Quality assessment
Direct and indirect methods & choosing appropriate methods for assessing completeness and qualitative accuracy of registration

Helge Brunborg

2nd TRAINING WORKSHOP
ESCAP. Bangko, 6-10 March 2017
Quality assurance:
The maintenance of a desired level of quality in a service or product, especially by means of attention to every stage of the process of delivery or production.

Quality assessment:
An evaluation of the extent to which a trial’s design and management are likely to have prevented systematic errors and biases.
Quality basic framework in the P&R

**Quality assurance**
- Encompasses each stage of CRVS operations
- All vital events are registered without duplication
- All related information is recorded
- Information is compiled, validated and processed
- Vital statistics are released in timely manner

**Quality assessment**
- Specific studies for specific questions
- Coverage of registration of vital events
- Accuracy of variables
- Overall functioning of sub-systems
- Can be ad hoc or regular exercises

Quality assessment methods

- Indirect methods → Demographic analysis
- Direct methods → Matching of records
Terminology

- **Coverage**: extent to which the population is served by the CRVS system as issue of access to the reporting system, and may be influenced by geography or other considerations such as the legal intent of the system, social or cultural influences.

- **Completeness**: measure the proportion of events in the population that are recorded by the CRVS: multiple ways exist to estimate the “true” number of events against which registered events are compared.

- **Content quality**: i.e., how complete and reliable (by variable) is the unit record data.
Estimating registration completeness

Most common method (indirect): Comparing number of *registered* and *expected* events (from census or population projections):

\[
\text{Birth completeness rate} = \frac{\text{Number of registered births within the year of occurrence}}{\text{Estimated number of live births within the year}} \times 100
\]

**Sources for denominator:**
- Census
- Population projection
- Population register

**Advantages:** Simple, inexpensive and fast

**Limitations:**
- Number of expected events not known, outdated or inaccurate
- No information about the *causes* of under-registration
What is record linkage?

- Merging records for the same entities from two or more data sources into a combined file
- Can be deterministic or probabilistic – or combined

Why record linkage?

- More data about the same entity (individual or event)
- Can check completeness of registration
- Can check the quality of the data in one source against another source
- Can estimate the number of events (or persons) not recorded in any of the data sources (capture-recapture)
Direct methods

Matching of records → Match registration records with records from an independent source
Direct methods: Matching

Birth registration with death registration
- limited to infant deaths
- can be carried out routinely

With administrative records
- a variety of sources can be used
- none is complete
- useful to detect certain type of underreporting
Direct methods: Matching

Lists from population censuses and surveys
- compiled from questions on births and deaths
- can lead to an estimate of completeness
- national or sub-national level

Dual records system
Use of two independent procedures to collect information on vital events:
(1) CR system
(2) A survey
Confronting the two sources
Record linkage to evaluate CR: Direct methods

Matching CR records with records from an independent source:
- Use of civil registration records
  - Example: Matching birth records with death records, particularly infant deaths
- Other administrative records, such as
  - School enrollment
  - Hospital records
  - Baptisms
  - Burials
  - Vaccinations
- Data from censuses and surveys on recent births and deaths
- *Questions* in censuses and surveys on registered births and deaths: This is considered to be an *indirect* method
Dual-records systems

- Using two independent sources to collect information on vital events from Civil Registration (CR) and Sample survey (SS):
  - Events recorded in both CR and SS
  - Events recorded in CR but not in SS
  - Events recorded in SS but not in CR
  - Events recorded in neither CR nor SS (capture-recapture)
Capture-recapture method

TABLE 1. Two-source model

<table>
<thead>
<tr>
<th>Source Y</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
<td>x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>a + b = Z_0</td>
</tr>
<tr>
<td>a + c = Y_0</td>
</tr>
<tr>
<td>N = a + b + c + x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimated values</th>
<th>Maximum likelihood estimator (MLE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unobserved cell:</td>
<td>^x</td>
</tr>
<tr>
<td>Completeness of source Y:</td>
<td>Y_e = bc/a</td>
</tr>
<tr>
<td>Completeness of source Z:</td>
<td>Z_e = a/(a + b) = a/Z_0</td>
</tr>
<tr>
<td>Total population:</td>
<td>N = a + b + c + (bc/a) or, (a + b)(a + c)/a</td>
</tr>
</tbody>
</table>

Completeness of Y = \( \frac{a+c}{a+b+c+x} \)

Completeness of Z = \( \frac{a+b}{a+b+c+x} \)

Conditions for doing record linkage

- Availability of micro data (data on individuals)
- In practice the records need to be in electronic format
- Existence of individual identifiers and variables that makes it possible to link records from different data sets

Examples:
- Unique personal identification number (PIN)
- Date of birth, Sex, Name(s), Place of birth, Names of parents, Place of residence, Religion, Nationality …

Need to design matching criteria
Advantages of direct methods to evaluate CR

- Measure the level of completeness, both at national and at local level
- Can indicate the sources of under- or overregistration
Challenges when doing record linkage

- Sources may not be independent
- Organizing the data in a common format to allow matching
- Identifying errors and cleaning the data
  - Misprints and other errors in PIN, DoB, names, etc
  - Changing PINs
  - Duplicates
    - Which is the most correct record?
  - Examples from BiH: Names, DoB, etc
- False matches
- Multiple candidates for matched pairs
- Undetected matches
- A low matching proportion may be due to errors in the data or that the variables are not specific enough
  - Examples: Only “Year of Birth” and not full “Day of Birth”
  - Some names are very common
  - Many people live in the same area
Mortality measures from sample-based surveillance: evidence of the epidemiological transition in Viet Nam
Nguyen Phuong Ho, Chalapati Rao, Damian G Hoy, Nguyen Duc Hinh, Nguyen Thi Kim Chuc & Duc Anh Ngo
Bull World Health Organ 2012;90:764–772

Analysis of completeness of vital statistics from civil registration systems
Chalapati Rao, Department of Global Health, Australian National University
Multiple sources with overlapping recall periods

Viet Nam 2008/9:
- civil registration
- health system
- peoples committee
- additional partial sources
Viet Nam 2009

Study population – 192 communes; 2.6 million pop
• Data sources
  - Commune Health Centre /Population department- (source 1);
  - Justice system (source 2)
  - Others: Farmer’s union, Womens group, aged care
• Manual matching at commune level, leading to reconciled list of unique events
• Relaxation of matching criteria (age, date of death) owing to inaccurate recording in either source (exercise of local judgement critical to the matching process)
• Unobserved cells computed from two-source analysis
• Reconciliation before ascertaining causes of death, hence reconciled data used as numerator for deriving completeness
• Completeness factor used to adjust life tables and later develop cause-specific
## Matching results

<table>
<thead>
<tr>
<th>Regions</th>
<th>Total in reconciled list</th>
<th>CHC (%)</th>
<th>Population Dep (%)</th>
<th>Justice system (%)</th>
<th>Other (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Ha Noi</td>
<td>2304</td>
<td>1723</td>
<td>1580 (75%)</td>
<td>1669 (72%)</td>
<td>720 (31%)</td>
</tr>
<tr>
<td>2 Thai Nguyen</td>
<td>1185</td>
<td>999</td>
<td>210 (85%)</td>
<td>183 (15%)</td>
<td>85 (7%)</td>
</tr>
<tr>
<td>3 Hue</td>
<td>2221</td>
<td>1768</td>
<td>1043 (78%)</td>
<td>1311 (59%)</td>
<td>777 (35%)</td>
</tr>
<tr>
<td>4 Ho Chi Minh</td>
<td>2453</td>
<td>435</td>
<td>571 (18%)</td>
<td>1871 (76%)</td>
<td>202 (8%)</td>
</tr>
<tr>
<td>5 Can Tho</td>
<td>1758</td>
<td>872</td>
<td>758 (49%)</td>
<td>1081 (43%)</td>
<td>535 (62%)</td>
</tr>
</tbody>
</table>

- A death could be recorded in more than one system
- = interdependence
<table>
<thead>
<tr>
<th>Sex-specific age group (in years)</th>
<th>Sample</th>
<th>a&lt;sup&gt;b&lt;/sup&gt;</th>
<th>b&lt;sup&gt;c&lt;/sup&gt;</th>
<th>c&lt;sup&gt;d&lt;/sup&gt;</th>
<th>x&lt;sup&gt;e&lt;/sup&gt;</th>
<th>Other source only</th>
<th>Deaths</th>
<th>Per cent completeness&lt;sup&gt;f&lt;/sup&gt; (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>1239937</td>
<td>2138</td>
<td>1984</td>
<td>1363</td>
<td>1265</td>
<td>215</td>
<td>Observed (a + b + c + additional) 6750</td>
<td>81.2 (74.1–87.1)</td>
</tr>
<tr>
<td>15–59</td>
<td>873727</td>
<td>903</td>
<td>873</td>
<td>597</td>
<td>577</td>
<td>92</td>
<td>Estimated (a + b + c + x)            2950</td>
<td>80.4 (72.2–80.3)</td>
</tr>
<tr>
<td>60–74</td>
<td>53985</td>
<td>453</td>
<td>414</td>
<td>274</td>
<td>250</td>
<td>38</td>
<td>1179</td>
<td>82.0 (74.9–87.9)</td>
</tr>
<tr>
<td>75+</td>
<td>22852</td>
<td>710</td>
<td>629</td>
<td>453</td>
<td>401</td>
<td>77</td>
<td>1869</td>
<td>81.7 (74.7–87.4)</td>
</tr>
<tr>
<td>Females</td>
<td>1309462</td>
<td>1572</td>
<td>1413</td>
<td>1026</td>
<td>922</td>
<td>181</td>
<td>Observed (a + b + c + additional) 4192</td>
<td>81.3 (74.4–87.1)</td>
</tr>
<tr>
<td>15–59</td>
<td>929773</td>
<td>373</td>
<td>350</td>
<td>251</td>
<td>236</td>
<td>56</td>
<td>Estimated (a + b + c + x)            1210</td>
<td>80.5 (72.5–87.1)</td>
</tr>
<tr>
<td>60–74</td>
<td>72999</td>
<td>342</td>
<td>271</td>
<td>213</td>
<td>169</td>
<td>41</td>
<td>867</td>
<td>83.0 (75.4–89.0)</td>
</tr>
<tr>
<td>75+</td>
<td>37684</td>
<td>812</td>
<td>734</td>
<td>539</td>
<td>487</td>
<td>80</td>
<td>2165</td>
<td>81.0 (73.9–87.0)</td>
</tr>
</tbody>
</table>

CI, confidence interval.

<sup>a</sup> Age- and sex-specific deaths deviate slightly from the totals reported in the text because 27 deaths had no age data.

<sup>b</sup> Number of deaths reported by the Commune Health Centre, the Commune Population and Family Planning Committee (CHC/CPFPC) and the Justice Department.

<sup>c</sup> Number of deaths reported by the CHC/CPFPC but not by the Justice Department.

<sup>d</sup> Number of deaths reported by the Justice Department but not by the CHC/CPFPC.

<sup>e</sup> Estimated number of deaths missing from CHC/CPFPC and Justice Department sources.

<sup>f</sup> Proportion of estimated deaths derived from the list obtained by reconciling the Justice Department and combined CHC/CPFPC lists. Derived with the following formula: \((a + b + c) ÷ (a + b + c + x) \times 100\)
<table>
<thead>
<tr>
<th>Data source</th>
<th>Per cent data completeness (95% CI)</th>
<th>Life expectancy at birth (95% CI) [e0]</th>
<th>Risk of death in children under 5 (deaths per 1000) [5q0]</th>
<th>Risk of death at ages 15–59 (deaths per 1000) [45q15]</th>
<th>Remaining years of life at age 60 [e60]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surveillance sample (unadjusted)</td>
<td>–</td>
<td>74.4 (74.0–74.8)</td>
<td>7.4</td>
<td>163</td>
<td>20.9</td>
</tr>
<tr>
<td>Surveillance sample (adjusted)a</td>
<td>81.1 (74.1–87.1)</td>
<td>70.4 (70.1–70.8)</td>
<td>24.6&lt;sup&gt;c&lt;/sup&gt;</td>
<td>199</td>
<td>19.4</td>
</tr>
<tr>
<td>Viet Nam census (unadjusted)</td>
<td>–</td>
<td>75.2 (75.0–75.4)</td>
<td>10.9</td>
<td>157</td>
<td>22.1</td>
</tr>
<tr>
<td>Viet Nam census (adjusted)b</td>
<td>65.6 (–)</td>
<td>68.8 (68.6–69.0)</td>
<td>16.5</td>
<td>230</td>
<td>17.9</td>
</tr>
<tr>
<td>WHO (2009)</td>
<td>NA (modelled)</td>
<td>69.8 (–)</td>
<td>24.6</td>
<td>173</td>
<td>17</td>
</tr>
<tr>
<td>UNPD (2005–2010)</td>
<td>NA (modelled)</td>
<td>72.3 (–)</td>
<td>No data</td>
<td>139</td>
<td>No data</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surveillance sample (unadjusted)</td>
<td>–</td>
<td>82.3 (82.0–82.7)</td>
<td>5.8</td>
<td>57</td>
<td>25.1</td>
</tr>
<tr>
<td>Surveillance sample (adjusted)a</td>
<td>81.3 (74.4–87.1)</td>
<td>78.7 (78.4–79.0)</td>
<td>22.5&lt;sup&gt;c&lt;/sup&gt;</td>
<td>71</td>
<td>23.6</td>
</tr>
<tr>
<td>Viet Nam census (unadjusted)</td>
<td>–</td>
<td>85.2 (85.0–85.6)</td>
<td>8.8</td>
<td>50</td>
<td>28.4</td>
</tr>
<tr>
<td>Viet Nam census (adjusted)b</td>
<td>57.8 (–)</td>
<td>77.8 (77.5–78.0)</td>
<td>15.7</td>
<td>86</td>
<td>22.4</td>
</tr>
<tr>
<td>WHO (2009)</td>
<td>NA (modelled)</td>
<td>74.5 (–)</td>
<td>22.6</td>
<td>107</td>
<td>19.8</td>
</tr>
<tr>
<td>UNPD (2005–2010)</td>
<td>NA (modelled)</td>
<td>76.2 (–)</td>
<td>No data</td>
<td>96</td>
<td>No data</td>
</tr>
</tbody>
</table>

CI, confidence interval; NA, not applicable; UNPD, United Nations Population Division; WHO, World Health Organization.

<sup>a</sup> Adjusted for data incompleteness and mortality in children under 5 years of age.

<sup>b</sup> Adjustment by the Preston-Coale method.
Conclusion

Completeness of death records was estimated to be 81%.

Viet Nam is undergoing the epidemiological transition.

Data are relatively complete but could be further improved through strengthened local collaboration. Medical certification for deaths in hospitals, and shorter recall periods for verbal autopsy interviews, would improve cause of death ascertainment.
Ethnic majority of the municipalities of Bosnia-Herzegovina

Record linkage of Voter’s list 1997 and Census 1991 by name, date of birth … to determine the ethnicity of persons registering to vote in 1997
Conditions for doing record linkage

- Availability of micro data (data on individuals)
- In practice the records need to be in electronic format
- Existence of individual identifiers and variables that makes it possible to link records from different data sets

Examples:
- Unique personal identification number (PIN)
- Date of birth, Sex, Name(s), Place of birth, Names of parents, Place of residence, Religion, Nationality …

- Need to design matching criteria
Advantages of direct methods to evaluate CR

- Measure the level of completeness, both at national and at local level
- Can indicate the sources of under- or overregistration
Challenges when doing record linkage

- Sources may not be independent
- Organizing the data in a common format to allow matching
- Identifying errors and cleaning the data
  - Misprints and other errors in PIN, DoB, names, etc
  - Changing PINs
  - Duplicates
    - Which is the most correct record?
  - Examples from BiH: Names, DoB, etc
- False matches
- Multiple candidates for matched pairs
- Undetected matches
- A low matching proportion may be due to errors in the data or that the variables are not specific enough
  - Examples: Only “Year of Birth” and not full “Day of Birth”
  - Some names are very common
  - Many people live in the same area
Examples of record linkage in Norway

- Births by parity (birth order) of mother (and father)
- Occupational mortality
- Occupational fertility
- School grades by education and immigration status of parents
- Crime by marital status
- Inheritance of pre-eclampsia
- Mortality by weight and height (for males)

- Income tax returns (spouse, children, income, wealth, ownership of property including housing, vacation homes and cars)
- Prescriptions
Other Indirect Methods
From Principles and Recommendations

(a) Comparison of trends
(b) Delayed registration
(c) Comparison with census data ("balancing equation")
(d) Comparisons of rates observed in similar populations or in previous periods
(e) Incomplete data methods: indirect techniques (UN Manual X)
(f) Questions in sample surveys on birth registration
(g) Other indirect assessment methods
Exercise

Discuss possible ways of estimating completeness of registration of births and/or deaths for your own country by

1. Indirect methods
2. Direct methods

Discuss the advantages and drawbacks of each method