المـركـز الإحـصـائي
لدول مجلس التعاون لدول الخليج العربية
GCC-STAT

Regional Workshop on Editing (Data Cleansing) of Statistics produced from Administrative Registers

Applying GCC Quality Framework – Case Study 2020 Census

Nancy McBeth

GCC-Stat

Content

- Quality Control and Quality Assurance
- Quality Management in NSOs in the GCC
- Case Study – Applying Quality Management to the 2020 Administrative Based Census

Quality Control and Quality Assurance

Why do we worry about Quality?

- Credibility and confidence in official statistics
- Reputation of the NSO as independent, objective source of trustworthy information
- Poor quality statistics used in decision making - Poor quality decisions – Poor outcomes

Quality Control

- Provides **regular and consistent checks** to **ensure data integrity, correctness and completeness**
- **Identifies and addresses errors/ substandard data**, either by fixing or by discarding it.
- Also used to ensure accuracy and completeness of program plans, agreements (e.g. memoranda of understanding) and contracts.
- **Short time focus** – identify and fix problems NOW
- Examples - physical checking of completed questionnaires and data **editing (cleansing)**

Quality Assurance

- Test the **accuracy and reliability of the processes**, according to agreed criteria.
- Aim to **provide confidence** in the quality of the product by **assessing the performance of a process** according to certain criteria.
- **Longer time horizon** - aims to fix identified problems in the next cycle of the activity – for example the next collection.
- Also used to provide **feedback on the previous phase** of the census cycle, including identifying areas for future improvement.
- Eg Quality audits, review performance measures and quality indicators (KPIs) after the survey.

Specific issues for Administrative Data

NSO does not have any direct control over the data or the administrative processes used to collect it

Quality Control checks for:

- **Differences in concepts** – leading to bias and coverage problems
- Possibility of having **missing items or missing records**
- Timeliness of the data (some or all of the data may arrive late)
- Punctuality – not all data relates to reference period
- Poor or limited metadata

Quality Assurance includes:

- Understanding the **Operational context** for the administrative collection and processes
- Understanding the **Quality Assurance conducted by agencies**
- Active two way **Communication** with the agencies supplying the data
- Preparing and publishing relevant **documentation for users**

Need for NSO to invest in systems and expertise to clean and combine the data to produce statistics.

Quality Management in NSOs in the GCC

Quality Management in National Statistical Office (NSO)

- In the context of a national statistical office:
- Quality is defined in terms of
  - **Quality Components (or Dimensions) and**
- Quality Management System is typically expressed in the form of
  - **Quality Assurance Framework**

Focus here – Quality Components (or Dimensions)

GCC Quality Dimensions

• Relevance	• الملاءمة والاستجابة لاحتياجات المستخدمين
• Accuracy	• الدقة
• Reliability	• الموثوقية
• Timeliness & Punctuality	• سرعة نشر النتائج ودقة مواعيد نشرها
• Accessibility & Clarity	• النفاذ إلى المعرفة والوضوح
• Comparability & Coherence	• الاتساق والمقارنة

Case Study – Applying Quality Management to the 2020 Administrative Based Census

Whether the available information sheds light on the issues of most importance to users.

- GCC statistics should meet the real needs of users – so belongs to the domain of statistics users; GCC statistics offices cannot establish it by themselves.

Implications for Census

Purpose of Census – small area or sub-population statistics. Statistics need to be available to meet these needs of users

Examples of Administrative data issues

- Are all the required topics (data basket variables) available in the registers?
- Do the concepts, definitions and classifications used in the registers meet requirements?
- Do the final Population and Address Registers cover all the country?

Accuracy الدقة

Whether a statistic correctly describes what it was designed to measure. May be described in terms of

- bias (systematic error) and variance (random error)
- major sources of error (e.g., coverage, sampling, nonresponse, response).

Implications for Census

Major sources of error. Undercoverage rate by region, Unit non-response rate, Item non-response rate

Examples of Administrative data issues

- Is the coverage of the population acceptable?
- Does each unit in the Population and Housing registers have a geographic code?

Reliability الموثوقية

Confidence that users place in GCC statistics - based on the image of the data producer. Built up over time. **Important aspect is trust in the objectivity of the data**. Implies **data are perceived to be produced professionally** in accordance with appropriate statistical standards, and that policies and practices are transparent.

Implications for Census

Data are not manipulated. Release not timed in response to political pressure.

Examples of Administrative data issues

How are differences between administrative concepts and statistical concepts (e.g. definitions of Usual Residence) managed? Eg, are differences in concepts explained in metadata?

Timeliness سرعة نشر النتائج

Difference between the reference point/ end of reference period and date on which the information becomes available.

- Typically involved in a trade-off against *accuracy*.
- *Timeliness* of information will influence its *relevance*.

Implications for Census

Timeliness of the First and Final Results. Do they meet user requirements?

Examples of Administrative data issues

- Are registers available for the census reference date? How will differences be managed? What statistical adjustments will be needed?
- Are the registers updated on a timely basis?

Punctuality دقة مواعيد نشر النتائج

Whether GCC statistics are **released in accordance with publication schedule** – target release dates.

Implications for Census

Ensuring Publication dates meets announced release dates

Examples of Administrative data issues

Are the registers available in a timely manner? (Often the most complex cases come last)

Accessibility النفاذ إلى المعرفة

Ease with which information can be obtained from GCC statistics offices. Includes, being able to determine if information exists, suitability of access methods

Implications for Census

- Can demand for small area data (e.g. detailed location information) be met through up to date tools (e.g. portal, interactive maps)?
- Can researchers access confidentialised unit record data?
- Is there Open Access to Portals, Website ? Or are there restrictions?

Examples of Administrative data issues

- Are the registers available and accessible to the NSO? Are registers available in a standard format?
- Can the registers be linked together? (Either through common Identifiers or through statistical IDs)

Clarity الوضوح

Availability of metadata necessary to interpret and utilize statistics appropriately. Covers the underlying concepts, variables and classifications, data collection and processing methodology, and indications of the accuracy of the published statistics

Implications for Census

- Results presented as absolute numbers as well as percentages
- Data are shown by territorial divisions: by statistical regions and by districts.
- Metadata provides information on definitions, known quality issues

Examples of Administrative data issues

- Is there metadata about each register?
- Does the metadata meet statistical requirements?
- Can the registers be linked together?

Whether **statistics can be successfully brought together with other statistics** within a broad analytic framework and over time.

Implications for Census

- Are Initial and Final results coherent?
- Are Census results coherent with other relevant statistics, e.g Census and Labour Force Survey, Population Census data are consistent with Vitals
- Are differences in concepts or definitions over time clearly stated in metadata

Examples of Administrative data issues

- Are the register sources internally coherent?
- Is the linked data from the registers coherent? Do the registers use international or regional standards?

Quality Dimensions in Census

Steps in Census Cycle

Quality Dimensions	Preparation	Collection	Processing*	Analysis	Dissemination	Evaluation
Relevance						
Accuracy and Reliability						
Timeliness and punctuality						
Accessibility and Clarity						
Coherence and Comparability						

*Includes Micro Editing, Coding, Macro Editing

	Primary Quality Role
	Secondary Quality Role

Underpinned by Planning/Monitoring (Project Management)

Types of Checks in Census

Source Checks		Checks on **Administrative Sources and Records** - Understand quality of administrative records – BEFORE using them in Census Checks on **Fieldwork Data and Processes** – identify and correct major field errors and provide feedback to field staff
Processing	Micro Editing	Checks on **Individual records and relationships between records** Aim to identify and correct errors and provide feedback to previous steps
	Macro Editing	Checks **Aggregate data** Aim to identify and correct errors before data is published
Evaluation		Checks **Aggregate data/Matching studies** Provides information to users

But no amount of editing can replace high quality enumeration (field or register based)

If systematic errors occur during data collection, editing cannot improve the quality of the data no matter how sophisticated the procedures.

Questions

Thank you