Australian National University

Contemporary methods for assessing coverage and completeness of CRVS systems with national examples

Chalapati Rao

Department of Global Health

Research School of Population Health

Outline

- Definitions of coverage and completeness
- Coverage
    - Types and examples
- Completeness
  - Methods
    - Aggregated data analysis (indirect methods)
    - Record linkage and matching (direct methods)
    - Typology of study designs / historical overview
  - Examples
    - Viet Nam, Indonesia, Oman

Indicators of generalizability

- Coverage: the population to which the CRVS laws/processes are applicable, or the population actually represented in the statistics

- Completeness: the proportion of vital events actually registered out of the total ‘true’ number of events in the ‘covered’ population

Given their inherently different purposes

all CR may not generate meaningful vital statistics; AND

All reliable vital statistics may not be based on civil registration data

However, a convergence is desirable

Dimensions of coverage

- Administrative coverage:
  - Populations defined in legal framework of civil registration
  - coverage of processes for compilation of vital statistics
  - E.g. coverage of laws or regulations regarding stillbirths / deaths among expatriates/refugees/ coverage of medical certification of cause of death

- Sample coverage: largely applicable to vital statistics reports
- While the broad goal of civil registration is universal coverage, establishment of vital statistics systems will require an incremental approach
- e.g the Chinese Ministry of Health vital statistics system covered 60 million in 1985, 110 million in 2000, and 230 million in 2012. Currently a nationally representative sample

- Sample coverage has importance for expanding CRVS systems, and should be closely monitored to track progress in establishment and development over time

Reporting coverage

- Reporting coverage:
- establishes the performance of the CRVS system in terms of the proportion of primary registration units submitting statistical reports for each reference period of time (month/quarter/year)

Statement 6: Number of registration units and level of reporting by place, India, States and UTs, 2014

Sl. No.	States/UTs	Registration Units (Number)			Level of Reporting (%)		
		Rural	Urban	Total	Rural	Urban	Total
	India	**247391**	**6291**	**265067**	**92.2**	**95.6**	**92.6**
	States						
1.	Andhra Pradesh	12922	109	13031	87.7	100.0	87.8
2.	Arunachal Pradesh	173	31	204	N.A	N.A	55.9
3.	Assam	602	93	695	100.0	100.0	100.0
4.	Bihar	9048	217	9265	34.7	79.7	35.8
5.	Chhattisgarh	N.A	N.A	11362	N.A	N.A	100.0

Reporting coverage

- Registration area with pop = 20,614
- Expected deaths (as per crude death rate 6.4/1000) = 132
- Reported deaths = 72
- Monthwise reporting pattern

Village id	1	2	3	4	5	6	7	8	9	10	11	12
1403011002									1		1	
1403011004			1				1	1				
1403011005		1	1	1	1			1	2	3		
1403011006						1						
1403011009						1						
1403011010		2		2	3		1	3	10	9	6	9
1403011011					1						1	
1403011012	1	1		1			1	1				
8												

Missing reports should be routinely followed up to ensure data quality

- Completeness = $\frac{number\ of\ registered\ events}{estimated\ number\ of\ total\ events} \times 100$

Analysis of completeness

- Aggregated data analysis (indirect methods)
- Types of aggregated analysis
- Comparisons of aggregated numbers with data from alternate sources
-
- E.g census enumerations; health service records, estimates derived from rates borrowed from other populations etc

- Overall, comparisons of aggregated data not a satisfactory method for evaluating completeness; since both sources may be incomplete; or in case of other populations, may have different age-structures/ epidemiological patterns of mortality, hence violating the comparison

Demographic analysis of aggregated data

- Indirect demographic analysis using models of population growth/ change to derive an expected number of deaths in the study population
- Observed vital events divided by expected deaths to derive proportion of completeness
- Models based on assumptions
  - accurate population counts;
  - no migration;
  - accurate age-reporting of population and deaths;
  - completeness invariant by age
  - In some methods – stable population (constant fertility and mortality in preceding decades)

- Difficult to fulfil assumptions; particularly in regard to accuracy of population counts, age-reporting, and migration
- Vastly differing measures from different methods, plus considerable uncertainty **(±25%)**

Record linkage and matching (direct methods)

- Involve linkage of individual records across different data sources, and are also referred to as dual record system studies; or matching studies

- Record linkage can be used for reconciling data across different sources, and as a basis for dual record system (DRS) analysis to estimate completeness

- DRS method can be defined as a method for estimating total population size (total deaths) when a full count of the total population is unavailable or unfeasible, but when there are two or more independent sources of information on individual members of the population

Conceptual basis

- Individuals are ‘captured’ from their record in one data source and ‘recaptured’ when the record for the same individual is matched in the second source

- Matching across key variables:
  - Personal details (UID/Name/age/sex)
  - geographical variables
  - Event details - Date of birth/death/registration

- Linkage produces 3 sets i.e Matched records; plus sets of unique records in either source

- Linkage allows data reconciliation to derive a larger set of empirical records than from either source

Computation

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, $(a + b)(a + c)/a$

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

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 DRS methods

- No ‘out-of-scope’ events in either source
  - All cases in each source are correctly diagnosed (true events)
  - All cases from each source are in the correct and same time-space frame
    - year of death/ address
    - Correct application of definitions of residence status
    - Study population is closed (no in/out migration)
- Homogeneity of capture probability in each source (in each data source each individual has equal probability of being captured)
  - No selective exclusion of specific sub groups - gender/age/ethnicity/geography/SES
- Independence of data sources (capture in one source does not influence capture in the second source)
- Accuracy of matching procedures and matching outcomes (no erroneous matches or erroneous non-matches)

Typology of data sources for record linkage studies

Type of data collection	Primary source[1]	Secondary source[2]	Remarks
Continuous recording systems			
Civil registration	Yes		• Optimal source • annual data on routine basis
Alternate registration	Yes	Yes	• Health system vital records e.g Vietnam, Fiji • Church records in Christian societies
Sample registration	Yes	Can serve as a secondary source for evaluating CRVS	• Best alternative to CRVS • Indian SRS (ref) • Chinese DSP (ref) • Bangladesh SVRS (ref)
Special registration	Yes	Can serve as a secondary source for evaluating CRVS or SRS	• E.g. Health and Demographic Surveillance Sites in several countries (INDEPTH Network) (ref)
Age based registers		Yes	• Maternal/child health • senior citizens /pensioners databases
Disease surveillance systems		Yes	• tuberculosis • cancers • injuries • stroke
Periodic data collections			
Census (total population)	Yes	Yes	• Optimal 2nd data source (national coverage)
National sample surveys		Yes	• Inter censal surveys • DHS program • WHO NCD surveillance (STEPS) surveys • UNICEF MICS surveys etc
Special surveys designed to assess completeness		Yes	• Evaluation surveys for sample/special registration • sporadic research based examples

1 = data source for which completeness needs to be evaluated

2 = data source which will be used to evaluate completeness of the primary source

Key concepts in DRS analysis

- There should be compatibility of data sources to minimize out of scope events
- Availability of multiple variables for matching
  - Enhances matching potential / validation of matching
- Assurance of data quality
  - Completeness and accuracy of all variables for each death record in each data source
- Matching procedures should be clearly defined
  - Manual / electronic / combination
  - Rules for matched cases – explicit rules vs implicit rules
  - Tolerable limits for specific criteria / deterministic matching / probabilistic matching
  - Mechanisms for field verification of matched/partially matched/ unmatched cases
- Analytical approach – reconciliation/DRS/hybrid approach
- Assessment of DRS conditions (potential for bias)
  - Description of design and data collection process / statistical evaluation
- Measure error of completeness estimate from sampling and bias
- Ethics and data confidentiality

Evaluating bias in completeness estimates

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

- RMSE of completeness estimate: RMSE = $\sqrt{variance + bias^2}$
- Three sources of bias
  - ‘out-of-scope-bias’: results in under estimate of true matches; leading to an ↓ underestimate of completeness; and ↑ overestimate of the vital rate
  - Response correlation bias (from communication/data sharing between sources i.e lack of statistical independence): results in overestimate of true matches; leading to an over estimate of completeness; and underestimate of the vital rate
  - Matching bias: expressed as the *net matching error* which is the difference between the erroneous matches and erroneous non matches.
    - Net matching error is positive = same effect as response correlation bias;
    - if net matching error is negative = effect as ‘out of-scope’ bias
  - ***Due to varying directions; net bias is usually less than any individual source of bias***

Historical review of record linkage completeness studies

Study type	Countries	Remarks
Special registration with periodic surveys	1960-1975 Pakistan, Egypt, Liberia, Malawi, Philippines, Columbia, Morocco, Turkey, Kenya 2006/07 Indonesia	• Time bound projects (-3 years) in listed countries during 1960-1975; USAID PGE program • Tested range of data collection e.g direct household contact; use key informants; combinations • Tested range of recall periods (1,3,6, 12 months) • Completeness; estimated by CD method (ranging from 53 to 90% settings); no 95% CI • Crude birth/death rates adjusted for completeness; no age-specific rates reported; • Indonesian studies in 2006-2007 as sentinel sites, later transformed into national SRS; completeness for 2006 by data reconciliation (no 95% CI); in 2007 by CD method (with 95% CI)
National sample registration with periodic surveys	India – SRS since 1970 Bangladesh-SVRS - 1980 China DSP since 1990 Indonesia since 2014	• India & Bangladesh – continuous recording in sample clusters with total coverage in routine 6 monthly surveys; data reconciliation used to measure mortality, completeness not routinely reported • China – continuous recording in sample clusters with triennial sample completeness surveys; completeness estimated by CD method, results reported with uncertainty intervals for • Indonesia – completeness survey of 2014 discarded due to data quality issues; new survey 2017
Civil registration with periodic data sources	Thailand (2006) Oman (2010) Philippines (2012-14)* Palestine (2017)*	• Thai study involved civil registration and intercensal survey; completeness by CD method, no 95%CI • Oman study involved civil registration and national census; completeness by CD method with 95% CI • Philippines and Palestine – civil registration and census (studies yet to be implemented)
Multiple sources with overlapping recall periods	Philippines 2006/7 Viet Nam 2008/9 Kiribati (2001-2009) Tonga (2000-2009)	• Philippines study – Civil registration; health system; parish records; CD method; with 95%CI by Max Lik Est • Viet Nam study – civil registration; health system; peoples committee plus additional partial sources; completeness by variant of CD method with 95% Ci (by bootstrapping method) • Kiribati – civil registration; health information system; reproductive surveillance, data reconciliation; no CI • Tonga –civil registration; health information system; completeness by CD method; No 95% CI
Civil registration with HDSS	South Africa 2006-09	• Civil registration and HDSS; electronic linkage with deterministic & probabilistic matching; completeness not measured due to 'out-of-scope' coverage

Example: Viet Nam 2009

- Study population – 192 communes; 2.6 million pop
- Data sources – Commune health station/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 cell 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 mortality estimates for burden of disease analysis

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*

Viet Nam 2009

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: $(a + b + c) \div (a + b + c + x) \times 100$.

- Hoa, N.P., Rao C et al., *Mortality measures from sample-based surveillance: evidence of the epidemiological transition in Viet Nam.* Bulletin of the World Health Organization, 2012. **90**(10): p. 764-772.

Example: Oman 2010

- Acknowledgement: This study was conducted by Dr Salah al Muzahmi as part of his PhD thesis titled: Mortality patterns in Oman: A demographic and epidemiological review. PhD awarded by University of Queensland, December 2015.

- Study covering entire population of Omani nationals (excl expats)

- Data sources – Health system death notifications 2010 (6036 deaths), Census 2010 (5400 deaths)
  - Census conducted on 18 Dec 2010 with one year recall of deaths including recording of date of death

- Three rounds of matching – electronic plus manual

- Analysis – capture-recapture adjustment of completeness of death notification data

Data quality – missing variables

Table 1 Missing/duplication of the primary variables.

Items	Birth and death notification system database	Census
Total records	6,039	5,400
Missing date of death	0	0^
Duplicates	3	19
Missing age	652	0
Missing sex	18	0
Missing governorate	457	0
Missing *Wilayat*	535	0
Missing nationality	18	0
Missing *Wilayat* and governorate	457	0
Records used in matching	6,036	5,381

^ Date of death in the census dataset is divided into three variables (year, month and day); there are 153 records with unknown day and month

Results of matching

FIRST ROUND

Table 14 Summary findings of the first phase of the matching process

	Records
Matched records in the first round	568 (9.5%)
Not matched from Death notification	5468
Missing age	500
Missing governorate	435
Missing *wilayat*	502
Missing village/locality	1022

SECOND ROUND

Table 15 Summary findings of the phase two of matching process

	Records
Matched according to age	2,983
Matched according to date of death	3,078
Matched according to gender	3,252
Matched according to *wilayat*/village	3,284
Total matched records on all variables	2,983 (49.5%)

THIRD ROUND

Table 17 Summary findings of the third round of matching process

	Records
Matched records after third corrections	4,819 (79%)
Not matched	1,217
Reasons for un-matched records*	
Missing age	192
Missing governorate	168
Missing *wilayat*/village	179
Under-recorded events in census	650

* Some records remained unmatched due to > 1 missing variable

Reasons for mismatch

- Variations in
- Spellings
- age
- address
- date of death

Correction strategy

- Corrected spellings, address variables,
- 5 year margin for age, if matched on other variables
- One month margin for date, if matched on other variables

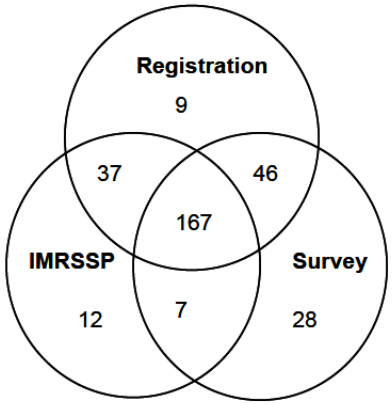
Correction strategy

- Field verification of variables for unmatched cases from health records
- 10 year margin for age for deaths above 65 years, if matched on other variables
- Two month margin for date, if matched on other variables

Example 3: Indonesia (2007)

- Central Java – record linkage/matching across three sources (health system, vital registration, independent survey)
- Independent survey and record linkage/matching conducted only in a sample of villages from the overall study population
- Completeness of health system data calculated as a proportion of total deaths obtained from the reconciled list of unique deaths

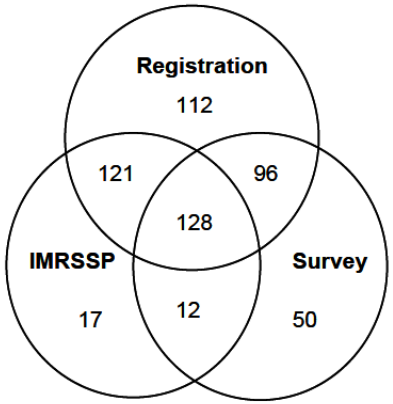
PEKALONGAN



Total deaths = 306

Completeness = 73%

SURAKARTA



Total deaths = 536

Completeness = 55%

Limitations of DRS methods

- Several conditions for record linkage difficult to fulfil (e.g. absence of out-of-scope events, homogenous capture probability; statistical independence of data sources,; accuracy of matching)

- potential bias in the completeness estimate due above

- Further, there is also sampling error / stochastic variation; which contribute to uncertainty in the completeness estimate

- In addition, there are considerable logistical challenges in implementing record linkages studies in terms of costs/ manpower/ technical challenges in matching, evaluation of bias etc

Strengths of DRS methods

- Essentially the major conditions / assumptions of record linkage and DRS methods are statistical as compared to the demographic assumptions for indirect techniques (related to underlying fertility/mortality/population growth patterns in the study population)
- The data collection procedures allow assessment of bias and error, hence enabling a more informed assessment of uncertainty of the completeness estimate
- Findings enable completeness assessment and also help identify systemic weaknesses in the registration system, including specific population sub groups
- Involvement of local staff in matching helps build awareness and capacity for strengthening registration
- Age specific measures of completeness
- Data reconciliation especially from additional fragmentary sources helps fill data gaps in cause of death information

Reasons for renewed interest

- Availability of computerised registration datasets as well as computerisation of periodic data collections (censuses, surveys); which will increase going forward
- Improved data quality of recorded variables used in linkage (name spellings; address variables, age, date of death etc)
- Increasing availability of Unique Identifiers which are invaluable for linkage
- Electronic linkage vastly reduces logistical challenges of manual matching
- Explicit rules and probabilistic approach using computerised datasets can be applied to test a range of scenarios and judge cut points for specific criteria
- Routine application of DRS method in India and China serve as robust examples of their general acceptability

Conclusions and recommendations

- Promote routine linkage, matching and reconciliation of data across different sources at local level, to augment completeness of civil registers [IRAN, BRAZIL]

- Hierarchy of study designs for record linkage completeness (based on sample size; potential for meeting condition of independence; cost considerations; sub group analysis)
    - CRVS with census based recall of deaths
    - CRVS with intercensal survey / nationally representative sample survey/special survey
    - SRS with periodic special surveys
    - Special registration in targeted surveillance sites with special surveys