

Improving survey design and estimation methods for parental survival histories

Katherine T Lofgren

A thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Public Health

University of Washington

2013

Committee:

Haidong Wang, PhD

Emmanuela Gakidou, PhD

Julie Knoll Rajaratnam, PhD

Program Authorized to Offer Degree:

School of Public Health – Department of Global Health

©Copyright 2013

Katherine T Lofgren

University of Washington

Abstract

Improving survey design and estimation methods for parental survival histories

Katherine T Lofgren

Chair of the Supervisory Committee:

Assistant Professor Haidong Wang

School of Public Health – Department of Global Health

Demographic transitions are occurring globally, prompting increased interest in late-adult aged mortality. To supplement information from vital registration systems and close knowledge gaps in information scarce settings, we have tested the use of parental survival data to estimate adult mortality. This research was conducted across four field sites in Tanzania, India (Uttar Pradesh and Andhra Pradesh), and the Philippines. Mortality information from parental survival histories was analyzed accounting for age heaping and missingness. Gakidou-King weights were applied to correct for survival bias in the survey sample. Age-specific mortality rates were calculated and compared with alternative sources of mortality information in the region to determine the validity of the estimation technique. The results show remarkable congruence between comparator sources of mortality information and the census and survey administered in this research project. Future surveys should implement direct parental survival modules to aid in the estimation of late-adult mortality levels and trends.

Acknowledgements

This thesis and my journey through graduate school would not have been possible without the academic and personal support of many individuals. Thank you to the Institute for Health Metrics and Evaluation for my training and education and especially my thesis committee for your guidance on this particular project.

To Haidong Wang: Thank you for being my advisor and my favorite formal demographer.

To Emmanuela Gakidou: Thank you for changing my life; this is the best opportunity I've ever had.

To Julie Rajaratnam: Thank you for all your patience and your curry dishes.

Thank you to Laura Dwyer-Lindgren for being by my side at work every day and teaching me how to code. Finally, thank you to my parents, brothers, and roommate Libby for all your love and support.

Background

As global populations continue to reach older ages, the accurate measurement of levels and trends in adult age groups becomes ever more essential as an indicator of population health. This includes both the measurement of mortality in the adult ages 15 to 60 as well as the measurement of late-adult mortality between the ages of 60 and 80. Measuring mortality in settings without reliable vital registration (VR) systems is challenging and requires robust estimation methods. In the context of under-5 mortality, birth history methods are a proven tactic to gain information about under-5 mortality when VR is unavailable or unreliable. Adult mortality trends are more difficult to estimate: survey respondents generally have more accurate information about their children compared to their siblings or parents.

The Population Health Metrics Consortium (PHMRC) project funded by the Bill & Melinda Gates Foundation as part of its Grand Challenges in Global Health Program aims to develop better methods to measure mortality through Grand Challenge 13 (GC13). GC13 is defined as the development of technologies that permit quantitative assessment of population health status [1]. Here we have specifically evaluated the performance of parental survival histories included as part of a household survey instrument as a means of estimating late-adult mortality. This research evaluates how questionnaire design can influence the accuracy of resulting mortality estimates as well as how parental survival histories perform relative to alternative, sibling survival histories.

Parental survival histories have been included in surveys and censuses dating back to the first proposal of the method by Henry in 1960 [2]. Typically, these modules ask survival status questions indirectly. For example, a respondent may be asked whether their parents were alive at the respondent's wedding or their 15th birthday. This information along with characteristics such as the age of the respondent are then used to estimate mortality rates [3–6]. It has generally been assumed that asking respondents

detailed questions regarding their parents' lives and deaths (i.e. ages, age at death, years since death) will result in poor quality data, producing unreliable mortality estimates.

In this research project we test this assumption, asking direct age and date of death information about the respondent's parents to estimate age specific mortality rates and the probability of death between age 60 and 80.

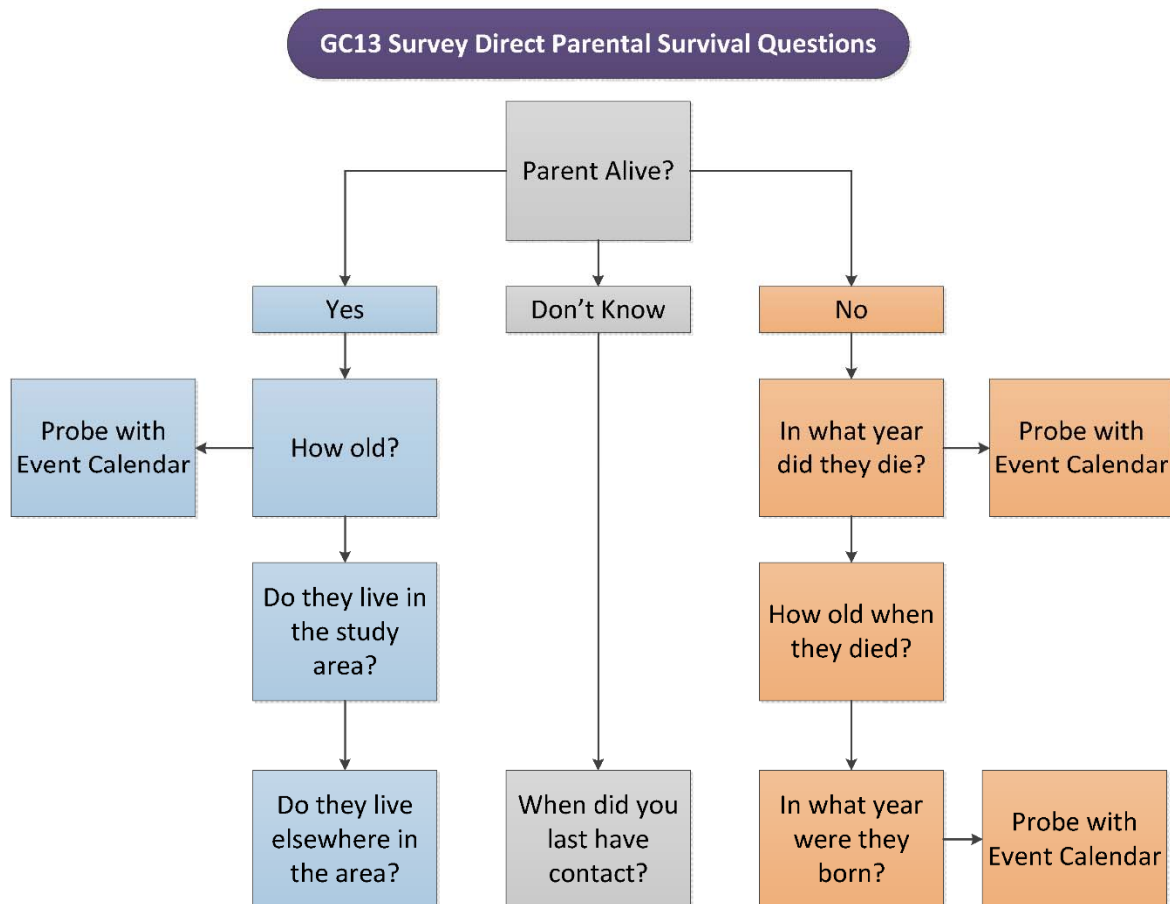
The GC13 Study

The GC13 study was conducted across four diverse field sites, Andhra Pradesh (India), Uttar Pradesh (India), Bohol (Philippines), and Pemba (Tanzania). Initially, a large scale census was conducted in each site, with sections in the survey instrument for household deaths recall, sibling survival histories, and parental survival histories. This census had four survey variants (A, B, C, and D) which asked questions in slightly different ways and orders to ascertain the optimal survey instrument. We fielded the direct parental survival instrument in variants C and D, asking questions about the survival of parents including age, age at death, and years since death. The parental survival questionnaire was asked of both male and female respondents.

In addition to the preliminary census, PHMRC researchers constructed alternative measures of mortality information in each of the field sites. These sources were used as comparators for the mortality estimates derived from the GC13 project. In three of the field sites (Pemba, Andhra Pradesh, and Uttar Pradesh) previous censuses were available for comparison using the Two Census Method which accounts for migration in the intercensal period [7, 8]. In the fourth site, Bohol, the Capture-Recapture Method was used to create a comparator source by triangulating amongst multiple sources reporting on deaths in the same population for a similar period [9]. Additionally, in Andhra Pradesh and Uttar Pradesh, surveillance data was available to serve as yet another comparator source of mortality information.

Based on a preliminary analysis conducted with the GC13 census parental survival data, it was determined that the direct form of the questionnaire should also be implemented in a modified survey instrument implemented at three of the four GC13 field sites, Bohol, Andhra Pradesh, and Pemba. The follow-up survey to the GC13 census included detailed event calendar probing and allowed the ability to answer questions in multiple ways. For example, if a respondent was unable to answer the age of death for a parent, they were able to answer in years ago that the parent died. **Figure 1** below shows the skip pattern and questionnaire flow for the GC13 survey implemented in Andhra Pradesh, Bohol, and Pemba.

Figure 1: Direct parental survival questionnaire implemented in the GC13 survey



All fieldwork for the mortality modules discussed in this paper has been completed. The data and results for both the parental survival histories from the GC13 census and survey are presented below.

Methods

Data

The sample size for each of the parental history modules conducted is shown below in **Table 1**. Of the four instrument variants administered as part of the GC13 Census, the direct parental survival module was only asked in variants C and D. Therefore, the sample sizes listed do not represent the census as a whole; rather only those respondents assigned to census variants C and D.

Table 1: Sample size for the census and survey by field site

Field Site	GC13 Census	GC13 Follow-up Survey
Andhra Pradesh	37198	4009
Bohol	15425	5172
Pemba	29644	4043
Uttar Pradesh	13297	N/A

Table 2 breaks down the respondents for the GC13 survey by sibship size. In Andhra Pradesh a survey respondent is most likely to report a family size of 3 total children (approximately 10% of the survey respondents) while in Bohol a respondent is most likely to report a family of 7 total children. In Pemba a sibship size of 9 was most frequently reported in the sample.

Table 2: Sibship size for survey respondents*

Siblings (including respondent)	Andhra Pradesh		Bohol		Pemba	
	Count	Percent	Count	Percent	Count	Percent
1	51	1.3	80	1.5	33	0.8
2	270	6.7	108	2.1	60	1.5
3	415	10.4	205	4.0	92	2.3
4	389	9.7	246	4.8	133	3.3
5	345	8.6	293	5.7	169	4.2
6	262	6.5	309	6.0	222	5.5
7	171	4.3	337	6.5	326	8.1
8	155	3.9	321	6.2	317	7.8
9	75	1.9	282	5.5	336	8.3
10	42	1.0	216	4.2	320	7.9
11	21	0.5	149	2.9	235	5.8
12	20	0.5	100	1.9	123	3.0
13	6	0.1	47	0.9	69	1.7
14	1	0.0	25	0.5	32	0.8
15	3	0.1	12	0.2	13	0.3
16	1	0.0	7	0.1	3	0.1
17					3	0.1
18			1	0.0	1	0.0
19					2	0.0
Missing	1782	44.4	2434	47.1	1554	38.4

*counts include both living and dead respondents

At the time that the survey was administered, a respondent from Andhra Pradesh was most likely to have 3 siblings still alive (including the respondent). Respondents from Bohol were most likely to have 5, while respondents from Pemba most frequently had a sibship of 7 remaining.

Missingness

The percent missingness across the key variables for the parental survival analysis is shown below in

Table 3. Interestingly, the survival status of the parents is extremely well reported, indicating that the death of a parent is a significant event that respondents across all four field sites recall very well. The age of living parents was also reported with less than 1 percent missingness in all of the sites fielded in the GC13 census and survey. There was greater missingness for variables regarding deceased parents.

Missingness in the age of deceased parents ranged from 1.3% (Pemba –Census) to 79.8% (Pemba – Survey), with potential survey administration errors in AP and Pemba for the Survey. Years since the death of a parent had missingness ranging from 0.8% to 25%. With the exception of problems in the Andhra Pradesh and Pemba survey, the missingness in the age (of both alive and dead parents) and the years since death variable had reduced missingness from the census to the follow-up modified survey instrument.

Table 3: Percent missingness by field site

	Field Site	Alive Status	Parent's age		Years since death
			Alive	Dead	
Census	Andhra Pradesh	0.1	0.6	17	9.8
	Bohol	2.2	0.2	3.4	2.3
	Pemba	0.1	0.3	1.3	0.8
	Uttar Pradesh	0.1	0.4	5.5	3.7
Survey	Andhra Pradesh	0.4	0.1	55.2	2.6
	Bohol	1.5	0.1	2.5	1.5
	Pemba	0.1	7.9	79.8	25

There are clear patterns in missingness by respondent characteristics; **Table 4** highlights the observed pattern of missingness in the Andhra Pradesh GC13 census by education level. This pattern was not only observed in AP, it was rather apparent to varying degrees across all of the field sites in the census and survey.

Table 4: Percent missingness by education level in the Andhra Pradesh census

Education Level	Alive Status	Parent's age		Years since death
		Alive	Dead	
No school	0.2	1.2	23.9	15.2
<Primary	0.2	0.8	15.5	7.8
Primary	0.1	0.6	12.3	6.8
Middle	0.1	0.4	10.3	4.3
Secondary	0.1	0.3	7.5	2.6
Sr. Secondary	0.1	0.1	5.5	2.3
Degree	0	0.2	3.8	2.1
Technical/vocational	0	0.2	5.2	2.3

To account for missingness multiple imputation was used [10]. Five imputed datasets were created; the results presented here are averages across all 5 of the imputed dataset for age of alive and deceased parents and the years since death of a parent.

Age and Date Heaping

It is common for survey respondents to give rounded estimates of dates rather than accurate numbers to the single age or year. This is referred to as heaping and is common in age or date reporting across demographic surveys [11]. In some cases, the analysis of data within 5 year bins rather than single years is a sufficient aggregation to reduce heaping [12]. In the GC13 project there was an observed preferential heaping every 10 years more so than at 5 years. This biased heaping pattern leads to a continued heaping pattern even in 5 year bins. **Figure 2 and 3** below show the observed age of the deceased parent and the data in 5 year bins.

Figure 2: Observed age of deceased parents pooled across all census sites

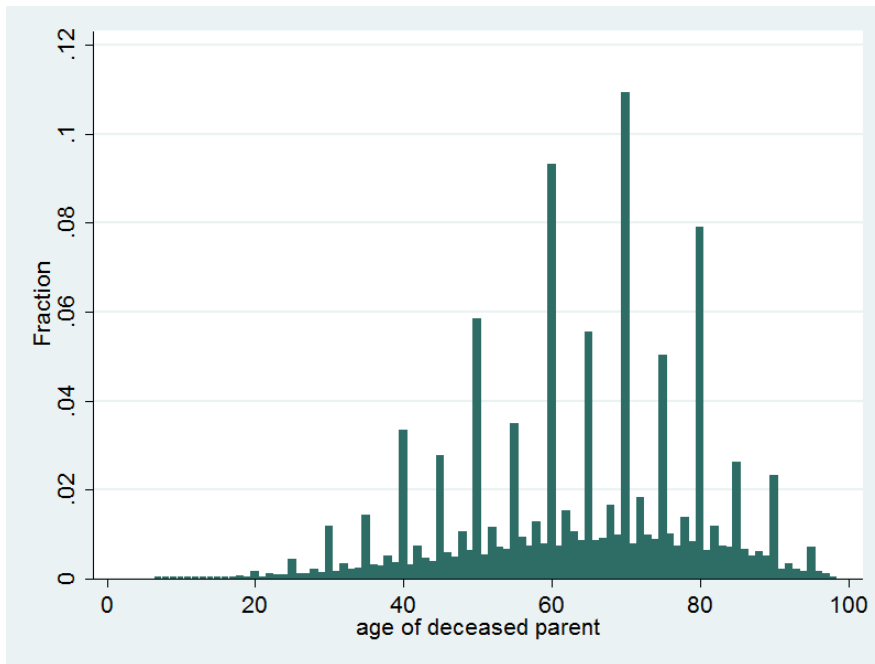
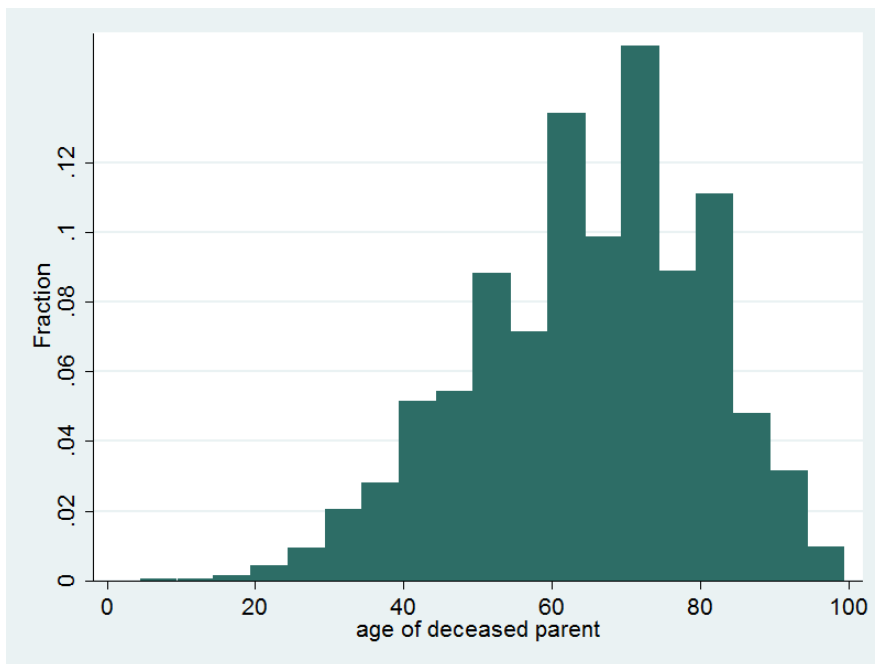


Figure 3: Observed age of deceased parents in 5 year age bins



We instead redistribute ages using a uniform distribution to smooth the age and date heaping. Heaped ages were redistributed centered on a 5 year window [13, 14]. This redistribution when combined with

analyzing the data in 5 year bins creates smoothed age or date trends that better fit our *a priori* beliefs about how ages and dates are distributed in the population (**Figure 4**).

Figure 4: Smoothed age of deceased parents in 5 year age bins

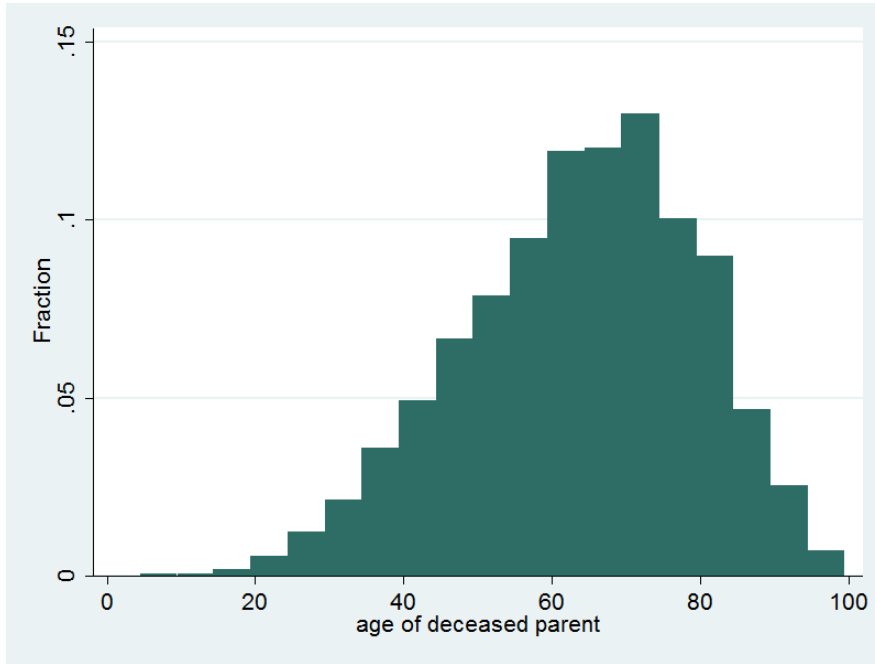


Table 5 below shows the age distribution of parents and years since death after multiple imputation has been performed and age heaping has been corrected with the smoothing algorithm described above.

Table 5: Age and years since death distributions post imputation and smoothing

Field Site	Age (Alive)		Age (Dead)		Years since death		
	mean	sd	mean	sd	mean	sd	
Census	Andhra Pradesh	53.7	13.5	64.3	15.7	18.7	15.5
	Bohol	59.9	12.9	65.0	16.1	19.1	16.0
	Pemba	58.0	13.3	63.8	14.7	18.1	13.7
	Uttar Pradesh	56.4	12.4	62.9	15.6	18.4	14.8
Survey	Andhra Pradesh	55.8	13.6	60.8	15.8	22.8	17.3
	Bohol	60.2	12.7	64.9	16.3	20.0	16.7
	Pemba	56.7	13.1	61.5	16.9	20.4	16.8

Survivorship Bias

Age-specific mortality rates were calculated by pooling across the 5 imputations and calculating the number of person years lived by 5 year age group and the death counts within each age group. Both the person-years lived and the death counts were corrected for selection bias in the sampled population. Selection bias refers to the underrepresentation of high-mortality sibships in the sampled population. Sibships with high rates of mortality are less likely to be selected because fewer siblings will be alive at the time of the survey to be sampled. Gakidou and King proposed a method to correct this bias by weighting the sibships represented in the survey sample (**equation 1**) [15]:

$$W_j = \frac{B_j}{S_j} \quad \mathbf{1}$$

Where B_j is the original sibship size and S_j is the number of siblings in sibship j who survive to the time of the survey. When each observation in the dataset is analyzed at the sibship level, as is the case with parental survival, this Gakidou-King (GK) weight is used to correct for survivorship bias at the sibship level.

Results and Discussion

Age-Specific Mortality Rates

The results from both the parental survival module in the GC13 census and household survey are promising. Especially in late-adult ages, above the age of 60, the mortality estimates derived from parental histories produce similar results to the field site comparator sources. **Figure 5a and 5b** show the GC13 census mortality estimates from the sibling and parental survival modules in addition to the dual-census and surveillance system comparator estimates for males and females respectively in Uttar Pradesh.

Figure 5a: Male age-specific mortality rate estimates in Uttar Pradesh

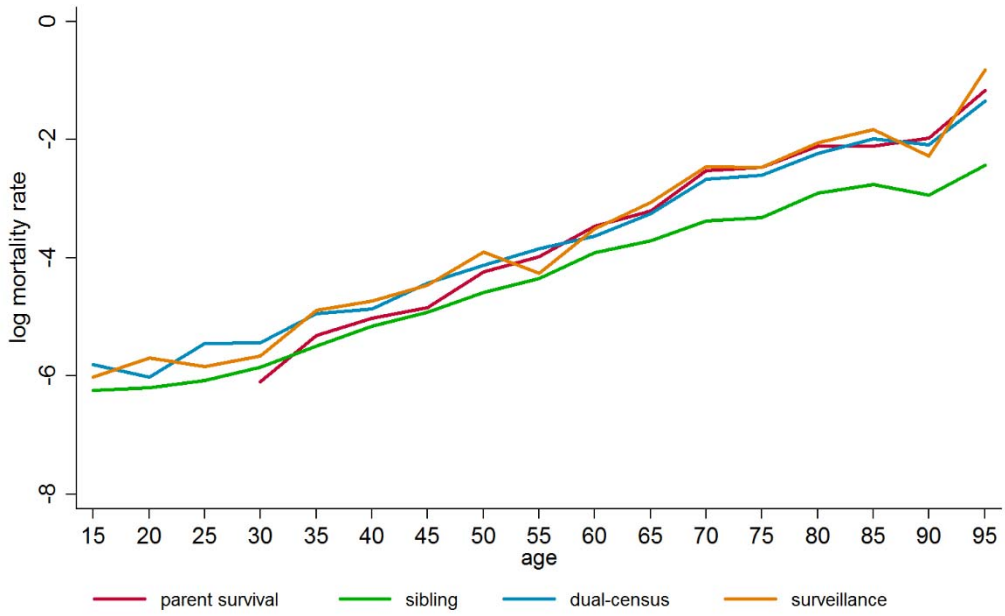
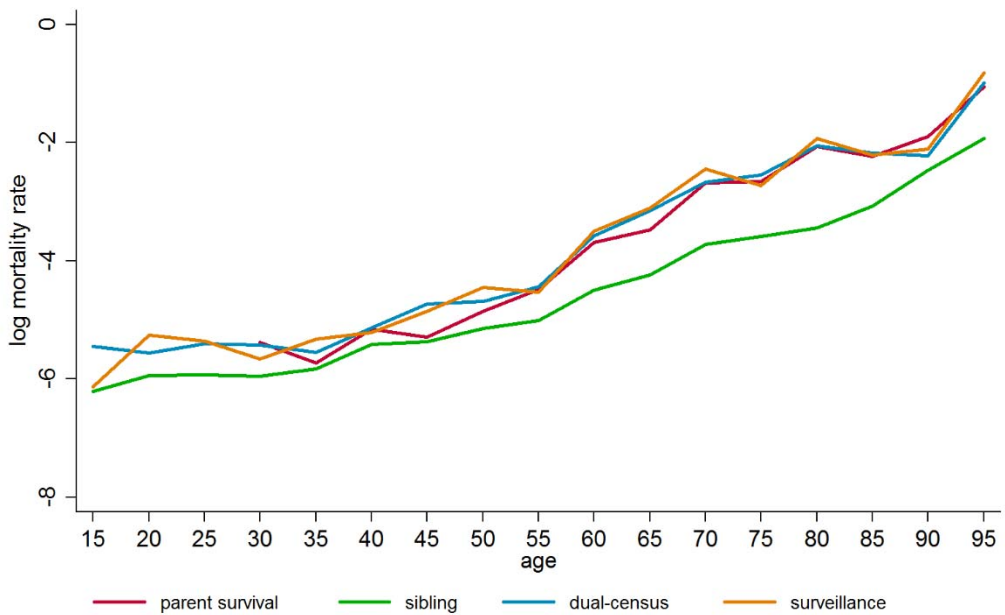


Figure 5b: Female age-specific mortality rate estimates in Uttar Pradesh



Figures 5a and 5b above show all four data sources lining up well, with parental-survival estimates informing older ages and sibling survival estimates extending estimates down to the age of 15. From

these figures, parental and sibling survival histories appear to be complimentary sources of mortality information. **Figure 6** below shows the results for both the GC13 census and survey parental survival histories along with the comparator sources and the sibling survival module from the census for females. In all sites the parental survival histories produce reasonable results with the context of the comparator sources.

Figure 6: Female age-specific mortality rate estimates across all field sites

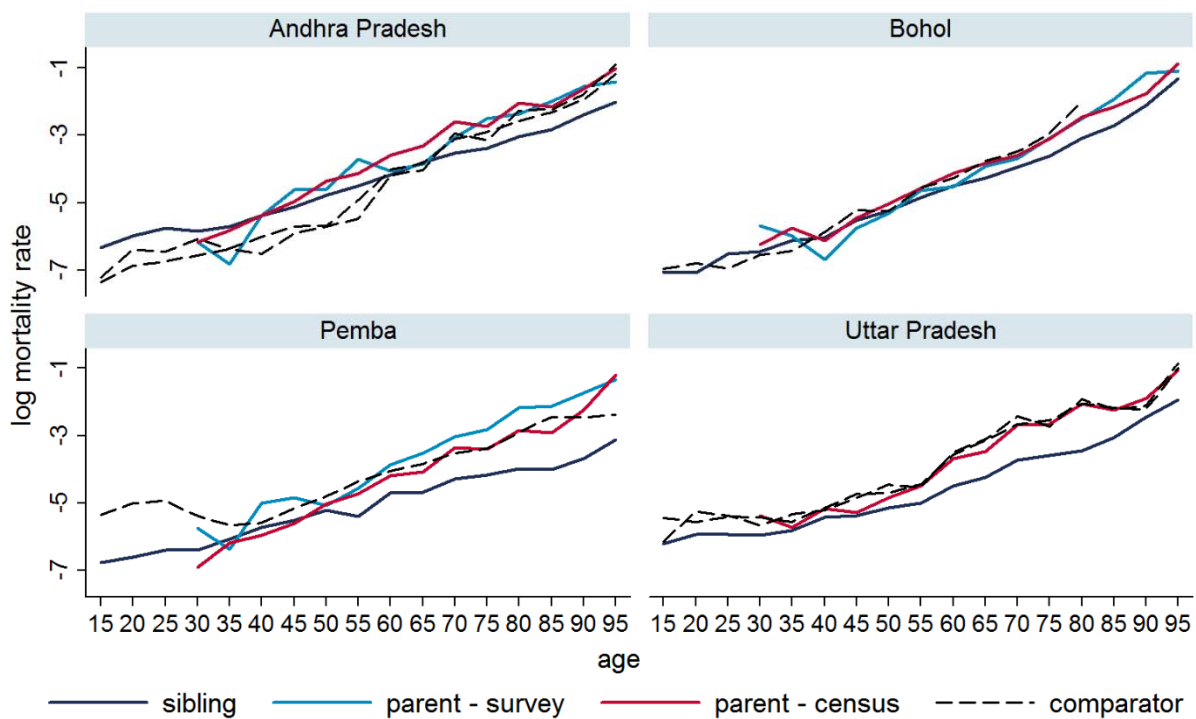


Table 6 below shows the relative error between the parental history estimates from the census and other comparator sources of information for ages 30 through 70. Comparator sources include the census sibling survival module, a dual census or capture-recapture census, and surveillance (AP and UP only). The relative error values fluctuate dramatically by age group revealing how difficult it is to get consistent age specific mortality estimates across ages. The comparators are considered our best options to create benchmarks but should not be considered a gold-standard estimate approximating

truth. The parental survival estimates clearly deviate from the sibling survival in older ages, reflecting the underestimations of older aged mortality by sibling survival histories. Although sibling survival histories may characteristically have recall bias in older ages, it has also been supposed that parental survival histories will perform worse at younger ages due to recall bias and the adoption effect [6, 16]. In general, there is a large gap in the age specific mortality estimates derived from parental survival estimates and sibling survival estimates at both young and old ages, supporting the need for both source of information as complimentary estimation techniques. The patterns for the other comparator sources are less clear with relative errors ranging from 1 percent in Uttar Pradesh (age 70) to as large as 277 percent in Andhra Pradesh (age 50).

Table 6: Relative error between parental survival estimates and comparator sources

Field Site	Age	Sibling Survival (%)		Comparator Census (%)		Surveillance (%)	
		Males	Females	Males	Females	Males	Females
Andhra Pradesh	30	21	27	25	7	74	48
	40	2	1	20	88	64	210
	50	25	53	82	277	78	285
	60	77	76	153	52	224	78
	70	150	150	104	67	87	41
Bohol	30	10	24	22	37	N/A	
	40	15	9	19	22		
	50	28	21	22	24		
	60	38	42	10	15		
	70	56	42	10	10		
Pemba	30	261	40	17	78	N/A	
	40	35	21	43	31		
	50	51	20	13	19		
	60	141	66	25	14		
	70	146	152	19	17		
Uttar Pradesh	30	22	78	48	5	36	32
	40	14	29	15	2	26	5
	50	42	34	10	16	28	33
	60	57	125	19	10	5	17
	70	134	185	15	1	7	21

The results shown above in **Figures 5a, 5b, and 6** are based on pooled person –years across time. It is also possible to disaggregate these mortality estimates by time-period. The advantage of these parsed estimates is more time-specific information regarding mortality levels. The disadvantage is smaller sample sizes that yield less stable estimates. Using the larger sample sized census data, **Figure 7a-d** show the results for the parental survival history broken down by time since the survey for late-adult age groups in Andhra Pradesh.

Figure 7a: Age-specific mortality estimates by time prior to the survey in Andhra Pradesh, age 60 to 64

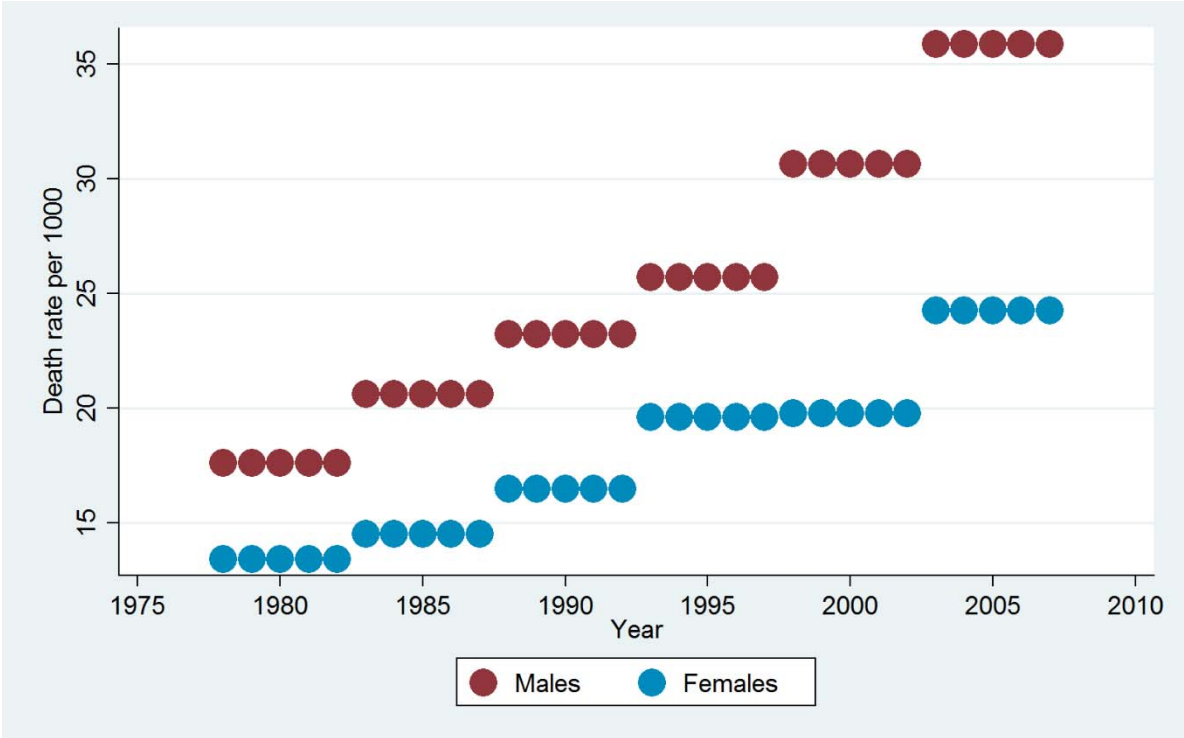


Figure 7b: Age-specific mortality estimates by time prior to the survey in Andhra Pradesh, age 65 to 69

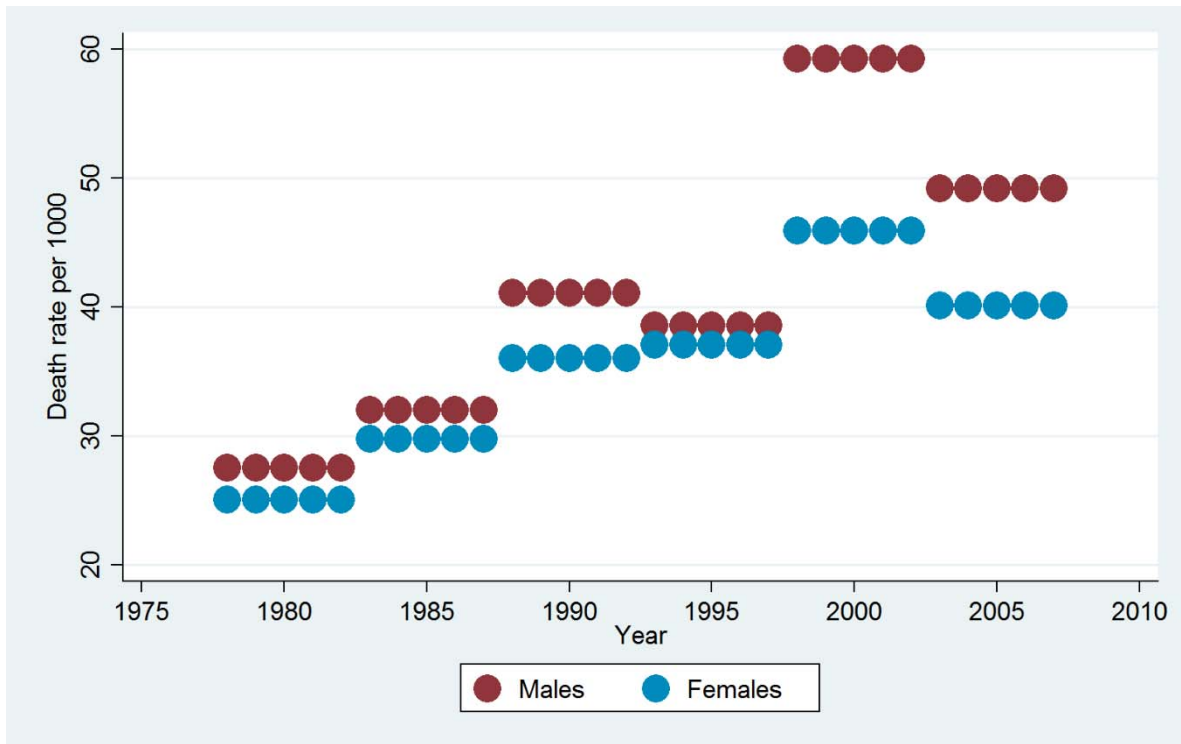


Figure 7c: Age-specific mortality estimates by time prior to the survey in Andhra Pradesh, age 70 to 74

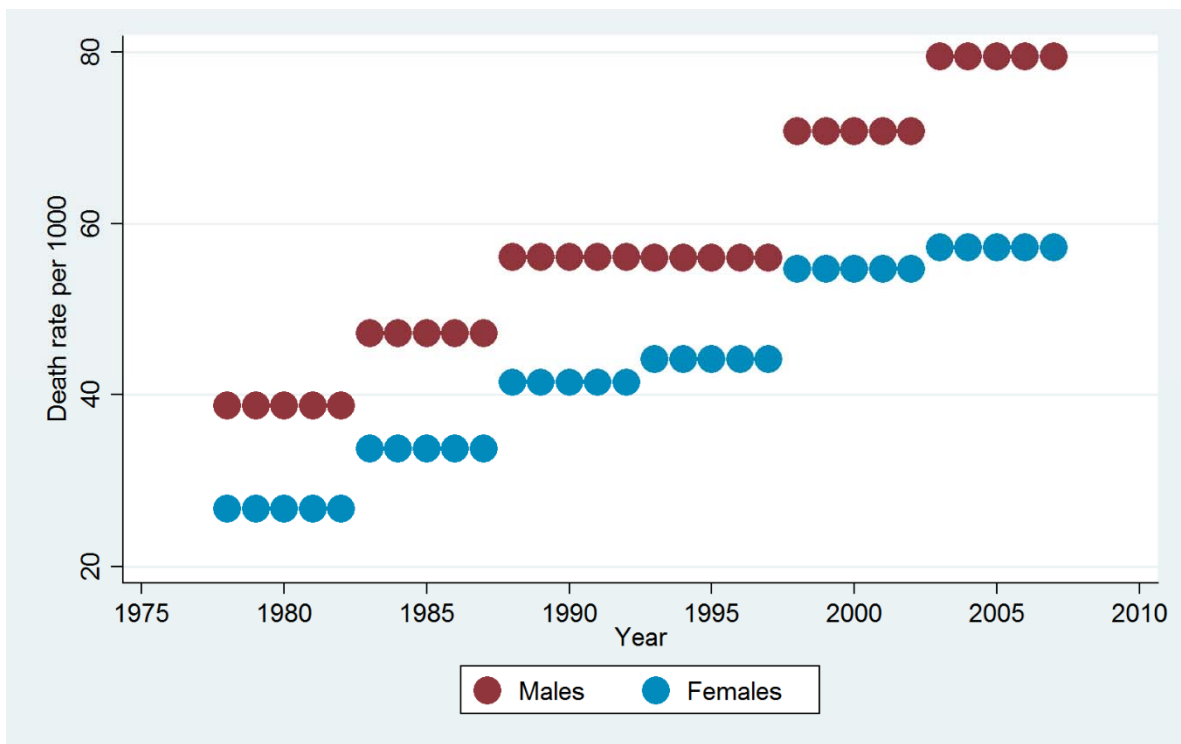
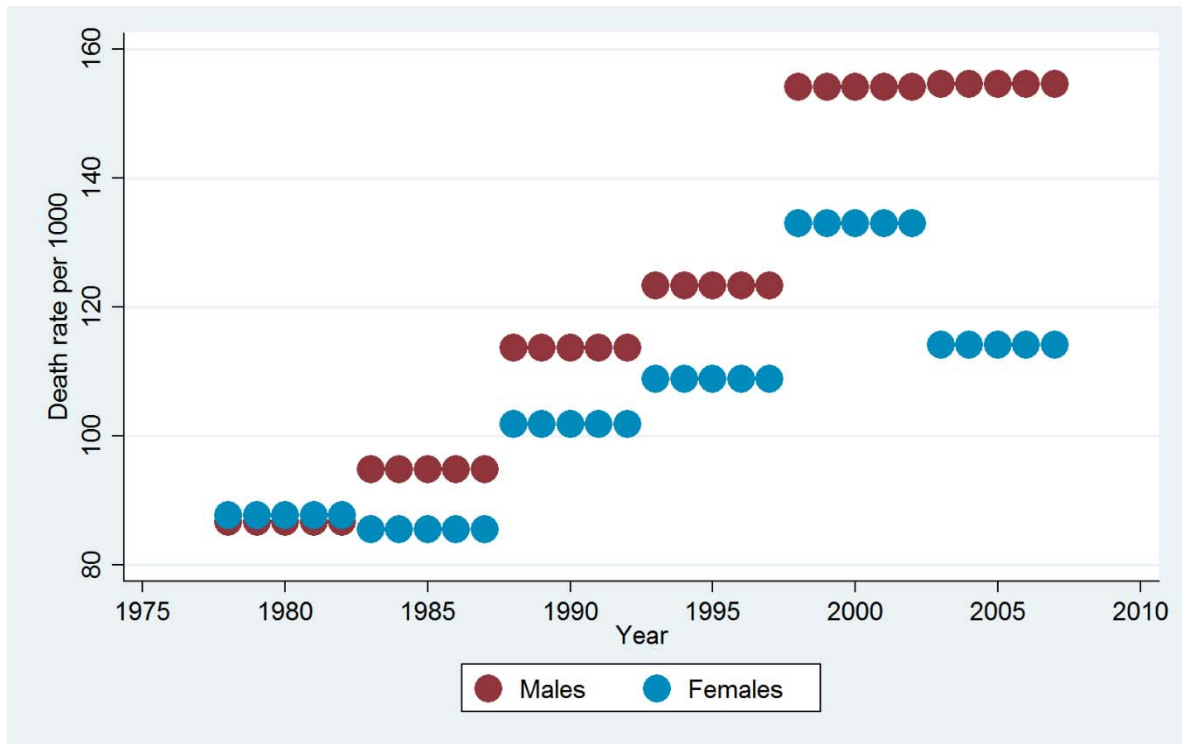


Figure 7d: Age-specific mortality estimates by time prior to the survey in Andhra Pradesh, age 75 to 79

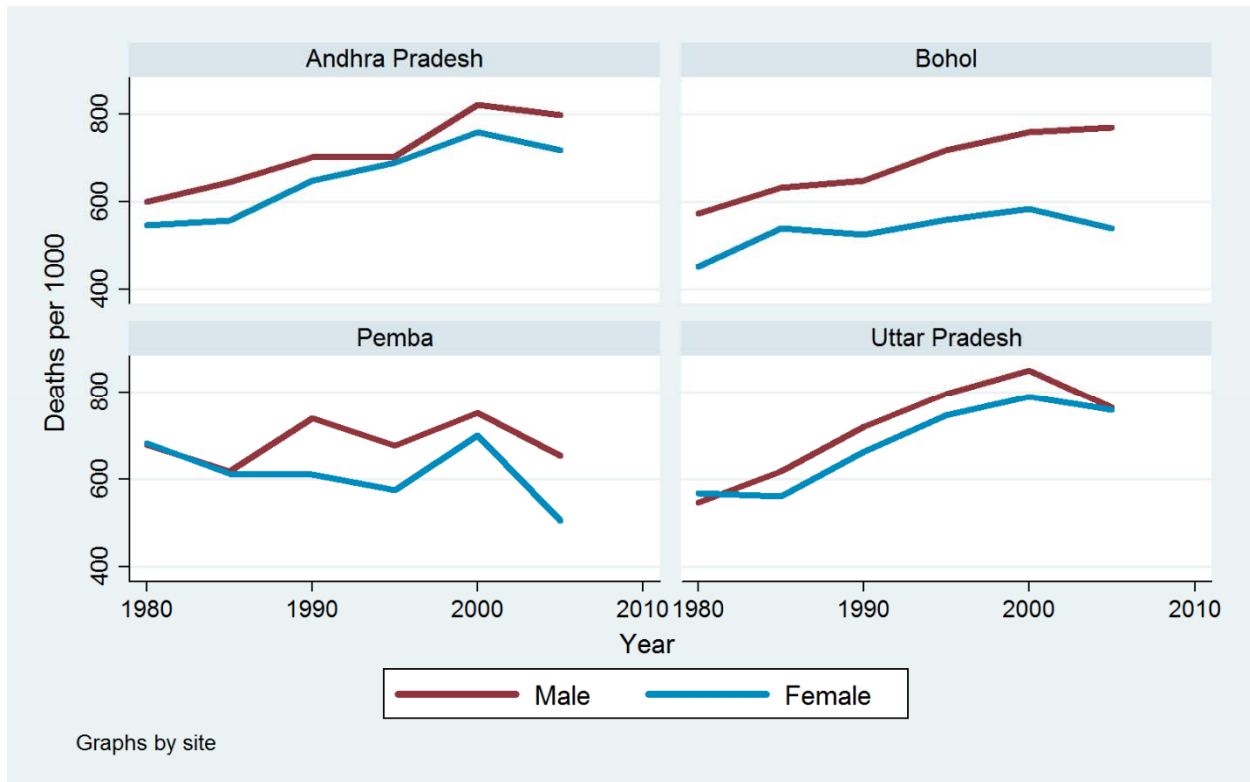


The clear trend across all of these figures is that mortality in these age-groups is dynamically changing over time. Further we see the age-specific mortality rates rising over time. Figure 7a-d emphasize the need for more effort in the measurement of late-adult ages and indicates the growing importance of these late-adult age mortality estimates as populations continue to undergo the demographic transition.

Probability of Death between 60 and 80 years (${}_{20}q_{60}$)

In addition to estimating age-specific mortality rates, survey estimates are often used to summarize the mortality risk experienced over an extended period of time. We have estimated the probability of death between ages 60 and 80 by converting age-specific mortality rates into a probability of death. **Figure 8** below shows the results of ${}_{20}q_{60}$ for each of the field sites in the census based on the parental survival histories.

Figure 8: Probability of death between age 60 and 80, by field site



By creating a summary measure like, ${}_{20}q_{60}$, we can communicate a myriad of mortality pressures in a population with one summary metric. The probability of death between the ages of 60 and 80 succinctly captures the mortality experience of a broad age group that policy makers are increasingly interested in providing health care for. This metric moves beyond the ages 15 to 60 that are typically estimated with the use of sibling survival methods. If parental survival histories were incorporated into more surveys, the estimates derived could provide valuable information for policy makers regarding the health of their late-adult aged population. Further, in the post MDG era as we look for new indicators of health, the measurement and continued tracking of progress in the health of 60 to 80 year olds is likely to increase.

Conclusions

The aim of this project was to develop better methods and survey techniques to measure mortality where vital registration systems are unreliable or non-existent. The results from the GC13 census and

survey show that direct parental survival questionnaires are sources of information that should be utilized in future survey instruments. The modules require relatively few additional questions, respondents have shown an ability to recall parental survival status well, and missingness and age heaping can be accounted for in data processing. Further, we observed decreasing missingness in parental survival data for respondents with higher education levels. As education, especially among women, continues to increase globally, increasingly literate populations are likely to report more accurate data on death and birth events than has been observed in the past.

As populations continue to age globally, the accurate measurement of mortality in ages above 60 is becoming increasingly important. In our comparisons with sibling survival histories, censuses, and surveillance systems, we have found that parental survival can provide a complimentary source of mortality information for these late-adult ages. Especially in information scarce settings, the addition of estimates derived from direct parental survival histories could strengthen and improve our understanding of the adult mortality landscape.

References

1. **Challenge 13: Developing Accurate Ways to Measure Health of Populations** [<http://www.grandchallenges.org/MeasureHealthStatus/Challenges/PopulationHealth/Pages/default.aspx>].
2. Henry L: **Mesure indirecte de la mortalité des adultes**. *Popul. Fr. Ed.* 1960, **15**:457–466.
3. Brass W, Hill K: **Estimating Adult Mortality From Orphanhood**. *Int. Popul. Conf.* 1973, **3**:111–23.
4. Blacker JG: **The estimation of adult mortality in Africa from data on orphanhood**. *Popul. Stud.* 1977, **31**:107–128.
5. Brass W, Bamgboye EA, London School of Hygiene and Tropical Medicine. Centre for Population Studies: *The time location of reports of survivorship : estimates for maternal and paternal orphanhood and the ever-widowed*. London: Centre for Population Studies, London School of Hygiene & Tropical Medicine, University of London; 1981.

6. Hill K, Trussell J: **Further Developments in Indirect Mortality Estimation.** *Popul. Stud.* 1977, **31**:313–334.
7. Preston SH, Coale AJ: **Age structure, growth, attrition and accession: A new synthesis.** *Popul. Index* 1982, **48**:217–259.
8. Preston SH, Bennett NG: **A census-based method for estimating adult mortality.** *Popul. Stud.* 1983, **37**:91–104.
9. Carter KL, Williams G, Tallo V, Sanvictores D, Madera H, Riley I: **Capture-recapture analysis of all-cause mortality data in Bohol, Philippines.** *Popul. Heal. Metrics* 2011, **9**:9.
10. Honaker J, King G, Blackwell M: *Amelia II.* 2011.
11. Aday LA, Cornelius LJ: *Designing and Conducting Health Surveys: A Comprehensive Guide.* John Wiley & Sons; 2011.
12. **HMD Overview** [<http://www.mortality.org/Public/Overview.php>].
13. Bendavid E, Seligman B, Kubo J: **Comparative analysis of old-age mortality estimations in Africa.** *PloS One* 2011, **6**:e26607.
14. Heitjan DF, Rubin DB: **Inference from Coarse Data Via Multiple Imputation with Application to Age Heaping.** *J. Am. Stat. Assoc.* 1990, **85**:304–314.
15. Gakidou E, King G: **Death by survey: estimating adult mortality without selection bias from sibling survival data.** *Demography* 2006, **43**:569–585.
16. Timaeus I: **Estimation of Adult Mortality from Orphanhood Before and Since Marriage.** *Popul. Stud.* 1991, **45**:455–472.