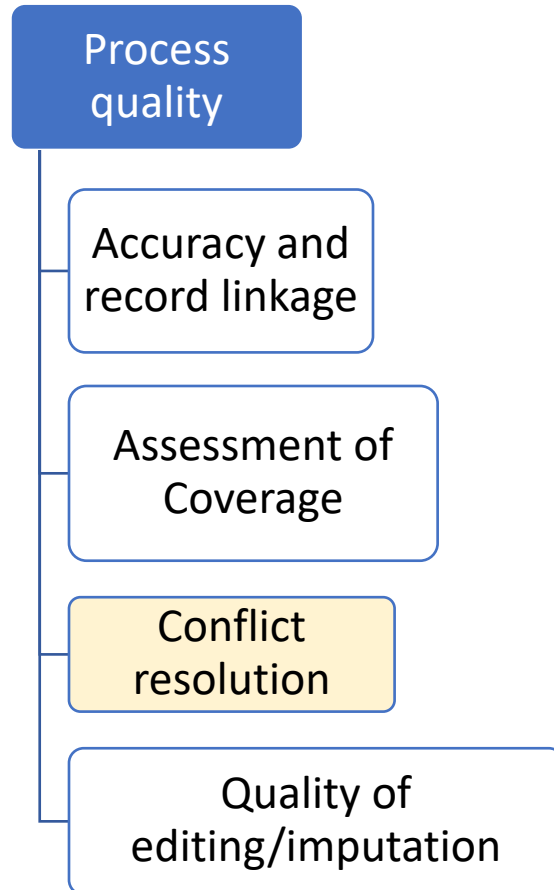
## Process quality- Assessment of the quality of statistical population register



- *Using signs of life*
  - An increasingly used tool to help minimise over-coverage is the so-called ‘signs of life’ (SOL) method based on a number of ‘rules’ that are adopted to ensure that only persons who are alive and meet a set of pre-defined residency criteria are included in the census
- *Using independent surveys and demographic analysis*
  - In addition to the signs of life approach there are several other methods that are available to estimate the coverage (and, indeed, content error) of censuses.
  - These include: simple quality assurance techniques such as internal consistency checks; demographic analyses; comparisons with data from other sources including previous censuses and/or current household surveys



## Process quality

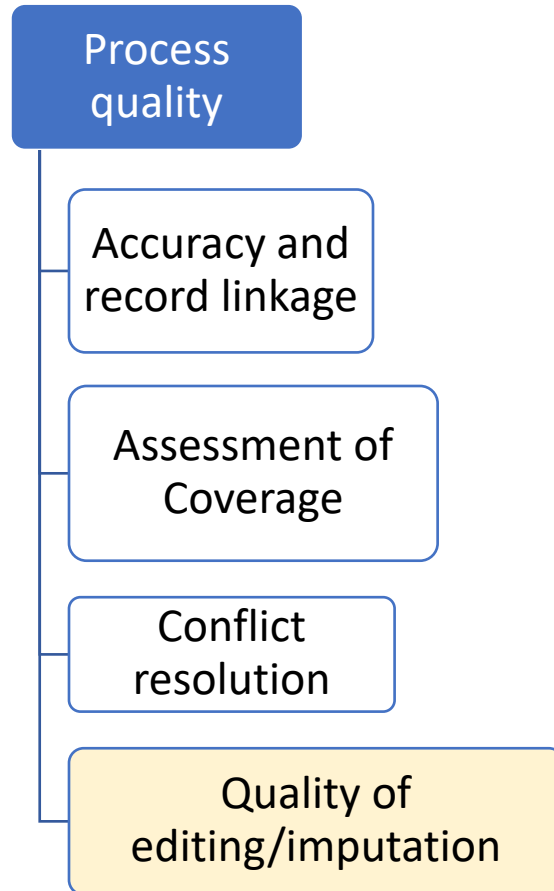


### *Conflict solutions*

- When using data from multiple sources, methods are required to assess the quality of variables where the same variable is reported in different registers
  - Then rules are necessary to determine which value is accurate
- Resolve conflicts in residence address recorded in different registers and identify the most accurate information
  - The NSO will need to decide to which address the census information should refer (perhaps by using, for example, the most recently reported address)



## Process quality



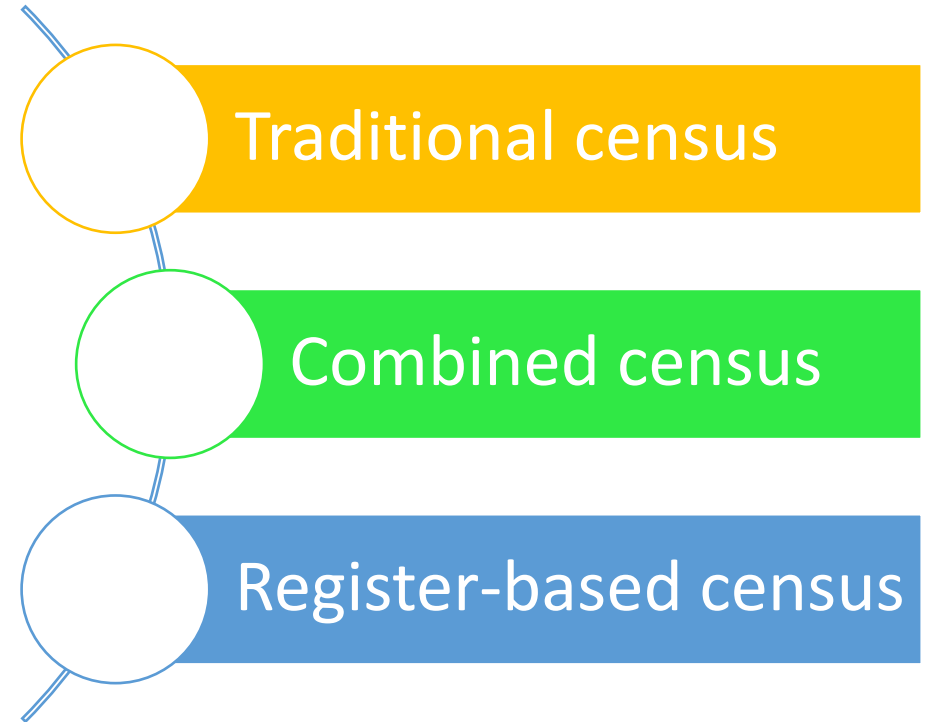
### *Assessing the quality of the editing and imputation process*

- Editing and imputation is an iterative process of reviewing the data in order to correct any errors resulting from invalid, inconsistent or implausible and for missing values
  - a number of indicators can be used, such as edit failure rate, adjustment rate, imputation rate, dissimilarity index



## Output quality

- Regardless of the type of data collection methodology used for the census, it is equally important for NSOs to assess the quality of the output
- It is particularly important for those NSOs that have moved to a register-based or combined census data collection methodology to assess the quality of outputs, to **determine whether or not the transition has affected the overall quality of the outputs**







## Conclusions

- ❑ Each country should plan the process of the transition based on
  - availability of administrative data sources
  - assessment of the quality of administrative data source and the quality of input data
  
- ❑ The transition should be planned gradually,
  - introducing more administrative data sources and variables each time, providing that the registers have been proven to be of good quality
  
- ❑ As a result of the transition, there may be some changes to definitions of variables, population bases and output classifications
  - The impact of these changes on the quality of statistical outputs should be assessed and the outcomes should be explained to users



## Reference documents

UNECE Guidelines for assessing the quality of administrative sources for use in censuses <https://unece.org/statistics/publications/CensusAdminQuality>

UNSD Handbook on Registers-Based Population and Housing Censuses [https://unstats.un.org/UNSDWebsite/statcom/session\\_53/documents/BG-3e-Handbook-E.pdf](https://unstats.un.org/UNSDWebsite/statcom/session_53/documents/BG-3e-Handbook-E.pdf)

## Thank you...

Email: [demircim@un.org](mailto:demircim@un.org)