

Quality issues in multisource statistics

Giovanna Brancato

“Methods and tools for quality measurement and assessment”

Istat

- Multisource Statistics and Quality
- Towards a common quality framework at European level
- Focus on some issues relevant in the multisource settings
- Conclusions

Multisource statistics

- Increase the potential information of the statistics
- Lower the statistical burden on the respondents
- Overcame some quality problems of the single-survey approach



Potentially ... more relevant, accurate and cost-effective statistics

- Additional process activities
- Traditional and new tailored methods for quality



Quality of Multisource Statistics: new and old challenges

Quality at the European Statistical System (ESS)

- ❑ European Statistics need to be highly harmonized and comparable
- ❑ EU Regulations set (also) quality requirements
- ❑ Eurostat has to be able to evaluate the quality
- ❑ EU member states data are transmitted accompanied with quality reports
- ❑ Need to keep the pace with the statistical production evolutions



ESSnet Quality of Multisource Statistics (2016-2019)

❑ **Three main goals** (Work Packages, WP):

1. Quality Guidelines for Multi-Source Statistics (WP1)
2. Quality Guidelines for Frames of Social Statistics (WP2)
3. Quality Measures and Computation Methods for output quality (WP3)

https://ec.europa.eu/eurostat/cros/content/essnet-quality-multisource-statistics-komuso_en

QUALITY IN OFFICIAL STATISTICS: MULTIPLE SOURCES

Statistical process



Validity (Specification)

**Frame &
Source**

Selection

**Measurement &
item missing**

Processing

Model

**T
I
M
E**

Output quality

Relevance

Accuracy & Reliability

Timeliness & Punctuality

Coherence & Comparability

Accessibility & Clarity

Type of error

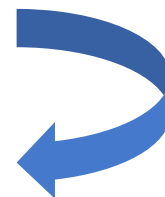
Key issues

- ❑ Relevance & validity/specification error
- ❑ Accuracy
 - measurement error
 - record linkage
 - model error
- ❑ Coherence
- ❑ Time

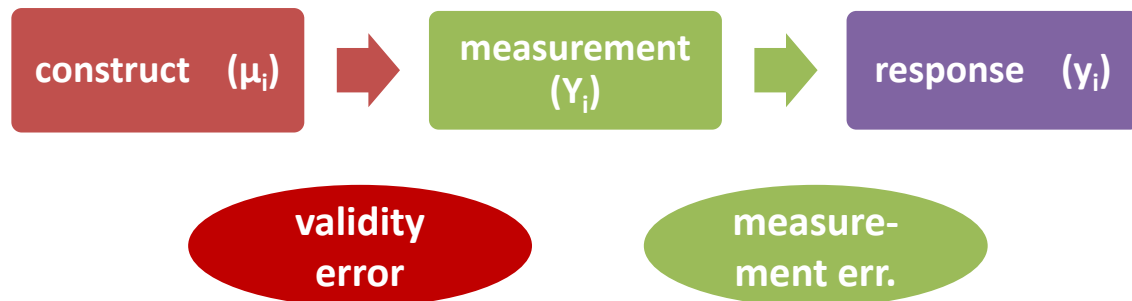


QUALITY ASSURANCE

QUALITY ASSESSMENT



Relevance: validity error

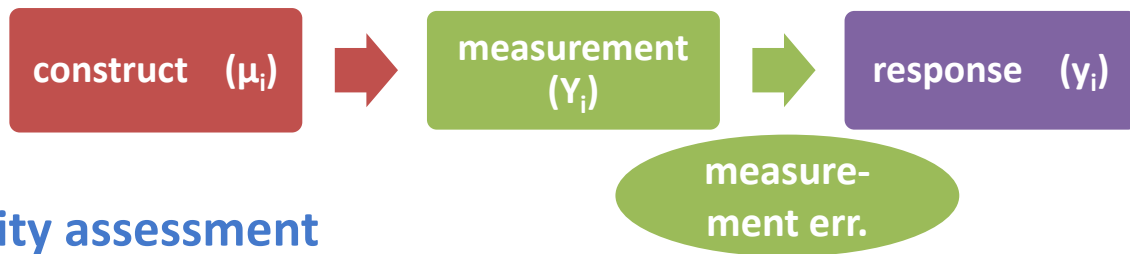


Groves et al. (2004)
Survey Methodology

Quality assurance

- ❑ conceptual analysis of administrative and statistical concepts and definitions
- ❑ harmonization of the measures from different sources
- ❑ legislation changes

Accuracy: measurement error



Quality assessment

- ❑ Multiple measurements of the same variable offer the opportunity to estimate the impact of measurement error (SEM, LCM)
- ❑ Experimental approaches
- ❑ Uncertainty estimation by bootstrapping
- ❑ Indicators [e.g. on impact of a new estimation strategy (multi-source) respect to the traditional sampling approach (Luzi *et al.*)]

Luzi O. *et al.* (2014). The new multiple-source system for Italian Structural Business Statistics based on administrative and survey data

Accuracy: record linkage error

Quality assurance

- ❑ Linkage key, blocking and sorting, choice of the linkage function

Quality assessment

- ❑ Gold Standard (subsample or records manually reviewed)
- ❑ Post-linkage data validation (gross comparisons), Comparing distributions in linked and unlinked data
- ❑ Sensitivity analyses
- ❑

SESSION SPE.9 RL & SM TECHNIQUES

Quality assurance

- ☐ Choice of the model
- ☐ Parsimony (goodness of fit vs. simplicity)
- ☐ Validity of model assumptions

Quality assessment

- ☐ Variance from the use of the model and bias from invalid model assumptions

Quality assurance

- ❑ Calibrating , Benchmarking, reconciling, ...
- ❑ Validation studies become part of the process

Quality assessment

- ❑ Descriptive measures
- ❑ The sources for comparisons might become part of the process

Internal coherence

Different sources not properly harmonized might produce estimates that are not internally coherent

Quality assurance

- ❑ Characteristics of the administrative data → *progressiveness*
- ❑ Continuously updated data sources → trade off accuracy-timeliness
- ❑ Need for harmonizing time reference of data coming from different sources
- ❑ Impact of changes in the legislation of the administrative data

Quality assessment

- ❑ Analyze the impact of different updates of the sources on final estimates → Revision analysis

Conclusions

Quality in multisource statistics can widely rely

- ❑ on the traditional approaches used for surveys:
 - the adoption of a quality control assurance system
 - the application of methods to estimate the potential impact of errors on the final results
- ❑ on the approaches used in other domains, e.g.:
 - psychometric field
 - benchmarking
 - revision analysis

Conclusions

The use of multiple sources share new and old challenges with the surveys:

- ❑ lack of full control on the data formation
- ❑ complex management activities
- ❑ usefulness of the information vs. its redundancy
- ❑ dominating errors, errors propagation, joint assessment
- ❑ still need for additional information for the assessment



... strive for controlling ex ante and evaluating ex post error sources

“There are only two mistakes one can make along the road to truth; not going all the way, and not starting” (*Buddha*)

Thank you for your attention!

brancato@istat.it