



Australian  
National  
University

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

**Analysis of completeness of vital statistics from civil  
registration systems**

Chalapati Rao

Department of Global Health

Research School of Population Health

- Background
- Statistical methods
- Assumptions
- Typology of study designs / historical overview
- Examples
  - Viet Nam, Indonesia, Oman
  - Other countries
- Limitations/advantages of record linkage studies

- Generalizability
  - Coverage
  - Completeness
    - Aggregated data analysis (indirect methods)
    - Record linkage and matching (direct methods)
- Accuracy
  - Reliability
  - Validity – particularly of registered causes of death
- Policy relevance
  - Timeliness
  - Sub national data availability (geographical disaggregation)

- Comparisons of absolute numbers of registered events / event rates for same population (e.g. registration unit) over time (monthly/quarterly/annually)
  - Useful as a monitoring tool on a routine basis, to promote reporting compliance
- Comparisons between populations with similar characteristics
- Comparisons of aggregated numbers with with data from alternate sources (e.g. census enumerations; health service records 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

- 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%)**

# Background

- Record linkage studies are being considered as an alternative to indirect demographic techniques to measure completeness of death registration
- Involve linkage of 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



- Completeness of either source could be computed as a proportion of the total reconciled deaths

### ALSO

- record linkage permits the application of another statistical procedure (based on certain conditions) to estimate deaths not captured by either source
- This estimate of missed events added to the reconciled deaths to derive an estimate of total deaths
- Subsequently, completeness of either source derived as a proportion of deaths recorded in it out of the estimate of the total deaths
- Other 'hybrid' models for estimating completeness, involving multiple data sources/partial data sources etc





# Computation

TABLE 1. Two-source model

		Source Y		Total
		Yes	No	
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.

- No 'out-of-scope' events in either source
  - All cases in each source are correctly diagnosed (true deaths)
  - 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)

- Continuous recording systems
  - Vital registration systems
  - Sample registration systems (India/China/Bangladesh/Indonesia)
  - Health system records / parish registers/ ‘population committee’ registers
  - Specific disease/program registers (TB, MCH), police records (injuries)
  - social sector or insurance databases
  - Special registration sites (HDSS/INDEPTH network) – **limited generalizability**
- ↑ **likelihood of ‘dependence’ between multiple continuous systems in same popln**
- Periodic/one-off cross-sectional data collection systems
  - Censuses / intercensal surveys
  - Periodic household surveys – DHS, STEPS, MICS, SES surveys
  - ‘completeness’ surveys – China/India/Bangladesh

Type of data collection	Primary source <sup>1</sup>	Secondary source <sup>2</sup>	Remarks
<b>Continuous recording systems</b>			
Civil registration	Yes		<ul style="list-style-type: none"> <li>• Optimal source</li> <li>• annual data on routine basis</li> </ul>
Alternate registration	Yes	Yes	<ul style="list-style-type: none"> <li>• Health system vital records e.g Vietnam, Fiji</li> <li>• Church records in Christian societies</li> </ul>
Sample registration	Yes	Can serve as a secondary source for evaluating CRVS	<ul style="list-style-type: none"> <li>• Best alternative to CRVS</li> <li>• Indian SRS (ref)</li> <li>• Chinese DSP (ref)</li> <li>• Bangladesh SVRS (ref)</li> </ul>
Special registration	Yes	Can serve as a secondary source for evaluating CRVS or SRS	<ul style="list-style-type: none"> <li>• E.g. Health and Demographic Surveillance Sites in several countries (INDEPTH Network) (ref)</li> </ul>
Age based registers		Yes	<ul style="list-style-type: none"> <li>• Maternal/child health</li> <li>• senior citizens /pensioners databases</li> </ul>
Disease surveillance systems		Yes	<ul style="list-style-type: none"> <li>• tuberculosis</li> <li>• cancers</li> <li>• injuries</li> <li>• stroke</li> </ul>
<b>Periodic data collections</b>			
Census (total population)	Yes	Yes	<ul style="list-style-type: none"> <li>• Optimal 2<sup>nd</sup> data source (national coverage)</li> <li>• Inter censal surveys</li> </ul>
National sample surveys		Yes	<ul style="list-style-type: none"> <li>• DHS program</li> <li>• WHO NCD surveillance (STEPS) surveys</li> <li>• UNICEF MICS surveys etc</li> </ul>
Special surveys designed to assess completeness		Yes	<ul style="list-style-type: none"> <li>• Evaluation surveys for sample/special registration</li> <li>• sporadic research based examples</li> </ul>

<sup>1</sup> = data source for which completeness needs to be evaluated

<sup>2</sup> = data source which will be used to evaluate completeness of the primary source

- **Scope of analysis** e.g national / sub national measures; by age; pop sub groups
- **Availability/choice of primary & secondary data sources**
- **Reference time period of analysis**
- **Matching process**
  - Manual/electronic
  - Deterministic/probabilistic/implicit rules
- **Statistical procedures**
  - Data reconciliation
  - Use of multiple parallel sources or partial data sources
  - DRS method ( 2source/multiple source models)
  - Hybrid models

- 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  $\downarrow$  underestimate of completeness; and  $\uparrow$  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***



- Periodic data collections (except censuses) are based on samples, and usually with cluster design
- Some study designs (e.g. DSP China) involves sampling in both data sources
- Sources of variance
  - Sample size
    - Measuring completeness for specific sub groups (sex, age, geography etc reduces the sample and therefore precision of the estimate
  - Cluster size and characteristics – need to account for design effect





- In 1949, CD proposed that SE of completeness =  $\sqrt{Nq_1q_2/p_1p_2}$
- Where N = total number of events estimated by the method (Table 1)
  - $p_1$  = the probability that an event is recorded in data source 1
  - $p_2$  = the probability that an event is recorded in data source 2
  - $q_1$  = the probability that an event is missed in data source 1
  - $q_2$  = the probability that an event is missed in data source 2
- Assuming that
  - There is true statistical independence between the two data sources, and zero matching bias or out-of-scope events; and no variance from sampling etc
- Subsequently, various scientists ((Seltzer & Adlakha 1973; Greenfield 1976; Chandrasekar & Deming 1981; Nour 1982, Ayhan 2000, El Khorazaty 2000) proposed methods to estimate bias arising from lack of statistical independence
- Nour (1982) illustrates computation with a practical example with data from Malawi; and El Khorazaty illustrates a practical example with Egyptian data for 1974/75

- Variations in design
- Matching all records from two sources of the study population – e.g sample registration system in India; Viet Nam study, Oman, Tonga
- Matching of records in only a sample of the study population – China, Thailand (2006), Indonesia, Malaysia (1995)
- Variations in method for computation of completeness
  - Data reconciliation after matching; no adjustment for cases potentially missed by both sources (Indian SRS; Tonga)
  - Data reconciliation after matching, with adjustment for potentially missed cases – Vietnam, Indonesia (Java)
  - Matching followed by adjustment, no data reconciliation – China, Thailand, Indonesia (other locations), Oman, Malaysia (1995)



# Historical review of record linkage completeness studies

Study type	Countries	Remarks
Special registration with periodic surveys	<u>1960-1975</u> Pakistan, Egypt, Liberia, Malawi, Philippines, Columbia, Morocco, Turkey, Kenya <u>2006/07</u> Indonesia	<ul style="list-style-type: none"> <li>• Time bound projects (-3 years) in listed countries during 1960-1975; USAID PGE program</li> <li>• Tested range of data collection e.g direct household contact; use key informants; combinations</li> <li>• Tested range of recall periods (1,3,6, 12 months)</li> <li>• Completeness; estimated by CD method (ranging from 53 to 90% settings); no 95% CI</li> <li>• Crude birth/death rates adjusted for completeness; no age-specific rates reported;</li> <li>• 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)</li> </ul>
National sample registration with periodic surveys	India – SRS since 1970 Bangladesh-SVRS - 1980 China DSP since 1990 Indonesia since 2014	<ul style="list-style-type: none"> <li>• India &amp; Bangladesh – continuous recording in sample clusters with total coverage in routine 6 monthly surveys; data reconciliation used to measure mortality, completeness <u>not</u> routinely reported</li> <li>• China – continuous recording in sample clusters with triennial sample completeness surveys; completeness estimated by CD method, results reported with uncertainty intervals for</li> <li>• Indonesia – completeness survey of 2014 discarded due to data quality issues; new survey 2017</li> </ul>
Civil registration with periodic data sources	Thailand (2006) Oman (2010) Philippines (2012-14)* Palestine (2017)*	<ul style="list-style-type: none"> <li>• Thai study involved civil registration and intercensal survey; completeness by CD method, no 95%CI</li> <li>• Oman study involved civil registration and national census; completeness by CD method with 95% CI</li> <li>• Philippines and Palestine – civil registration and census (studies yet to be implemented)</li> </ul>
Multiple sources with overlapping recall periods	Philippines 2006/7 Viet Nam 2008/9 Kiribati (2001-2009) Tonga (2000-2009)	<ul style="list-style-type: none"> <li>• Philippines study – Civil registration; health system; parish records; CD method; with 95%CI by Max Lik Est</li> <li>• 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)</li> <li>• Kiribati – civil registration; health information system; reproductive surveillance, data reconciliation; no CI</li> <li>• Tonga –civil registration; health information system; completeness by CD method; No 95% CI</li> </ul>
Civil registration with HDSS	South Africa 2006-09	<ul style="list-style-type: none"> <li>• Civil registration and HDSS; electronic linkage with deterministic &amp; probabilistic matching; completeness not measured due to 'out-of-scope' coverage</li> </ul>

## 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

Table 1. Age- and sex-specific observed and estimated deaths<sup>a</sup> and completeness of mortality data, Viet Nam, 2009

Sex-specific age group (in years)	Sample	a <sup>b</sup>	b <sup>c</sup>	c <sup>d</sup>	x <sup>e</sup>	Other source only	Deaths		Per cent completeness <sup>f</sup> (95% CI)
							Observed (a + b + c + additional)	Estimated (a + b + c + x)	
<b>Males</b>	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)
<b>Females</b>	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.

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



# Adjusted mortality indicators

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]
<b>Males</b>					
Surveillance sample (unadjusted)	–	74.4 (74.0–74.8)	7.4	163	20.9
Surveillance sample (adjusted) <sup>a</sup>	81.1 (74.1–87.1)	70.4 (70.1–70.8)	24.6 <sup>c</sup>	199	19.4
Viet Nam census (unadjusted)	–	75.2 (75.0–75.4)	10.9	157	22.1
Viet Nam census (adjusted) <sup>b</sup>	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
<b>Females</b>					
Surveillance sample (unadjusted)	–	82.3 (82.0–82.7)	5.8	57	25.1
Surveillance sample (adjusted) <sup>a</sup>	81.3 (74.4–87.1)	78.7 (78.4–79.0)	22.5 <sup>c</sup>	71	23.6
Viet Nam census (unadjusted)	–	85.2 (85.0–85.6)	8.8	50	28.4
Viet Nam census (adjusted) <sup>b</sup>	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.

<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.

<sup>c</sup> WHO estimate.



## 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



Table 1 Variables by source

Variable	BDNS database	Census 2010 database
Notification number	✓	
Reported institution	✓	
Name of deceased	✓	
Name/tribe name of applicant*	✓	✓
<u>Governorate/region</u>	✓	✓
<u>Wilayat (district)</u>	✓	✓
<u>Town/village</u>	✓	✓
Locality or compound		✓
<u>Sex</u>	✓	✓
<u>Date of death</u>	✓	✓
<u>Age at death</u>	✓	✓
Date of birth	✓	

\* The applicant for death registration, as well as the census respondent, is assumed to be from the same household and tribe as the deceased. Hence the tribe name of the deceased would be the same as the tribe name of the BDNS applicant as well as the census respondent. Hence, the tribe names were used in the matching process.

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 <sup>^</sup>
Duplicates	3	19
Missing age	652	0
Missing sex	18	0
Missing governorate	457	0
Missing <i>Wilayat</i>	535	0
Missing nationality	18	0
Missing <i>Wilayat</i> and governorate	457	0
Records used in matching	6,036	5,381

<sup>^</sup> 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

## 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 <u>wilayat</u>	502
Missing village/locality	1022

### Reasons for mismatch

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

## 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 <u>wilayat</u> /village	3,284
Total matched records on all variables	2,983 (49.5%)

### 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

## 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 <u>wilayat</u> /village	179
Under-recorded events in census	650

\* Some records remained unmatched due to > 1 missing variable

### 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



# Final Results

**Table 18 Overall completeness of reporting of deaths census 2010**

		Yes	No	Total
Death notification system	Yes	4,819	1,217	6,036
	No	562	142	644
	<b>Total</b>	<b>5,381</b>	<b>1,359</b>	<b>6,740</b>

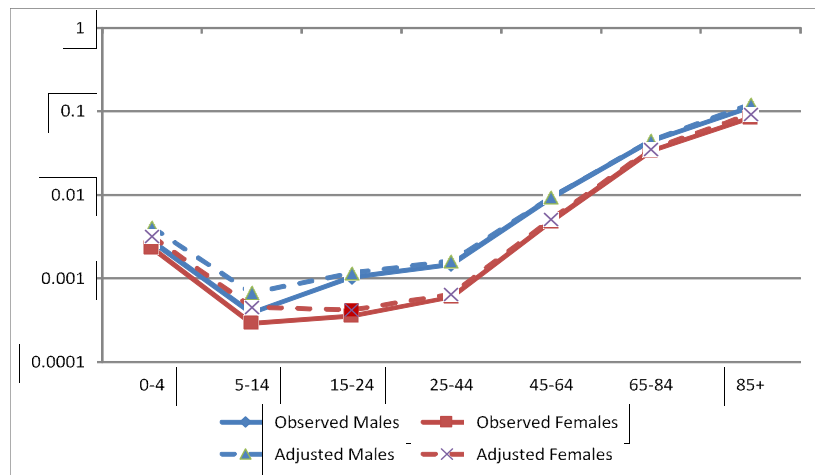
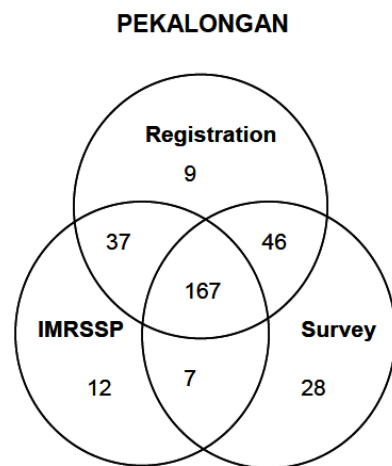


Figure 1: Log plot of Age and sex specific death rate (Observed vs Adjusted), Oman 2010

Sex	Governorate	Completeness rate (95% CI)	Adjusted LE in years (95% CI)
Males	Ad Dhakhliah	92 (90 - 95)	73.7 (72.4 - 74.9)
	<u>Ad Dhahira</u>	<u>86 (83 - 91)</u>	72.1 (70.4 - 73.9)
	Al Buraymi^	81 (71 - 91)	81.0 (77.1 - 84.7)
	<u>Dhofar</u>	<u>87 (82 - 91)</u>	75.3 (73.5 - 77.0)
	Musandam^	83 (72 - 95)	83.6 (77.5 - 89.6)
	<u>Muscat</u>	<u>87 (84 - 90)</u>	74.2 (73.1 - 75.2)
	N Al Batinah	93 (91 - 95)	73.0 (71.9 - 74.0)
	N. Sharqiyah	98 (86 - 93)	74.1 (72.3 - 75.8)
	S Al Batinah	91 (89 - 94)	73.3 (72.0 - 74.6)
	S. Sharqiyah	92 (89 - 95)	77.0 (75.0 - 79.0)
<b>Total</b>	<b>90 (89 - 91)</b>	<b>73.7 (73.3 - 74.2)</b>	
Females	Ad Dhakhliah	91 (88 - 94)	78.8 (77.4 - 80.2)
	Ad Dhahira	87 (82 - 93)	82.1 (79.7 - 84.3)
	Al Buraymi^	84 (72 - 97)	83.4 (79.2 - 87.6)
	<u>Dhofar</u>	<u>88 (83 - 93)</u>	80.2 (78.4 - 82.0)
	Musandam^	67 (46 - 87)	79.6 (76.1 - 83.2)
	<u>Muscat</u>	<u>82 (78 - 86)</u>	80.3 (79.1 - 81.5)
	N Al Batinah	97 (95 - 98)	80.6 (79.3 - 82.0)
	N. Sharqiyah	90 (86 - 95)	79.3 (77.5 - 81.3)
	S Al Batinah	90 (86 - 93)	81.5 (80.0 - 83.2)
	S. Sharqiyah	89 (84 - 94)	86.3 (84.1 - 88.4)
<b>Total</b>	<b>89 (88 - 90)</b>	<b>80.0 (79.5 - 80.4)</b>	

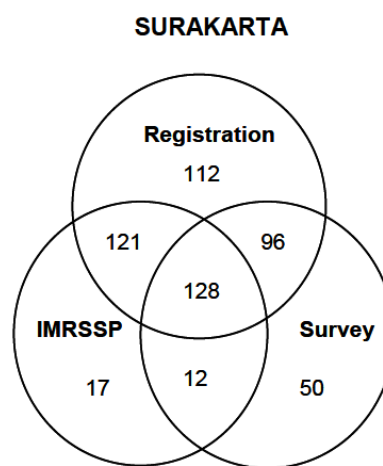
## Example 3: Indonesia (a)

- 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



Total deaths = 306

Completeness = 73%



Total deaths = 536

Completeness = 55%

## Example 3: Indonesia (b)

- Lampung/Gorontalo (2007-2008) – two data sources - health system records of facility and community deaths, and an independent survey
- Independent survey in a sample of villages from the overall study population, recall of deaths over two years, record linkage/matching across the two sources
- Analysis using capture recapture methods completeness computed as a proportion the total deaths including the estimated unobserved deaths

Survey characteristic	Lampung	Gorontalo
Number of villages included in validation survey	8	18
Total number of households	10 240	9 225
Survey population	36 117	35 184
Deaths common to IMRSSP and survey datasets	306	316
Unique deaths in survey dataset	150	145
Unique deaths in IMRSSP dataset	204	99
Estimate of deaths missed by both sources	100	45
Estimated completeness of IMRSSP data, % (95%CI)	67.1 (64–70)	68.5 (66–71)

IMRSSP = Indonesian Mortality Registration System Strengthening Project;  
CI = confidence interval.



## Limitations of DRS methods

- In PGE studies, 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)
- These occur as a result of the
  - nature of the events (e.g deaths in low SES strata less likely to be registered);
  - nature of data collection processes (passive or active)
  - Quality of data collected in each source
- All the above lead to potential bias in the completeness estimate
- Further, there is also sampling error / stochastic variation; which contribute to uncertainty in the completeness estimate
- In addition, there were considerable logistical challenges in implementing record linkages studies in terms of costs and manpower, as well as technical challenges in matching, evaluation of bias etc

- 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 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
- Potential to improve data quality of recorded variables used in linkage (name spellings; address variables, age, date of death etc)
- Wider use and recording 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

- Develop an efficient study design based on a careful choice of alternatives
  - E.g existing routine data sources vs special data collection
  - Scope of desired outcome measures (e.g by age, geography etc)
- Establish a clear understanding of data collection procedures to evaluate potential for and degree of bias
- Conduct a thorough analysis and evaluation of completeness estimates alongwith margins of error

- 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



- Focus on computerisation of all data sources
- Inclusion of relevant variables in all future potential data sources
- Emphasis on data quality (name spellings; address variables; accuracy of age, date of death; and where available Unique ID numbers)
- Promote follow up of electronic linkage with field verification of sample of matched/partially matched pairs and unmatched cases (to assess net matching error)
- Use all available evidence and methods to assess for bias and error in completeness; and where possible, conduct sensitivity analysis applying different methods
- Completeness estimates should be presented with margins of error, to assess impact on mortality indicators



- Several countries with near complete data, but limited quality on causes of death
  - Thailand, Egypt, Malaysia, Brazil etc
- Need for validation studies using hospital death records / discharge records
- Manual / electronic linkage studies to validate and correct registration diagnoses of causes of death
- Record linkage to verify specific variables of ethnicity / indigenous status / occupation etc for more detailed analysis of fertility & mortality patterns

Causes of death	Medical record diagnoses																
	HIV/AIDS	Liver cancer	Lung cancer	Other malignant	Diabetes	Hypertens	IHD	Oth heart	Stroke	Pneum	COPD	liver dis	Oth digest	Renal	RTA	All others	Total
Septicaemia	44	2	3	3	53	6	8	3	55	38	16	27	19	47	2	144	470
Ill defined	16	6	7	5	27	16	75	36	25	14	39	10	14	13	9	135	447
Stroke			1		7	1	4	5	203					1	9	31	262
IHD	1		2		26	5	138	9	3	2	3		3	6		16	214
Pneumonia	40		3		9	1	4	2	25	44	21	7	1	10	3	37	207
All other external causes					1	1	2	1	25	1					93	61	185
Genitourinary diseases	1	1		1	37	24	2	3	3	1	1	5	2	58		17	156
Lung cancer		1	85	6					1		4					5	102
Road traffic accidents								1							91	0	92
Liver diseases	2	2			1		2		2			63	2	1		11	86
HIV/AIDS	79											1				3	83
Other malignant neoplasms	1	14	3	24						2				1		34	79
COPD	1		2		2		3	3	2	3	54			2		5	77



## Misclassification - hospital deaths: Malaysia 2013

	MRdiagnosis																		
VRdiagnosis	TB	Septic	AIDS	Oth inf	Oth canc	Diab	Hypert	Isch HD	Oth heart	Stroke	Pneum	COPD	Oth GIT	Renal	Road trfc	Falls	Oth Inj	All others	Grand Total
TB	<b>29</b>		6			1					1	1						4	42
Septicaemia	8	<b>34</b>	12	16	7	50	2	14	8	19	43	17	16	37	5	2	2	99	391
Oth inf dis	1	1	5	<b>31</b>		5		3	1	2	4	1		2				5	61
other cancers		1			<b>105</b>			1		1	1			1				20	130
Diabetes				1		<b>35</b>	1	3	1		4		2	2	1			3	53
Isch heart dis	1			1	2	36	9	<b>522</b>	11	10	18	9	13	12	4	3	1	32	684
Other/ill def heart dis	4		2	1		4	10	37	<b>62</b>	8	3	2	2	6		3	1	34	179
Stroke	2	1		3		2	2	13	7	<b>341</b>	7	1	3	5	13	4	1	20	425
Pneumonia	13	3	13	15	6	42	5	22	19	86	<b>291</b>	67	20	18	8	5	4	94	731
Oth/ill def resp dis	4	2	1	1	1	4	1	7	4	77	15	13	2	5	4	2	2	42	187
Renal diseases	4	1	2	2	1	33	7	10	1	11	7		3	<b>94</b>	7	4		21	208
Ill defined diseases				1		4	2	22	2	4	5		5	1	2	2	2	13	65
Other/ill def injuries	1				1	3			2	6	1		1		94	7	<b>27</b>	24	167
All others	10	4	54	6	39	64	7	27	10	71	33	66	72	21	202	12	10	1010	1718
Grand Total	77	47	95	78	162	283	46	681	128	636	433	177	139	204	340	44	50	1421	<b>5041</b>

If hospital underlying causes accurate, then numbers would be concentrated along diagonal

Net under-reporting in VR – Stroke, diabetes, COPD, TB, AIDS, road traffic injuries



THANK YOU