1991

The Effect of Age Misreporting in China on the Calculation of Mortality Rates at Very High Ages

Ansley J. Coale

Shaomin Li
sli@odu.edu

Original Publication Citation
Coale, A., & Li, S. (1991). The effect of age misreporting in China on the calculation of mortality rates at very high ages. *Demography*, *28*(2), 293-301. https://doi.org/10.2307/2061281

Demography, Vol. 28, No. 2, May 1991

The Effect of Age Misreporting in China on the Calculation of Mortality Rates at Very High Ages

Ansley J. Coale
Office of Population Research
Princeton University

Shaomin Li
AT&T Business Operations Analysis
Basking Ridge, NJ

When mortality rates by age are calculated from recorded deaths and enumerated populations, rates at higher ages are typically in error because of misstated ages. Mortality rates for China in 1981 have been calculated from the number of deaths in 1981 in each household recorded in the 1982 census, and from the census population back-projected one year. Because age was determined from date of birth, and because persons of the Chinese culture have very precise knowledge of date of birth, the mortality rates even at high ages should be unusually accurate. This expectation is fulfilled for most of China, but severe misreporting of age is found in a province that contains a large minority of a non-Han nationality, which lacks precise knowledge of date of birth. Although the province contains only 1.3% of China's population, male death rates above age 90 for all of China are distorted seriously by the erroneous data from this location.

When death rates at each age are calculated from the number of registered deaths and the number of persons derived from a census, the resultant mortality rates at higher ages are frequently an understatement of the true rates. Such understatement is to be expected in populations in which the records of age in censuses are inexact, and also in which the recorded age at death of decedents is not precise. Ages of older persons as recorded in censuses (and of older decedents as registered) are especially subject to error, and most often lead to a net relative overstatement of the number of persons (and often of deaths) that increases with age. The typical effect on calculated mortality rates at advanced ages is understatement: true death rates above age 95 or 100 (and often above 80 or 85) are higher than the directly calculated rates.

In the United States, there are errors in the ages recorded both in the register of deaths and in censuses; mortality rates at higher ages as calculated directly from registered deaths and numbers of persons derived from the census are lower than true figures. Even when death rates are based on data taken from the Medicare rolls, they are too low. Ages of the very old are overstated in the census, ages at death are overstated in the register of deaths, and ages recorded for very old enrollees in Medicare are also too high (Coale and Kisker 1990). The number of centenarians is especially prone to exaggeration. In England and Wales, where misreporting of age at most ages is infrequent, the number of centenarians recorded in the 1971 census was estimated as about two times too large because some persons less than 100 years old overstated their ages (Thatcher 1981).

In many less developed countries, exact knowledge of one's own age or the age of other members of the household is rare. Thus the ages recorded in a census are rough estimates by the respondent, by the enumerator, or by a third party. The existence of inaccuracy is evident in the recorded sequence of the number of persons by age. Far too many persons are recorded at ages preferred by the respondents or the enumerators, such as even numbers, and especially numbers divisible by 10. The number of persons recorded at age 70 is as much as ten or twelve times as great as the average of the numbers at ages 69 and 71. When age is severely misreported because it is not known accurately, overstatement of the age of older persons is extensive. For example, the census of Iraq in 1957 showed 14.4 times as many persons at age 70 as the mean number at ages 69 and 71; the corresponding ratio in West Germany in 1961 was 1.004. Iraq, with a total population of 6.3 million, listed 5,759 centenarians; Germany, with a population of 54 million, listed only 400.

One result of inaccurate listing of age is a systematic understatement of death rates at older ages, an understatement that increases with age. "Smoothing" of the distributions of deaths and population does not yield valid approximate death rates. As noted by Rosenwaike and Preston (1984), age misreporting at higher ages in Puerto Rico accounts for the very low death rates above age 75 or 80, for the very high expectation of life at 65, and for a substantial overstatement of expectation of life at birth. The official mortality rates in Puerto Rico are higher than in the United States at younger ages but become lower at older ages. This "crossover," whereby populations subject to high mortality in the early part of life have low mortality at later ages, has been noted widely, but in most instances the mortality of the less privileged population is understated at older ages, so that the "crossover" is the consequence of faulty data. (Coale and Kisker 1986).

In this article, the death rates at each age in China based on the deaths at each age in 1981 and the number of persons at each age in 1982 as listed in the 1982 census are examined for evidence of distortion caused by overstatement of age among the elderly.

Generally Accurate Reporting of Age in the Censuses of China

Respondents in China, even if illiterate, can supply a precise date of birth (according to the traditional calendar) for themselves or their children. The Chinese calendar consists of a cycle of "animal years" that repeats every 60 years. (There are 12 animals, and each animal has five different qualities). Each year is composed of either 12 or 13 lunar months. Because there are 29.5 days in a lunar month, there are about 12.4 such months in a year, and a 12-month year would become increasingly out of synchronization with the solar year of 365.25 days. The Chinese astronomers therefore designed a calendar in which, at irregular intervals (sometimes two years, sometimes three), an extra lunar month is inserted between the two New Years that bound an animal year. As noted below, these "leap years" of 13 lunar months create a slight systematic distortion in the recorded age distribution in censuses of China. The precise date that respondents can supply includes the animal year, the lunar month, and the day of the month. Some of our Chinese students tell us that a parent or a grandparent informed them of the *hour* of their birth. The date of birth, which of course never changes, is significant in making daily decisions throughout one's life, and can be recalled with no effort. In designing the questionnaire for censuses or surveys in China, it is important to ask for date of birth in the lunar calendar, and to compute age by converting the date to the Western calendar and subtracting from the date of the census or survey. If the questionnaire asks the individual's age, the Chinese system of reckoning age makes the response ambiguous. In China a person is counted as one year old on the day of birth, and one year older with each new year.

Age was determined from date of birth according to the Chinese calendar in the national censuses of 1953, 1964, and 1982. Nevertheless, in these censuses slightly too many persons were recorded at ages divisible by 10. We calculated an index of the deviation of the number at each age from a smooth sequence by taking the ratio of the number at each age to a two-stage moving average (the five-term average of a five-term average); with no heaping, the mean index at ages divisible by 10 from 40 to 90 would be about 1.0; for males the mean index at these ages was 1.085, 1.040, and 1.018 respectively in the three censuses. The corresponding average for males enumerated in India in 1971 was 3.637, indicating extensive errors in reported ages. In short, there is very little preferential reporting of ages ending with a zero in the Chinese censuses, compared to censuses in South Asia, Africa, or Latin America. There is, however, a minor but intriguing form of heaping that characterizes the reported age distributions in China and in some other East Asian populations. The translation of the number of persons born in a past one-year period according to the Chinese calendar into the number born in a one-year period in the Western calendar can be biased because a Chinese year containing 13 lunar months is 8.3% longer than a year containing 12 months. If each lunar year is considered equivalent to the Western year with which it overlaps, too many births may be ascribed to a Western year considered equivalent to a year with 13 lunar months. In 1982 the median index of deviation of persons with an age implying birth in a 13-month lunar year was 1.035; the index was .982 for those whose birth was in a 12-month year. This minor heaping on "leap years" suggests accurate reporting of dates in the Chinese calendar but imperfect conversion to the Western solar calendar.[1]

The 1982 census included information on deaths that had occurred in 1981, the year preceding the census. The deaths are classified by gender and single years of age. Each household reported deaths of members of the household that occurred in 1981, giving the decedent's date of death and date of birth. Mortality rates (and life tables) for 1981 are derived from the number of deaths thus ascertained, and the number of persons at each age in 1981 is calculated from these deaths and from the number of persons one year older in 1982. The listing of deaths in 1981 shows slightly more "heaping" on ages divisible by 10 than does the listing of the population in 1982. The average index of deviation for male deaths in 1981 at decadal ages from 40 to 90 is 1.052, compared to 1.018 for male persons in 1982. Evidently, relatives of the decedent knew the decedent's date of birth less precisely than living persons knew their own date of birth.

Less Accurate Reporting of Age among Some Minority Populations

The accurate reporting of date of birth in China is a characteristic of the population of the dominant Han culture. About 68 million of the 1 billion persons enumerated in 1982 belonged to non-Han minorities. Some of these minorities, such as Mongolians and Koreans, share with the Han an accurate knowledge of date of birth in the lunar calendar; other minorities do not. We do not have access to a tabulation by single years of age of the minority nationalities from the census, but the age distribution of these nationalities compiled in a 1/1000 sample of the population in a fertility survey in 1982 provides at least an indication of minority age heaping. The index of deviation was calculated for the combined age distribution of the minority populations in eight provinces (Ginghai, Ningxia, Inner Mongolia, Yunnan, Xinjiang, Guangxi, Guizhou, and Gansu). The minority populations totaled 51,858 in the sample. The average deviation index at ages 40, 50, 60, 70, and 80 was 1.318; for a particular minority, the Wei (also called the Weiwuer or Uyghur) it was 2.234.

The 5.96 million members of the Wei minority are concentrated in Xinjiang Province, where they constitute 46% of the province total. The average value of the deviation index at decadal ages from 40 to 90 for males in this province is 1.647, compared to 1.018 for all China. When the male population of Xinjiang is subtracted from the all-China population, the average deviation index for these ages is 0.999; thus heaping on ages divisible by 10 in China arises from the very strong heaping in this province, largely from one minority population.

Overstatement of Age at Very Old Ages in China

When age at death is recorded precisely, deaths above age 110 are extremely rare. For example, in extended periods of experience in France, Italy, the Netherlands, Sweden, Switzerland, and Japan, there were only two deaths above 110 (at 111 and 112; Coale and Kisker 1986). The highest age at death in 1981 in China recorded in the 1982 census was 135; the highest age of a living person in 1982 was listed at 130. The 130-year-old was in Xinjiang Province. As noted earlier, overstatement of age is general among the very old when knowledge of age is imperfect. There were 144 males listed as over 110 years of age in China; 121 were in Xinjiang and another 15 in four other provinces with the highest proportion of minorities whose cultures are not related closely to the Han. Five provinces with the highest fraction of these minorities (Xinjiang, Guanxi, Qinghai, Ningxia, and Yunnan) contained less than 9% of the population of China but 93% of males listed as over 110. In nine provinces with a total population of 217 million (mostly persons of the Han culture), nobody, male or female, was listed as over 110. It is very probable that the true number of persons over age 110 in China was zero, and that those so listed are simply instances of exaggerated age.

Age overstatement is not limited to listings above 110, or even 100. When knowledge of age is imperfect, typically too many persons are listed over 95, 90, 85, and even above lower limits. For example, in the provinces of China the median ratio of the number of males listed as over 95 to the number listed as over 70 is 4.54 per 10,000; the corresponding ratio is 74.5 times as large in Xinjiang, and is more than 2.5 times the median in the other four provinces with the largest fraction of non-Han-related minorities. When age is thus overstated, death rates calculated from deaths and population as listed by age are too low at the upper range of the age distribution. Even when overstatement of age is about the same for deaths and for population, it assigns the death rate at age x to $x+d$, where d is the extent of overstatement of age. Mortality rates decline by about 10% per year of age among the older population; therefore if d is about five years, the death rate at $x+d$ would be understated by about 40%. The male death rate calculated at 95 to 99 years of age in Xinjiang in 1981 is only 29% of the median in the 28 provinces for which a rate can be determined. (Deaths in 1981 were not asked in Tibet.) The male rates in the other four provinces with the largest proportion of non-Han-related minorities ranged from 46% to 73% of the median, and (with Xinjiang) were the five lowest mortality rates in this age range.

Mortality in Xinjiang Province and in the Other Provinces Combined

Particularly extreme misreporting of age in Xinjiang has been indicated by the high degree of age heaping, by the presence of the person with the highest reported age, and by

84% of the males in China reported as age 110+ in a province with only 1.3% of the total population. One consequence of the invalid allocation by age of deaths in 1981 and of persons in 1982 in Xinjiang is a major distortion in the sequence of death rates obtained by dividing the number of deaths listed at each age in 1981 by the number of persons at each age constructed from the number listed as one year older in 1982. In Figure 1, the directly calculated death rates from 55 to 100 for males in Xinjiang are plotted, together with the rates for all of China and for China without this province. The distorted nature of the death rates in Xinjiang is quite evident: the rates at ages 60, 70, 80, and 90 are much above neighboring ages, and the rates at 59, 69, 79, and 89 are much below. The rates at ages divisible by 10 are doubly inflated by heaping on the number of deaths reported at these ages in 1981 and by understatement of the denominators, derived from the numbers of persons listed as 61, 71, 81, and 91 in 1982. At the ages ending in 9, the number of deaths is understated, the denominator is inflated, and the death rates actually seem to be more severely displaced than at ages ending in 0.

The second obvious distortion in the male death rates in Xinjiang is the near absence of a continued upward trend in death rates above age 80. The contrast with the death rates for China without this province is especially conspicuous in regard to these features. There is very little irregularity in the mortality rates for the remainder of China, and those rates continue to rise quite smoothly. In fact, the sequence of rates for males in China without Xinjiang is strong evidence of the feasibility of calculating valid death rates to very high ages, without adjustment, from data in the 1982 census. This approach is feasible because of the unusual accuracy of reported ages in China — if the effect of very inaccurate ages supplied by a minority that does not share the Han knowledge of date of birth can be kept out of the calculations.

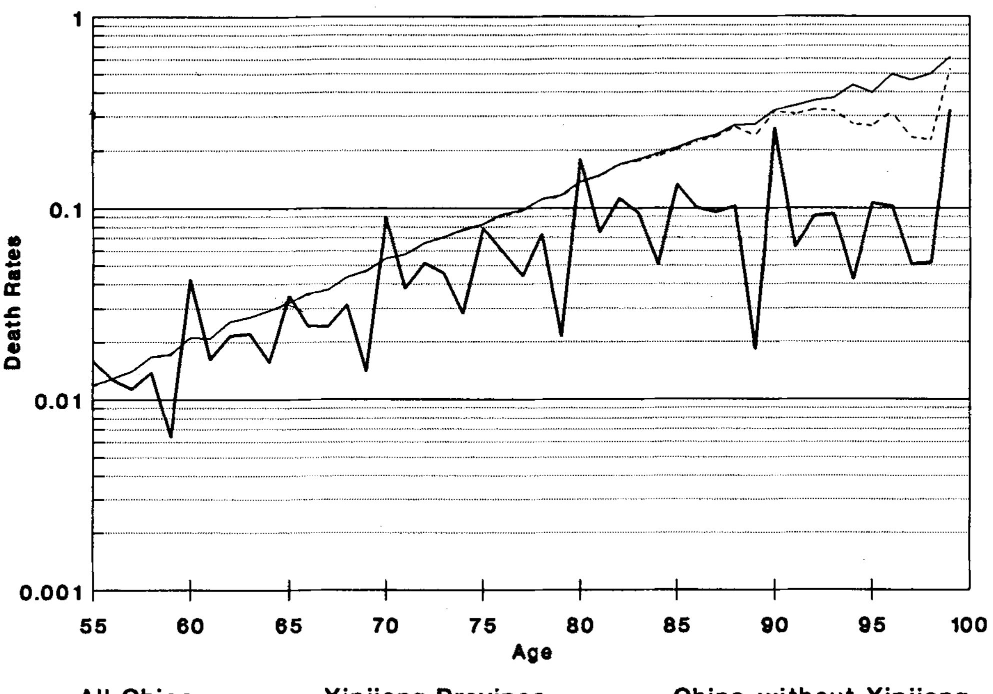

Figure 1. Death Rates, Chinese Males, 1981

The mortality schedule above age 55 for males in all China compared with the schedule for males in China without Xinjiang province can be seen in Figure 1. On this logarithmic scale there is no visible difference until above age 85; from 90 to 98, however, the death rates in all China decline, and the rates continue to increase when Xinjiang is omitted. How can the inclusion of a province with only 1.3% of the population of China distort the mortality rates so severely above age 90? It can do so because age overstatement is so extensive in Xinjiang that the number of males listed at ages 95 to 99 in this province is 47% of the number in all China. The greatly understated death rates in Xinjiang above age 90 distort the national death rates because Xinjiang contains so many younger men with exaggerated ages and low death rates, especially at stated ages 95 to 99. Males in Xinjiang account for nearly half of the males 95 to 99 years of age in all China.

Because of the biases often found in the data from which death rates at very old ages are derived, and because of inevitable random variation at ages where the number of persons and the number of deaths become very small, the construction of mortality rates and life tables often includes the use of a mathematical function to "close out" the set of death rates above some high age such as 85 or 90. Standard formulae such as those of Gompertz or Makeham are employed to estimate the continued increase in mortality rates above the age limit where the directly calculated rates are accepted. A recently introduced method for closing out death rates above age 85 or 90 fits the recorded rates for males and females in Sweden, 1978-1982, quite closely — more closely, in fact, than the official rates, which are closed out by a traditional formula. The new procedure accepts as valid the directly calculated rates (smoothed slightly to minimize random variation) up to age 80 or 85, and also accepts the rate of increase with age in the death rate at 80 or 85. It then ascribes a linear change in the rate of increase in mortality with age from 80 (or 85) to 110, a change chosen to yield an annual death rate of 1.0 (or an average monthly rate of .083) from 110 to 111. The linear change in the rate of increase in the death rate duplicates the pattern in populations with exceptionally accurate data; the limit for the mortality rate from 110 to 111 is consistent with the presence of few survivors above 110 (Coale and Kisker 1990).

When the mortality rates for China without Xinjiang are closed out by this procedure, the adjusted rates differ very little from the directly calculated rates, even above age 90, as in Sweden (see Figure 2). In contrast, when the mortality rates for all China are closed out, the adjusted rates above 85, and especially above 90, are distinctly higher than the directly calculated rates (see Figure 3). The adjusted rates for all China are only slightly different from the unadjusted (or the adjusted) rates for China without Xinjiang. In other words, the effect of severe errors in reported ages in one province with less than 2% of the total population on the calculated death rates at high ages can be minimized either by omitting the data from this province or by using a suitable procedure for closing out death rates above age 80.

Discussion

In many populations, including those in most less-developed countries but also in the United States, the death rates of the very old population are in error (typically too low) when calculated from official data, because of misstated ages in the official sources.

Because of the unusual accuracy of the reported date of birth in China (in the majority Han population and among minority nationalities who have the same belief in the importance of date of birth and who use the same calendar), death rates for ages as high as 90 to 100 can be calculated from the listed age of persons in 1982 and from the listed age at death of decedents in 1981, both from the 1982 census. The death rates at high ages are subject to serious distortion from misreported ages if the data for males in all China are

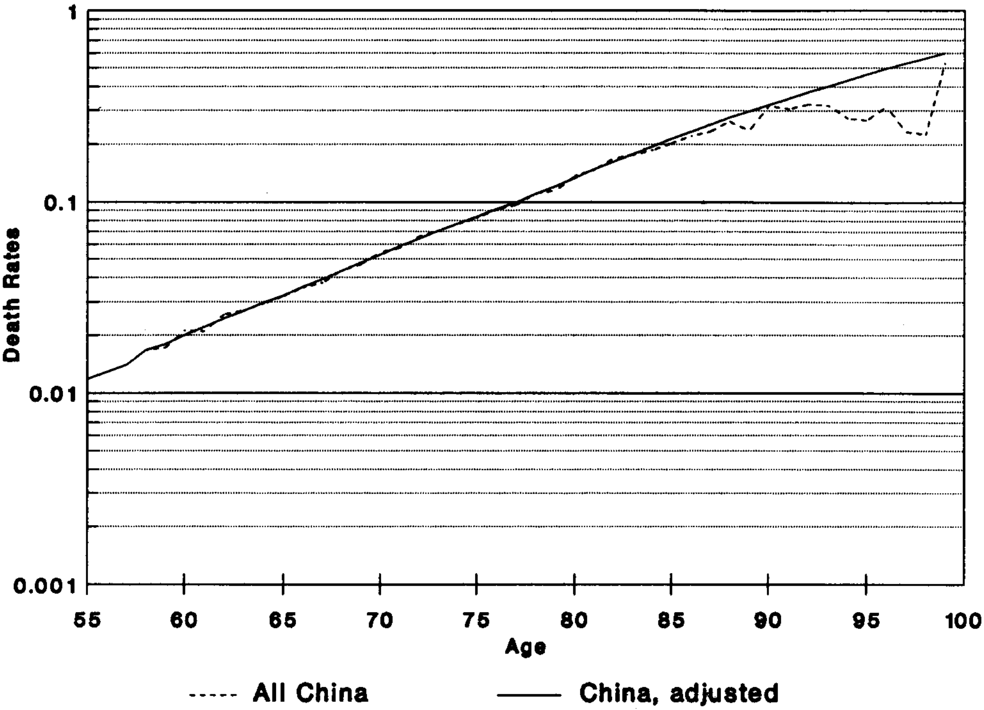

Figure 2. Death Rates, Chinese Males, 1981

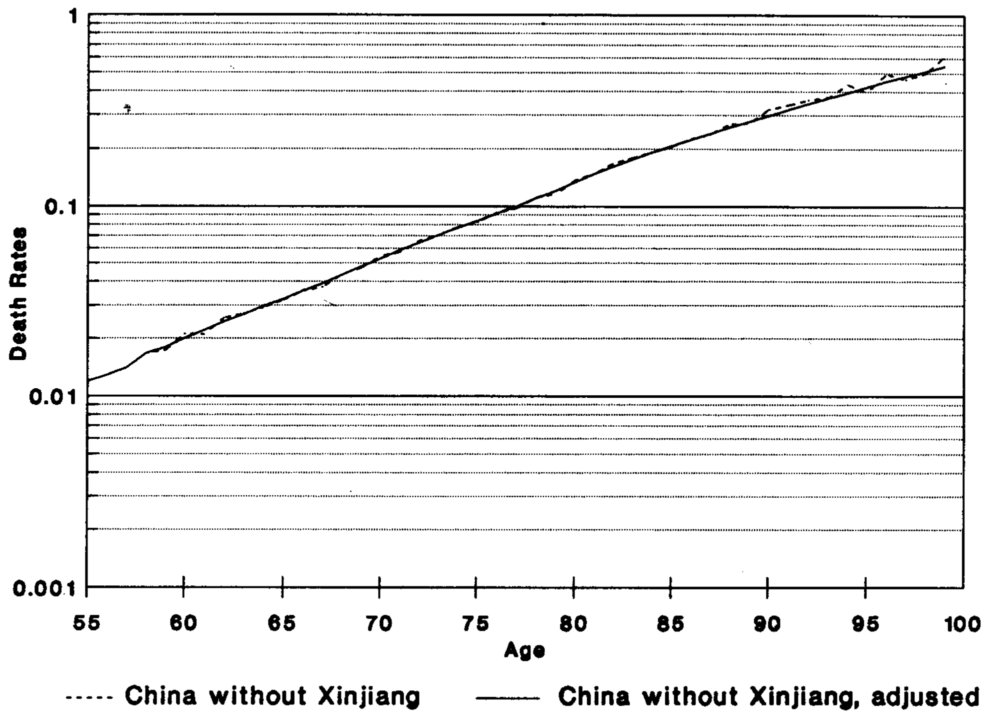

Figure 3. Death Rates, Chinese Males, 1981

used, but they escape such distortion if the data from one province (Xinjiang) are omitted. The death rates calculated directly for Xinjiang are clearly not credible, but the severe age heaping in the listing of deaths and population in this province has only a minimal effect on the death rates for all China until very high ages because Xinjiang's population is small. Increasingly above age 85, age exaggeration makes the listed male population in Xinjiang a substantial fraction of the total, and the erroneous death rates contaminate the rates for all China.

The expectation of life at birth for Chinese males in 1981 as calculated from the 1982 census data was 66.43 years; for China without Xinjiang it was only trivially different at 66.40 years. The numbers surviving to ages 90, 95, and 100 of 100,000 newborn males according to the life table for all China were 3,421, 731, and 151; without Xinjiang the numbers would be 3,191, 509, and 43. Remaining expectation of life falls from 4.05 years to 3.84 at age 85 and from 3.09 years to 2.15 at 95 when Xinjiang is omitted. The evident ignorance of date of birth among the non-Han minorities in Xinjiang makes essentially useless the male death rates in that province derived directly from the 1982 census, and the distortion cannot be remedied by smoothing and routine procedures for closing out at higher ages. Significant distortion caused in the all-China death rates is limited to ages above 85; it can be remedied either by omitting the province or by using an appropriate closing procedure.

The major lesson to be learned from this research is that data on ages listed in censuses, surveys, or registers of a population must be scrutinized critically, even when there are reasons to suppose that the data are accurate. Accuracy of most of the data does not mean that all of the data are accurate; as William Brass said, all data are guilty until proved innocent. Our detective work has shown that although recorded ages of the majority Han population in China are innocent of extensive error, the recorded ages of some minority populations are guilty of severe misreporting, and consequently the calculated death rates of one province are too far from reality to be useful as an indication of mortality in the province. The recorded number of males in this province at high ages is so highly inflated that its distorted old age death rates contaminate the national rates.

Notes

[1] Another distortion in the age distribution in the 1982 census is an apparent reduction in the tabulated number at ages above 100, relative to the number likely to have been reported by the respondents. The proportion reported as surviving to age 100 in 1964 from 11 years younger in 1953 is more than twice the proportion reported as surviving to age 99; the proportion reported as surviving to 100 in 1982 from 18 years younger in 1964 is less than one-third the proportion surviving to age 99. The numbers listed at age 100 are respectively 1.72 and 1.4 times the average at ages 98 to 102 in 1953 and 1964, and only .55 times the average in 1982. The number reported in 1982 at age 100 is below a five-term average in every province but one. The age distribution within the span over age 100 is similar in the three censuses (except for a reduction in age heaping at 100, 105, and 110, and in the proportion over 110 in 1982), but the fraction of the population above 95 at each age from 100 to 110 is much lower in 1982. Even in the male population of Xianjiang Province, where overstatement of age is so strong, the listed numbers at ages 97, 98, and 99 are 786, 941, and 689, followed by 84, 55, and 56 at ages 100 to 102. One can only speculate that in 1982 responses which would lead to an age reported as over 100 might have been altered so as to reduce the numbers listed at ages above 100.

One possibility is misuse of the process by which age was finally assigned. The respondent supplied a date of birth, almost certainly a date in the lunar calendar for persons over age 100. This date then was translated into the Western calendar by the enumerator, or by someone within the State Statistical Bureau, and was subtracted from 1982.5 to determine the age to enter in the census. One can imagine that when the person assigning the age encountered a Western date before 1882.5, he may have thought it a mistake and converted it to a twentieth-century date before 1982.5. After all, dates before 1882.5 were very rare: only 3.8 per million of the population was listed as above 100 years of age. This oddity implies that a death rate calculated for the age interval above 100 is not readily made commensurate with death rates calculated for ages below 100.

Acknowledgments

Single-year age distributions for China and for each province from the 1982 census, and complete life tables for China in 1981, are listed in the official *Yearbook of China's Population*, 1985 (*Zhongguo Renkou Nianjian*, 1985).

Single-year mortality rates and a complete life table for each province in 1981 were generously supplied by Professor Jiang Zheng-hua of the Population Research Unit of Xian Jiatong University.

The age distribution of selected minorities in the household population of the 1/1000 fertility survey of 1982 was tabulated at the Population Institute at the East-West Center in Honolulu as part of collaborative research between the Institute and the State Family Planning Commission.

References

Coale, A.J. and E.E. Kisker. 1986. "Mortality Crossovers: Reality or Bad Data?" *Population Studies* 40:389-401.

———. 1990. "Defects in Data on Old-Age Mortality in the United States." *Asia and Pacific Population Forum* 4:1-31.

Rosenwaike, I. and S.H. Preston. 1984. "Age Overstatement and Puerto Rican Mortality." *Human Biology* 56:503-25.

Thatcher, A. R. 1981. "Centenarians." *Population Trends* 25:11-14.