Get every one in the picture

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

Source: SESSION 15. QUALITY ASSESSMENT AND ASSURANCE IN THE CIVIL REGISTRATION AND VITAL STATISTICS SYSTEM. Expert Group Meeting on Management and Evaluation of Civil Registration and Vital Statistics Systems, New York, UNHQ, 20-24 February 2017

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):

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

Sources for denominator:

- Census
- Population projection
- Population register

Advantages: Simple, inexpensive and fast

Limitations:

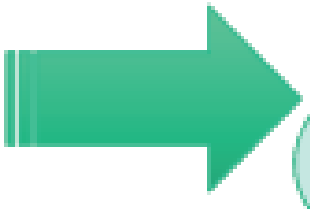
- Number of expected events not known, outdated or inaccurate
- No information about the *causes* of under-registration

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 indpendent 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

		Source Y: Yes	Source Y: No	Total
Source Z	Yes	a	b	$a+b=Z_0$
	No	c	x	
	Total	$a+c=Y_0$		$N=a+b+c+x$

Estimated values		Maximum likelihood estimator (MLE)
Unobserved cell:	$\hat{x}$	bc/a
Completeness of source Y:	$\hat{Y}_c$	$a/(a+b)=a/Z_0$
Completeness of source Z:	$\hat{Z}_c$	$a/(a+c)=a/Y_0$
Total population:	$\hat{N}$	$a+b+c+(bc/a)$ or, $(a+b)(a+c)/a$

$$\text{Completeness of Y} = \frac{a+c}{a+b+c+x}$$

$$\text{Completeness of Z} = \frac{a+b}{a+b+c+x}$$

- Hook, E.B. and R.R. Regal, *Capture-recapture methods in Epidemiology: Methods and limitations*. Epidemiologic Reviews, 1995. **17**(2): p. 243-64.

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

Exmple of record linkage of CR events: Viet Nam 2009

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

Expert Group Meeting on Management and Evaluation of CRVS systems, New York, 20-24 February, 2017

- 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

Australian National University

Matching results

	Regions	Total in reconciled list	CHC	⟺	Population Dep	Justice system	Other
1	Ha Noi	2304	1723 (75%)		1580 (69%)	1669 (72%)	720 (31%)
2	Thai Nguyen	1185	999 (85%)		210 (18%)	183 (15%)	85 (7%)
3	Hue	2221	1768 (78%)		1043 (47%)	1311 (59%)	777 (35%)
4	Ho Chi Minh	2453	435 (18%)		571 (23%)	1871 (76%)	202 (8%)
5	Can Tho	1758	872 (49%)		758 (43%)	1081 (62%)	535 (30%)

- *A death could be recorded in more than one system*
- ⟺ *= interdependence*

Table 1. **Age- and sex-specific observed and estimated deaths[a] and completeness of mortality data, Viet Nam, 2009**

Sex-specific age group (in years)	Sample	a[b]	b[c]	c[d]	x[e]	Other source only	Deaths		Per cent completeness[f] (95% CI)
							Observed (a + b + c + additional)	Estimated (a + b + c + x)	
Males	1 239 937	2138	1984	1363	1265	215	5700	6750	81.2 (74.1–87.1)
15–59	873 727	903	873	597	577	92	2465	2950	80.4 (72.2–80.3)
60–74	53 985	453	414	274	250	38	1179	1391	82.0 (74.9–87.9)
75+	22 852	710	629	453	401	77	1869	2193	81.7 (74.7–87.4)
Females	1 309 462	1572	1413	1026	922	181	4192	4933	81.3 (74.4–87.1)
15–59	929 773	373	350	251	236	56	1030	1210	80.5 (72.5–87.1)
60–74	72 999	342	271	213	169	41	867	995	83.0 (75.4–89.0)
75+	37 684	812	734	539	487	80	2165	2572	81.0 (73.9–87.0)

CI, confidence interval.

[a] Age- and sex-specific deaths deviate slightly from the totals reported in the text because 27 deaths had no age data.

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

[c] Number of deaths reported by the CHC/CPFPC but not by the Justice Department.

[d] Number of deaths reported by the Justice Department but not by the CHC/CPFPC.

[e] Estimated number of deaths missing from CHC/CPFPC and Justice Department sources.

[f] Proportion of estimated deaths derived from the list obtained by reconciling the Justice Department and combined CHC/CPFPC lists. Derived with the following formula: $(a + b + c) \div (a + b + c + x) \times 100$

Table 2. **Summary sex-specific measures of mortality based on WHO, UNPD and Viet Nam census data for the 16 study provinces, Viet Nam, 2009**

Data source	Per cent data completeness (95% CI)	Life expectancy at birth (95% CI) [e0]	Risk of death in children under 5 (deaths per 1000) [5q0]	Risk of death at ages 15–59 (deaths per 1000) [45q15]	Remaining years of life at age 60 [e60]
Males					
Surveillance sample (unadjusted)	–	74.4 (74.0–74.8)	7.4	163	20.9
Surveillance sample (adjusted)[a]	81.1 (74.1–87.1)	70.4 (70.1–70.8)	24.6[c]	199	19.4
Viet Nam census (unadjusted)	–	75.2 (75.0–75.4)	10.9	157	22.1
Viet Nam census (adjusted)[b]	65.6 (–)	68.8 (68.6–69.0)	16.5	230	17.9
WHO (2009)	NA (modelled)	69.8 (–)	24.6	173	17
UNPD (2005–2010)	NA (modelled)	72.3 (–)	No data	139	No data
Females					
Surveillance sample (unadjusted)	–	82.3 (82.0–82.7)	5.8	57	25.1
Surveillance sample (adjusted)[a]	81.3 (74.4–87.1)	78.7 (78.4–79.0)	22.5[c]	71	23.6
Viet Nam census (unadjusted)	–	85.2 (85.0–85.6)	8.8	50	28.4
Viet Nam census (adjusted)[b]	57.8 (–)	77.8 (77.5–78.0)	15.7	86	22.4
WHO (2009)	NA (modelled)	74.5 (–)	22.6	107	19.8
UNPD (2005–2010)	NA (modelled)	76.2 (–)	No data	96	No data

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

[a] Adjusted for data incompleteness and mortality in children under 5 years of age.

[b] Adjustment by the Preston-Coale method.

Conclusion

Completeness of death records was estimated to be 81%.

Viet Nam is undergoing the epidemiological transition.

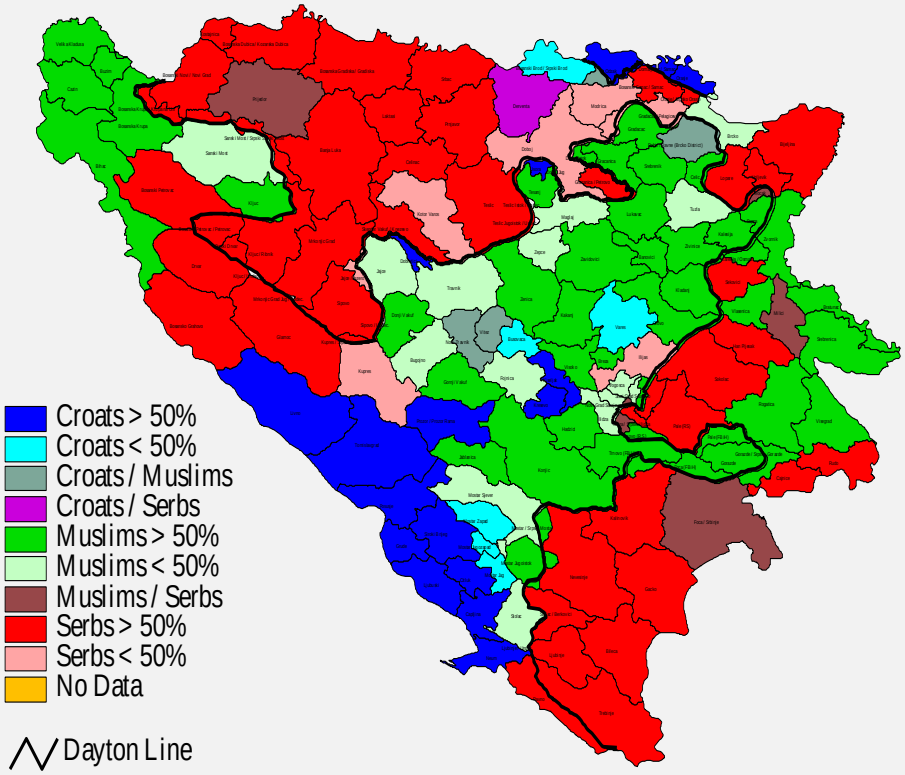
Data are relatively complete but could be further improved through strengthened local collaboration.

Medical certification for deaths in hospitals, and shorter recall periods for verbal autopsy interviews, would improve cause of death ascertainment.

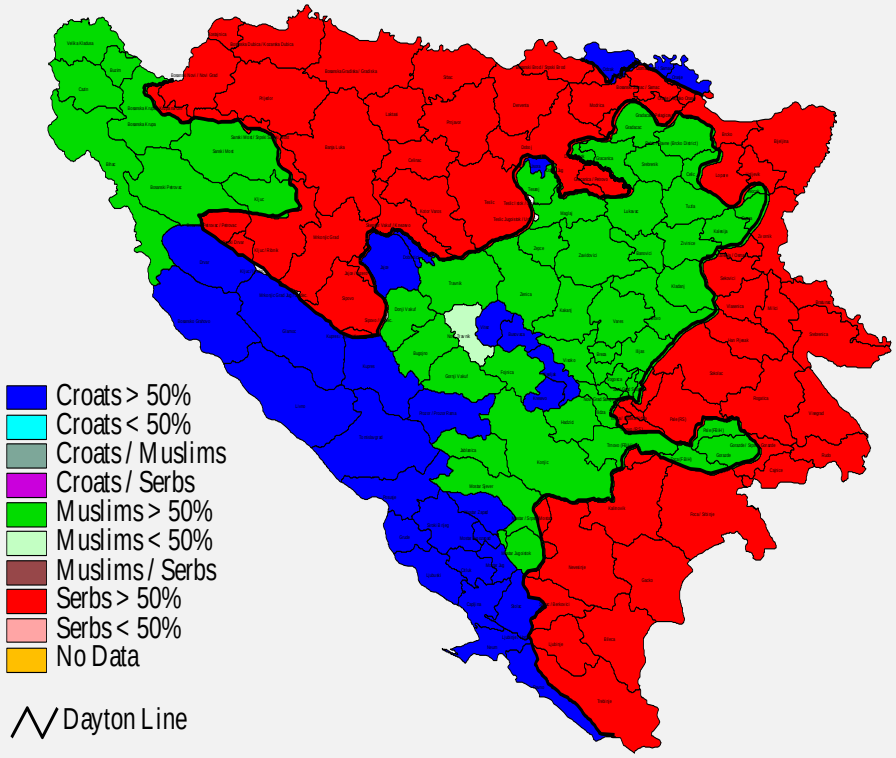
Ethnic majority of the municipalities of Bosnia-Herzegovina

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

Census 1991



Voters' list 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 erors 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, owership of property includiong 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 simila 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 mehods
2. Direct methods

Discuss the advantages and drawbacks of each method