

P. N. Mari Bhat

Age Misreporting and its Impact on Adult Mortality Estimates in South Asia*

ESTIMATES of adult mortality in developing countries usually depend upon intercei analyses of cohort survival or upon civil registration data converted into life tables. E types of analyses rely heavily upon one of the most suspect pieces of data, an individual age. In much of South Asia, Africa and Latin America, a large fraction of people do not know their true age, and what gets reported in a census or a survey is an educated guess mad either the enumerator or the informant. Any systematic biases in reported ages transmitted to estimates based on them. For example, death rates at older ages are wi suspected of being biased downwards in many developing countries by age exaggerat In this paper we will focus on patterns of age misstatements in South Asia, when errors are particularly pronounced, and their effect on estimates of adult mortality dei from intercensal analyses. In contrast toearlierstudieson this issue, which relied uponnn life tables or on assumed age-patterns of mortality, we attempt to use demographic iden to infer patterns of age misreporting and demonstrate through simulation typical bia: mortality estimates these may give rise to. Discussion below makes extensive use of I data, but conclusions of the study could easily be generalized to whole of South Asi;

Age Misreporting Patterns in India

Among the developing countries India has the distinction of conducting an uninter series of decennial censuses from 1871. Until recently, age distributions of the popi recorded in them were the sole basis of estimating levels of adult mortality. Unforti

* This paper is based on a chapter of my doctoral dissertation submitted to the University of Pennf in 1987. An expanded version of the paper that included a section on Latin America, was presented to th Seminar on Mortality Transition in East Asia and South Asia at Beijing, China during August 29-Sept 1988. I am thankful to my co-authors, Aimee Dechter and Samuel Preston, for permlting me to publish t independently. I am especially indebted to Samuel Preston, who was also the Chairman of my dis committee. Thanks are also due to the participants of the IUSSP seminar, especially, Shiro Horiuchi. Al and Alberto Palloni for helpful comments and suggestions.

1. See for example, R. Retherford and G. M. Mirza, Evidence of Age Exaggeration in Demographic of Pakistan, *Population Studies*. 36 (1982): 257-70; S. Horiuchi and A. J. Coale, A Simple Equation for t the Expectation of Life at older Ages, *Population Studies*. 36(1982): 317-26; and A. J. Coale and E. Mortality Crossovers: Reality or Bad Data?, *Population Studies*. 40 (1986): 389-402.

age distributions recorded in Indian census suffer from gross misreporting of age. The single-year age distributions show strong preference for ages ending in digit 0 and 5. The Whipple's index computed from them falls usually in the neighbourhood of 300 and the Myers' index is typically over 60. Although grouping of ages in quinquennial intervals reduces considerably the errors from digital preference, its effects can still be seen at higher intervals in the form of concentration of persons in intervals that include an age ending in zero.

That one should exercise great caution in using recorded age distributions for estimating adult mortality is forcefully brought out by a study by Dandekar.² She matched reported ages of residents of a small Western Indian town who were enumerated in both the 1951 and 1961 censuses. From her data it is possible to compute the percentage of the survivors of the population reported in ages x to $x + 5$ in 1951 reporting in ages $x + 10$ to $x + 15$ in 1961:

		<i>Reported age interval in 1951</i>										
		0-4	5-9	10-14	15-19	20-2-1	23-29	JO-34	35-39	40-44	45-49	All
Males		80	56	53	57	52	38	24	30	32	28	48
Females		75	32	51	48	38	35	31	32	23	23	40

In 1961, only 48 percent of male survivors and 40 percent of female survivors of the 1951 population were found in age intervals ten year older than what they reported in 1951. Clearly, with such data it is precarious to construct a life table by comparing the 'cohort' sizes enumerated at two censuses.

These data also show that the consistency of reported ages decreases sharply as a person's age advances. While more than three-quarters of 0-4 year olds reported their ages in the 10-14 interval in the next decennial census, the proportion of 40-44 year olds reporting as 50-54 in the following census was less than one-third. Proximity to the date of birth, and rapid physiological changes with age are probably responsible for the greater accuracy of reporting at younger ages. Dandekar's data also suggest that while both underreporting and overreporting tendencies are higher at older ages, underreporting tends to be somewhat stronger among both males and females. This inference, however, is not supported by the evidence reviewed below.

Inconsistencies between Sample Registration System and Censuses

Until recently, little independent information existed to check the validity of estimates of vital rates derived from census age distributions. For the 1970s, however, estimates from the Sample Registration System (SRS) provide a viable alternative. The SRS is a dual-record system with continuous enumeration of vital events by a local enumerator and an independent survey by an investigator—supervisor. The half-yearly survey serves the dual purpose of

1. K. Dandekar. Age Reoporting in the iwo Recent Censuses: A case study. *Indian Population Rulletin*. 3 (1986): 41-50. For a detailed review of the results of this study, see P. N. Mari Bhat. Mortality in India: Levels, Trends and Patterns. *Ph.D. Dissertation*. University of Pennsylvania (Michigan. University Microfilms. 1987).

providing an independent check on births and deaths recorded by the enumerator, and of updating the baseline information on population and housing.

Unfortunately, the census and the SRS do not agree on several counts. The discrepancies were particularly pronounced during the 1970s. For 1971-81, the SRS showed an average rate of natural increase of 2 percent per annum, whereas the 1971 and 1981 censuses showed a rate of 2.25 percent. The difference could not be explained by immigration.³ If there was an improvement in the enumeration in 1981, it was not revealed by the post-enumeration checks.⁴ While there are considerable sampling errors in the SRS state-level estimates, at the national level these are very minimal.⁵ Most likely there was either an underestimation of crude birth rate or an over-estimation of the crude death rate in the SRS.

An attempt was made by the present author to answer this question by assessing the completeness of SRS death registration through indirect techniques.⁶ The test took two forms. In the first instance, the SRS age-specific death rates were combined with age distributions from the censuses to produce an intercensal age structure of deaths. This distribution was used in conjunction with age-specific growth rates implied by the census to derive the completeness of death reports using a variable/- version of the Preston-Coale technique.⁷ The results suggested a 10-17 percent overestimation of death rates in the SRS, but confirmed birth rates recorded by the source. In the second instance, the same age-specific growth rates were used but the death distribution was derived by applying SRS age-specific death rates to the age distribution of the population also of the SRS. This exercise showed that death registration is very nearly complete, but birth reports are 7-8 percent incomplete.

The two exercises provided such vastly different results because the two sources differed with respect to the reported age distributions of the population. In particular, the SRS age distributions show significantly lower proportions of the population at older ages than those in the censuses (see Table 1). Therefore, with the same set of age-specific growth rates, the older age distributions from the censuses imply lower levels of fertility and mortality. To decide which set of results should be accepted, we must inquire why the reported age distributions of the population differ in the two sources.

After eliminating other possible reasons, it was concluded that the discrepancy arose from qualitative differences in age reporting between SRS baseline surveys and decennial

3. Actually there was significant outmigration of Indians to Gulf countries of the Middle East during 1971-81. See Bhat. *ibid.*

4. According to the post-enumeration checks, under enumeration was 1.7 percent in 1971 and 1.0 percent in 1981. See Registrar General of India. *Report on Post Enumeration Check*. Census of India 1981. Series 1. Paper 4 of 1982 (New Delhi. 1983).

5. Combining the estimated upper limit of the SRS annual crude birth rate with the lower limit of the crude death rate [see Registrar General of India. *Sampling Variability of SRS Vital Rates*. Occasional Paper 2 of 1981. (New Delhi. 1981)]. We find that upper limit of the annual rate of natural increase can be about six percent higher than the point estimate. In the decennial average of the point estimates, however, this error should be only one-tenth of the above, or just 0.6 percent. But the discrepancy between the census and the SRS growth rates is as much as 12 percent.

6. See Bhat. *loc. cit.* in In. 2.

7. S. H. Preston. Use of Direct and Indirect Techniques for Estimating the Completeness of Death Registration Systems. In: *Data Rimes for Mortality Measurement*. Population Studies. No. 84 (United Nations. New York. 1984).

TABLE I: COMPARISON OF AGE DISTRIBUTION OF THE POPULATION FROM CENSUS AND SAMPLE REGISTRATION SYSTEM, INDIA, 1971 AND 1981

<i>Age interval</i>	<i>Census 1971</i>	<i>SRS 1971</i>	<i>Census 1981</i>	<i>SRS 1981</i>
Males				
0-14	41.75	41.32	39.49	38.21
15-34	30.10	31.27	32.11	34.35
35-59	22.09	22.22	21.99	21.95
60+	6.06	5.19	6.41	5.49
Total	100.00	100.00	100.00	100.00
Females				
0-14	41.99	41.11	39.65	37.84
15-34	31.14	32.22	32.45	33.99
35-59	20.85	21.15	21.33	22.19
60+	6.02	5.51	6.58	5.98
Total	100.00	100.00	100.00	100.00

Note: Figures for 1971 are excluding Bihar and West Bengal and those of 1981 are excluding Assam. *Source:* Census and SRS publications of relevant years.

censuses. Census of India is a gigantic operation, wherein some deterioration in the quality of collected data and accentuation of common forms of biases are to be expected. In comparison SRS baseline surveys in various states were probably better managed, and employed more trained and local enumerators. Furthermore, the longitudinal character of the SRS probably helped to ensure proper recording of the baseline information, and enabled corrective steps to be taken if errors were detected in subsequent rounds. Moreover, the census-based distributions imply an implausible overrecording of deaths. The SRS age distributions are therefore likely to be closer to the true distributions.

Data on the age structure of the population are yet to become available from the 1991 census. However, the provisional population totals from the census have indicated a growth rate for 1981 -91 almost exactly identical to that registered by the SRS. But if the census age distribution continues to be different from the SRS—which it should if our explanation is correct—intercensal analyses of census data for the 1980s would produce estimates of crude birth and death rates significantly lower than the SRS.

Patterns of Net Transfers Due to Age Misstatement

Suppose we accept the results that suggest SRS death reports were nearly complete during the 1970s but its birth rates are 7-8 percent under registered. This would imply that the census age distributions, which suggested overenumeration of deaths in the SRS, are in error. The implicit pattern of net errors in the census distributions can be studied by constructing an age distribution that is consistent with the estimated levels of SRS complete-

ness, using the generalized stable population relationships.⁸ A detailed account of the procedures used in estimating the true age distribution of India's population is described elsewhere.⁹ In brief, the proportion of the population in each five-year age interval was derived using the following approximation applicable to any closed population.

$${}_5C_x \approx \frac{\exp[-5 \sum_0^{x-5} {}_5\Gamma_a - 2.5 {}_5\Gamma_x] \frac{{}_5L_x}{l_0}}{\sum_0^x \exp[-5 \sum_0^{x-5} {}_5\Gamma_a - 2.5 {}_5\Gamma_x] \frac{{}_5L_x}{l_0}}$$

where ${}_5\Gamma_x$ is the growth rate of the population in age interval x to $x + 5$, and ${}_5L_x$ is the person-years lived in the same age interval in a corresponding stationary population with l_0 annual births. For India, the requisite age-specific growth rates were computed from the 1971 and 1981 census totals after some minor adjustments for international migration. The female population of the 1971 was raised by 0.8 per cent to allow for the likelihood of improved enumeration in the 1981 census. The survival probabilities (${}_5L_x/l_0$) were taken from the intercensal life table constructed from the SRS data.

The age distribution constructed in the above manner is shown in Table 2. This age distribution can be used to study the pattern of errors in the average of the age distributions of 1971 and 1981 censuses (also shown in Table 2).¹⁰ Figure 1 compares the estimated age distribution with the average distribution computed directly from the censuses. In the left panel, the cumulative proportion of the population under a given age that is in excess of the true proportion has been plotted. It can be seen that proportion of the population reported under an age is invariably lower than the true proportion, implying a net upward transfer of persons across each specific age. This bias appears to be particularly strong in the male population.

However, the shortage of persons *under* an age in recorded distributions may not be due solely to the transfer of population from lower ages to upper ages. Such a deficit could also have resulted from the underenumeration of children. A possibility of as much as ten percent undercount at childhood ages has been suggested by a SRS-census record matching study.¹¹ It is, therefore, safer to compare cumulative distributions above an age in relative terms. This comparison is made in the right panel of Figure 1. It can be observed that the upward bias in the proportion of the population reported above an age increases steady with age, reaching an excess of 50 percent in males by age 70. Females show a similar but muted pattern of increase.

8. See S. H. Preston and A. J. Coale. Age Structure, Growth, Attrition and Accession, *Population Index*. 48 (1982): 217-259.

9. See Bhat. *loc. cit.* in th. 2.

10. Strictly speaking, the estimated age distribution for 1971-81 refers to the age distribution of person-years lived during the period. Hence it is appropriate to compare it to the mean of 1971 and 1981 age distributions.

11. See Registrar General of India. *Age and Life Tables Based on One PerCent Sample*. Paper 2 of 1974 (New Delhi. 1974).

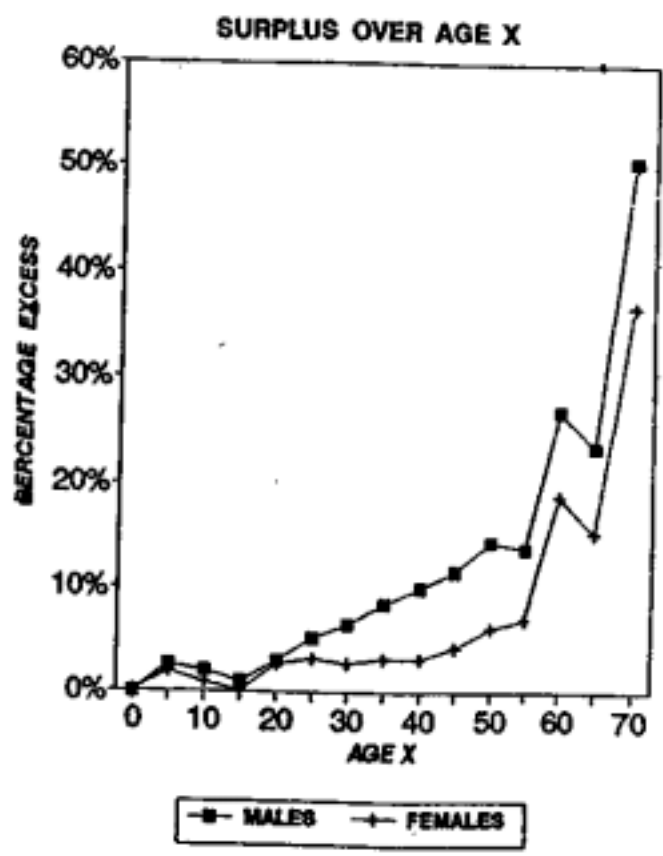
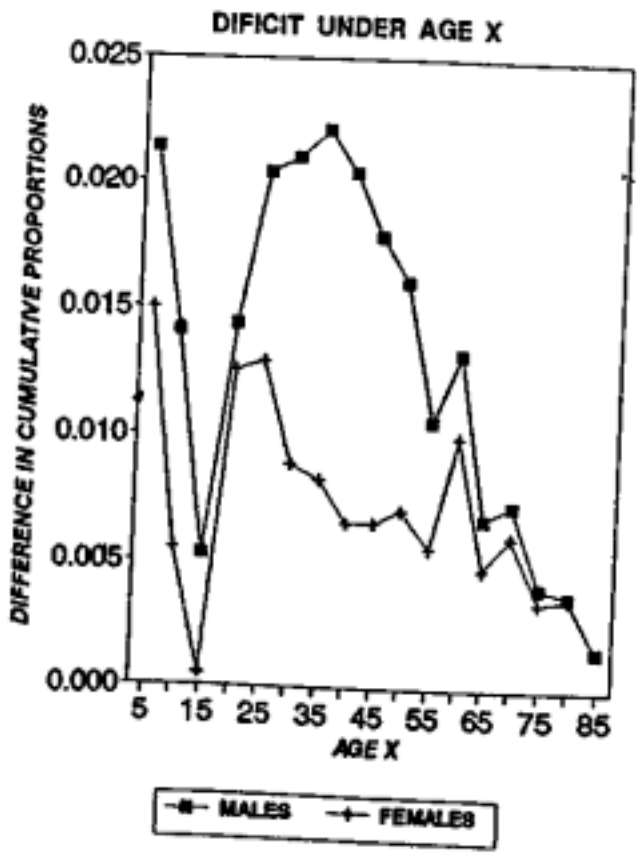


Fig. 1. Estimated Distortions in Cumulative Age Distributions, India, 1971-81

TABLE 2: RECORDED AND ESTIMATED AGE DISTRIBUTIONS, INDIA. 1971-8M

Age interval	Recorded age distribution #		Estimated age distribution §	
	Male	Female	Male	Female
0-4	13.20	13.72	15.34	15.22
5-9	14.32	14.38	13.59	13.43
10-14	12.94	12.41	12.06	11.91
15-19	9.43	8.96	10.34	10.17
20-24	8.08	8.52	8.68	8.55
25-29	7.31	7.77	7.37	7.37
30-34	6.33	6.59	6.44	6.53
35-39	5.90	5.96	5.73	5.79
40-44	5.27	5.05	5.02	5.05
45-49	4.43	4.21	4.25	4.26
50-54	4.03	3.67	3.48	3.52
55-59	2.48	2.39	2.75	2.83
60-64'	2.72	2.69	2.07	2.17
65-69	1.36	1.39	1.42	1.52
70-74	1.17	1.19	0.84	0.93
75-79	0.44	0.46	0.41	0.48
80-84	0.38	0.41	0.16	0.20
85+	0.21	0.24	0.05	0.07
Total	100.00	100.00	100.00	100.00

'Excluding the eastern states of Assam, Bihar and West Bengal.

Mean of 1971 and 1981 census figures, adjusted slightly for migration.

§ Taken from Bhat, *loc. cit.* in fn 2.

We should be cautious in concluding from the above that the tendency to exaggerate age becomes stronger as a person gets older. Such an inference is not immediately permissible because comparisons of cumulative distributions suggest only the direction of the *net* transfer across an age, which is sensitive to the slope of the age distribution at the point of transfer. For example, a bell-shaped pattern of age reporting centred on one's true age would produce a pattern of net overstatement of age, increasingly visible as age advances, because more people will be transferring upwards from lower age intervals than downwards from higher intervals. Even though individuals' age reporting may be unbiased, the resulting age distribution would be biased towards higher ages.

Matrix of Gross Transfers from Age Misstatement

In order to obtain some idea about the distribution of the reported age about its true value, we should first compute a matrix of gross transfers from each age due to age misreporting. Such a matrix was computed by the author using the biproportional adjustment algorithm.¹² The method essentially involved iteratively adjusting rows and columns of an initial guess

12. See P. N. Mari Bhat. Estimating Transition Probabilities of Age Misstatement. *Demography*. 27(1990): 149-63.

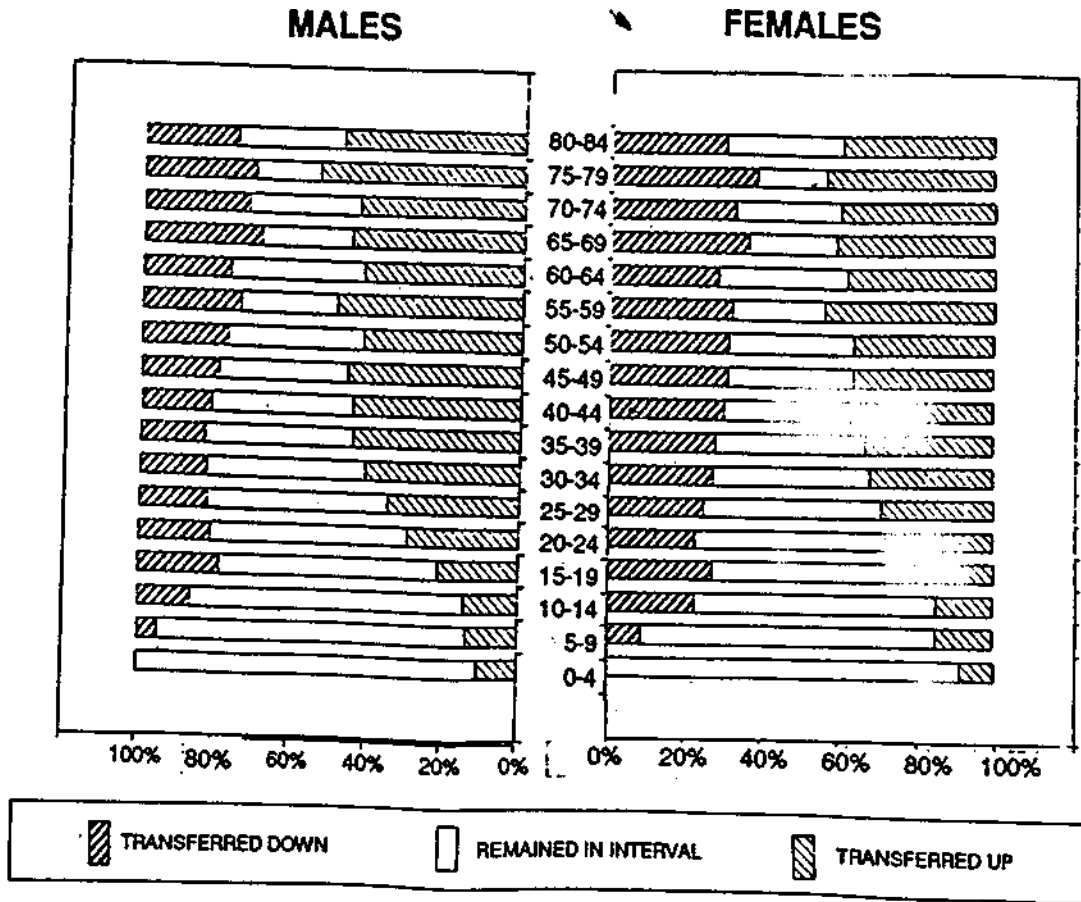


Fig. 2. Estimated Propensities for Age Overstatement and Age Understatement by Respodent's Age, India, 1971-81

matrix such that when the resultant transition matrix is applied to the estimated true age distribution, the output is the recorded age distribution. Willekens had earlier employed the same technique to estimate migration probabilities from one region to another from incomplete data.¹³

In applying the biproportional adjustment algorithm to estimate the matrix of age misstatement, I began with a matrix of gross transfers that was symmetric about its diagonal, and an error variance that increased with age as the Dandekar's study had suggested. It is important to note that, although the initial matrix was symmetric about the diagonal, this condition does not get imposed on the final estimates. The information on the symmetry/ asymmetry (i.e., the skewness) is contained in the known marginal distributions of the matrix, and probabilities of gross transfers are appropriately modified in the iteration phase. The final estimate of the age misstatement matrix derived in this manner is shown in Table 3 for males and females separately. For easy comprehension, the horizontal stacked bar chart in Fig. 2 shows the estimated percent of population who were reported in their true age interval, the percent who were transferred to ages below their true interval, and the percent who were transferred to ages above their own interval. It can be seen that the estimated coefficients retain two important properties implied by the Dandekar's study, namely, proportion of persons reporting in their true age interval decreases steadily with age (indicated by the pyramidal structure in the middle of the bar chart) and compared to males, lower proportion of females are reported in their true interval. But, unlike the initial matrix, the final estimates are not symmetric about the diagonal. It appears that in their youth, more persons overstate their age than understate it (indicated by the pyramid's leftward tilt). Men exaggerating their age outnumber those understating it by 20-30 percent. Among women this difference is considerably smaller. At no age does it exceed 10 percent. At older ages, for both men and women, contrary to common wisdom, the propensity to exaggerate age does not appreciably increase. However, results at older ages are sensitive to the assumed spread of reported ages around the true value; the more scattered are the reported ages, the lower are the estimated propensities of age exaggeration.¹⁴

Why do ages of men, and to a lesser extent those of women, tend to be exaggerated in Indian censuses? The answer could be found in the position of elder members in traditional societies wherein power and prestige grow steadily with one's age, not only within the family but also in the community at large. Under such circumstances, a natural tendency may exist to overstate one's age. At the same time, two conflicting tendencies may explain why reported ages of women have smaller upward biases than those of men. In a census, the age of the married woman is often estimated on the basis of the reported age of her husband (usually the head of the household), and an assumed age difference between spouses. This may explain why the reported ages of women may have larger error variance, a property displayed by the estimated age misstatement matrices shown in Table 3. However, the net errors tend to be smaller in the case of females because the upward bias in the age of the

13. See F. Willekens, *Multidimensional Population Analysis with Incomplete Data*. In: K. C. Land and A. Rogers (eds). *Multidimensional Mathematical Demography* New York. Academic Press. 1982.

14. See Bhat, *loc. cit.* in In 12.

TABLE 3: ESTIMATED AGE MISSTATEMENT MATRIX FOR MALES AND FEMALES, INDIA, 1971-81

Reporting age interval	True age interval																True mean age*				
	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79		80-84	85+		
0-4	0.898	0.053																		2.8	
5-9	0.102	0.812	0.138																	7.6	
10-14		0.135	0.721	0.213																12.6	
15-19			0.140	0.578	0.187	0.001														17.5	
20-24				0.209	0.520	0.176	0.001													22.0	
25-29					0.293	0.477	0.174	0.006												26.6	
30-34						0.344	0.418	0.163	0.012											31.3	
35-39						0.002	0.399	0.392	0.171	0.021										36.2	
40-44							0.007	0.408	0.374	0.183	0.035									41.1	
45-49								0.031	0.379	0.341	0.192	0.048								46.0	
50-54									0.063	0.386	0.360	0.212	0.069	0.001						50.7	
55-59										0.069	0.283	0.256	0.164	0.067	0.004					55.4	
60-64											0.131	0.377	0.352	0.242	0.107	0.011				59.8	
65-69												0.107	0.251	0.241	0.165	0.079	0.013			64.1	
70-74													0.165	0.323	0.297	0.204	0.107	0.026		68.1	
75-79															0.122	0.196	0.174	0.126	0.074	71.9	
80-84																0.004	0.218	0.311	0.283	0.218	75.7
85+																	0.014	0.222	0.472	0.682	80.5

Females																				
0-4	0.908	0.090									3.0									
5-9	0.092	0.760	0.227								8.0									
10-14		0.150	0.623	0.273	0.001						12.8									
15-19			0.150	0.501	0.227	0.003					17.6									
20-24				0.227	0.492	0.246	0.013				22.3									
25-29					0.278	0.462	0.261	0.027			27.3									
30-34					0.001	0.286	0.408	0.252	0.042		32.3									
35-39						0.004	0.301	0.391	0.258	0.061	37.3									
40-44							0.017	0.295	0.357	0.248	0.075	42.1								
45-49								0.035	0.285	0.329	0.236	0.086	0.002	46.9						
50-54									0.058	0.296	0.327	0.232	0.101	0.006	51.5					
55-59										0.067	0.236	0.245	0.180	0.091	0.010	56.0				
60-64											0.127	0.329	0.337	0.263	0.138	0.027	60.3			
65-69												0.108	0.224	0.235	0.178	0.113	0.026	64.6		
70-74													0.155	0.284	0.279	0.243	0.138	0.045	68.6	
75-79														0.002	0.108	0.170	0.186	0.138	0.093	72.4
80-84															0.014	0.206	0.235	0.305	0.260	75.9
85+																0.019	0.196	0.393	0.603	80.7

*Derived by applying the age misstatement matrix to India's estimated age distribution of 1971-81 (see Table 2).

Source: Bhat, *loc. cit.* in fn. 12.

husband is matched partially by a tendency to overstate the age difference between the husband and wife, since having a younger wife may have a prestige value.

Sensitivity of Intercensal Procedures to Patterns of Age Misstatement

Assuming that the above estimated age misstatement matrices are correct, simulations can be used to test the sensitivity of intercensal procedures to typical patterns of age misstatement. First, a population is chosen and projected forward for ten years under assumed fertility and mortality conditions. Next, age misstatement matrices are used to simulate errors in the age distributions at the beginning and the end of the projection interval. Various intercensal procedures are then applied to the 'recorded' distributions to see how far estimates deviate from the assumed true conditions. Below we describe the results of simulation exercises done to assess the effects of age misreporting on forward and backward projection methods for the estimation of adult mortality, and on the procedure developed by Preston and Bennett.¹⁵

Forward and Backward Projection Methods

The population used in the simulation was a female stable population of the West model family with an annual growth rate of two percent, and a life expectancy at birth of 45 years.¹⁶ Its fertility and mortality levels were unchanged during the projection period. The age misstatement matrix of females shown in Table 3 was employed in simulating age errors. When this matrix was applied to the stable population of the study, 58 percent of the population remained in its true quinquennial age interval, 23 percent moved to higher intervals, and 19 percent transferred to lower intervals. Thus population on the whole moved upward than downward, signifying age exaggeration. It should be noted that the female distortion pattern used here shows significantly lower propensities of age exaggeration than the male pattern, which would have produced biases in mortality estimates even larger than what is described below.

The transfer coefficients of Table 3 are capable of simulating only errors from age misstatement. But it is known that age distributions are also disfigured by age-selective underenumeration. For example, as noted above, a record matching study of the SRS and the 1971 census had suggested that as many as ten percent of children in the 0-4 interval were omitted in the census." The above matrices do not reflect age-selective omission. To make the simulation realistic, we reintroduced this error by removing 10 percent of the population from the 0-4 age interval at the beginning and the end of projection.

15. See S. H. Preston, and N. G. Bennett. A Census-Based Method for Estimating Adult Mortality. *Population Studies*. 37 (1983): 91-104.

16. See A. J. Coale, P. Demeny and B. Vaughan. *Regional Model Life Tables and Stable Populations*. Second Edn., New York. Academic Press. 1983.

17. See Registrar General of India, *loc.cit.* in In. 11.

We applied forward and backward projection methods to estimate levels of adult mortality from the simulated age distributions. As the name implies, the forward projection method consists of projecting the first census age distribution forward using a family of model life tables, and identifying which life table produces the observed population above a given age in the second census.¹⁸ Because of errors in the data and in assumptions regarding the proper mortality family, levels implied by different cumulative age segments are not usually identical. A common rule of thumb is to take the mean or median of estimates corresponding to cohorts aged 0+ to 40+ in the first census. The backward projection method describes the inverse process: the age distribution of the second census is reverse survived using model life tables and compared with sizes of the cohorts enumerated the first census. Applications of the latter method are rare. That the two methods can provide significantly different estimates in the presence of errors in the data was not widely known until recently."

In Figure 3, estimates of the mortality level derived from the application of the above procedures are plotted against age segments used in the projection. To facilitate comparison, mortality levels implied by different age segments are expressed in terms of life expectancy at age five (e^5). In addition to the simulation results, estimates derived from the application of the procedures to female age distributions of the 1961 and 1971 censuses of India are shown.²⁰ As the stable population used in the simulation belonged to the West family, model life tables of this family were used in the simulated projections. For India, life tables of the South family were used.

Figure 3 shows that age variations in the estimates of e , produced under the simulation are very similar to those of Indian females for 1961 -71. Since the misstatement patterns were inferred from the 1971 and 1981 age distributions, a relatively constant pattern of age reporting errors is indicated. The simulation results suggest that, even though systematic overstatement of age is mild among females, life expectancy at age 5 is overstated by about two years when forward projection is used and by 3-4 years when backward projection is used. The mean of the first nine estimates of e^5 is 54.9 years in forward projection and 56.5 years in backward projection, whereas the true value of e , is only 52.9 years. Even larger biases are likely when the more distorted age structure of males are employed, as shown below. Note that application of the methods to actual Indian data preserves not only the age pattern of error reflected in the simulations but also the substantially higher estimates of e , produced by backward projection than by forward projection.²¹

18. See A. J. Coale and P. Demeny, *Manual IV: Methods for Estimating Basic Demographic Measures from Incomplete Data*. Population Studies No. 42, New York, United Nations, 1967.

19. See A. Palloni and R. Kominski, Estimation of Adult Mortality Using Forward and Backward Projections, *Population Studies*. 38 (1984): 479-93.

20. The 1971 census figures were adjusted for migration, slight variation in the intercensal period, and an estimated undercount of 0.8 percent.

21. We observed that the estimates from the backward projection methods are sensitive to the open interval used in the calculation. The results quoted above employed 75+ of the second census as the last age interval. If, on the other hand, 85+ had been used as the last age category, e_5 estimates from backward projection would have been higher, on average, by about two years. However, this made little difference to the estimates of the forward projection.

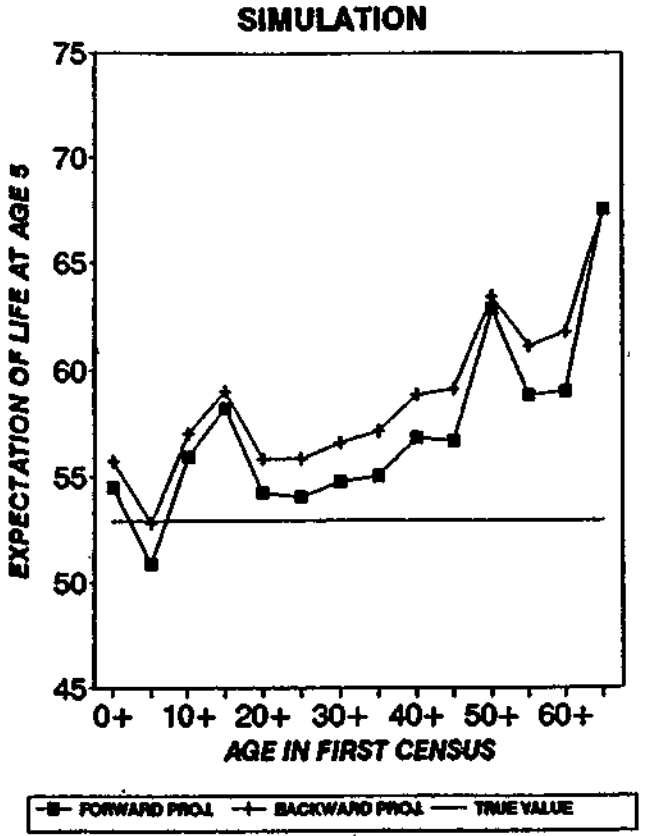
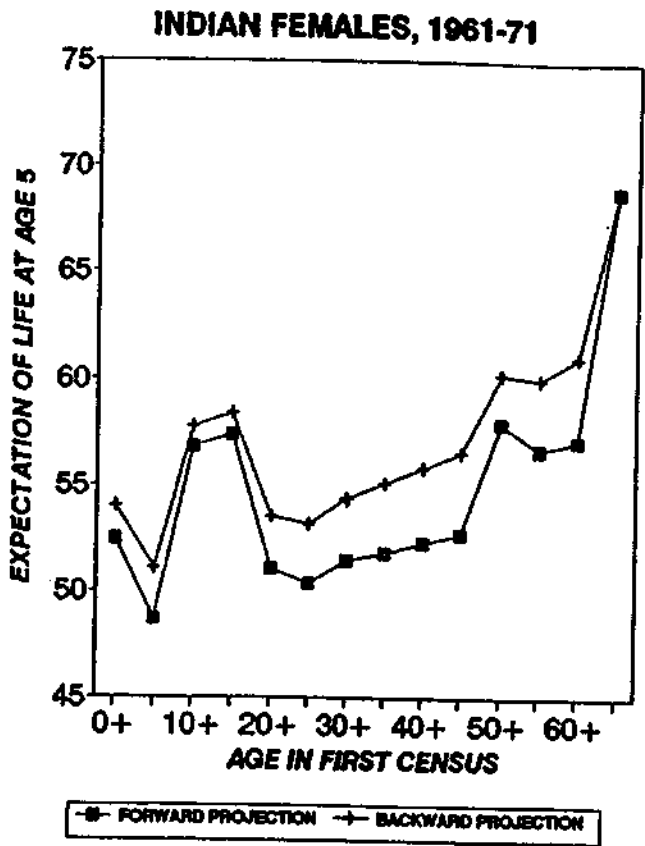


Fig. 3. Estimates of Female Expectation of Life at Age Five from Forward and Backward Projection Methods

Preston-Bennett Method

Preston and Bennett have suggested a method of estimating adult mortality levels from two enumerations of the population using generalized stable population relationships.²² The method involves the use of average age-specific growth rates of the population recorded for the intercensal period, and the average age distribution of the population for the same period. In effect, the method uses the age-specific growth rates to convert the age structure of the observed population into the age structure of a stationary population having current mortality. However, there are some ambiguities as to how the average age structure of the population for the intercensal period should be computed from the observed data. Initially the method suggested was to average (preferably geometric) populations returned at the same age at the two censuses. When the population is nearly stable, this procedure is flawless. However, when a population's age structure has large 'craters' from past fluctuations in fertility and mortality, the method is inexact. Coale has proposed an iterative method of computing the mean age distribution from the interpolation of cohort sizes enumerated at two dates." His procedure requires accurate age data in single years. An approximate method of estimating intercensal age distribution through cohort interpolations when age distributions are available in the conventional quinquennial intervals from censuses taken at ten-year or five-year apart has been developed recently by Bhat.²⁴ In exercises reported below, we have employed this method (designated as the cohort interpolation method) and the more conventional method of averaging the populations enumerated at the same age at two censuses (designated as the period interpolation method). In both cases we have employed the growth rates computed from the usual exponential formula.

In order to test the performance of the computational procedures in both stable and destabilized populations, two test populations were utilized in the simulation. Test population A is the same stable population considered earlier. Test population B has initially the estimated age distribution of Indian females for the period 1971-81 (see Table 2). It was projected forward in five-year intervals with the West model life table of level 17, and a total fertility rate of 4.5 in the initial five years, and 3.5 in succeeding five years. The actual age distributions of both hypothetical populations at the beginning and end of the projection period were distorted using the age misstatement matrix of Indian females shown in Table 3. The greater under enumeration of children in censuses was simulated by removing ten percent of children in the age interval 0-4 at both ends of the projection interval.

Table 4 reports estimates of the expectation of life at various ages for the simulated populations. In each case, two sets of figures are reported, one using the period interpolation method and the other using the cohort interpolation technique. The estimates of life expectancy at birth are made by assuming that one has knowledge of the true number of intercensal births.

22. See Preston and Bennett. *loc. cit.* fn. 15.

23. A. J. Coale, Life Table Construction on the Basis of two Enumerations of a Closed Population. *Population Index*, 50(1984): 193-213.

24. See Bhat, *loc. cit.* in fn. 2.

TABLE 4: IMPACT OF AGE MISSTATEMENTS ON LIFE EXPECTANCY ESTIMATES OF PRESTON-BENNET METHOD: COMPARISON BY METHOD OF INTERPOLATION

<i>Estimates of Life Expectancy</i>					
Age	True values	<i>With no distortions</i>		<i>With estimated distortions of Indian females</i>	
		<i>Period</i>	<i>Cohort</i>	<i>Period</i>	<i>Cohort</i>
Test Population A					
0	45.00	45.04	45.07	45.76	45.50
5	52.90	52.71	52.96	54.77	53.03
10	49.32	49.25	49.20	47.42	48.96
15	45.34	45.38	45.38	47.73	46.92
20	41.59	41.63	41.64	45.99	44.13
25	38.05	38.03	38.04	38.97	39.57
30	34.55	34.54	34.52	35.11	35.30
35	31.10	31.08	31.05	32.07	31.83
40	27.63	27.61	27.58	28.55	28.58
45	24.13	24.11	24.10	25.69	25.37
50	20.59	20.63	20.68	22.07	22.33
55	17.24	17.27	17.38	19.99	19.33
60	14.05	14.12	14.29	16.29	15.96
65	11.23	11.25	11.45	11.94	12.81
70	8.67	8.67	8.90	11.14	10.53
75	6.52	6.41	6.60	8.63	8.76
80	4.74	4.49		8.84	
Root mean square error*		0.06	0.12	2.00	1.52
Test Population B					
0	60.00	59.47	59.99	60.77	61.21
5	61.73	63.92	62.45	67.07	63.07
10	57.40	57.64	57.36	56.59	57.85
15	52.87	53.03	52.84	56.99	55.45
20	48.52	48.68	48.40	54.75	52.03
25	44.29	44.06	44.11	48.09	46.51
30	40.11	39.63	39.97	41.28	41.43
35	35.95	35.58	35.88	37.59	37.29
40	31.83	31.64	31.79	33.56	33.45
45	27.75	27.67	27.73	30.33	29.70
50	23.74	23.70	23.76	26.16	26.18
55	19.92	19.82	19.97	23.92	22.66
60	16.28	16.23	16.40	19.31	18.65
65	12.97	12.84	13.11	14.57	15.01
70	10.00	10.00	10.12	13.49	12.36
75	7.48	7.41	7.43	10.89	10.24
80	5.44	5.14		10.69	
Root mean square error*		0.59	0.20	3.16	2.15

* For ages 0 to 75

When there are no age distortions, both interpolation methods produce estimates of life expectancy very close to its true value. In the case of the stable population, the estimates based on period interpolation are slightly superior. But even the estimates from cohort interpolation, which assume the piece-wise linearity of survival function at adult ages, have a margin of error no greater than one fifth of a year. In the case of the destabilized population, estimates based on cohort interpolation are nearer to the true levels. In the example shown, estimates from the period interpolation have an average error of 0.6 years in life expectancy while those using cohort interpolation have an average error of 0.2 years. Destabilization of vital rates produces ripples in the age distribution that progress through time in a cohort-specific manner, and this process is clearly captured better by the cohort interpolation method, notwithstanding its linearity assumption.

When age errors are present in the data, both methods give estimates significantly different from the true values, although, estimates from the cohort interpolation show smaller deviations. With the distortion pattern of Indian females, estimates of life expectancy are higher than the true value at most ages. This bias in the estimates is clearly seen in Fig. 4. Though the estimates from cohort interpolations show a somewhat smoother age trend, the overall levels of mortality they imply are not much different from those derived by the period interpolation procedure. Both methods produce life expectancy estimates for ages 5 to 50 that are, on average, four percent higher than the true values. This magnitude of error is roughly the same as that observed with forward projection. Estimates typically worsen as age advances. At age 50, life expectancy estimates from period interpolation have an upward bias of 7-10 percent, and those from the cohort interpolation have an upward bias of 8-10 percent. At age 60, the upward bias increases to 16-19 percent and 14-15 percent, respectively. These biases of course, reflect the female age misreporting pattern. With the male pattern they would have been even greater.

Compared to the estimates at older ages, the biases in the estimates of life expectancy at birth (e_0) are surprisingly small. The estimates of e_0 are too high by only 1 to 2 percent, also the period and cohort interpolation methods give very similar estimates. Note that errors in the estimates of e_0 are small despite the fact that the simulations assumed that ten percent of 0-4 year old age group were not counted in both the 'censuses'. The insensitivity of the estimate of expectation of life at birth to age errors results from the fact that age reporting errors that distort the survival function at older ages in one way distort the survival probabilities at younger ages the other way. An upward transfer of persons at adult ages would result in the overestimation of life expectancy at these ages; however, in the estimate of life expectancy at birth there would be a compensating error from understatement of survival chances from birth to adult ages. Unfortunately, the estimate of e_0 cannot be derived without independent information on intercensal births.

Age Misreporting and Sex Differentials in Mortality

India is one of the few areas in the world where female mortality is said to exceed male mortality.²⁵ Although males do outnumber females in India, until recently the population's

25. See, for example, P. Visaria, *The Sex Ratio of the Population of India*. Census of India 1961. Monograph No. 10, New Delhi, 1971 and M. A. El-Badry, Higher Female than Male Mortality in some Countries of South Asia: A Digest. *Journal of the American Statistical Association*. 64 (1969): 1234-12244.

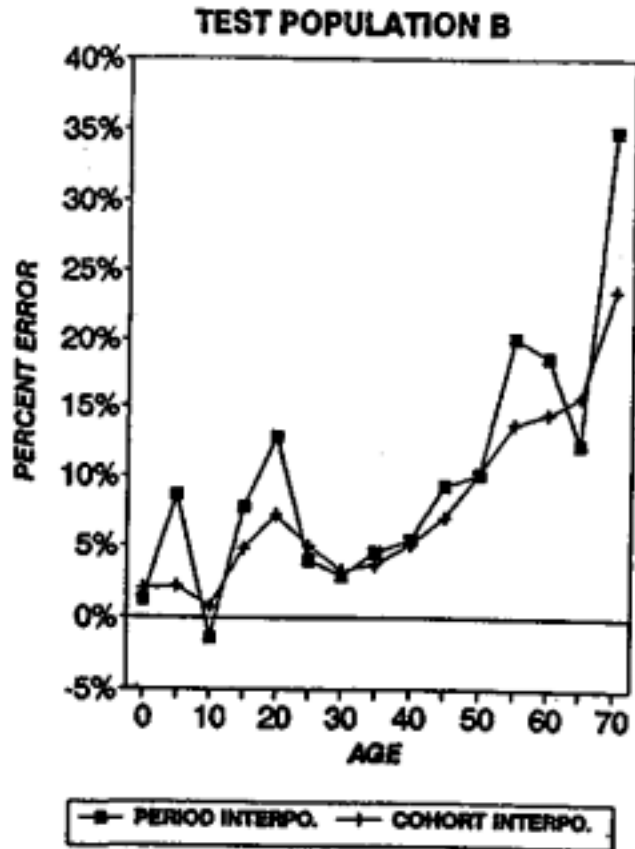
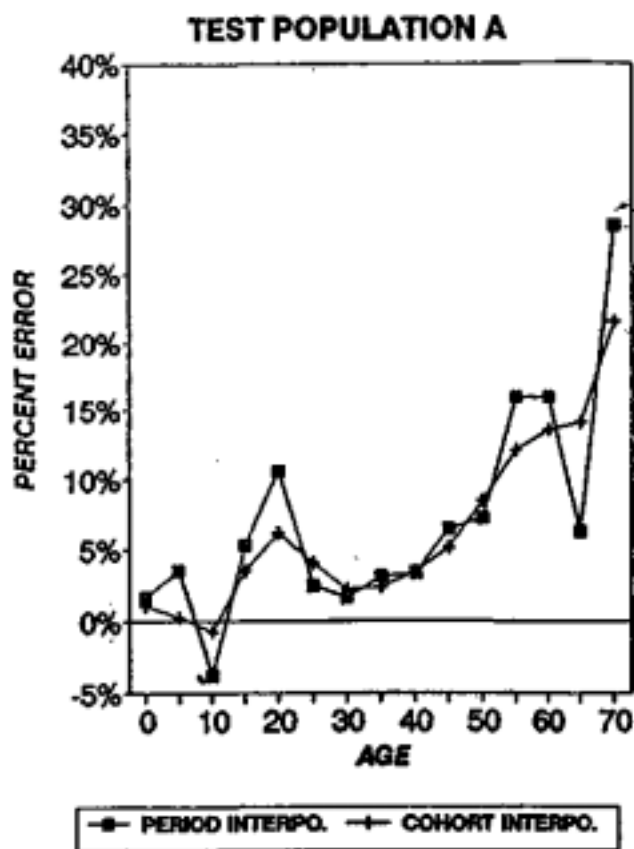


Fig. 4. Percentage Error in Estimates of Life Expectancy Derived from Preston- Bennett Procedure in Presence of Age Misstatement

sex ratio was well within the margin of variation in the sex ratio at birth found in human populations. This fact suggests that the claim that female mortality is perceptibly higher than male is dubious, particularly if there is greater underenumeration of females in censuses. Nevertheless, conventional intercensal analyses do indicate that males enjoy greater longevity. What might lie beneath this apparent contradiction?

Although not well founded, there is a general belief that age exaggeration is more common among males than females. Our analysis of the census and SRS data for 1971-81 did support such a belief (see Fig. 1). In light of the above result that age exaggeration leads to the underestimation of adult mortality, it seems possible that the reported female disadvantage in India has been exaggerated by greater exaggeration of age among males. Consider, for example, the sex-specific variation in estimates of e_5 derived from the application of the forward projection method to 1961 and 1971 census data (see the left panel of Fig. 5). One observes that male advantage in longevity is particularly pronounced in young adult ages. This pattern has sometimes been taken as a sign of unusually high maternal mortality among younger women. The average estimates of e_5 differ by nearly 2.5 years between the two sexes (54.9 in males and 52.5 in females).

In order to show that this difference may be an artifact of age reporting errors, we again employ simulations. The male age misstatement matrix shown in Table 3 was applied to the same stable population for which the female pattern was applied, and estimates of e_5 were made from the distorted age distributions. The right panel of Fig. 5 compares male and female e_5 estimates derived from the forward projection method. It can be seen that our simulation is able to replicate the sex-specific patterns observed in the estimates for 1961-71. The average of the first nine values under the male pattern is 57.0, while under the female pattern it is 54.9. But the true value of e_5 in both cases is 52.9. Thus a difference of about two years in sex-specific estimates of e_5 could arise from greater exaggeration of age among males.

The sex patterns of age misstatement used in the above simulation were based on SRS mortality and intercensal age-specific growth rates for the 1970s. It is desirable to have an independent confirmation of the indicated sex variation in the pattern of age misstatement. Such a confirmation can be obtained from orphanhood data. If patterns of age misstatement are similar among the sexes, then male and female respondents of the same age should report the same proportion orphaned, since they were born on the same year, and thus their parents were exposed to the mortality of the same ages and time spans.²⁶

For India, orphanhood data do not exist to make this comparison. But such data were collected in a 1974 retrospective survey of fertility and mortality in Bangladesh.²⁷ The left panel of Fig. 6 shows the observed difference in orphanhood proportions of males and

26. In theory some difference can occur if the probability of parents survival is not independent of children's survival, and the degree of dependency varies by the sex of the progeny. The significance of this is probably trivial, however. It is also possible for the reported incidence of orphanhood to be somewhat lower among males if male children are more likely to be adopted, and the survival of adopted parents are reported instead of the true parents. Such misreporting, however, cannot explain the age pattern of the sex difference we observe in the Bangladeshi data.

27. See J.G. C. Blacker. *Report of the 1974 Bangladesh Reproductive Survey of Fertility and Mortality*. Vol. 2. London. Ministry of Overseas Development. 1977.

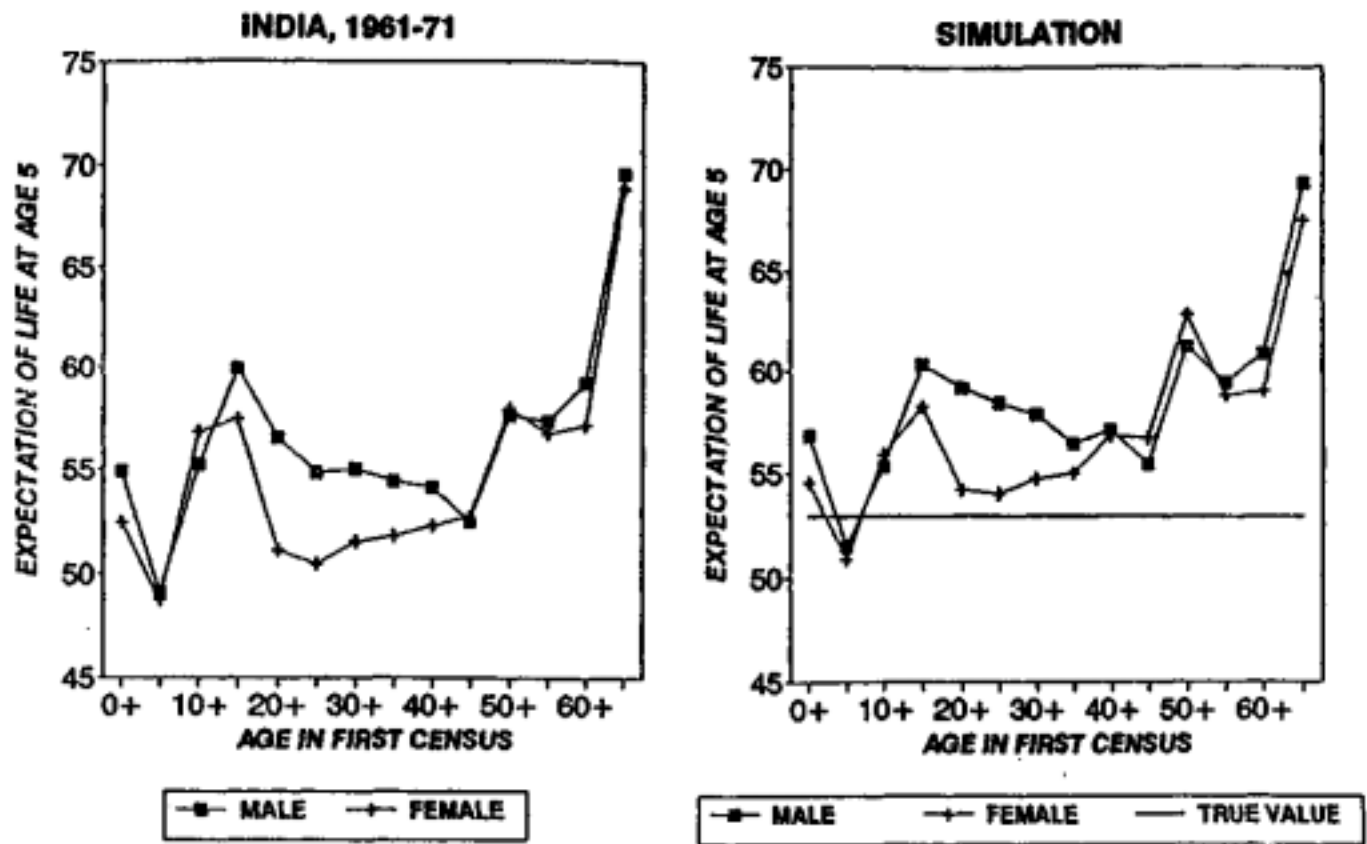


Fig. 5. Comparison of Estimates of Male Male and Female Expectation of Life at Age Five
Derived from Forward Projection Method

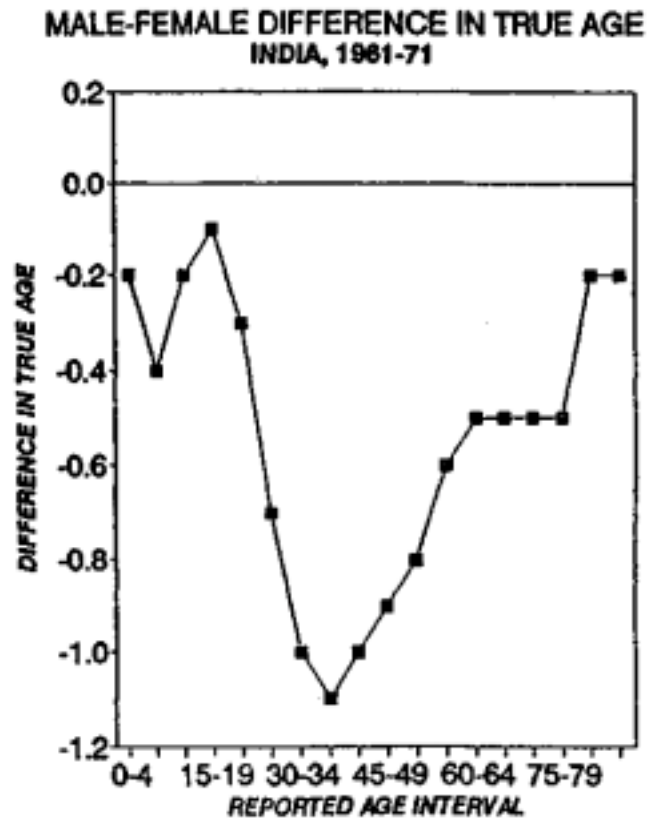
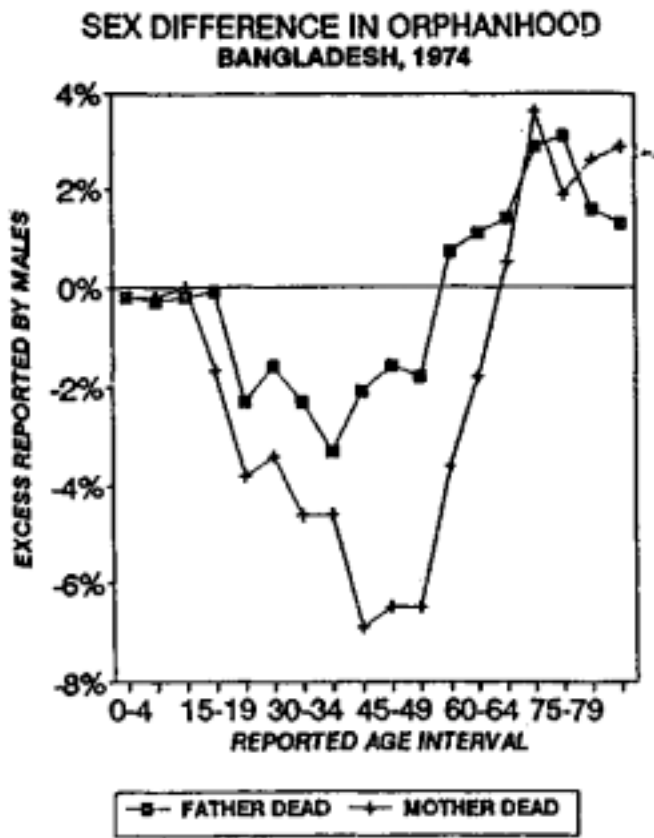


Fig. 6. Discrepancies in Orphanhood Proportions Reported by Males and Females and their Relation to Age Misreporting

and paternal orphanhood are significantly lower among males than females at ages 15 to 55, and are higher thereafter. Could this pattern reflect differences in age reporting between men and women? From the age misstatement matrices presented in Table 3 it is possible to compute the true mean age of the persons reported in a given age interval. The right panel of Fig. 6 shows a plot of the estimates of the true age differences of males and females who are reported at the same age. Clearly, the graph has the same shape as that of the reported sex differences in orphanhood. At younger adult ages, men are comparatively younger than women reporting at the same age, just as a higher fraction of their parents are reported as living. At advanced ages, however, this pattern is less pronounced, as are sex differences in reported orphanhood. Thus our estimates of sex differences in age misstatement patterns are consistent with the distortions in orphanhood proportions reported by males and females in Bangladesh.

Conclusion

From the observed inconsistencies in the estimates of adult mortality derived from the intercensal analyses and the Sample Registration System, we deduce that proportions of population in older ages returned in Indian censuses are artificially inflated. Although systematic age overstatement is probably the main culprit, even a random pattern of age misreporting can cause an upward bias in the age distribution because of rapidly tapering numbers in advanced ages. As age in is reporting alters the slope the age distribution, a simple smoothing of the distribution will not adequately correct the distortions in the data.

The use of age distributions with artificially inflated population at older ages produces estimates of life expectancy that are typically biased upward. Our simulations with an inferred pattern of gross transfers due to age misreporting suggests that estimates of life expectancy at age five derived from popular techniques of intercensal analysis are biased upward by about two years for Indian females and four years for Indian males. They suggest that sex differentials in Indian adult mortality derived from intercensal techniques may largely be spurious, resulting from greater age overstatement among males. This suggestion receives indirect support from orphanhood data in Bangladesh.

An important corollary of our findings is that as age reporting improves in the future, gains in life expectancy in South Asia would be understated, and in sex differentials in mortality would be overstated.