

Non-sampling errors in survey data on mortality & fertility: implications for demographic estimation in unsettled environments

Bruno Masquelier^{1,2}

¹Université catholique de Louvain, Belgium

²Institut National d'Études Démographiques (INED)

Civil Registration in Refugee Settings: Technical Meeting UN
ESCWA - December 2016



Background

In countries where vital registration systems remain incomplete, survey and census reports are a key source of mortality and fertility estimates, and are sometimes used to validate VR data.

- ▶ Censuses and large-scale surveys include questions on **recent births & deaths**, **orphanhood**, and **children ever born/surviving**.
- ▶ Full & summary **birth histories** in population-based surveys are the main avenue for estimating recent trends in mortality (and fertility).
- ▶ **Sibling survival histories** (SSH) collected in DHS and other survey programs are increasingly used to generate estimates of adult mortality (e.g. Global Burden of Disease Study).

What are the **main reporting errors** in these retrospective reports and how do these errors affect our ability to assess the quality of VR data?

Objectives

1. Present the main strategies available to retrospectively collect data on mortality and fertility in the absence of CRVS, with a focus on conflict-affected areas.
 -  how are non-sampling errors evaluated and adjusted for?
2. Highlight some challenges in the assessment of completeness and quality of vital registration based on surveys/censuses.
3. Put forward a number of suggestions to improve mortality measurement through population-based surveys.
 -  collecting parental survival histories, revising sibling survival histories, introducing VAs in surveys, adding questions on places of death/death certification.

Reports on recent household deaths

Widely used in national censuses and large-scale surveys around the world. Usually restricted to residents of the household, asked to the head of household, and referring to the 12 months prior to data collection:

- ▶ Age and sex of the deceased
- ▶ Date of death
- ▶ Place of death
- ▶ Main cause of death according to the respondent (mostly to identify pregnancy-related deaths, and sometimes violent deaths).
- ▶ In some (rare) cases, death registration (eg. Senegal 2013, Jordan 2015, Nicaragua in 2005, ...).

Ideally, one should also know **the dates of arrival and departure of HH members**.

Questionnaire used in the Jordanian census

Deaths Among Household Members									
No, skip to section four .2					Yes, cont. .1		Dose any of the household died during the late 24 months? 301		
302	303	304	305	306	307	308	309	310	311
					<u>ONLY For Females</u> Ages 13-54 years, at the time of death				
Serial number of deceased person	Name of deceased person	Sex of deceased 1.Male 2.Female	Age at the time of death	Is the event of death has been registered? 1. Yes 2. No 8. Don't know	Marital status at death 1. never married, yes go to next person if last go to next section 2.Married 3.Divorced 4.Widowed 5.Separated	Was she pregnant at time of death? 1.yes, go 311 2.No 8.Don't, know	Did she die during delivery? 1.yes, go to 311 2.No 8.Don't, know	Was the death during the 42 days of the birth? 1.yes, go to 311 skip to next individual else, go to next section 2.No 8.Don't, know	Was death due to an accident? 1. Yes 2. 2. No
x		x	xx	x	x	x	x	x	
.1									
.2									

(Note: deaths of Syrians would need to be identified from the nationality of the surviving household members).

Most frequent non-sampling errors in data on HH deaths:

1. omissions of deaths,
2. misunderstandings related to the reference period,
3. systematic misstatement of ages at death and heaping,
4. difficulties in identifying residents of the household, leading to multiple reporting or underreporting,
5. household dissolution after the death of a household member.

⇒ Adjustments for incompleteness are possible based on **death distribution methods** (GGB, SEG or both), but these methods are underpinned by strict assumptions (Moultrie et al., 2013): migration is negligible or known, completeness of death reporting invariant by age, etc.

✗ Not recommended in unsettled environments.

Recent household deaths in large-scale surveys in conflict-affected areas

Examples of surveys on Iraqi deaths:

- ▶ Burnham et al. (2006) on household deaths in the period 2002-2006, and Roberts et al. (2004) (2002 to 2004) before them.
- ▶ Iraq Family Health Survey Study Group (2008) (March 2003 through June 2006)
- ▶ Hagopian et al. (2013) in Iraq in the 2003–2011 period: every household head was asked in 2011 about births and deaths since 2001. Secondary data sources were used to correct for out-migration.

Example of the questionnaire used by Hagopian et al. (2013)

13. For each death noted in question 12, please provide details: information from Key Informant. *If the remember age of death but not year use Events Calendar. For cause of death (d & g) select the choice that most closely matches the description provided by the person—Do not read out the options.*

Death no.	Sex M/F	Date of death? month & year	What was the cause of death? (codes)	May I see the death certificate?	Was the death war related? Y/N/DK	If war-related, What was the cause? (codes)	If war related, Who do think was responsible? Codes	Was fatal event at the house or > 1km away?
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
1								
2								
3								
4								
5								

(d) Under 18 causes of death: 1=diarrhea, 2=respiratory, 3=pre-term, 4=neonatal causes, 5=injury (not war), 6=injury (war), 7=neonatal tetanus, 8=cancer/tumor 9=other, 10=don't know,

(d) 18 and over Adult causes of death: 11=cardiovascular, 12=cancer, 13=injury (not war), 14=injury (war), 15=lung disease, 16=liver disease, 18=maternal, 19=kidney condition, 20=other, 21=don't know

(g) Causes of violent death related to war: 1=road accident that was war related, 2=gunshot, 3=car bomb, 4=airstrike, 7=other explosion, 8=other war injury, 9=don't know

(h) Responsible parties for war deaths: 1=criminals, 2=Iraq police (security forces), 3=Iraq army, 4=coalition forces, 5=militias, 6=unknown, 7=other

(j) Death certificate: 1=able to see the death certificate, 2=told there is a death certificate, but did not see or 3=death certificate not available.

Iraq Population Demographic Survey version of 15 March 2011

Page 2 of 5

Note: causes of death are established based on household reports, with a brief list. Would VA be feasible?

Additional questions worth considering for HH deaths

1. Some attempts to confirm deaths by asking to see the death certificates (Roberts et al., 2004; Hagopian et al., 2013).
2. Use of national events calendar and age/birth-year charts to assist with recalling dates of birth or death (Hagopian et al., 2013).
3. Question on places of death: the reliability of registration data is probably lower for deaths occurring outside health facilities.
4. Was the death reported to UNHCR?

Sibling survival histories

Used in more than 100 Demographic and Health Surveys (DHS). Respondents are asked to list all their maternal siblings and report their survival status, sex, ages at survey or at death, and the dates of death.

1104	What was the name given to your oldest (next oldest) brother or sister?	(1) _____	(2) _____	(3) _____	(4) _____	(5) _____	(6) _____
1105	Is (NAME) male or female?	MALE 1 FEMALE 2	MALE 1 FEMALE 2	MALE 1 FEMALE 2	MALE 1 FEMALE 2	MALE 1 FEMALE 2	MALE 1 FEMALE 2
1106	Is (NAME) still alive?	YES ... 1 NO ... 2 GO TO 1108 DK ... 8 GO TO (2)	YES ... 1 NO ... 2 GO TO 1108 DK ... 8 GO TO (3)	YES ... 1 NO ... 2 GO TO 1108 DK ... 8 GO TO (4)	YES ... 1 NO ... 2 GO TO 1108 DK ... 8 GO TO (5)	YES ... 1 NO ... 2 GO TO 1108 DK ... 8 GO TO (6)	YES ... 1 NO ... 2 GO TO 1108 DK ... 8 GO TO (7)
1107	How old is (NAME)?	<input type="text"/> GO TO (2)	<input type="text"/> GO TO (3)	<input type="text"/> GO TO (4)	<input type="text"/> GO TO (5)	<input type="text"/> GO TO (6)	<input type="text"/> GO TO (7)
1108	How many years ago did (NAME) die?	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

With additional questions to identify pregnancy-related deaths.

Sibling survival histories

- ▶ Event/exposure-type rates; limited modeling is required.
- ▶ Frequent, up-to-date estimates with relatively large sample of deaths (considering the survey sample)
- ▶ Not sensitive to migrations, not dependent on model life tables
- ▶ Widely collected, standardized in DHS, and similar format to birth histories,
- ▶ A single survey can give a certain sense of past trends in mortality,
- ▶ Assumptions are not as strong as in other types of indirect methods, but some modeling is required to reduce the scatter around age-specific mortality rates.

Selection biases

- ▶ No information is available for sibships without a surviving member: sibships with high mortality are under-represented
- ▶ Low mortality sibships are over-represented because the experience of the respondent's siblings is counted multiple times when more than one sibling might be interviewed.
- ▶ The respondents themselves are not counted in the denominator.

Trussell and Rodriguez (1990) showed mathematically that these three limitations neutralize each other, provided that :

1. all siblings in the sampling frame are interviewed,
2. the respondents are not included,
3. there is no association between mortality and sibship size.

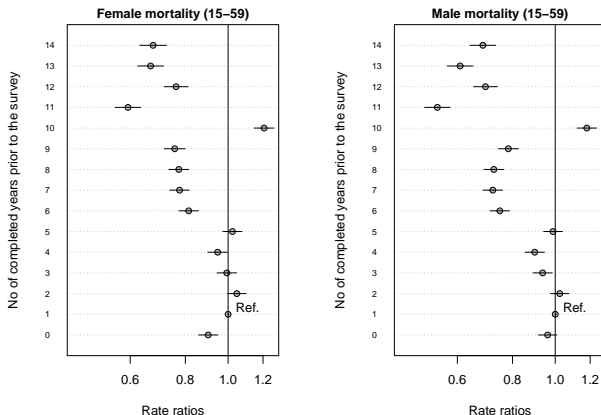
This was confirmed by microsimulations (Masquelier, 2013).

Recall biases

Recall biases in sibling histories include the following (Helleringer et al., 2014a):

- ▶ **List errors:** omissions of (deceased) siblings or inclusion of others who are not maternal siblings;
- ▶ **Vital status errors:** reporting that a deceased sibling is alive;
- ▶ **Age and date errors:** missing data on current ages, ages at deaths and the timing of death (usually not a serious concern), heaping and systematic misstatement of ages at death, displacement of deaths.

Estimated rate ratios associated with the number of completed years prior to the survey, estimated by pooling together all overlapping periods in 80 DHS conducted in 28 sub-Saharan countries. Sc: Masquelier et al. (2014)



Similar adjustment in the GBD 2013 (here, not country-specific and not linear). Deaths and exposure time that predate each survey by 15 years or more are often discarded.

Use of sibling histories in conflict-affected areas

- ▶ Hagopian et al. (2013) on mortality in Iraq associated with the 2003–2011 war and occupation: all adults asked in 2011 about the survival of their siblings → mortality rates computed for 2003-2011.
- ▶ Iraq Family Health Survey Study Group (2008) on mortality in Iraq from 2002 to 2006: reasonable consistency with the estimates from recent HH deaths.
- ▶ Other secondary analysis of survey data in conflict-affected settings: Obermeyer et al. (2008) on violent mortality in 13 countries from the World Health Survey programme, de Walque and Verwimp (2010) on the Rwandan genocide based on the 2000 DHS.

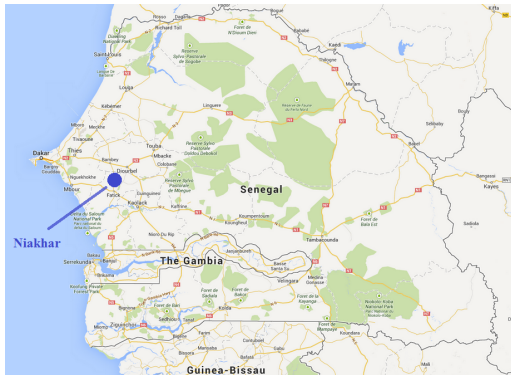
A validation study of sibling histories

We developed a new SSH questionnaire, the siblings' survival calendar (SSC)

- ▶ including additional recall cues designed to limit omissions of siblings,
- ▶ adopting a life calendar format to improve reporting of ages and dates.

We tested the SSC questionnaire against a standard DHS-type questionnaire during a randomized controlled trial (RCT) conducted in a demographic surveillance site (Niakhar, Senegal) (Helleringer et al., 2014b).

A validation study of sibling histories (2)



- ▶ located 120kms southeast of Dakar
- ▶ started in 1962, covering about 43 000 inhabitants
- ▶ low levels of educational attainment, high levels of mobility
- ▶ update visits now 3 times a year
- ▶ mother ID number established at first entry or at birth

Siblings' survival calendar

- ▶ Standardized script used to sensitize respondents to the issue of misreporting
- ▶ List of maternal siblings in the order that they come to mind
- ▶ Non-specific prompting (reading back the list of siblings)
- ▶ Additional recall cues:
 1. Any other deceased siblings?
 2. Any other siblings that migrated ?
 3. Any other siblings with a different biological father ?

Based on results from a pilot study in another HDSS in Senegal (Bandafassi) (Helleringer et al., 2014a).

Approaches to mortality estimation in countries with limited VR systems

YEAR	DURATION	NATIONAL OR PERSONAL EVENTS	RESPONDENT				YEAR	DURATION	SIBLINGS								
			BIRTHS <i>N</i> = respondent's DOB <i>First names</i> = children's DOB	UNION(S) <i>DV</i> = divorce <i>SP</i> = separation <i>W</i> = widowhood	RESIDENCE	SCHOOLING			#1	#2	#3	#4	#5	#6	#7	#8	#9
									M F DK	M F DK	M F DK	M F DK	M F DK	M F DK	M F DK	M F DK	M F DK
1953	60						Prior to 53	>									
1954	59						1953	60									
1955	58						1954	59									
1956	57						1955	58									
1957	56						1956	57									
1958	55						1957	56									
1959	54						1958	55									
1960	53	<i>Independence</i>					1959	54									
1961	52						1960	53									
1962	51						1961	52									
1963	50						1962	51									
1964	49						1963	50									
1965	48						1964	49									
1966	47						1965	48									
1967	46						1966	47									
1968	45	<i>Student Strike</i>					1967	46									
1969	44	<i>Death of N' BACKE</i>					1968	45									
1970	43						1969	44									
1971	42						1970	43									
1972	41						1971	42									
1973	40	<i>Drought</i>					1972	41									
							1973	40									

1. Political or sporting events used as landmarks.
2. Events affecting the respondent in terms of residence, marriages, births and schooling
 - ▶ Help to anchor the reporting of siblings' events.
3. Sibling section with sex, dates of births and deaths

Study participants and data collection

- ▶ Individuals who have ever been members of the HDSS, aged 15 to 59, with one known sibling in the HDSS dataset.
- ▶ All sibships **with at least one adult death** were selected (n=592) with two respondents (to assess the inter-sibling reliability) when available.
- ▶ 500 additional sibships in which all siblings were still alive were selected at random.

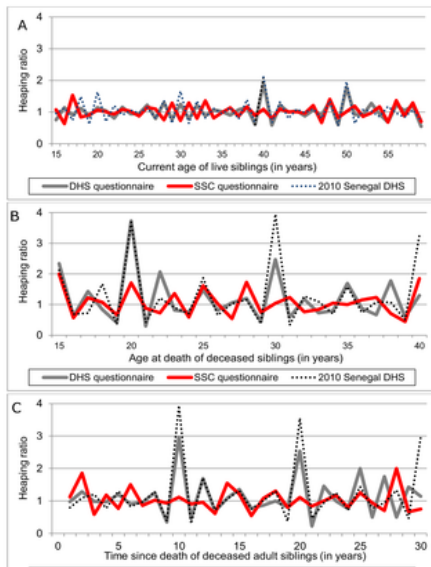
- ▶ The study team was made of 8 interviewers who had previously been included in DHS data collection in Senegal.
- ▶ Absent residents and migrants were followed-up in Dakar, Mbour and their suburbs, and within 50 miles of the HDSS.

Data quality

- ▶ Low proportions of missing data (< 1%), with significantly lower proportions of missing ages at survey in the SSC.
- ▶ Much less heaping in the SSC than in the DHS.

Index :

$$\frac{N(a)}{0.5 \times (N(a+1) + N(a-1))}$$



Reporting of deaths

Sensitivity is the proportion of concordant among sibships with at least one adult death.

Significant improvement in the reporting of adult **female** deaths.

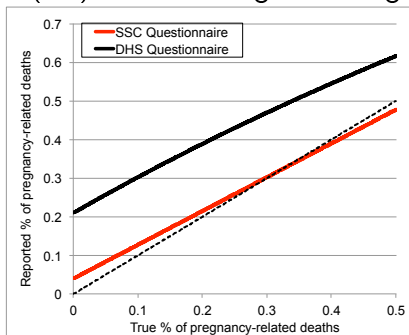
Sensitivity	DHS	SSC	DHS	SSC	p-Value
Adult female deaths	75/99	85/95	75.6	89.6	0.027
Adult male deaths	173/201	170/200	85.8	85.1	0.848
Specificity	DHS	SSC	DHS	SSC	p-Value
Adult female deaths	466/510	456/485	91.3	94.0	0.146
Adult male deaths	364/408	340/380	89.4	89.3	0.973

Specificity is the proportion of concordant among sibships with no adult death.

No evidence of more frequent "additions" (cousins, etc.), i.e., it did not increase the number of false reports of deaths.

Proportion of deaths due to maternal causes

In the Niakhar HDSS, SSH collected using the DHS questionnaire significantly overestimated the proportion of pregnancy-related (PR) deaths among women aged 15-49.



- ▶ PR deaths were more likely to be reported (96%) than non-PR deaths (71%).
- ▶ PR deaths were correctly classified as such (93%) but 20% of non PR deaths were classified as PR.

The SSC questionnaire yielded almost unbiased estimates of the proportion of deaths due to pregnancy-related causes.

Deaths from external causes

Large bias also observed in the proportion of deaths due to external causes in sibling histories (here we could not separate the questionnaires).

	Sibships with adult death from external causes		Sibships with adult death from other causes		p-value
	Reported/exp.	% (95% CI)	Reported/exp.	% (95% CI)	
All deaths	34/35	97.1 (82.2, 99.6)	498/568	87.7 (84.7, 90.1)	0.093

Deaths due to external causes were more frequently reported.

Sensitivity was low, while specificity was high.

		Sensitivity		Specificity	
	Correct/ reported	% (95% CI)	Correct/ reported	% (95% CI)	p-value
All deaths	20/34	58.8	479/498	96.2	<0.001

The likely direction and extent of bias in survey estimates of the proportion of deaths due to external causes will vary across populations. If the true proportion is above 10%, then survey data will most likely underestimate the proportion.

Orphanhood method

Questions on parental survival are widely used in large-scale surveys and censuses since the mid-1960s to estimate adult mortality.

- ▶ No information is collected on the ages of surviving parents or the ages at deaths, and the timing of deaths is unknown
- ▶ Proportions of surviving parents are converted into life table probabilities, using the age of the respondents as a proxy for the duration of exposure. Several methods exist (Brass and Hill, 1973; Timæus, 1992).

Main reporting errors:

- ▶ Adoption bias (fostered orphans being reported as non-orphans), more pervasive among young respondents are for mothers.
- ▶ Misreporting of the ages of respondents.

Seldom used in conflict-affected areas:

- ▶ Hirschman et al. (1995) included orphanhood questions in the 1991 Vietnam Life History Survey, with questions on current ages, ages at death, and the timing of deaths. Parental and sibling histories were highly consistent.

Collecting data on ages at death and dates of death would be useful for estimating mortality among older adults (but has seldom been tested so far).

Estimating fertility/mortality from birth histories

Relatively straightforward based on the DHS or MICS instruments.
The main non-sampling errors include:

- ▶ Shifting births to the more distant periods to avoid specific questions on recent births.
- ▶ Omissions of births and deaths (especially in the distant past)

DHS contain questions on the registration of births in the household questionnaire (e.g. Jordan 2012 DHS → 2017 DHS?), but ...

- ▶ limited to children who are alive at the time of the survey.
- ▶ is based on reports from the HH deaths, probably mixing birth notification/registration, or other institutional documents.

Estimating fertility from censuses

Questions on births in the past 12 months or date of last child born alive are widely used (but these were not asked in the last Jordanian census).

- ▶ Such reports are affected by reference period errors, and omission of neonatal deaths.
- ▶ In the absence of data on fertility, one can also use the "own children" method, using the child's age and mother's age to estimate a series of annual fertility rates (with additional mortality estimates).

Conclusion

Birth registration

- ▶ Population-based surveys usually provide reliable estimates of fertility and allow for a direct estimation of birth registration completeness (→ 2017 Jordan DHS?)
- ▶ An additional quantitative survey could shed light on the specific barriers to full registration (with specific questions on birth notification, birth certification, and dates of these events, place of birth, etc.)

Death registration

- ▶ Given the uncertainty around mortality estimates, not clear that a new population-based survey would allow a proper evaluation of the completeness of death registration.
- ▶ Child mortality estimates remain uncertain, and adult mortality are probably not reliable enough to validate VR data based on a single survey.

- W. Brass and K. Hill. Estimating adult mortality from orphanhood. In International Union for the Scientific Study of Population, editor, *Proceedings of the International Population Conference: Liège*, volume 3, pages 111–123, 1973.
- G. Burnham, R. Lafta, S. Doocy, and L. Roberts. Mortality after the 2003 invasion of Iraq: a cross-sectional cluster sample survey. *Lancet*, 368(9545):1421–1428, Oct 2006.
- D. de Walque and P. Verwimp. The Demographic and Socio-economic Distribution of Excess Mortality during the 1994 Genocide in Rwanda. *Journal of African Economies*, 19(2):141–162, 2010.
- A. Hagopian, A. Flaxman, T. Takaro, S. Esa Al Shatari, J. Rajaratnam, S. Becker, A. Levin-Rector, L. Galway, B. Hadi Al-Yasseri, W. Weiss, C. Murray, and G. Burnham. Mortality in Iraq Associated with the 2003-2011 War and Occupation: Findings from a National Cluster Sample Survey by the University Collaborative Iraq Mortality Study. *PLoS Med*, 10(10):e1001533, 10 2013.
- S. HELLERINGER, G. Pison, M. Kanté, G. Duthé, and A. Andro. Reporting errors in survey data on adult mortality: results from a record linkage study in Senegal. *Demography*, 51(2):387–411, 2014a.
- S. HELLERINGER, G. Pison, B. Masquelier, AM Kanté, L. Douillot, G. Duthé, Ch. Sokhna, and V. Delaunay. Improving the quality of adult mortality data collected in demographic surveys: Validation study of a new siblings' survival questionnaire in niakhar, senegal. *PLoS Med*, 11(5):e1001652, 05 2014b.
- C. Hirschman, S. Preston, and V. M. Loi. Vietnamese Casualties During the American War: A New Estimate. *Population and Development Review*, 21(4):pp. 783–812, 1995.
- Iraq Family Health Survey Study Group. Violence-Related Mortality in Iraq from 2002 to 2006. *The New England Journal of Medicine*, 358(5):484–493, 2008.
- B. Masquelier. Adult Mortality from Sibling Survival Data: A Reappraisal of Selection Biases. *Demography*, Volume 50, Issue 1:207–228, 2013.
- B. Masquelier, G. Reniers, and G. Pison. Divergences in mortality trends in sub-Saharan Africa: survey evidence on the survival of children and siblings. *Population Studies*, 68(2):161–177, 2014.
- T. Moultrie, R. Dorrington, A. Hill, K. Hill, I. Timæus, and B. Zaba. *Tools for demographic estimation*. Paris: International Union for the Scientific Study of Population, 2013.
- Z. Obermeyer, C. Murray, and E. Gakidou. Fifty years of violent war deaths from Vietnam to Bosnia: analysis of data from the world health survey programme. *BMJ*, 336:1482–1486, 2008.
- L. Roberts, R. Lafta, R. Garfield, J. Khudhairi, and G. Burnham. Mortality before and after the 2003 invasion of Iraq: cluster sample survey. *Lancet*, 364(9448):1857–1864, 2004.
- I. Timæus. Estimation of adult mortality from paternal orphanhood: a reassessment and a new approach. *Population Bulletin of the United Nations*, 33:47–63, 1992.
- J. Trussell and G. Rodriguez. A note on the sisterhood estimator of maternal mortality. *Studies in Family Planning*, 21 (6):344–346, 1990.