Training course on demographic evaluation of age and sex data

Part 2(i)

27 May – 3 June 2022
Contents of Part 2(i)

• Introduction to data analysis
• Types of error
• Sources of error
• The assessment-analysis cycle
• Types of assessment
• Modes of assessment
Introduction to data analysis

• The UN *Principles and Recommendations* observe that a census involves not only collecting and compiling data on a population, but also the evaluation, analysis and dissemination of those data.

• Ideally, we would want data from censuses and surveys to have been so well collected and compiled that the processes of evaluating and analysing the data are rendered both trivial and mechanical.

• The reality is unfortunately somewhat different, especially in developing countries.
Why do things go wrong?

• “In a perfect world, data would always be complete, accurate, current, pertinent and unambiguous. In the real world, data is generally flawed on some or all of these dimensions” (Feeney, 2003)

• Demographic data collection is not a completely scientific exercise in that the data is collected **by** individuals, **on** individuals
Types of error

- **Coverage** – “faithfulness of the correspondence between the records that constitute the data set and the statistical aggregate the data set represents”
  - Not all people or people in a given area / with given characteristics are even interviewed or enumerated

- **Content** – “completeness and accuracy of the information contained in the data set”
  - Answers given in response to a question / set of questions do not conform to the actuality of the situation
Sources of error

• Coverage
  • Censuses and surveys
    • Poor enumerator training and management
    • Seasonal or climatic factors
    • War, insurrection, violence
    • Linguistic or other operational difficulties
    • Systematic exclusions (e.g. nomads, homeless)
  • Vital registration
    • Lack of ‘reach’ of government services (registration offices)
Sources of error

• Content
  • Censuses and surveys
    • Poor enumerator training and management
    • Poor, or incorrect, translation
    • Badly phrased, incoherent, or complex questions
    • Technological obstacles (e.g. scanning)
    • Changing definitions (e.g. urban/rural)
    • Reference period error
    • Post-collection errors (e.g. editing and imputation)
  • Vital registration
    • Badly designed data collection forms
    • Poorly-trained administrative or medical staff
The assessment-analysis cycle

• In data assessment we focus (almost exclusively) on completeness and accuracy.
• Data assessment is often thought of as a precursor to data analysis.
  • Sometimes this is a useful distinction; but in most demographic work, the assessment of error, and the drawing of substantive conclusions are different sides of the same coin (Feeney)
Evaluation and analysis

• ... is neither an exact science, nor a linear process
Evaluation and assessment

• Before starting with the analysis (estimation), one needs to have some initial understanding of the data
  • In broad terms, are the data consistent with expectations?
    • Projections from a previous census
    • Comparability with other data sources (MICS; DHS …)
  • What aspects of the data are known \textit{a priori} to be problematic?
    • Coverage
    • Completeness
Estimation

Evaluation and Assessment of Data

Evaluation and Assessment of Estimates

Estimation
Evaluation and assessment of estimates

• Common checks
  • Are the estimates that have been derived consistent with previous estimates?
  • Are the levels and trends implied plausible?

• Less common checks
  • In the context of overall population dynamics, do the estimates result in sensible macro-level trends
    • E.g. if fertility is indicated to have fallen rapidly over time, do population projections over that period bear this trend in fertility out?

• Need to revisit the data assessment, and estimates, in a recursive procedure
Evaluation and analysis

Evaluation and Assessment of Data

Evaluation and Assessment of Estimates

Estimation
Evaluation and assessment: Process

• Before you start:
  • What other data might be used as comparators?
  • Understand your data
    • Instrument design
    • Data collection design (sampling frames for surveys, for example)
    • Logistics, planning, field reports
  • Understand something of the history that might have had a bearing on your data
    • Boundary changes
    • Wars or famines
    • Politics …
Types of assessment

• “External consistency”
• “Internal consistency”
Assessment: External consistency

• ‘External consistency’ is the evaluation of one’s data against data from an external source
  • E.g. comparison of age-sex structure in your data with that from the UNWPP

• A particular form of this is possible when your data is drawn using a sampling frame determined from another source
  • It follows, then, that the basic population characteristics in the sample should be (virtually) identical to those in the population used to determine the sample
  • E.g. a DHS using a sampling design determined by a recent census
Assessment: External consistency

• If the data are intended to be representative of some defined population, compare the observed age and sex structure against other data for the same population
  • Some differences are normal (sampling error), but overall, the populations should ‘look the same’
  • E.g., a DHS which is sampled off a census should have almost exactly the same age distribution of women as the census from which it is drawn
Assessment: External consistency

- E.g. ZA DHS 1998, distribution of women 15-49
  - Are these the same population?

  - Beware of drawing firm conclusions from proportional data
Assessment: Internal consistency

• “Internal consistency”, on the other hand, involves evaluation of the data as to the coherence within the data
  • Starting points are usually determined by prior knowledge (‘common sense’)
  • Typically exposes coverage and content errors simultaneously
    • Can be hard to distinguish which is which without further investigation (particularly with respect to external consistency)
Assessment: Internal consistency

• Preliminary checks
  • Head count with other estimates (same ball-park?)
  • Urban and rural; subnational populations
  • Average household size
  • No. of single-person households

• Check for high incidence of illogical or improbable outcomes
  • ... men who have given birth ...
  • ... divorced infants ...
  • ... 10 year olds with PhDs, or earning >$100k p.a.
Modes of assessment

• The most important determinants of all three demographic processes (fertility, mortality, migration) that determine a population’s size and structure are **age and sex**

• It follows that, in any demographic data evaluation exercise, attention must first be directed at understanding these aspects of the data
Reminder!

• If you have any queries or questions on this material, please email Thomas.Moultrie@un.org, copied to escap-crvs@un.org

• I will do my best to answer questions, either by return of email, or in the first plenary session on 27 May 2022