Population Studies, 53 (1999), 165–177
Printed in Great Britain

Effects of age misreporting on mortality estimates at older ages

SAMUEL H. PRESTON, IRMA T. ELO AND QUINCY STEWART

Abstract. This study examines how age misreporting typically affects estimates of mortality at older ages. We investigate the effects of three patterns of age misreporting – net age overstatement, net age understatement, and symmetric age misreporting – on mortality estimates at ages 40 and above. We consider five methods to estimate mortality: conventional estimates derived from vital statistics and censuses; longitudinal studies where age is identified at baseline; variable-r procedures based on age distributions of the population; variable-r procedures based on age distributions of deaths; and extinct generation methods. For each of the age misreporting patterns and each of the methods of mortality estimation, we find that age misstatement biases mortality estimates downwards at the oldest ages.

Even if mortality conditions do not improve in the twenty-first century, a substantial majority of individuals born in the world today will live to age 65. Accordingly, conventional measures of mortality, such as life expectancy at birth, are heavily influenced by mortality conditions above age 65. Unfortunately, knowledge of these conditions is quite imprecise. Mortality data at older ages are beset with a variety of problems in many populations, the most important of which is age misstatement.

Coale and Kisker (1986) have presented a useful analysis of mortality data and age distributions in countries with good and with poor quality data. These groups are distinguished by the presence or absence of population registers or long-standing civil registration systems and by their scores on indicators of data quality, such as age heaping. The authors show that recorded death rates at advanced ages in countries where data quality is poor tend to be below those found in countries with good data, a reversal of the ordering at younger ages. In addition, the reported proportions of the population at older ages appear implausibly high in countries with poor data quality. Coale and Kisker interpret these findings as reflecting a widespread exaggeration of age in such countries. In his monograph on age misreporting, Ewbank (1981) also considers age exaggeration at older ages to be one of the most common forms of age misreporting.

In this paper, we examine how age misreporting can affect estimates of mortality at older ages. We investigate the effects of three patterns of age misreporting: net age overstatement, net age understatement, and symmetric age misreporting. We consider the effect of these misreporting patterns on five types of mortality estimates: conventional estimates from vital statistics and censuses; longitudinal studies where age is identified at baseline; variable-r procedures based on age distributions of the population; variable-r procedures based on age distributions of deaths; and extinct generation methods. For each of the age misreporting patterns and each of the methods of mortality estimation, we find that age misstatement biases mortality estimates downwards at advanced ages. Thus, the downward bias in death rates at older ages may not be solely attributable to age exaggeration.

AGE REPORTING PATTERNS AT OLDER AGES: INDIVIDUALS VERSUS POPULATIONS

Coale and Kisker argue that the predominant effect of age misreporting on population estimates at older ages is to increase spuriously the number and fraction of the population at very old ages. Inflation of the population at advanced ages has been inferred using indirect methods in Pakistan (Retherford and Mirza 1982), Latin America (Dechter and Preston 1991), Puerto Rico (Rosenwaike and Preston 1984), the Soviet Union (Garson 1991), and England between 1841 and 1931 (Lee and Lam 1983).

Direct investigations of age reporting show a more mixed pattern. Gibril (1975) uses records maintained by the British Medical Council in the Gambia to show a net upward bias in age reporting in the 1973 census, a bias that increases with age. An investigation using birth and marriage records and collateral sources in an Ecuadoran village shows that a high fraction of the population reported to be aged 70+ in a household census had their ages overstated (Mazess and Forman 1979). Such a pattern is not, however, conclusive evidence that individuals' ages are, on average, overstated. In order to make such a claim, one would need to examine the distribution of reported ages among persons of a particular true age, rather than the distribution of true ages among persons of a particular reported age.

Four studies that construct the age reporting matrix in this fashion fail to reveal a net overstatement of age among older people. Ortega and García (1986) examine age reporting in the 1984 Costa Rican census. For the two regions investigated they show a symmetrical pattern of age misreporting, with a slight downward bias in age reporting above age 70 (their age reporting matrix is also presented in Dechter and Preston 1991). Using either historical calendars or age assessments among cohort contemporaries, Caldwell and Igun (1971) reveal a marked tendency for ages reported in the 1969 Nigerian census to be understated at ages 50+, especially for males. Preston, Elo, Rosenwaike and Hill (1996) trace African Americans reported to be age 60+ at death back to census records when these individuals were children and into records of the Social Security Administration. The authors show a clear tendency for ages at death to be understated, especially for females. Kestenbaum (1992) shows a similar pattern for Hispanics and non-Hispanic whites in Texas and Massachusetts.

The indirect evidence that seems invariably to suggest that there are too many very old people in censuses, surveys, and death statistics would appear on the face of it to be incompatible with evidence that ages are, on average, often understated. The two types of evidence, however, are readily reconciled. In all investigations of which we are aware, age misreporting occurs in both directions. But because age distributions taper off rapidly at older ages, the base for upward transfers into an age category is much larger (sometimes three or four times larger) than the base for downward transfers out of it. Even though the net direction of age misreporting is downwards, individuals who inappropriately move upwards into an age interval can outnumber those who move downwards out of it. Just such a mechanism has been demonstrated among African American decedents (Preston, Elo, Rosenwaike, and Hill 1996: Table 8).

A key element affecting the degree of bias in reported age distributions is the slope of the true age distribution. The steeper the slope, the more likely it is that a particular pattern of bi-directional age misreporting will bias upwards the number of persons reported in an older age interval. The slope of the age distribution of the population at age a is (Preston and Coale 1982):

$$\frac{d \ln N(a)}{da} = -r(a) - \mu(a), \qquad (1)$$

where $N(a)da$ is the number of persons aged a to $a + da$, $r(a)$ is the growth rate of the population aged a and $\mu(a)$ is the death rate at age a. Thus, the higher the growth rate and the mortality rate at older ages, the steeper the slope of the age distribution will be and the more severe will be the potential inflation of the older population from age misreporting. Growth rates at ages 65+ are now quite high throughout most of the world as a result of past mortality declines (Horiuchi and Preston 1988).

The slope of the age distribution of deaths is

$$\frac{d \ln D(a)}{da} = \frac{d \ln N(a)}{da} + \frac{d \ln \mu(a)}{da}, \qquad (2)$$

where $D(a)$ is the true number of deaths at age a. Since death rates are invariably rising with age at older ages [i.e., $d \ln \mu(a)/da$ is positive], the slope of the age distribution of deaths will be less negative than the slope of the age distribution of the population. Above the modal age at death, the age distribution of deaths will in general be less affected – less upwardly biased – by a particular pattern of age misreporting than will the age distribution of the population. Death rates estimated in the conventional way, with deaths in the numerator and population counts in the denominator, would thus in general be biased downwards at older ages by a bi-directional pattern of age misreporting that applied to both deaths and population counts.

SIMULATIONS OF THE EFFECTS OF AGE MISREPORTING ON MORTALITY ESTIMATES

To investigate the influence of age misreporting on measures of mortality at older ages, we perform a series of simulations. We assume that we know the true age distributions of the population and of deaths and then distort these distributions by an assumed pattern of age misreporting. We then identify the effect of such misreporting on mortality estimates made in one of five ways.

The population and mortality conditions that we simulate are approximately those of Latin American females during 1995–2000. The life expectancy of females at birth in this region during this period is estimated to be 72.4 years (United Nations 1995). This is an intermediate mortality level between those of more developed regions and those of most

Table 1. *Base population, death rates, and number of deaths by five-year age groups*

Age group (x to $x + n$)	Population ($_nN_x$)	Death rate ($_nM_x$)	Deaths ($_nD_x$)
40–44	13,339,000	0.00248	33,081
45–49	10,798,000	0.00382	41,248
50–54	8,667,000	0.00581	50,355
55–59	7,270,000	0.00892	64,848
60–64	6,062,000	0.01434	86,929
65–69	5,022,000	0.02440	122,537
70–74	3,774,000	0.04221	159,301
75–79	2,530,000	0.07286	184,336
80–84	1,576,000	0.11863	186,961
85–89	609,355	0.19288	117,532
90–94	146,875	0.31080	45,649
95–99	16,920	0.49531	8,381
100+	705	0.78615	554
Total	59,811,855		1,101,712

Note: The population and mortality conditions are approximately those of Latin American females in the period 1995–2000.

Table 2. *Age-misreporting matrix for African-American female decedents, 1985*

	Percentage reporting in 5–year age bracket that is				
True age	2 below true age bracket	1 below true age bracket	1 above same age bracket	2 above true age bracket	true age bracket
40–44	0.00	0.00	99.87	0.03	0.00
45–49	0.00	1.27	98.67	0.05	0.00
50–54	0.03	2.11	97.76	0.10	0.00
55–59	2.16	4.50	92.98	0.35	0.00
60–64	4.29	6.90	88.20	0.61	0.00
65–69	6.42	9.29	83.42	0.87	0.00
70–74	6.42	10.88	81.31	1.02	0.37
75–70	4.63	13.31	80.60	0.93	0.53
80–84	6.72	13.50	77.57	1.76	0.45
85–89	9.23	14.29	68.91	5.93	1.64
90–94	9.23	14.29	68.91	5.93	1.64
95–99	11.79	7.27	75.07	5.87	n.a.
100+	0.00	0.00	100.00	n.a.	n.a.

Note: Percentages may not add up to 100 due to rounding.

of the developing world. We have modelled the age-specific death rates using "West" female level 22, for which life expectancy at birth is 72.5 years (Coale and Demeny 1983). The Coale and Demeny model life tables provide age-specific death rates up to a terminal age category 100+. We maintain this degree of detail in our simulations. The population age structure is that of Latin American females in 1995 (United Nations 1994). The published age structure terminates at ages 80+. We have extended the age distribution to age 100+ by combining the model life table with the growth rate projected for Latin American females between 1995 and 2000 at ages 80+ (.0372). This growth rate is assumed to apply at age 80 and above. The parameters of this population are presented in Table 1.

The patterns of age misreporting that we have chosen to use are based upon the pattern identified in the study of African American decedents with which we were involved (Preston, Elo, Rosenwaike, and Hill 1996). This study compared ages reported on death certificates in 1985 to ages for the same individuals in U.S. Censuses of 1900, 1910, and 1920 and to ages reported for them in records of the Social Security Administration, for which age validation standards are relatively strict. Based on these three sources of data, we constructed a 'true' age at death for each individual and compared it to the age reported on the death certificate itself. The age reporting matrix in five-year age intervals for African-American females that was derived from this study is presented in Table 2[1]. The extent of age misreporting in this matrix – the proportion of persons who are reported in an incorrect five-year

Table 3. *Effect of various age reporting patterns on age distributions of the population and deaths*

Age group (x to x+n)	True	Age misreporting pattern: Understatement		Symmetric		Overstatement	
	${}_nN_x^T$	${}_nN_x^U$	Ratio*	${}_nN_x^S$	Ratio*	${}_nN_x^O$	Ratio*
40–44	13,339,000	13,475,307	1.01	13,356,783	1.00	13,238,260	0.99
45–49	10,798,000	10,997,889	1.02	10,883,895	1.01	10,769,902	1.00
50–54	8,667,000	9,066,041	1.05	8,851,005	1.02	8,635,970	1.00
55–59	7,270,000	7,508,696	1.03	7,244,461	1.00	6,980,227	0.96
60–64	6,062,000	6,081,871	1.00	5,907,999	0.97	5,734,128	0.95
65–69	5,022,000	4,754,397	0.95	4,785,321	0.95	4,816,245	0.96
70–74	3,774,000	3,555,130	0.94	3,690,557	0.98	3,825,984	1.01
75–79	2,530,000	2,346,678	0.93	2,578,306	1.02	2,809,935	1.11
80–84	1,576,000	1,360,629	0.86	1,600,355	1.02	1,840,081	1.17
85–89	609,355	484,036	0.79	620,278	1.02	758,515	1.24
90–94	146,875	145,668	0.99	221,426	1.51	295,189	2.01
95–99	16,920	31,405	1.86	60,669	3.59	89,934	5.32
100+	705	4,107	5.83	10,797	15.31	17,487	24.80
Total	59,811,855	59,811,854		59,811,854		59,811,854	
	${}_nd_x^T$	${}_nd_x^U$	Ratio*	${}_nd_x^S$	Ratio*	${}_nd_x^O$	Ratio*
40–44	33,081	33,612	1.02	33,226	1.00	32,841	0.99
45–49	41,248	43,171	1.05	42,093	1.02	41,016	0.99
50–54	50,355	55,898	1.11	52,941	1.05	49,983	0.99
55–59	64,848	74,208	1.14	68,050	1.05	61,892	0.95
60–64	86,929	98,517	1.13	89,889	1.03	81,260	0.93
65–69	122,537	128,618	1.05	120,417	0.98	112,215	0.92
70–74	159,301	167,694	1.05	157,447	0.99	147,199	0.92
75–79	184,336	186,287	1.01	182,639	0.99	178,991	0.97
80–84	186,961	168,338	0.90	177,922	0.95	187,506	1.00
85–89	117,532	92,770	0.79	104,628	0.89	117,473	1.00
90–94	45,649	39,877	0.87	51,086	1.12	61,308	1.34
95–99	8,381	10,926	1.30	17,294	2.06	23,663	2.82
100+	554	1,795	3.24	4,080	7.36	6,365	11.48
Total	1,101,712	1,101,712	1.00	1,101,712	1.00	1,101,712	1.00

* Ratio of misreported population and deaths to true population and deaths.

age interval – is similar to that found in Costa Rica, below that reported in Nigeria, and much below that reported in Gambia.

As noted earlier, this age reporting matrix reveals age misreporting in both directions, with a net downward bias for all age intervals below 100. In order to represent the effects of patterns of net overstatement, such as that revealed by Gibril in Gambia, we have simply reversed the elements in a row on either side of the main diagonal. That is, we have converted the matrix into one embodying net overstatement of age, while maintaining the proportions of persons in a particular age group who misstate their age by a certain number of intervals.

Finally, in order to investigate the effects of a symmetric pattern of age misreporting such as that identified by Ortega and García in Costa Rica, we have maintained the observed proportions with correctly stated ages in each age interval (the elements on the diagonal) but assumed that those with misstated ages were equally likely to have overstated or understated ages. The proportions with ages misstated by one five-year interval were averaged together at a particular age and that average proportion was used for both understatement and overstatement propensities; the same procedure was used for persons with ages misstated by two intervals.

EFFECT OF AGE MISREPORTING ON THE AGE DISTRIBUTIONS OF POPULATION AND DEATHS

All of the methods of mortality estimation that we will consider are based on the age distribution of the population, the age distribution of deaths, or combinations thereof. Table 3 displays how the three misreporting patterns that we have chosen affect these two age distributions. The age distributions are graphed in Figure 1a, b.

Relative to the true age distribution of the population, the effect of all three misreporting patterns is to increase the numbers of persons reported to be above a particular age. The age at

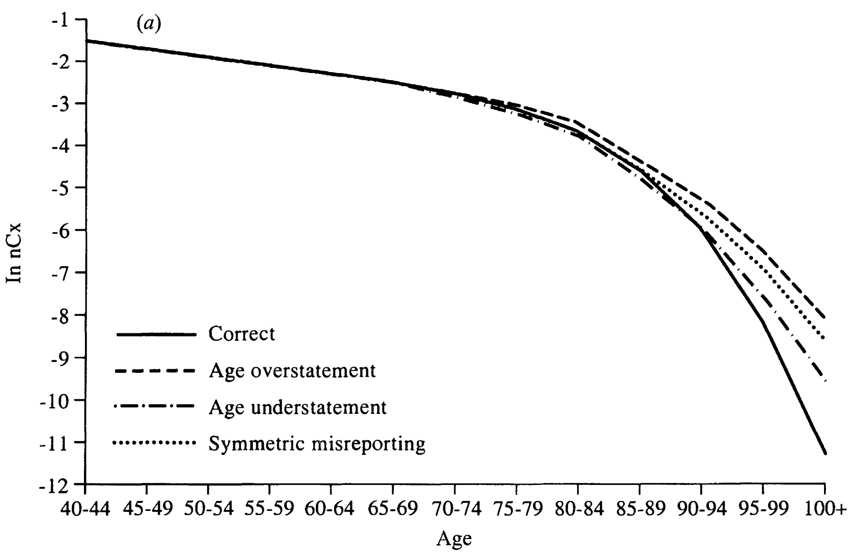

Figure 1a. Population age distributions produced by age misreporting.

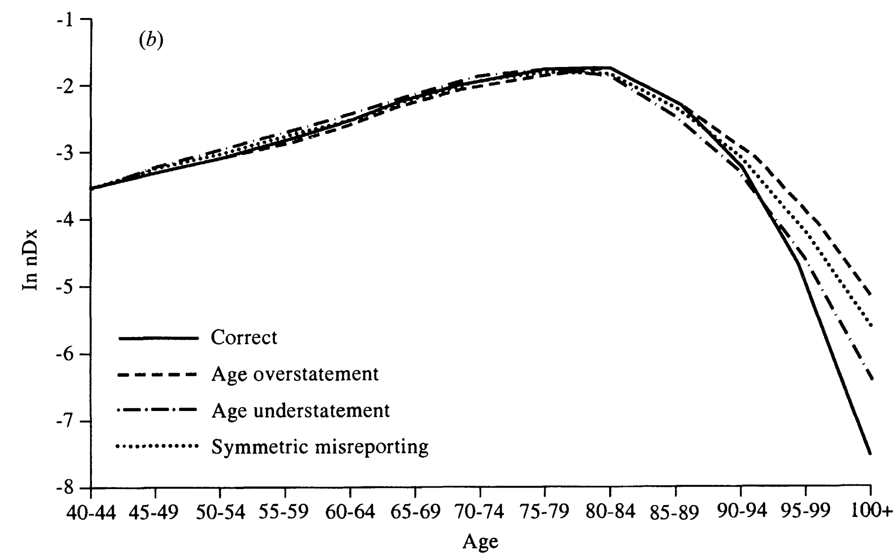

Figure 1b. Age distributions of deaths produced by age misreporting.

which the inflation begins varies with the misreporting pattern. It begins at ages 70–74 for the net overstatement pattern, at 75–79 for the symmetric age-misreporting pattern, and at 95–99 for the pattern of net age understatement.

The effects of age misreporting on the death distribution are similar but, as the earlier discussion would indicate, more muted. One important difference is that the true age distribution of deaths in the simulated population does not begin sloping downwards until after the age interval 80–84. As a result, an inflation in the number of deaths does not begin until age 80–84 for the pattern of age overstatement, 90–94 for the symmetric age-misreporting pattern, and 95–99 for understatement of age.

An age-specific death rate computed in the conventional manner is simply the ratio of the number of deaths at a particular age to the number of persons living at that age. So the difference between the effects of age misreporting on deaths and its effects on the population will affect recorded death rates, even though the underlying reporting pattern is the same for populations and deaths. Table 4 and Figure 2 demonstrate that all three forms of age misreporting reduce recorded age-specific death rates below their actual values at ages above 85–89. The error occurs at nearly all ages in the age overstatement pattern and becomes larger as age advances. It begins at ages 75–79 in the symmetric age-misreporting pattern; at younger ages, death rates are actually too high because the misreporting pattern either reduces the number of deaths by proportionally less than it reduces the

Table 4. *Effect of age-misreporting patterns on age-specific death rates calculated in the conventional manner*

Age group (x to x+n)	True	Age-misreporting pattern: Understatement		Symmetric		Overstatement	
	${}_nM_x^{\,T}$	${}_nM_x^{\,U}$	Ratio*	${}_nM_x^{\,S}$	Ratio	${}_nM_x^{\,O}$	Ratio*
40–44	0.00248	0.00249	1.01	0.00249	1.00	0.00248	1.00
45–49	0.00382	0.00393	1.03	0.00387	1.01	0.00381	1.00
50–54	0.00581	0.00617	1.06	0.00598	1.03	0.00579	1.00
55–59	0.00892	0.00988	1.11	0.00939	1.05	0.00887	0.99
60–64	0.01434	0.01620	1.13	0.01521	1.06	0.01417	0.99
65–69	0.02440	0.02705	1.11	0.02516	1.03	0.02330	0.95
70–74	0.04221	0.04717	1.12	0.04266	1.01	0.03847	0.91
75–79	0.07286	0.07938	1.09	0.07084	0.97	0.06370	0.87
80–84	0.11863	0.12372	1.04	0.11118	0.94	0.10190	0.86
85–89	0.19288	0.19166	0.99	0.16868	0.87	0.15487	0.80
90–94	0.31080	0.27375	0.88	0.23071	0.74	0.20769	0.67
95–99	0.49531	0.34790	0.70	0.28506	0.58	0.26311	0.53
100+	0.78615	0.43702	0.56	0.37788	0.48	0.36399	0.46

* Ratio of death rates based on misreported population and death data to those based on accurate data.

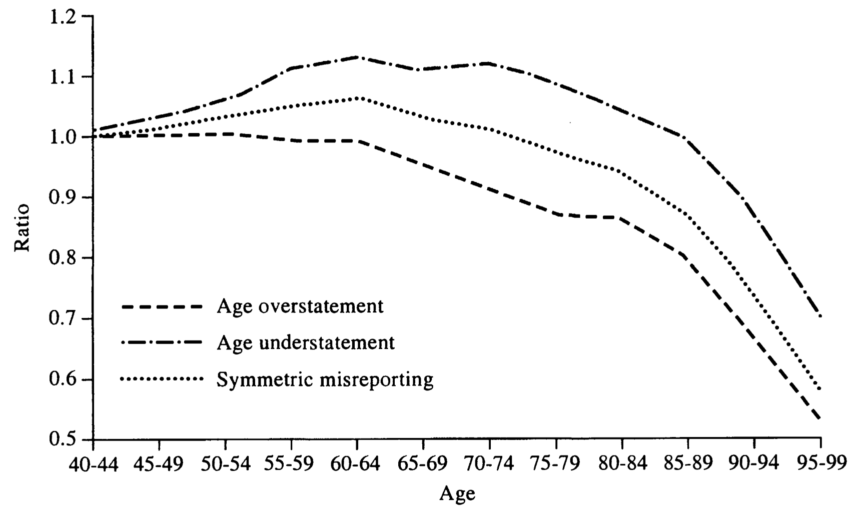

Figure 2. Ratio of estimated to true age-specific death rates.

number of individuals, or raises the number of deaths by proportionally more. The same is true for the age understatement pattern at ages below 85–89.

In the simulations presented above, we have assumed that age misreporting propensities are identical in population and death data. It is possible, however, that the degree of age misreporting differs between the two data sources. Reports of ages at death by relatives may be less accurate than ages recorded in census and survey data, which are typically reported by the individuals themselves or by the head of the household. On the other hand, incentives for age mis-statement which may exist for living individuals are unlikely to play a role in the reporting of ages for decedents. Therefore, the relative degree of bias that may exist in the two data sources is not immediately obvious.

To examine the effect of differential age misreporting in population and death data, we conducted additional simulations in which we assumed that age misreporting propensities in the population data are only half as large as those in the data on deaths. We again employ the three age-misreporting patterns described above. Our conclusions about the impact of age misreporting on mortality estimates at the oldest ages are unaffected. In the presence of net overstatement of age, mortality rates are again underestimated at nearly all ages. With symmetrical age misreporting, age-specific death rates are below their true values beginning in the age interval 70–74, and at ages 80–84 and above with net understatement of age. The differences between the true rates and those estimated in the presence of age misreporting are,

Table 5. *Effect of age-misreporting patterns on five-year death probabilities ($1–S_i$) in prospective studies*

Initial age group (i)	True	Age-misreporting pattern					
		Understatement		Symmetric		Overstatement	
	$1–S_i^T$	$1–S_i^U$	Ratio*	$1–S_i^S$	Ratio*	$1–S_i^O$	Ratio*
40–44	0.01548	0.01557	1.01	0.01553	1.00	0.01548	1.00
45–49	0.02353	0.02419	1.03	0.02383	1.01	0.02346	1.00
50–54	0.03574	0.03802	1.06	0.03684	1.03	0.03561	1.00
55–59	0.05569	0.06164	1.11	0.05861	1.05	0.05535	0.99
60–64	0.09052	0.10113	1.12	0.09541	1.05	0.08934	0.99
65–69	0.14959	0.16300	1.09	0.15283	1.02	0.14280	0.95
70–74	0.24694	0.26611	1.08	0.24518	0.99	0.22573	0.91
75–79	0.37168	0.39248	1.06	0.35886	0.97	0.33078	0.89
80–84	0.53431	0.54304	1.02	0.50138	0.94	0.47058	0.88
85–89	0.70973	0.69794	0.98	0.64379	0.91	0.61006	0.86
90–94	0.86216	0.80920	0.94	0.73701	0.85	0.69989	0.81
95–99	0.95852	0.85263	0.89	0.79912	0.83	0.78044	0.81

S_i = probability of surviving next five years for someone in age interval *i*.
* Ratio of death probabilities based on misreported data to those based on accurate data.

however, somewhat less pronounced at the oldest ages in this case than when age misreporting propensities are identical in the population and death data (results not shown).

All forms of age misreporting considered here thus produce a 'crossover' between the true age-specific death rate function and the reported function, even when a net understatement of age occurs at the individual level. The ages at which the crossover occurs under a pattern of net age understatement are 80–84 or 85–89, which are approximately the ages at which death rates for blacks and whites have crossed in the United States in recent years (e.g. National Center for Health Statistics 1985). Preston, Elo, Rosenwaike, and Hill (1996) conclude that the black/white crossover is attributable to age misreporting in the data for blacks, but their demonstration relied upon relatively unconventional variable -r methods. Clearly, such a crossover can also be produced by standard procedures.

EFFECT OF AGE MISREPORTING ON MORTALITY ESTIMATES USING LESS CONVENTIONAL PROCEDURES

A. Prospective studies

One piece of evidence that is sometimes said to support the validity of the black/white mortality crossover in the United States is that survival probabilities for blacks and whites also cross over when mortality is assessed in a prospective study and survival experience is linked to ages recorded at baseline (Manton, Stallard, and Wing 1991; Manton and Stallard 1997). Such results have been documented based on both national and subnational data sources (see Elo and Preston 1997). That such a data system will also produce distorted mortality rates in the presence of age misreporting in the baseline survey should be evident from the fact that the age distribution of the population at baseline is already deformed. It is possible, however, that distortions in this case would be small because ages are recorded at younger ages, when age misreporting is typically less severe.

To investigate the effect of age misreporting on mortality estimates derived from prospective studies, we use the three patterns of age misreporting described above. Each misreporting pattern, when applied to the true age distribution of the population, will produce a distorted baseline age distribution, which can then be cross-classified by the true population by age. In this fashion, we can identify the true ages of all persons in a particular reported age category and can assign to them their proper survival probabilities. Aggregating across the true survival probabilities of persons reported at a particular age interval indicates what the reported survival probabilities would be in a prospective study when age misreporting occurs. We have performed these simulations for both five-year and ten-year prospective studies. The results are so similar that only the five-year results need be presented.

Table 5 and Figure 3a show that the three patterns of age misstatement produce different biases in the death probabilities estimated in prospective studies. These effects are similar to the biases shown above for the conventional method. When the pattern of net age overstatement prevails, the measured probabilities of death are reduced below their true values at nearly all ages. The reason is that the population reported in an age interval in the baseline survey is invariably younger than the true population in that interval. This bias increases with age as both the slope of the age distribution and the tendency to misreport age grow. As five-year probabilities of death increase towards unity, the proportionate bias eventually begins to diminish.

Both the symmetric and age-understatement patterns produce crossovers between reported and actual death probabilities. The symmetric pattern creates reported probabilities of death that are biased downwards starting at the age interval 70–74, while the understatement pattern creates such biases starting at ages 85–89. Even with net understatement of age at the individual level, there are many more people moving into the age interval 85–89 from below than from above, biasing downwards the true ages of persons who are reported in this interval. Thus, as is the case with conventional procedures, prospective studies confronted with any of the age-misreporting patterns considered here would estimate probabilities of death that are too low beginning at ages 85–89.

B. Mortality estimates based on the age distribution of the population

We now turn our attention to two "indirect" methods of estimating mortality that are based upon variable-r procedures (Preston and Coale 1982). In one case mortality estimates are inferred directly from the age distribution of the population with no reference to deaths; in the other case inferences are made from deaths with no reference to population counts. The replacement for the missing ingredient in both cases is a set of age-specific growth rates. The reason why such substitutions are possible is indicated by equation (1) above, which demonstrates an identity between the growth rate of the population, the death rate, and the slope of the age distribution. From any two of these elements, the third can be inferred.

Preston and Bennett (1983) describe how information on the age distribution of the population and age-specific growth rates can be converted into a life table (see also United Nations 1983). We use their approach. In particular, we use the following formula to develop the ${}_5L_x$ column of the life table:

$$ {}_5L_x = {}_5\overline{N}_x \exp\left[5 \times \sum_{0}^{x-5} {}_5r_a + 2.5\,{}_5r_x\right], \qquad (3)$$

where ${}_5r_a$ is the growth rate in the interval a to $a+5$ and ${}_5\overline{N}_x$ is the mean number of persons alive in the age interval x to $x+5$ during the period of observation. The set of ${}_5r_a$'s is derived by projecting the true age distribution shown in Table 3 forward by five years and calculating the implied age-specific growth rate during the 5-year period. These growth rates are used in all applications of the procedure. We then calculate the mean number of persons in an age interval during the period, ${}_5\overline{N}_x$, as the geometric mean of the population at the beginning and end of the period.

This procedure yields precisely the correct life table ${}_5L_x$ values (to an arbitrary scalar) when all data are accurate and censuses are separated by five years. A complete life table, however, also requires l_x values. Preston and Bennett provide a rule of thumb for calculating them: $l_x = 1/10({}_5L_x + {}_5L_{x-5})$. This procedure, which assumes that l_x changes linearly over a 10-year period, proved inadequate at the higher ages where substantial curvature in l_x is visible. Instead, we have assumed that the l_x function is a second-degree polynomial over the age interval $x-5$ to $x+10$. This assumption produces the following general formula for calculating l_x:

$$ l_x = -0.5 l_{x-5} + 0.25\,{}_5L_x + 0.05\,{}_5L_{x+5}. \qquad (4)$$

In order to introduce age misreporting into the procedure, we have distorted the values of ${}_5N_x$ by the ratio of reported population to true population for each misreporting pattern shown in Table 3. The effect on mortality estimates is shown in Table 6 and Figure 3b. In this case, our basic mortality index is life expectancy at age x, used by Preston and Bennett as a convenient way of summarizing a column of ${}_5L_x$ values. The "true" value of life expectancy presented is that calculated from the Preston/Bennett method applied to undistorted data. In all calculations, life expectancy at age 100 is set at 1.21 years.

For all misreporting patterns, life expectancy at age 85 and above is overestimated relative to its true value. The overestimation of life expectancy – underestimation of mortality – begins at the earliest age, age 45, when age overstatement is present, at age 60 when age misreporting is symmetric, and at age 85 with age understatement. Thus, this method of mortality estimation also produces under-estimates of mortality at advanced ages when any of the forms of age misreporting simulated here is present. The reason is simply that the ratio of population at older ages to the population at younger ages, which is the basis of the implied probabilities of survival, is too high.

Table 6. *Effect of age-misreporting patterns on estimates of life expectancy derived from population age distributions and age-specific growth rates*

		Age-misreporting pattern					
	True	Understatement		Symmetric		Overstatement	
Age (x)	e_x^T	e_x^U	Ratio*	e_x^S	Ratio*	e_x^O	Ratio*
45	31.49	30.59	0.97	31.38	1.00	32.19	1.02
50	27.01	25.67	0.95	26.66	0.99	27.60	1.02
55	22.75	20.87	0.92	22.20	0.98	23.66	1.04
60	18.67	17.36	0.93	18.96	1.02	20.67	1.11
65	14.88	14.00	0.94	15.41	1.04	16.87	1.13
70	11.46	11.33	0.99	12.35	1.08	13.31	1.16
75	8.55	8.07	0.94	8.98	1.05	9.80	1.15
80	6.23	6.00	0.96	6.59	1.06	7.05	1.13
85	4.33	4.60	1.06	5.21	1.21	5.64	1.30
90	2.99	4.20	1.40	4.84	1.62	5.15	1.72
95	2.11	2.39	1.13	2.76	1.31	2.96	1.40

* Ratio of life expectancy based on misreported data to true life expectancy at a given age.

C. Mortality estimates based on the age distribution of deaths

The estimation method based on the age distribution of deaths and age-specific growth rates to be considered next was developed by Preston and Elo (Preston, Elo, Rosenwaike, and Hill 1996). They note that the age distribution of deaths in any population can be mapped directly onto the age distribution of deaths in the underlying life table by means of the age-specific growth rate function. A more roundabout method for performing the same mapping was earlier developed by Bennett and Horiuchi (1984). Preston and Elo use the following formula:

$$ {}_5d_x = {}_5D_x \exp\left[5 \times \sum_{0}^{x-5} {}_5r_a + 2.5\,{}_5r_x\right], \qquad (5)$$

where ${}_5D_x$ is the number of deaths observed at ages x to $x+5$ and ${}_5d_x$ is the number of deaths in the underlying life table (to a scalar that depends on the age at which estimation begins).

In order to implement this approach, we use the age distribution of deaths presented in Table 3 for each of the misreporting patterns. The growth rates to be used should be centered on the date at which the death distribution is available. This centring is accomplished by forward-projecting and back-projecting the true age distribution in Table 1 by the true life table and calculating the age-specific growth rate over the ten-year period.

The Preston/Elo procedure produces the age distribution of deaths in a life table. In order to complete the life table, it is necessary to know the value of ${}_5a_x$, the mean years lived in an age interval by those dying in the interval. This calculation is made by assuming that the number of deaths by age follows a second-degree polynomial over the 15-year period centred on the interval x to $x+5$. This assumption yields the following formula for ${}_5a_x$:

$$ {}_5a_x = \left(-\tfrac{5}{24}\,{}_5d_{x-5} + 2.5\,{}_5d_x + \tfrac{5}{24}\,{}_5d_{x+5}\right) / {}_5d_x \qquad (6)$$

As before, life expectancy at age 100 is set at 1.21 in all calculations.

The effects of age misreporting on estimates of mortality, again presented in the form of life expectancy values, are presented in Table 7 and Figure 3c. In this case, all estimated values of life expectancy are too high at ages 80 and above. For the patterns of age overstatement and symmetrical age misreporting, the values are too high at all ages. The errors introduced by age misreporting in this method are generally somewhat smaller than those introduced into the procedure based on the age distribution of the population. This result reflects the fact that the age distribution of deaths is less distorted by a particular pattern of age misreporting than is the age distribution of the population (Table 3).

D. Mortality estimates based on extinct generation methods

Finally, an alternative method of estimating mortality at older ages is to track deaths by age in a birth cohort until the last member of the cohort has died. One can then reconstitute the size of the cohorts and death rates at all previous ages. This procedure is called the "extinct generation" method of estimating mortality (Vincent 1951). It is the basis of two recent monographs on old age mortality by Kannisto (1994, 1996).

To apply this method, we first project forward our simulated population, starting at age 30–34, over a

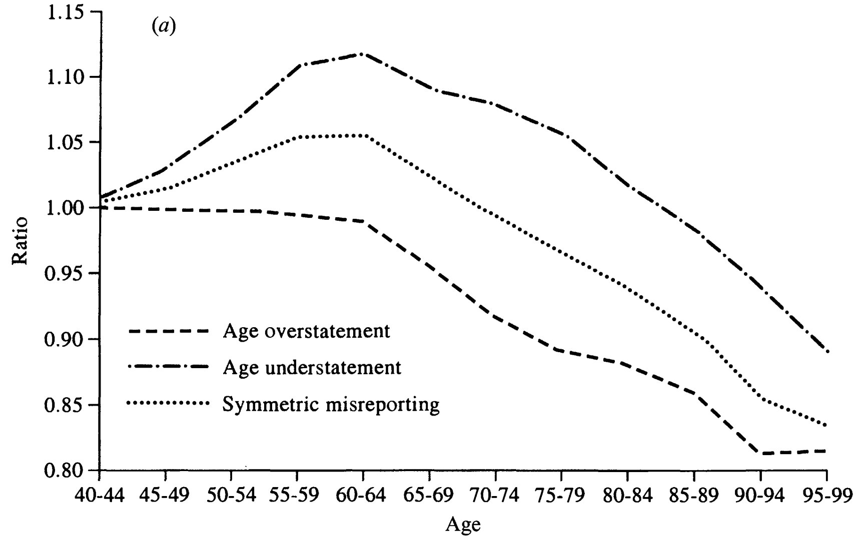

Figure 3a. Ratio of estimated to true five-year probabilities of dying.

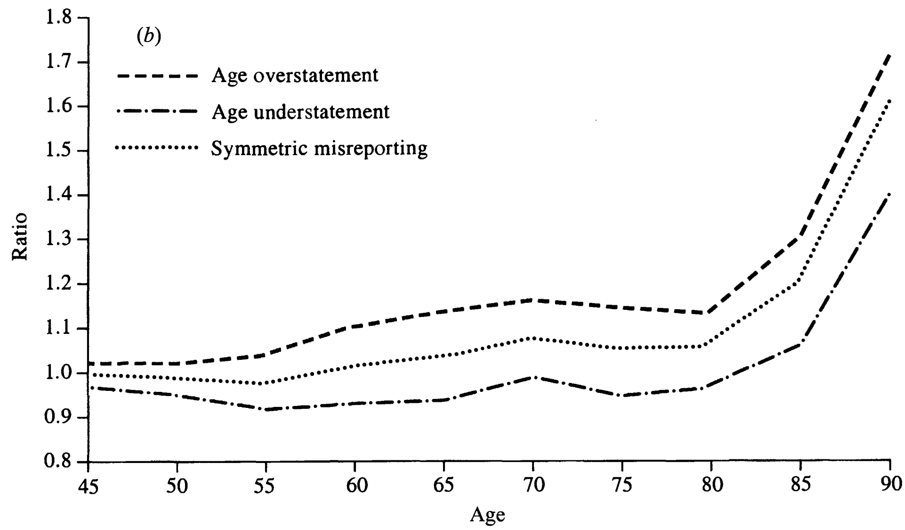

Figure 3b. Ratio of estimated to true life expectancy: Variable-r based on population.

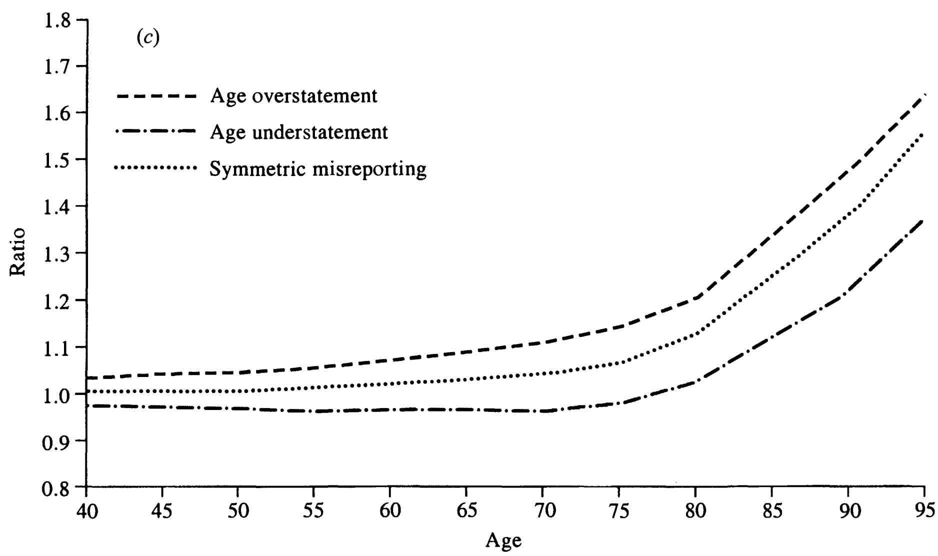

Figure 3c. Ratio of estimated to true life expectancy: Variable-r based on deaths.

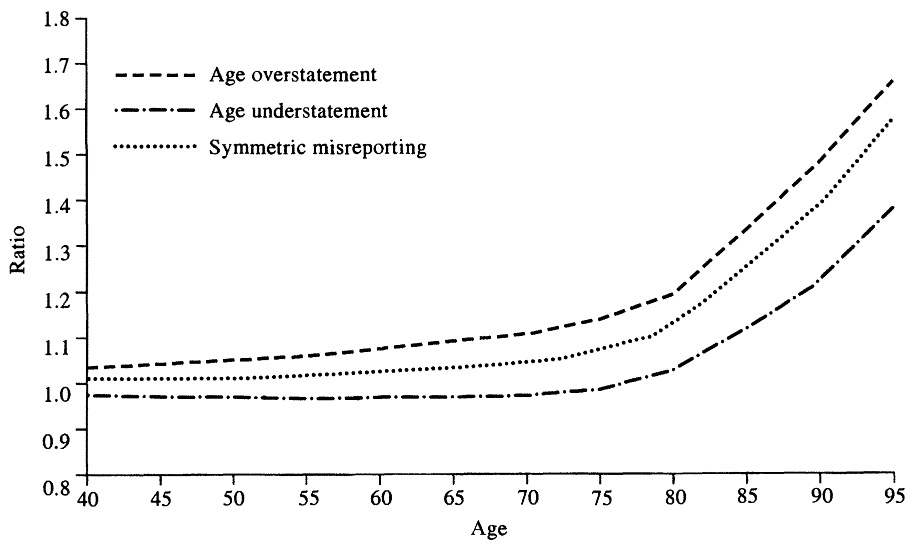

Figure 3d. Ratio of estimated to true life expectancy: Extinct generation method.

Table 7. *Effect of age-misreporting patterns on estimates of life expectancy derived from age-distributions of deaths and age-specific growth rates*

		Age-misreporting pattern					
	True	Understatement		Symmetric		Overstatement	
Age (x)	e_x^T	e_x^U	Ratio*	e_x^S	Ratio*	e_x^O	Ratio*
40	36.12	35.26	0.98	36.41	1.01	37.49	1.04
45	31.54	30.68	0.97	31.83	1.01	32.91	1.04
50	27.10	26.26	0.97	27.40	1.01	28.48	1.05
55	22.80	22.04	0.97	23.15	1.02	24.20	1.06
60	18.72	18.08	0.97	19.13	1.02	20.10	1.07
65	14.90	14.42	0.97	15.38	1.03	16.24	1.09
70	11.49	11.13	0.97	12.01	1.04	12.78	1.11
75	8.58	8.43	0.98	9.18	1.07	9.80	1.14
80	6.14	6.32	1.03	6.95	1.13	7.41	1.21
85	4.38	4.92	1.12	5.49	1.25	5.85	1.34
90	2.98	3.63	1.22	4.10	1.38	4.39	1.47
95	1.86	2.55	1.37	2.89	1.56	3.04	1.64

* Ratio of life expectancy based on misreported data to true life expectancy at a given age.

60-year period by the true life table survival rates ($_5L_{x+5}/_5L_x$). We then apply the true death rates ($_5M_x^T$) to the estimated true populations, $_5N_x^T$, to obtain the true number of deaths by age (($_5D_x^T$) = $_5N_x^T \times {}_5M_x^T$) over the 60-year period. We then count the deaths recorded to the cohort age 40–44 at baseline over its lifetime, and, based on these deaths, estimate cohort life expectancies at age x. Values of $_na_x$ are derived by the procedure described in the previous section.

To estimate the effects of age misreporting on extinct generation estimates, we apply the three age misreporting matrices to the true counts of deaths as follows:

$$_5D_x^R = {}_5D_{x-10}^T * P_{x-10,x} + {}_5D_{x-5}^T * P_{x-5,x} + {}_5D_x^T * P_{x,x} + {}_5D_{x+5}^T * P_{x+5,x} + {}_5D_{x+10}^T * P_{x+10,x} \quad (7)$$

where $_5D_x^R$ is the number of deaths reported in the age interval x to x+5 and $P_{i,j}$ is the probability that a true death in age interval i would be reported in age interval j. The values of the $P_{i,j}$'s change with age according to the three misreporting matrices discussed above. In this fashion, we obtain the number of deaths recorded for the cohort age 40–44 at baseline over its lifetime in the presence of age misreporting, on the assumption that age misreporting patterns are constant over time.

The estimates are presented in Table 8 and Figure 3d. All age misreporting patterns produce overestimates of life expectancy beginning at age 80. The distortions are less pronounced with age understatement than with either of the other two age misreporting patterns, although in every case life expectancies at the oldest ages are substantially

Table 8. *Effect of age-misreporting patterns on estimates of life expectancy derived from extinct generation methods*

		Age-misreporting pattern					
	True	Understatement		Symmetric		Overstatement	
Age (x)	e_x^T	e_x^U	Ratio*	e_x^S	Ratio*	e_x^O	Ratio*
40	36.07	35.28	0.98	36.42	1.01	37.48	1.04
45	31.49	30.71	0.98	31.84	1.01	32.89	1.04
50	27.05	26.29	0.97	27.42	1.01	28.45	1.05
55	22.77	22.09	0.97	23.18	1.02	24.18	1.06
60	18.68	18.13	0.97	19.15	1.03	20.08	1.07
65	14.88	14.48	0.97	15.41	1.04	16.23	1.09
70	11.47	11.18	0.98	12.03	1.05	12.75	1.11
75	8.56	8.47	0.99	9.19	1.07	9.76	1.14
80	6.24	6.42	1.03	7.02	1.13	7.45	1.19
85	4.40	4.95	1.12	5.51	1.25	5.87	1.33
90	3.01	3.67	1.22	4.16	1.38	4.45	1.48
95	1.86	2.58	1.38	2.94	1.58	3.10	1.66

* Ratio of life expectancy based on misreported data to true life expectancy at a given age.

overstated. Both symmetric age misreporting and age overstatement produce life expectancy values that are too high at every age, with the distortions becoming more pronounced as age advances.

The pattern of distortions produced by age misreporting in extinct generation estimates is virtually identical to that produced in the variable -r method based upon deaths (compare Tables 7 and 8). The reason is that the variable -r method is designed to produce an estimate of what a cohort's death distribution would be based on current mortality conditions; it converts period deaths into cohort deaths by means of a growth correction. The extinct generation procedure that we have simulated simply extended current mortality conditions into the future. So it is not surprising that the two methods produce very similar results. We have elected to display both sets of results because to many analysts the procedures will appear to be radically different in nature.

SUMMARY

Based on evidence derived from population-level data, it has been commonly assumed that age exaggeration is common at advanced ages. Individual-level data suggest, however, that other forms of age misreporting, such as age understatement, are common and that misreporting generally occurs in both directions.

In this paper, we have examined how three different patterns of age misreporting – age overstatement, age understatement, and symmetric age misreporting – affect mortality estimates. We consider three types of methods commonly used in developed countries as well as two methods that are typically applied to developing country data. We find that, regardless of the method employed, all three forms of age misreporting lead to underestimates of mortality at the oldest ages. The age at which the distortion begins varies according to the type of age misstatement that is present. In the presence of net overstatement of age, all mortality estimates that we have considered are too low beyond age 55. When age misreporting is symmetric in direction, all mortality estimates are too low beyond age 75. Even when a pattern of net understatement of age is present at the individual level, all forms of mortality estimation that we have considered produce estimates that are biased downwards at ages 85 and above.

Other age misreporting patterns and different demographic structures will, of course, produce different patterns of distortion. However, the distortions seem likely to vary more in magnitude than in direction. The results of our simulations suggest that analysts should be careful not to accept recorded death rates for older persons at face value unless there is good reason to believe that the incidence of age misreporting is minimal. Death rates of older persons are especially vulnerable to distortion both because age misreporting is unusually common at these ages and because any particular misreporting pattern will produce distortions that are amplified by the severe slope of the age distribution itself.

NOTES

Samuel H. Preston, Irma T. Elo and Quincy Stewart are at the Population Studies Center, University of Pennsylvania, Philadelphia, PA 19104. This research is supported by a grant from the National Institute of Aging, AG10168-02. We are grateful to Bert Kestenbaum for comments on an earlier draft.

[1] The age misreporting matrix obtained by Preston, Elo, Rosenwaike, and Hill (1996) applied to ages 65 and above. The

authors extended the matrix to younger ages by modeling age misreporting as a function of sex, state of birth, and proportion of individuals in their birth cohort who had achieved 0–4 years of schooling. In addition, a slight modification was made to the 1985 misreporting matrix to eliminate an irregularity. There was an evident tendency to report 1900 as the year of birth in 1985, thus inflating misreporting into the age category 80–84. Hence, the authors reduced the propensity to move into this age group by borrowing values from adjacent age groups (see Preston, Elo, Foster, and Fu 1998 for details on these procedures).

REFERENCES

Bennett, N.G. and S. Horiuchi. 1984. "Mortality estimation from registered deaths in less developed countries", *Demography* **21**: 217–34.

Caldwell, J.C. and A.A. Igun. 1971. "An experiment with census-type age enumeration in Nigeria", *Population Studies* **25** (2): 287–302.

Coale, A.J. and P. Demeny. 1983. *Regional Model Life Tables and Stable Populations*. 2nd ed. New York: Academic Press.

Coale, Ansley J. and Ellen Kisker. 1986. "Mortality crossover: Reality or bad data, *Population Studies"* **40** (3): 389–401.

Dechter, Aimée and Samuel H. Preston. 1991. "Age misreporting and its effects on adult mortality estimates in Latin America", *Population Bulletin of the United Nations* **31/32**:1–17.

Elo, Irma T. and Samuel H. Preston. 1997." Racial and ethnic differences in American mortality at older ages", in L. Martin and B. Soldo (eds.), *Racial and Ethnic Differences in the Health of Older Americans.* Washington, DC: National Academy of Sciences, pages 10–42.

Ewbank, Douglas C. 1981. *Age Misreporting and Age Selective Underenumeration: Sources, Patterns, and Consequences for Demographic Analysis*. Committee on Population and Demography, Report No. 4. Washington, DC: National Academy Press.

Garson, Lea Keil. 1991. "The centenarian question: Old-age mortality in the Soviet Union, 1897 to 1970", *Population Studies* **45** (2): 265–278.

Gibril, M.A. 1975. Some reporting errors in the 1973 Gambian Census. M. Sc. Thesis, University of London.

Horiuchi, S. and S.H. Preston. 1988. "Age-specific growth rates: The legacy of past population dynamics", *Demography* **25**: 429–40.

Kannisto, Väinö. 1994. *Development of Oldest-Old Mortality, 1950–1990: Evidence from 28 Developed Countries*. Odense Monographs on Population Aging, 1. Odense, Denmark: Odense University Press.

Kannisto, Väinö. 1996. *The Advancing Frontier of Survival.* Odense Monographs on Population Aging, 3. Odense, Denmark: Odense University Press.

Kestenbaum, B. 1992. "A description of the extreme aged population based on improved Medicare enrollment data", *Demography* **29**: 565–80.

Lee, R. and D. Lam. 1983. "Age distribution adjustments for English censuses, 1821 to 1931", *Population Studies* **37** (3): 445–464.

Manton, Kenneth G., Eric Stallard, and Steve Wing. 1991."Analyses of black and white differentials in age trajectory of mortality in two closed cohort studies", *Statistics in Medicine* **10**: 1043–1059.

Manton, Kenneth G. and Eric Stallard. 1997. "Health and disability differences among racial and ethnic groups", in L. Martin and B. Soldo (eds.), *Racial and Ethnic Differences in the Health of Older Americans.* Washington, DC: National Academy of Sciences, pages 43–105.

Mazess, Richard B. and Sylvia H. Forman. 1979. "Longevity and age exaggeration in Vilcabamba, Ecuador", *Journal of Gerontology* **34** (1): 94–98.

National Center for Health Statistics. 1985. *U.S. Decennial Life Tables for 1979–81*. Vol. 1, No. 1. Washington, DC: U.S. Government Printing Office.

Ortega, A. and V. García. 1986. Estudio experimental sobre la mortalidad y algunas características socioeconómicas de las personas de la tercera edad: informe de la investigación efectuada en los cantones de Puriscal y Coronado, del 3 al 20 de Junio de 1985. Serie A, No. 1048. San José, Costa Rica: CELADE.

Preston, S. and A. Coale. 1982. "Age structure, growth, attrition and accession: A new synthesis", *Population Index* **48** (2): 217–59.

Preston, S. H. and N.G. Bennett. 1983. "A census-based method for estimating adult mortality", *Population Studies* **37** (1): 91–104.

Preston, Samuel H., Irma T. Elo, Ira Rosenwaike, and Mark Hill. 1996. "African-American mortality at older ages: Results of a matching study", *Demography* **33** (2): 193–209.

Preston, Samuel H., Irma T. Elo, Andrew Foster, and Haishan Fu. 1998. "Reconstructing the size of the African American population by age and sex: 1930–1990", *Demography* **35** (1):1–21.

Retherford, R.D. and G.M. Mirza. 1982. "Evidence of age exaggeration in demographic estimates for Pakistan", *Population Studies* **36** (2): 257–270.

Rosenwaike, Ira and Samuel H. Preston. 1984. "Age overstatement and Puerto Rican longevity", *Human Biology* **56** (3): 503–525.

United Nations. 1983. *Manual X: Indirect Techniques for Demographic Estimation*. Population Study No. 81. New York: United Nations.

United Nations. 1994. *The Sex and Age Distribution of the World Population: The 1994 Revision*. Population Study No. 144. New York: United Nations.

United Nations. 1995. *World Population Prospects: The 1994 Revision*. Population Study No. 145. New York: United Nations.

Vincent, P. 1951. "La mortalité des vieillards", *Population* **6**: 181–204.