

In Y. Zeng, D.L. Poston, D.A. Vlosky, and D. Gu. (eds.). *Healthy Longevity in China: Demographic, Socioeconomic, and Psychological Dimensions* (Pp61-78). Dordrecht, The Netherlands: Springer.

Chapter 4

Reliability of Age Reporting among the Chinese Oldest-old in the CLHLS Datasets¹

ZENG YI

*Center for Study of Aging and Human Development, Medical School of Duke University
Center for Healthy Aging and Family Studies/ China Center for Economic Research, Peking University*

DANAN GU

Center for Study of Aging and Human Development, Medical School of Duke University

Abstract

This chapter evaluates age reporting among the oldest-old, especially centenarians, in the Chinese Longitudinal Healthy Longevity Survey (CLHLS) based on comparisons of various indices of elderly age reporting and age distributions of centenarians in Sweden, Japan, England and Wales, Australia, Canada, China, the U.S.A., and Chile. The analyses demonstrate that age reporting among the oldest-old interviewees (Han and six minority groups combined) in the 22 provinces in China where the CLHLS has been conducted is not as good as that in Sweden, Japan, and England and Wales, but is relatively close to that in Australia, more or less the same as that in Canada, better than that in the U.S.A., and much better than that in Chile. As indicated by the higher density of centenarians, age exaggeration exists in the six ethnic minority groups in the 22 Han-dominated provinces, although we cannot rule out and quantify the potential impacts of past mortality selection and better natural environmental conditions among these minority groups. We find that the age exaggeration of minorities in the CLHLS may not cause substantial biases in demographic and statistical analyses using the CLHLS data, since minorities consist of a rather small portion of the sample (6.8 percent at baseline and 5.5 percent in the grand total sample of the 1998, 2000, and 2002 waves).

Keywords: Age exaggeration, age heaping, age misreporting, age reporting, animal year, Australia, Canada, Chile, China, census data, density of centenarian, distributions of centenarians, England and Wales, ethnic minority, Han Chinese, Han-dominated provinces, Japan, Kannisto-Thatcher Database, late childbearing, lunar calendar, Myer's Index, pre-designed, ratio of centenarian, ratio index, Sweden, the oldest-old, the U.S.A., Western calendar, Whipple's Index

¹ The research reported in this Chapter is supported by The National Institute on Aging grant (R01 AG023627-01) and National Natural Science Foundation of China key project grant (70533010). We thank very much the helpful comments provided by Dudley Poston, Heather Booth and Zhongwei Zhao.

1. INTRODUCTION

This chapter examines the quality of age reporting among the oldest-old in the Chinese Longitudinal Healthy Longevity Survey (CLHLS). Assessments of data quality pertaining to disability, cognitive reliability and validity, proxy use, non-response rate and data incompleteness, sample attrition, and logical consistency in the CLHLS may be found in Chapter 3 of this volume and in Zeng et al. (2001).

Age reporting is a crucial issue in the study of healthy longevity. Age exaggeration will cause an underestimation of mortality rates at higher ages (Coale and Li 1991). The literature has established that accurate age reporting is generally a feature of developed societies with a few exceptions (Coale and Kisker 1986, 1990; Ewbank 1981; Seltzer 1973; Thatcher 1981). On the other hand, although age reporting among elderly persons in most developing countries is poor mainly due to age exaggeration (Dechter and Preston 1991; Retherford and Mirza 1982; Rosenwaike and Preston 1984), previous studies have shown that the quality of age reporting appears to be relatively good in several developing countries where date of birth has longstanding astrological significance such as among the Han majority in China (e.g., Coale and Li 1991; Wang et al. 1998), Korea (Jowett and Li 1982) and some other countries or regions (Knodel and Chayovan 1991).

Specifically, Coale and Li (1999) have shown that Han Chinese (and some minorities who over the years have been culturally and residentially integrated with the Han) tend to use the Chinese lunar calendar (for older generations) and the Western calendar (for younger generations) plus animal year to remember their birthdays². This is important for Chinese people because the precise date of birth is significant in making decisions on important life events such as matchmaking for marriage, date of marriage, and the date to start building a house, among other events. Therefore, Han Chinese, even if

² Although the Chinese animal year cycle is 12, there are no preferences for reporting ages, which are a multiple of 12. Some people may prefer their children to be born in a particular animal year (preference varies with region and time period), which may cause the birth rate in such a year to be somewhat higher than that in other years. But once born, Chinese people remember their actual animal year precisely, according to Chinese cultural tradition. Combining the animal year and a Chinese calendar such as the *Gan Zhi*, which reports the year or the years since the establishment of the ruling period of an emperor (e.g. *Guang Xu* year) or the Republic of China (*Ming Guo* year), for old people, and the Western calendar for younger people generally helps Han Chinese people to accurately report their birth date.

illiterate, can usually provide a reliable date of birth for themselves or for their close family members (Coale and Li 1991).

In the CLHLS data collection efforts we employed user-friendly forms for converting the reported birth dates of the Chinese lunar calendar into the Western calendar. The CLHLS asks for date of birth (rather than age directly) and computes the respondent's age after the survey by subtracting it from the date of the survey, because the Chinese system of calculating nominal age may make the response ambiguous.³ Other information relevant to the date of birth such as genealogical record, ID card, and household registration booklet were also collected in the CLHLS to validate the sampled elder's age. The interviewers and supervisors also check the parents' age, sibling's age, and the children/grandchildren's age of the sampled person, and the age of the sampled person at marriage and at birth, and so forth, to further validate age reporting. An additional question was designed for each interviewer to provide his/her judgment on the validity of the sampled person's age in the interviewer section in the CLHLS questionnaire. If the sampled person reported her/his age to be over 105, the interviewer was instructed to obtain additional evidence or concurrence from the local residential committee and local aging committee. If any inaccuracy in the reported age or any other logical problem in the questionnaire was found, a re-interview or phone call regarding specific questions was conducted.⁴

We examined the quality of age reporting in the CLHLS (1998) through comparisons with Sweden (1970-79), Japan (1970-79), England and Wales (1970-79), Australia (1970-79), Canada

³ According to the Chinese nominal age system, a person is counted as one year old on the day of birth, and one year older with each Chinese new year's day so that the nominal age is exaggerated by one or two years as compared with the actual age. If one simply asks for age, someone may respond in nominal age and some others may provide actual age, which will result in false age records. During the logical checks on the completed questionnaires and during data processing, we discovered that some interviewers did not follow the instructions but simply put the nominal age in the blank cell of the questionnaire where they are supposed to convert the reported birth date into age. The interviewers obtained the nominal age either from the list of centenarians provided by the local aging committee, which sometimes consider the nominal age 100+ as qualified ages for issuing centenarians' subsidies, or from conversations with the interviewees, although there is no question for directly asking age in the questionnaire. In these cases, the survey team corrected the nominal ages by the correct ages converted from the reported birth dates.

⁴ For example, in 2002 the age reporting of 46 persons was inconsistent with what was recorded in the 2000 data set. The survey team re-visited them and corrected the errors case by case.

(1960-69), the U.S.A. (1960-69), and Chile (1980-89), all with similar life expectancies at age 65⁵ as compared to the 22 surveyed provinces in China in 1998⁶. The inter-country comparisons are based on the indices of age heaping, age-specific percentage distributions of centenarians, age ratios of centenarians proposed by Booth and Zhao (Chapter 5 in this Volume), and the density of centenarians. It is commonly believed among international demographic experts in the aging fields that the data quality of age reporting among oldest-old people is the best in Sweden and Japan, very good in England and Wales, and in Australia, acceptably good in Canada, not so good but acceptable for academic research in the U.S.A. and Chile. For example, the Kannisto-Thatcher Database on Old Age Mortality consists of 31 countries, including Sweden, Japan, England and Wales, Australia, Canada, the U.S.A., and Chile, as well as the Singapore Chinese. The inclusion of these 31 countries was based on careful data quality evaluations by Kannisto and Thatcher: they determined that these data were sufficiently reliable and detailed for their demographic and comparative analysis on oldest-old mortality using uniform methods and measurements (Kannisto 1994). The Kannisto-Thatcher database classified these 31 countries into four relative categories, as follows: good quality, acceptable quality, conditional acceptable quality, and weak quality. Sweden, Japan, and England and Wales were in the “good quality” category. Australia was in the “acceptable quality” category. The data quality of Canada and U.S. Whites was also considered to be generally acceptable; however, the oldest-old data of United States non-whites (and the U.S. as a whole) and Chile were the least reliable, although they passed Kannisto-Thatcher’s careful data evaluations and were therefore entered into the 31-country

⁵ Unlike Booth and Zhao (2007, Chapter 5 in this volume) who used Swedish data from 1943-52, data from England & Wales in 1950-1955, and Japanese data in 1962-66 to compare to the 1998 CLHLS data in 1998 according to the closeness of Chinese life expectancy at birth around 1998, we selected the periods for inter-country comparisons based on the closeness to the Chinese life expectancy at age 65 ($e_{65} = 17$) around 1998. We used this approach because period life expectancy at age 65 is more directly and closely related to age distributions at the oldest-old ages, but period life expectancy at birth largely depends on the mortality of infants, children, and young adults, which is not directly and closely relevant to the age distribution at the oldest-old ages. Therefore, an inter-country comparative analysis aimed at examining age reporting through constricting the age distributions among the oldest-old would be more robust if we use the closeness of life expectancy at age 65 instead of the life expectancy at birth as a criterion for selecting the periods of the data sources.

⁶ The data for Sweden, E&W, Australia, Canada, and U.S.A. are from Human Mortality Database (<http://www.mortality.org>). The data for Chile are from the Kannisto and Thatcher Database on Old Age Mortality (<http://www.demogr.mpg.de>).

database of old age mortality (Kannisto 1994). Note that Chile was the only developing country (in addition to the data on the Singapore Chinese) which was entered into the Kannisto-Thatcher Database on Old Age Mortality.

2. AGE HEAPING

Given the specially designed CLHLS target-sampling procedure aimed at interviewing approximately equal numbers of males and females at each single age category from age 65 to age 99 (see Section 3 in Chapter 2 of this volume for more details), we could not use the age distribution of all interviewees in the CLHLS to examine age heaping. Therefore, data from the 2000, 1990, and 1982 censuses for the 22 CLHLS provinces were used to investigate age heaping as a general context for an analysis of CLHLS age reporting. Table 1 presents the percent distribution of sampled Han (93.2 percent at baseline and 94.5 percent in the three waves combined) and the six ethnic minority groups (6.8 percent at baseline and 5.5 percent in the three waves combined) in the CLHLS, as well as the Whipple's Index and Myer's Index, both of which are conventional measurements of age heaping based on census data. The census data indicate that there is no age heaping among the Han and the six minority groups in the 22 CLHLS provinces.

---Table 1 is about here---

Since both the Whipple's Index and the Myer's Index do not focus on the very old ages, one may reasonably question its validity for identifying age heaping problems among the oldest-old (see Chapter 5 in this volume). Coale and Li (1991:395) proposed an index to measure age heaping (digit preference) by deviation of the ratio of the number at each age to a two-stage moving average (the five-term average of a five-term average) from a perfect standard without any age heaping. Employing Coale and Li's method, we computed the average ratio at ages 85-105 for the 22 provinces in China in the 2000 census, Sweden, Japan, England and Wales, Australia, Canada, the United States, and Chile for selected periods in which each country had a female life expectancy at age 65 which was close to

that in China in 2000. We then calculated single-age-sex-specific and average odds for the ratio of each country compared to that of Sweden in the 1970s, considering the Swedish age distribution as perfect, that is, without any age heaping. The results shown in Table 2 indicate that the quality of age reporting at the oldest-old ages in terms of age heaping measured by the ratio index proposed by Coale and Li for the 22 provinces of China is not as good as that in Sweden, Japan, England and Wales, and Australia, but it is similar to that in Canada, it is better than that in the U.S.A., and it is much better than that in Chile. So, we conclude that there is little age heaping among the oldest-old in the 22 provinces in China where the CLHLS was conducted.

----- Table 2 is about here-----

3. AGE DISTRIBUTION AND AGE RATIOS AMONG CENTENARIANS

Age exaggeration may still exist even if there is no age-heaping because it is possible that people at very high ages may systematically tend to over-report their ages. It is, therefore, worthwhile to explore the quality of age reporting among the centenarians in the CLHLS in comparison with Sweden, Japan, England and Wales, Australia, Canada, U.S.A., and Chile, using the single-year-age-specific percentage distributions of centenarians. We tried to interview all centenarians in the sampled cities and counties of the 22 CLHLS provinces. The age distribution of the interviewed centenarians in our survey should be therefore compatible with the national age distribution in the 22 provinces in China if there are no substantial age exaggerations in the CLHLS survey. Note that we purposely tried to have approximately equal numbers of male and female octogenarians and nonagenarians at each age from 80 to 99, who resided nearby the centenarians, and their age and sex were pre-designed based on the centenarians' code numbers which are randomly assigned. Thus, it makes no sense to compare the age distributions of octogenarians and nonagenarians interviewed in our survey to that of other countries. Rather, it is sufficient to compare the age distribution of the Chinese centenarians to that of centenarians in other countries for assessing the quality of age reporting of the oldest-old in our survey.

There is no reason to suspect substantive age exaggerations among elders below the age of 100, if the age reporting of centenarians is acceptably good. This is because age exaggeration is much more likely among centenarians than among elders who are younger than age 100 and who live in the same area and share the same cultural traditions.

Single-year age-specific death rates are around 0.5 at ages 100 and over. Such extremely high mortality rates have dominated the shape of the age distribution of centenarians, which means that the effects of differentials in cohort size are minor. Therefore, the age distributions of centenarians of the European and Japanese populations, which have the highest data quality, look very much alike. Figure 1 is a comparison of the percentage age distributions of centenarians among Sweden, Japan, the U.S.A., Chile and the CLHLS (including the Han and minorities in the 22 provinces). Table 3 presents detailed numerical results for the five countries depicted in Figure 1 plus England and Wales, Australia, and Canada. The comparisons shown in Figure 1 clearly demonstrate that the age reporting among the Chinese centenarians in the 22 provinces is relatively close to, but not quite as good as, that in Sweden and Japan, but better than that in the U.S.A. (especially for males), and much better than that in Chile. Additional data in Table 3 show that the age reporting of the male centenarians in the CLHLS is relatively close to, but not as good as, that in England and Wales, more or less the same as that in Australia, and slightly better than that in Canada; the age reporting of the female centenarians in the CLHLS is not as good as that in England and Wales, and somewhat worse than that in Australia and Canada.

--- Figure 1 and Table 3 are about here ---

Similar to the percentage age distribution of centenarians but with a stronger assumption of ignoring the cohort size effect, the age ratio of centenarians is defined as the number of persons at each age over the age of 100 divided by the number of persons at age 100 (see Chapter 5 in this volume). The results listed in Table 4 show that, if we consider Sweden as the best standard, the age reporting among centenarians measured by the age ratios of male centenarians in the CLHLS is not as good as

that in Japan and England and Wales, more or less the same as that in Australia, slightly better than that in Canada, better than that in the U.S.A., and much better than that in Chile; the age ratios of female centenarians in the CLHLS are not as good as those in Japan, England and Wales, and Australia, somewhat worse than that in Canada, better than that in the U.S.A., and much better than that in Chile.

--Table 4 is about here---

4. DENSITY OF THE OLDEST-OLD AND CENTENARIANS

Another way of addressing the issue of age reporting among the oldest-old is to imagine that if the age exaggeration at very old ages is significant in a population, the reported proportion of the oldest-old persons among all elders and the total population would be relatively large, compared with other populations with accurate age reporting. As shown by Coale and Kisker (1986), the proportions of those aged 95 or over among all elders aged 70 or over in the 23 countries studied by them with good data quality were all less than six per thousand. This proportion in the 28 countries with poor data because of age exaggeration by old persons extends from one percent to 10 percent (Coale and Kisker 1986: 398). The proportions of male and female Han Chinese aged 95 or over among those aged 70 or over in 1990 in all of China are 0.76 per thousand and 2.18 per thousand, proportions that are rather close to those of their Swedish counterparts (Wang et al. 1998).

Coale and Kisker (1986: 389-390) plotted values of e_{70} (life expectation at age 70) against values of l_{70}/l_5 (the conditional survival probability from age 5 to age 70), for the female populations in countries or regions with good data⁷. A very close relation between the e_{70} and l_{70}/l_5 values among countries or regions with good data is evidenced by a third-degree polynomial curve fitted by least squares (Figure 1 in: Coale and Kisker 1986: 389). The plotting of the e_{70} against l_{70}/l_5 values for the

⁷ Countries or regions with good data include Sweden, Austria, Belgium, Czechoslovakia, Denmark, England, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Luxembourg, Netherlands, Norway, New Zealand, Scotland, Switzerland and Taiwan, as indicated by Coale and Kisker (1986).

female populations in countries with poor data⁸ all lie far above the polynomial curve fitted to the data from the countries with good data (Figure 2 in: Coale and Kisker 1986: 390). Wang et al. (1998) computed the ratio of the e_{70} against l_{70}/l_5 values for the Han Chinese female population in 1990. The Han Chinese ratio is almost exactly on the third-degree polynomial curve fitted to the data of populations with good age reporting.

Table 5 presents census data on the number of centenarians per million population, per million elderly aged 65 and above, and per million oldest-old aged 90 and above in the 22 provinces and other selected countries/periods with more or less the same female life expectancy at age 65 as in the 22 CLHLS provinces in China. There were 13.8, 191, 17,056 centenarians per one million population, per one million elderly aged 65+ and per one million oldest-old aged 90+ in the 22 provinces (including both Han and minorities) in China in 2000, as compared to 22.9, 155, 9,618 in Sweden in 1970-79, and 30.6, 218, and 12,360 in England and Wales in the 1970s, respectively. The census data in Table 5 demonstrate that the density of centenarians, which can be considered as one kind of quality of age reporting indicator, in the 22 provinces in China is not as good as that in Sweden and Japan, but is relatively close to England and Wales and Australia, more or less the same as that in Canada, better than that in the U.S.A, and much better than that in Chile. However, the density of centenarians among the six minorities combined in the 22 provinces is substantially higher (worse) than that in Sweden, Japan, England and Wales, Australia and Canada, and moderately higher (worse) than that in the U.S.A., but lower (better) than that in Chile.

---Table 5 is about here---

The analysis presented above and in other studies including those by Coale and Li (1991), Wang et al. (1998) and Zeng et al. (2001) show that age reporting among the Han Chinese elderly, which is the population that constitutes the majority (about 92 percent) in China, is acceptably good.

⁸ The countries with poor data include Bolivia, Costa Rica, El Salvador, Guatemala, Honduras, Malaysia, Mexico, Panama, Peru, Philippines, Sri Lanka and Thailand, as indicated by Coale and Kisker (1986).

However, Coale and Li (1991) found that age reporting among some ethnic minorities, especially the Uyghur ethnicity in the Xinjiang autonomous region, was seriously biased with age exaggeration, leading to abnormal age patterns of mortality at old ages for China as a whole; once the Xingjiang data were excluded, the Chinese age pattern of mortality at old ages became normal as compared to other countries with accurate age reporting. We know of only one published study on age misreporting of China's ethnic minorities (Poston and Luo, 2004). Using Whipple's and Myers methods of evaluating the presence of age misreporting, it showed little evidence of age heaping among most of China's minorities in 2000. The Uyghur nationality was the major exception, showing a preference for digits 0 and 5, and an avoidance of digits 1, 3 and 9. Among most of the other minority groups, and among the Han, there was little if any evidence of age misreporting.

The summary indices in Table 5 show that the number of centenarians per million persons of all ages, per million elderly aged 65+, and per million oldest-old aged 90+ among the six minority groups in the 22 provinces, is much higher than that in Han Chinese and other populations with accurate age reporting (Sweden, Japan, England and Wales, Australia, and Canada), higher than that in the U.S., and somewhat lower than that in Chile. This fact leads us to seriously suspect that the oldest-old, especially centenarians, of the six minority groups in the CLHLS might well be exaggerating their ages.

On the other hand, we also suspect that the higher density of the centenarians among the six minority groups in the 22 CLHLS provinces may be partially due to differential mortality and natural environmental selection among these heterogeneous populations. Very old people of many of the ethnic minorities (except for Koreans⁹) in the 22 provinces have suffered poor living conditions and have had inadequate medical care in the past decades, and were more likely to be living in remote mountainous areas where the natural environment was more likely to have been well protected. Their

⁹ Koreans in China have an even higher socioeconomic status than Han Chinese and their age reporting could be as good as that of the Han, but again, we have no detailed data with which this can be verified.

misery in the past has been much worse than that of their counterpart cohorts among the Han Chinese, and among their cohorts in Sweden, Japan, England and Wales, Australia, Canada, and the U.S.A., which could result in a mortality selection where minority persons who survived to very old ages were more robust in genetic and other biological characteristics, while those who were frail died before reaching old age. It is well known that much higher mortality rates at young ages in disadvantaged populations can produce relatively more robust old people as compared to those from advantaged populations (e.g. Coale and Kisker 1986; Horiuchi and Wilmoth 1998). Thus, the mortality rates among the minority oldest-old in the CLHLS might be lower because there are more robust survivors among them. The possible past mortality selection plus better natural environmental conditions in the minority areas in the 22 provinces might partially contribute to the higher density of centenarians among the six minority groups. It is, however, also highly possible that some minority oldest-old in the CLHLS survey areas indeed exaggerated their ages, which is our educated best guess. But, unfortunately, we do not have adequate data to quantitatively decompose the impacts of possible factors of age exaggerations, past mortality selection, and better natural environmental conditions on the higher density of centenarians among the six minority groups in the CLHLS areas.

5. IMPACTS OF POSSIBLE AGE EXAGGERATIONS AMONG THE SIX MINORITY GROUPS ON THE ANALYSES USING THE CLHLS DATA

Given the likelihood of age exaggeration among the six minority groups in the CLHLS data set and our inability to quantify the degree thereof, CLHLS data users may reasonably ask the following question: what will be the effects of the minority populations in the CLHLS dataset on statistical analysis, assuming that the impacts of past mortality selection and better natural environmental conditions are all negligible and that the higher density of centenarians among minorities is solely or largely caused

by age exaggeration¹⁰? To address this important question, we conducted several additional analyses on the “association of late childbearing with health and healthy survival at the oldest-old ages,” following exactly the same methodology, statistical modeling procedures and using the same dependent and independent variables as those used in Zeng and Vaupel (2004); however, in these new analyses we relied only on the CLHLS Han data, excluding the data for the ethnic minorities. We then compared the results of these additional analyses with the results of the original analyses in Zeng and Vaupel (2004), where they used the entire CLHLS data set including both the Han and minority data.

We listed the additional and the original estimates in parallel positions with additional estimates in parentheses in Tables 6, 7, 8, and 9. It is clear that the two sets of parameter estimates including and excluding the six minority groups in the 22 Han-dominated provinces are fairly close to each other, indicating that the same qualitative conclusions concerning the association of late childbearing with health and healthy survival at oldest-old ages in China can be drawn based on the two sets of estimates. We also note that 66.7 percent, 28.7 percent, and 4.6 percent of the total number (216) of the estimates of the odds ratios and relative risks in the additional analysis excluding minorities indicate a slightly weaker, the same, and slightly stronger positive association of late childbearing with health and healthy survival at the oldest-old ages, respectively, compared to the original estimates including minorities. These results suggest that the minority groups in the 22 CLHLS provinces may indeed exaggerate their ages, compared to the Han, although we are not able to quantify and decompose the possible effects of age exaggeration, past higher mortality selection and better natural environmental conditions among the minority groups.

The total number of estimates of the odds ratios and relative risks which are statistically significant in the additional analyses excluding the minority data is reduced by 13.5 percent (perhaps at least partially due to the 5.5-6.8 percent reduction of the sample size), but the direction and conclusion

¹⁰ We appreciate very much that Booth and Zhao (Chapter 5 in this volume) thoughtfully raised this important question, which helped us to conduct careful comparative assessments to address the issue.

of the association of late childbearing with health and healthy survival at the oldest-old ages remains unchanged, compared to the analyses using both Han and minority data. This fact leads us to believe that the inclusion of the six minority groups in the CLHLS may not cause substantive bias in demographic and multivariate statistical analyses. This is mainly because the minority groups consist of a rather small portion of the samples: 6.8 percent of the total sample in the baseline survey and 5.5 percent of the grand total sample of the 1998, 2000, and 2002 waves combined¹¹.

Gu and Dupre (Chapter 6 in this volume) also show that possible age exaggeration among minority oldest-old in the CLHLS does not significantly affect the outcome for estimates of indicators such as ADL, MMSE, self-reported health, as well as the estimates of other covariates in multivariate statistical models using either cross-sectional or longitudinal CLHLS data.

6. CONCLUDING REMARKS

This chapter has evaluated patterns of age reporting among the oldest-old, especially centenarians, in the Chinese Longitudinal Healthy Longevity Survey (CLHLS) based on comparisons of the various indices of elderly age reporting and age distributions of centenarians among the countries of Sweden, Japan, England and Wales, Australia, Canada, China, the U.S.A., and Chile. The analyses demonstrate that age reporting among the oldest-old interviewees (Han and the six minority groups combined) in the 22 CLHLS provinces is not as good as that in Sweden, Japan, and England and Wales, but is relatively close to that in Australia, more or less the same as that in Canada, better than that in the U.S.A, and much better than that in Chile. As indicated by the high density of centenarians, however, age exaggeration does exist in the six ethnic minority groups, although we cannot rule out and quantify the potential impacts of past mortality selection plus better natural environmental conditions among

¹¹ In each follow-up wave of the CLHLS, the interviewees who died or were lost to follow-up were replaced by new interviewees with the same gender and age as those died or were lost to follow-up. We matched the gender and age only; there was no ethnicity requirement in the replacement process. The Han are the large majority in survey areas. Thus, it is usually much easier to find a replacement who is Han with the required gender and age, which is why new interviewees are more likely to be Han. This explains why the percent of minority groups in the CLHLS in later waves was somewhat lower than at baseline.

these minority groups. We suggest that age exaggeration of minorities in the CLHLS does not likely cause substantial biases in demographic and statistical analyses using the CLHLS data, since, as noted above, minorities comprise a rather small portion of the sample¹². Of course, we must keep in mind that the minority groups in the CLHLS study areas might exaggerate their ages and some Han Chinese oldest-old might also exaggerate (to a lesser extent) their ages. We thus need to exercise caution in our analyses and interpretations of findings.

REFERENCES

- Booth, H., and Z. Zhao. (Chapter 5 in this volume). Age reporting in the CLHLS: a re-assessment.
- Coale, A.J., and E. Kisker (1986), Mortality crossovers: reality or bad data? *Population Studies* 40, pp.389-401
- Coale, A.J., and S. Li (1991), The effect of age misreporting in China on the calculation of mortality rates at very high ages. *Demography* 28(2), pp. 293-301
- Dechter, A., and S.H. Preston (1991), Age misreporting and its effects on adult mortality estimates in Latin America. *Population Bulletin of the United Nations* 31/32, pp.1-16
- Ewbank, D.C. (1981), *Age misreporting and age-selective underenumeration: sources, patterns, and consequences for demographic analysis*. Washington, DC: National Academy Press
- Gu, D, and M.E. Dupre. (Chapter 6 in this volume). Assessment of reliability of mortality and morbidity in the 1998-2002 CLHLS waves

¹² It may be worthwhile to note that it is necessary to include the six minority groups in the CLHLS surveys to ensure adequate representation of minority groups in the study areas. U.S. studies have a similar requirement.

- Horiuchi, S., and J.R. Wilmoth (1998), Deceleration in the age pattern of mortality at old ages. *Demography* 35(4), 391-412
- Jowett, A.J., and Y.Q. Li (1982), Age heaping: Contrasting pattern from China. *GeoJournal* 28(4), pp. 427-442
- Kannisto, V (1994), *Development of oldest-old mortality, 1950-1990: Evidence from 28 developed countries*. Odense: Odense University Press
- Knodel, J., and N. Chayovan (1991), Age and birth date reporting in Thailand. *Asian and Pacific Population Forum* 5(2-3), pp.41-76
- Mantel, N. and W. Haenszel (1959), Statistical aspects of the analysis of data from retrospective studies of diseases. *Journal of the National Cancer Institute* 22, pp. 719–748
- Poston, D.L., Jr., and H. Luo (2004), Zhongguo 2000 nian shaoshu minzu de nian ling dui ji he shu zi pian hao (Age structure and composition of the Chinese minorities in 2000). *Zhongguo Shaoshu Minzu Renkou (Chinese Minority Populations)* 19 (3), pp. 9-15
- Retherford, R.D., and G.M. Mirza (1982), Evidence of age exaggeration in demographic estimates for Pakistan *Population Studies* 36(2), pp. 257-270
- Rosenwaiké, I., and S.H. Preston (1984), Age overstatement and Puerto Rican longevity. *Human Biology* 56(3), pp.503-525
- Seltzer, W. (1973), *Demographic data collection: a summary of experience*. New York: Population Council
- Thatcher, A.R. (1981), Centenarians. *Population Trends* 25, pp.11-14
- Wang, Z., Y. Zeng, B. Jeune, and J.W. Vaupel (1998), Age validation of Han Chinese centenarians. *GENUS - An International Journal of Demography* 54, 123-141

Zeng, Y., and J.W. Vaupel (2004), Association of late childbearing with healthy longevity among the oldest-old in China. *Population Studies* 58(1), pp. 37-53

Zeng, Y., J.W. Vaupel, Z. Xiao, C. Zhang, and Y. Liu (2001), The healthy longevity survey and the active life expectancy of the oldest old in China. *Population: An English selection* 13 (1), pp.95-116

Table 1. Ethnic composition and the age heaping indices

Ethnic group	Percent of the Sample in the 1998 CLHLS (%) (N=8,805)*	Percent of the Sample in the 1998-2002 pooled CLHLS (%) (N=19,890)*	Whipple's Index in census			Myer's Index in census		
			1982	1990	2000	1982	1990	2000
Han	92.75	94.03	101.5	100.5	101.1	1.48	2.85	2.04
Zhuang	4.41	3.48	100.1	102.1	104.3	2.79	2.25	2.88
Hui	1.31	1.01	101.4	102.4	105.7	1.81	2.71	2.69
Yao	0.57	0.42	101.1	101.1	102.8	3.58	2.28	2.50
Korea	0.11	0.07	103.2	104.3	104.1	1.96	2.33	1.96
Manchu	0.33	0.52	100.1	105.3	102.9	2.57	3.13	1.50
Mongolia	0.03	0.03	99.7	104.0	102.8	2.56	2.45	2.31

Notes: (1) There are around 0.5% of respondents who are other ethnic minorities or whose ethnic identities are missing. They are excluded from the data in the second and third columns.

(2) Based on the United Nations' criteria, Whipple's Index: <105 very good, 105-110 good, 110-125 so-so, >125 poor. Myer's Index: <10 good, 10-20 so-so, >20 poor.

(3). 1982 and 1990 data are cited from Zeng et al. (2001).

Table 2. Mean of the ratios of the number at each age to a two-stage moving average, ages 85-105

	Females		Males		Both Sexes	
	MR	D%	MR	D%	MR	D%
China, 22 provinces	0.793	5.4	0.753	3.9	0.782	5.1
Sweden, 1970s	0.753	0.0	0.725	0.0	0.744	0.0
Japan, 1970s	0.728	-3.3	0.689	-5.0	0.717	-3.6
England & Wales, 1970s	0.780	3.6	0.711	-1.9	0.768	3.3
Australia, 1970s	0.777	3.2	0.756	4.3	0.771	3.7
Canada, 1960s	0.789	4.8	0.783	8.0	0.786	5.7
USA, 1960s	0.823	9.3	0.809	11.6	0.818	10.0
Chile, 1980-90	0.860	14.2	0.849	17.1	0.857	15.2

Note: The age-specific number of centenarians for 22 provinces in China in 2000 is estimated based on the age-specific percentage distribution in the CLHLS in 1998 and the total number of centenarians from the 2000 census (the 2000 census publication does not include age-specific numbers of centenarians).

Table 3. Comparison of percentage age distribution of centenarians, between CLHLS and selected countries

	CLHL S 1998	CLHLS(Han) 1998	Sweden 1970s	Japan 1970s	England & Wales 1970s	Australi a 1970s	Canada 1960s	USA 1960s	Chile 1980s
Female e_{65}	17.2	NA	17.4	16.6	16.4	16.9	16.7	16.7	17.6
Males									
100	41.9	42.7	50.8	46.4	50.3	41.0	37.9	33.8	23.3
101	25.4	25.2	25.8	23.9	26.0	24.7	24.1	21.7	15.9
102	14.0	13.9	11.9	13.8	12.9	15.5	15.2	14.3	11.4
103	6.7	6.6	6.2	8.2	6.4	9.5	9.1	9.8	9.4
104	4.8	4.6	3.3	3.8	2.7	5.0	5.6	6.8	9.4
105	3.1	3.1	1.5	1.6	1.0	2.2	3.6	4.7	10.2
106	2.5	2.4	0.4	1.2	0.4	1.1	2.1	3.3	7.1
107	0.4	0.4	0.2	0.7	0.1	0.7	1.3	2.4	5.4
108	0.4	0.2	0.0	0.2	0.1	0.3	0.8	1.8	4.3
109	0.6	0.7	0.0	0.2	0.0	0.1	0.4	1.4	3.6
Females									
100	37.2	37.7	46.0	47.3	44.5	43.4	42.8	35.4	30.4
101	24.3	24.3	26.5	25.9	25.7	25.7	25.6	22.8	20.6
102	14.8	14.8	14.1	13.3	14.4	14.8	14.6	14.7	13.7
103	9.5	9.5	7.4	6.8	7.9	8.0	7.9	9.5	9.8
104	5.3	5.0	3.2	3.5	4.1	4.0	4.2	6.3	7.9
105	3.0	3.0	1.6	1.7	2.0	2.1	2.5	4.1	6.8
106	2.2	2.1	0.7	0.8	0.9	1.0	1.4	2.8	4.4
107	1.9	1.9	0.3	0.4	0.4	0.5	0.6	2.0	3.0
108	1.0	1.0	0.1	0.2	0.2	0.3	0.3	1.4	2.1
109	0.8	0.8	0.1	0.0	0.1	0.2	0.1	1.1	1.4

Table 4. Comparison of ratio of number of centenarians at each age to age 100, the CLHLS and selected countries

	CLHLS	CLHLS(Han)	Sweden	Japan	England & Wales	Australia	Canada	USA	Chile
Age	1998	1998	1970-79	1970-79	1970-79	1970-79	1960-69	1960-69	1980-90
Males									
100	10000	10000	10000	10000	10000	10000	10000	10000	10000
101	6050	5907	5076	5155	5169	6022	6360	6430	6800
102	3350	3264	2348	2964	2566	3770	4004	4239	4900
103	1600	1554	1212	1774	1268	2307	2394	2888	4033
104	1150	1088	644	814	538	1212	1472	2018	4033
105	750	725	303	355	205	530	942	1393	4367
106	600	570	76	250	75	280	565	975	3067
107	100	104	38	146	25	172	331	725	2333
108	100	52	0	42	11	84	205	527	1833
109	150	155	0	42	5	20	97	405	1533
Mean of the absolute deviation as compared to Sweden									
101-104	717.4					1007.9	1237.6	1573.9	2621.6
	(4)	633.3 (3)	0.0	356.7 (2)	118.2 (1)	(5)	(6)	(7)	(8)
105-109	256.7								2543.3
	(5)	237.9 (4)	0.0	83.6 (2)	25.3 (1)	133.7 (3)	344.7 (6)	721.8 (7)	(8)
101-109	461.4							1100.5	2578.1
	(4)	413.6 (3)	0.0	205.0 (2)	66.6 (1)	522.2 (5)	741.5 (6)	(7)	(8)
Females									
100	10000	10000	10000	10000	10000	10000	10000	10000	10000
101	6541	6452	5762	5464	5774	5918	5978	6427	6795
102	3978	3922	3066	2801	3231	3413	3413	4140	4519
103	2549	2530	1605	1439	1779	1840	1835	2677	3216
104	1429	1332	690	745	916	919	979	1775	2618
105	798	808	353	362	442	484	586	1167	2233
106	602	554	161	179	205	230	330	790	1453
107	518	494	64	92	95	108	150	557	983
108	266	269	16	36	39	77	69	402	684
109	210	210	16	5	12	35	17	307	449
Mean of the absolute deviation as compared to Sweden									
101-104	843.1								1506.0
	(6)	778.2 (5)	0.0	196.1 (2)	143.8 (1)	241.7 (3)	270.4 (4)	973.7 (7)	(8)
105-109	357.0							522.5	1038.3
	(6)	345.1 (5)	0.0	17.1 (1)	38.2 (2)	65.0 (3)	108.3 (4)	(7)	(8)
101-109	573.0								1246.1
	(6)	537.6 (5)	0.0	96.6 (2)	85.1 (1)	143.5 (3)	180.3 (4)	723.0 (7)	(8)

Note: The numbers in the parentheses are the rank of the mean of absolute deviation of the ratio of each age as compared to that of Sweden. The smaller the rank number is, the better the quality the age reporting is supposed to be.

Table 5. Comparison of density of centenarians among selected ethnicities and countries

	Female	Centenarians per million population		
	e_{65}	among		
	Age 65	Age 0+	65+	90+
All Han Chinese in whole				
China	17.3	12.7	175	16,124
22 Provinces (Han & minorities)	17.2	13.5	184	16,441
Six minorities in the 22 provinces	18.4	33.7	524	32,672
Sweden, 1970-79	17.4	22.9	155	9,618
Japan, 1970-79	16.6	5.2	67	7,538
England & Wales, 1970-79	16.4	30.6	218	12,360
Australia, 1970-79	16.9	15.4	214	13,476
Canada, 1960-69	16.7	20.0	258	15,601
USA, 1960-1969	16.4	32.1	337	21,189
Chile, 1980-1990	17.6	34.9	662	33,464

Notes: The Chinese data are from the 2000 Census publications released by National Bureau of Statistics of China (2003a, 2003b). The Chinese Female life expectancies at age 65 are estimated from published data by NBSC (2003a, 2003b), adjusted for mortality rate underreporting at ages 94 and over.

Table 6. Effects (odds ratios) of late childbearing on health status of the oldest-old in China, based on multivariate logistic regression applied to data collected at 1998 baseline survey, adjusted for covariates of demographic characteristics, family support, social connections, and health practice

-- Estimates not in the parenthesis are based on the whole CLHLS data set including both Han and the six minority groups cited from Zeng and Vaupel (2004); Estimates in the parenthesis are based on the CLHLS data set excluding the six minority groups

Focused Covariates (category in parentheses is the reference group)	ADL disabled	MMSE impaired	Self-reported bad health	Depression symptoms
Women (0 birth after age 35)				
1 birth after age 35	0.890 (0.940)	0.949 (0.966)	0.923 (0.971)	0.854 (0.788*)
2 births after age 35	0.939 (0.984)	0.835# (0.836)	0.906 (0.960)	0.827 (0.763#)
3+ births after age 35	0.765* (0.857)	0.772* (0.768*)	1.031 (1.074)	0.626** (0.580**)
Men (0 birth after age 35)				
1 birth after age 35	0.954 (1.017)	1.111 (1.003)	1.002 (1.021)	0.846 (1.141)
2 births after age 35	0.879 (1.025)	0.937 (1.010)	0.858 (0.980)	0.757 (1.002)
3+ births after age 35	0.880 (0.977)	0.791 (0.873)	0.939 (0.981)	0.711 (0.805)

Notes:

(1) #, $p < 0.10$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

(2) Differences between results including and excluding minorities for births after age 40 are minor, too. They are not listed in Tables 6, 7, 8, and 9, due to space limit but available upon request.

Table 7. Effects of late childbearing (relative risks and odds ratios) on survival and healthy survival of the oldest-old in China between 1998 and 2000 based on multivariate Cox proportional hazards and ordinal logistic regression models

-- Estimates not in the parenthesis are based on the whole CLHLS data set including both Han and the six minority groups cited from Zeng and Vaupel (2004); Estimates in the parenthesis are based on CLHLS data set excluding the six minority groups

Models and Focused covariates	Relative risk of Cox hazards models survival analysis		Odds ratios of ordinal logistic regression healthy survival analysis	
	Women	Men	Women	Men
Model I (0 birth after age 35)				
1 birth after age 35	0.941 (0.937)	0.910 (0.903)	0.914 (0.874#)	1.103 (0.979)
2 births after age 35	0.850* (0.893#)	1.077 (1.097)	0.758*** (0.764***)	0.930 (1.008)
3+ births after age 35	0.722 *** (0.749***)	0.747 *** (0.742***)	0.697*** (0.678***)	0.769** (0.738**)
Model II (0 birth after age 35)				
1 birth after age 35	0.941 (0.941)	0.919 (0.922)	0.977 (0.926)	1.209# (1.100)
2 births after age 35	0.858* (0.912)	1.100 (1.138)	0.847# (0.838#)	1.045 (1.183)
3+ births after age 35	0.715 *** (0.754***)	0.800* (0.807*)	0.766* (0.753*)	0.895 (0.936)
Model III (0 birth after age 35)				
1 birth after age 35	0.949 (0.953)	0.919 (0.931)	0.967 (0.925)	1.157 (1.076)
2 births after age 35	0.861 * (0.915)	1.116 (1.163)	0.828* (0.833#)	1.066 (1.208)
3+ births after age 35	0.720 *** (0.763**)	0.821 # (0.841#)	0.747** (0.748**)	0.889 (0.965)
Model IV (0 birth after age 35)				
1 birth after age 35	0.977 (0.980)	0.947 (0.940)	1.003 (0.949)	1.169 (1.096)
2 births after age 35	0.902 (0.955)	1.181 (1.2136#)	0.854 (0.850)	1.111 (1.218)
3+ births after age 35	0.767 ** (0.797*)	0.889 (0.884)	0.791* (0.787*)	0.944 (1.014)

Notes: (1) The category in parentheses in the column of “Models and Focused covariates” is the reference group in each case. (2) Covariates in Model I are late childbearing plus demographic variables of age, gender, residence, education, and ethnicity. Model II is Model I plus covariates of family support and social connection. Model III is Model II plus covariates of health practices. Model IV is Model III plus covariates of health conditions. (3) The number of degree of freedom for Model I to Model IV is 9, 17, 21, 25 respectively. (4) #, p<0.10; *, p<0.05; **, p<0.01; ***, p<0.001.

Table 8. Ratio of survivorship (RS) of elders who had 1+, 2+, or 3+ births after age 35 to those who did not have such late births

-- Estimates not in the parenthesis are based on whole CLHLS data set including **both Han and** the six minority groups **cited from Zeng and Vaupel (2004)**; Estimates in the parenthesis are based on CLHLS data set excluding the six minority groups

$P_1(x)$	Ages 100-105 vs. 80-85				Ages 100-105 vs. 90-95				Ages 90-95 vs. 80-85			
	$P_1(80-85)$	$P_1(100-105)$	RS	P	$P_1(90-95)$	$P_1(100-105)$	RS	P	$P_1(80-85)$	$P_1(90-95)$	RS	P
Men												
1+ births	69.0(69.0)	71.0(70.1)	1.10(1.1)	0.449(0)	72.3(71.7)	71.0(70.1)	0.93(0.93)	0.596(0)	69.0(69.0)	72.3(71.7)	1.17(1.14)	0.106(0.203)
2+ births	50.1(49.5)	54.4(53.8)	1.19(1.2)	0.125(0)	49.9(49.3)	54.4(53.8)	1.20(1.20)	0.121(0)	50.1(49.5)	49.9(49.3)	0.99(0.99)	0.913(0.926)
3+ births	30.2(29.6)	35.6(34.3)	1.28(1.2)	0.036(0)	32.7(32.6)	35.6(34.3)	1.14(1.08)	0.281(0)	30.2(29.6)	32.7(32.6)	1.12(1.15)	0.233(0.164)
Women												
1+ births	49.8(48.8)	58.8(57.6)	1.44(1.4)	0.000(0)	51.3(50.3)	58.8(57.6)	1.35(1.34)	0.000(0)	49.8(48.8)	51.3(50.3)	1.06(1.06)	0.489(0.497)
2+ births	24.6(23.7)	34.2(32.7)	1.59(1.6)	0.000(0)	28.2(26.6)	34.2(32.7)	1.32(1.34)	0.001(0)	24.6(23.7)	28.2(26.6)	1.2(1.17)	0.059(0.134)
3+ births	8.9(8.0)	18.1(16.6)	2.27(2.3)	0.000(0)	12.6(11.1)	18.1(16.6)	1.53(1.59)	0.000(0)	8.9(8.0)	12.6(11.1)	1.48(1.44)	0.005(0.019)

Notes: (1) $P_1(x)$: Percentage of elders who gave birth after age 35 among those aged x. (2) p: Significance level of chi-square tests based on method of Mantel and Haenszel (1959) to test whether there is a statistically significant difference of survivorship between those with the fixed attribute and those without it.

Table 9. Ratio of healthy survivorship (RHS) of elders who had 1+, 2+, or 3+ birth after age 35 to those who did not have such late births

-- Estimates not in the parenthesis are based on whole CLHLS data set including both Han and the six minority groups cited from Zeng and Vaupel (2004); Estimates in the parenthesis are based on CLHLS data set excluding the six minority groups

$P_1(x), \pi(x+n)$	Ages 100-105 vs. 80-85				Ages 100-105 vs. 90-95				Ages 90-95 vs. 80-85			
	$P_1(80-85)$	$\pi(100-105)$	<i>RHS</i>	P	$P_1(90-95)$	$\pi(100-105)$	<i>RHS</i>	P	$P_1(80-85)$	$\pi(90-95)$	<i>RHS</i>	P
Men												
1+ births	69.0(69.0)	71.4(70.3)	1.12(1.06)	0.506(0.717)	72.3(71.7)	71.4(70.3)	0.95 (0.93)	0.778(0.700)	69.0(69.0)	72.3 (71.2)	1.17 (1.11)	0.151(0.358)
2+ births	50.1(49.5)	56.3(55.7)	1.28(1.28)	0.103(0.114)	49.9(49.3)	56.3(55.7)	1.29 (1.29)	0.097(0.111)	50.1(49.5)	49.2 (48.4)	0.96 (0.96)	0.709(0.669)
3+ births	30.2(29.6)	39.3(37.5)	1.50(1.43)	0.010(0.029)	32.7(32.6)	39.3(37.5)	1.34 (1.24)	0.069(0.196)	30.2(29.6)	31.5 (31.4)	1.07 (1.09)	0.556(0.445)
Women												
1+ births	49.8(48.8)	60.5(58.2)	1.54(1.46)	0.000(0.000)	51.3(50.3)	60.5(58.2)	1.45 (1.37)	0.000(0.003)	49.8(48.8)	51.9 (50.2)	1.09 (1.06)	0.891(0.597)
2+ births	24.6(23.7)	36.7(34.1)	1.78(1.66)	0.000(0.000)	28.2(26.6)	36.7(34.1)	1.48 (1.42)	0.000(0.002)	24.6(23.7)	29.0 (26.7)	1.25 (1.17)	0.045(0.188)
3+ births	8.9(8.0)	20.7(18.6)	2.68(2.61)	0.000(0.000)	12.6(11.1)	20.7(18.6)	1.81 (1.82)	0.000(0.000)	8.9(8.0)	13.9 (11.6)	1.67 (1.51)	0.001(0.018)

Notes: (1) $P_1(x)$: percentage of elders who gave birth after age 35 among those aged x. (2) $\pi(x+n)$: percentage of elders who gave birth after age 35 among those age x+n and ADL independent. (3) p: Significance level of Chi-square tests based on method of Mantel and Haenszel (1959) to test whether there is a statistically significant difference of healthy survivorship between those with the fixed attribute and those without it.

Figure 1. A comparison of the age distributions of centenarians among Sweden, Japan, U.S.A., Chile and CLHLS (including Han and minorities in the 22 CLHLS provinces)

