Quality issues specific to administrative data

Regional Training

Producing Register-based Population Statistics in Developing Countries

Tehran, 27 – 31 October 2013

Presented by Harald Utne, Statistics Norway

harald.utne@ssb.no

Quality in official statistics

Definition of quality in statistics

- Product quality
  - Relevance
  - Accuracy
  - Timeliness and punctuality
  - Comparability and coherence
  - Accessibility and clarity

- Process quality
  - Best methods
  - Cost efficiency
  - Low response burden

Process quality

– Use of registers is cost efficient
– Use of registers = no additional response burden

Product quality (1)

Accuracy

- Registers normally have good quality for administrative purposes
- Improving accuracy by combining data from several registers
  - Editing for statistical purposes

Relevance

- Register data is based on administrative definitions that may differ from statistical definitions
  - Units, coverage, variables, time references etc
- "We have the right answers, but can we answer the right questions?"
- "The authorities picture of the world?"
- Combining data from different registers to improve relevance
- In some cases: additional data collection is necessary

Product quality (2)

Timeliness and punctuality

- Production time sometimes longer than for statistical surveys
  - Administrative process may take time (example: taxation data)
  - Delay in updating of registers
  - Data extraction: Necessary to wait some weeks or months and even longer…..

Comparability and coherence
  - Building a coherent register-based statistical system
  - Harmonising with statistics based on other sources
    - Dutch experiences

- Accessibility and clarity
  - Almost independent of data sources used

Register-based statistics compared to statistical surveys

- Costs (++)

- Response burden (++)

- Relevance (-)
  - Not all variables are included in registers
  - Less direct control over data content

- Accuracy (0)

- Timeliness (-)?

Quality assessment for register-based data

Framework by Statistics Austria

Four dimensions/indicators

1. Documentation
2. Pre-processing
3. External source
4. Quality of imputation

1-3 applied to basic data, 4 to final statistical data

Documentation

- Assess quality of documentation
- Questionnaire answered by data owners
- Monitoring quality in data collection, legal base etc
- Carried out for each register variable of interest

Pre-processing

- Concerned with formal errors in raw data
  - Range errors, item non-response, missing ID-number,
- Ratio of usable records to total number of records
- Carried out for each register variable of interest

External source

- Comparison with external source at micro level
  - Data from sample survey
  - Ratio of congruent values to total number of linked records

Quality of imputation

- Applied on final statistical data
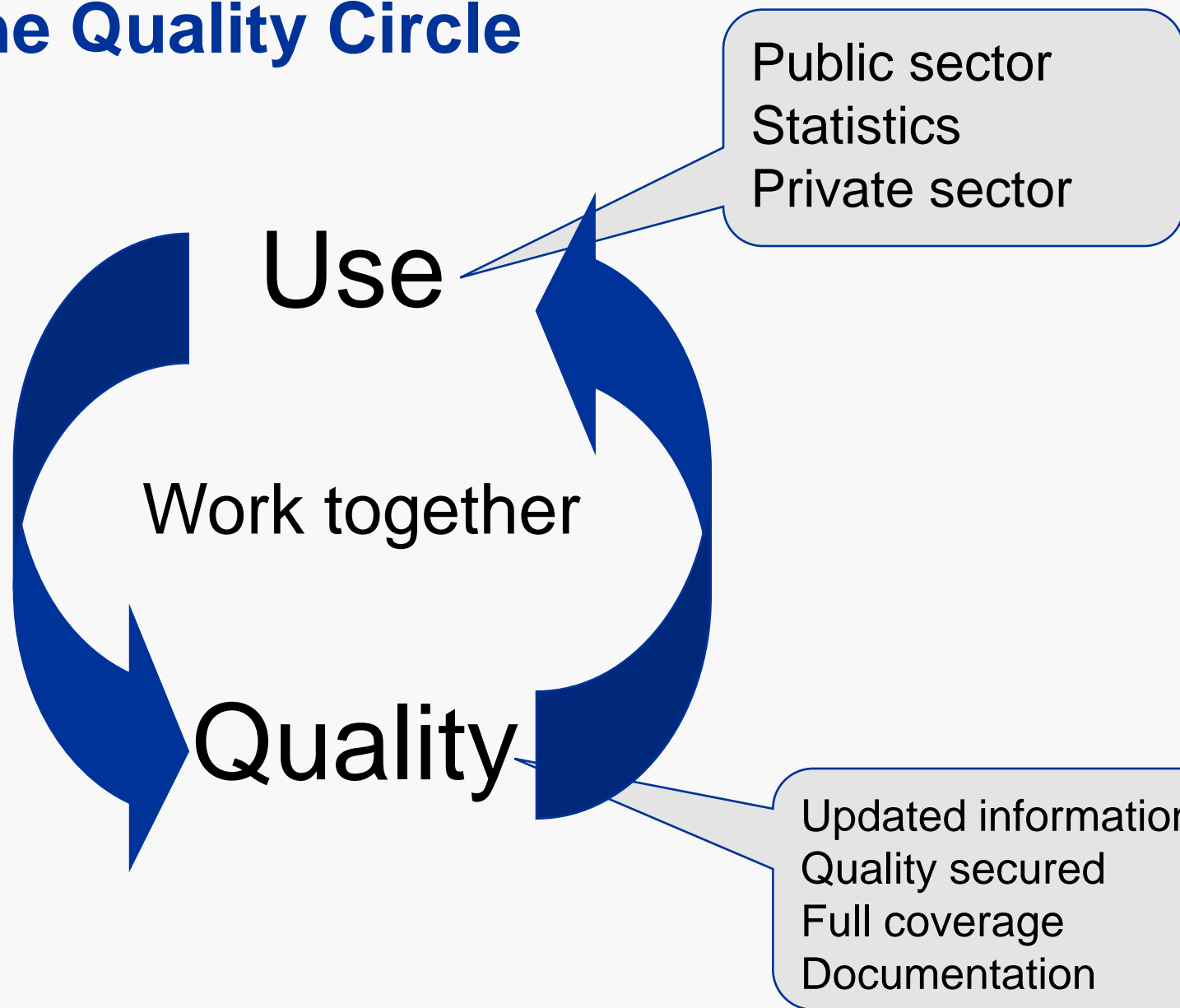- Quality of predictors and hit-ratio of imputation models

Use of quality indicators

- Evaluate each attribute at each level of processing
- Useful in comparing different register
    - What is the best register for variables of interest?
- Compare quality over time

Statistics Austria is constructing an aggregated quality index based on all four dimensions

http://www.unece.org/stats/documents/2013.10.census1.html, paper 28

Thank you for your attention!