Mortality Analysis of China’s 2000 Population Census Data: A Preliminary Examination*

Li Shuzhuo and Sun Fubin

Abstract

Mortality analysis is an important part of the demographic analysis of census data. Using the tabulated census data released by census authority, this paper examines the quality of census data, calculates the mortality indicators of the 2000 population census, and analyses the levels and patterns of reported mortality in the last decade in China. The paper finds no obvious age preference and age heaping. However, underreporting in

LI Shuzhuo is a Professor at the Institute for Population and Economic Studies, School of Management, Xi’an Jiaotong University, China. He received a Ph.D. from Xi’an Jiaotong University. His research interests include mortality analysis, gender and child survival, and reproductive health in China.

SUN Fubin is an associate professor at the Institute for Population and Economic Studies, School of Management, Xi’an Jiaotong University, China. He received a M.Sc. degree from the London School of Economics, and a master’s degree and a Ph.D. degree from Xi’an Jiaotong University. His research interests include population studies, urbanization and human resources management. He has published widely in Zhongguo renkou kexue (Population Science of China), Renkou yanjiu (Population Research), Asia-Pacific Population Journal and various other academic journals.

*An earlier version of this paper was presented at the Conference on Chinese Population and Socioeconomic Studies: Utilizing the 2000/2001 Round Census Data, 19–21 June 2002, Hong Kong, and the Workshop on Chinese Census 2000, 22–23 August 2002, Seattle. We are very grateful to participants in the two workshops, especially to the two anonymous reviewers, for their insightful and constructive comments on earlier drafts of this paper.
age and death counts is found to be serious, especially among young children and young adults, resulting in a lower reported mortality than the actual figure. Due to the more hazardous occupations of male migrant labourers in urban areas, females have experienced a more rapid decline in mortality, especially among young adults in urban areas. However, excess female infant mortality is increasing, especially in rural areas. The direction of further thorough analysis of mortality data from the 2000 census is also discussed.

Introduction

China conducted its fifth population census in 2000, setting 1 November as the standard reference date. Compared with the four previous censuses, this census made several innovations, such as a new definition of the usual residence, new definitions of shi (city), zhen (town) and xian (county), and an improved design of the census questionnaire with its many new census questions, as well as changes in data entry techniques.1 By international standards, the census has been officially claimed as a success.2 However, the 2000 census is the first one implemented in the transitional market economy in China, and many unprecedented challenges strongly influenced the census implementation and quality. In fact, the quality of the 2000 census is not regarded as comparable with that of the previous four censuses. This is evidenced in an underreporting rate of 1.81% in the post-enumeration survey for the total population in the 2000 census, while the corresponding underreporting rate in the 1990 census was only 0.06%.3 Several studies have examined the implementation and quality of the 2000 census, and found that many changes and factors contributed to the increase in underreporting, for example, a huge increase in the floating population, separation of people from their household registrations in urban areas, changes in people’s attitudes towards cooperation, and possibly, in some areas, the intervention of local officials.4

Enumeration of deaths is an important part of a population census. Analysis of reported death data is also an important component of demographic studies to indicate the level and pattern of mortality, as well as to reflect socioeconomic change and improvement in people’s living standards. However, given the relative complexity of the 2000 census implementation and the relatively poor quality of the 2000 census data, this paper does not attempt to conduct a thorough mortality study of it. Instead, as a first step in the mortality analysis of the 2000 census data, it aims to
present the levels, patterns and changes in reported mortality, describe quality of age and death reports, and discuss possible ways of carrying out further mortality analysis in the future.

Data and Methods

Data

Data used in this paper are tabulated population data provided by the State Statistical Bureau. The registered total population in this census is 1,242.61 million, 23.22 million fewer than the total population of 1,265.83 million that is formally promulgated by the State Statistical Bureau. The total population figure is adjusted based on the underreporting rate of 1.81% obtained in the post-enumeration survey. Since there is no formally released information about the age and sex distribution of the undercounted population, this paper uses only the data of the registered population and population deaths.

Methods

Population deaths in the 2000 census refer to those who died between 1 November 1999 and 31 October 2000. To generate life tables corresponding to this period and consequently analyse mortality levels and patterns in the 2000 census, population by age and sex on 1 May 2000 needs to be estimated first. For this purpose, the self-iteration method developed by Jiang et al. is used in this paper, with some modifications. In addition, the cohort intercensal survival ratios between the 1990 census and the 2000 census are calculated and analysed to examine quality of population data in the 2000 census. Since the standard reference time is different between the two censuses, the population on 1 July 2000 needs to be estimated first based on the population on 1 November 2000 in the 2000 census, so as to calculate the 10-year cohort survival rates. The following describes the main methods used in this paper to estimate the reported mortality.

Estimation of Population on 1 May 2000

The self-iteration method was first developed by Jiang et al. in analysing the 1982 census data, and then modified in 1992 for analysing the 1990 census data. This iteration method is used in this paper to estimate the
population on 1 May 2000. However, since this method was specifically proposed for the analysis of the 1982 population census, here some modifications are made to fit the special situation of the 2000 census.

For simplification, let $PN_a$ represent the registered population aged $a$ on the census standard reference date of 1 November 2000, $PM_a$ represent population aged $a$ on 1 May 2000, $PL_a$ represent population aged $a$ on 1 November 1999. Furthermore, let $D_a$ represent the registered population deaths aged $a$ in the period of one year prior to 1 November 2000. Finally, let $SR_a$ represent the survival rates for population aged $a$ on 1 November 1999 to population aged $a+1$ on 1 November 2000. To start the iteration, the initial values of $SR_a$ are set as $SR_a^0$. Hence, the initial values of $PL_a$ could be calculated as follows:

$$PL_a^0 = PN_{a+1}^0 / SR_a^0$$  \hspace{1cm} (1)

Then,

$$PM_a^0 = (PL_a^0 + PN_a^0) / 2$$  \hspace{1cm} (2)

Therefore, the mortality rate of population aged $a$ is calculated as follows:

$$m_a^0 = D_a / PM_a^0$$  \hspace{1cm} (3)

Now, a single-year life table could be calculated based on these age-specific mortality rates $m_a^0$, and the average population aged $a$, represented by $L_a^0$, could be obtained. With the assumption of linear population age distribution, the second round survival rates are calculated as:

$$SR_a^1 = L_{a+1}^0 / L_a^0$$  \hspace{1cm} (4)

Next, these results from equation (4) could be used in equation (1) above, and then the iteration could be continued until it converges to a given accuracy. At the point of convergence, the obtained $PM_a$ is exactly the estimated population aged $a$ on 1 May 2000.

**Generation of Life Tables**

The population by age and sex on 1 May 2000, which is exactly the population at the middle of the time period for the registered deaths, is obtained by the above method. This estimated population could be used as the average population for calculating the age-specific death rates by sex. Obviously, with these age and sex-specific mortality rates, a standard
method could be used to calculate the probabilities of death by age and sex and then to generate single-year age group life tables. For convenience, the single-year life tables are converted into five-year abridged life tables. As a result, the life tables based on reported deaths of the 2000 census are finally generated for China as a whole, as well as for shi, zhen, and xian, respectively.

It should be noted that due to the nonlinear distribution of infant deaths, a special formula below is used to convert mortality rate aged 0 represented by $m_0$ to probability of death for infants represented by $q_0$:

$$q_0 = 1 - \exp (-m_0 [0.9539 - 0.5509m_0]) \quad (5)$$

Estimation of Population on 1 July 2000

The standard reference date is 1 November for the 2000 census and 1 July for the 1990 census. Thus it is difficult to calculate directly the 10-year cohort intercensal survival rates by five-year age group between these two censuses. Therefore, the population on 1 July 2000 needs to be estimated from the registered population and population deaths by age and sex in the 2000 census. Figure 1 illustrates the relationship between the

Figure 1: The Diagram for Estimating the Population on 1 July 2000 from the Registered Census Population on 1 November 2000

![Diagram](image-url)
estimated population on 1 July 2000 and the registered population in the census on 1 November 2000. It is noted that the survival rates are calculated only for the national population, and international migration is largely neglected.

For simplification, let $P_{N_a}$ represent the registered population aged $a$ on 1 November 2000, and $P_{J_a}$ represent the estimated population aged $a$ on 1 July 2000. Further, let $D_a$ represent the registered population deaths aged $a$ in the time period of one year prior to 1 November 2000. Then the following equation is used for estimating $P_{J_a}$ from $P_{N_a}$:

$$P_{J_a} = \frac{2}{3} P_{N_a} + \frac{1}{3} P_{N_{a+1}} + \frac{5}{18} D_a + \frac{1}{18} D_{a+1}$$  \hspace{1cm} (6)

With the estimated population $P_{J_a}$ on 1 July 2000, the intercensual survival rates from 1990 to 2000 could be calculated. It needs to be pointed out that since the registered population deaths are possibly underreported in the 2000 census, the population estimated above in equation (6) might also be underestimated, which would obviously affect the intercensual survival rates. In addition, the armed forces are not included in calculating the survival rates, which may affect the survival rates for young adults.

**Results and Discussion**

**Age Reporting**

Before analysing mortality levels and patterns, it is necessary to evaluate the census age reporting in order to obtain some preliminary ideas of data quality. This work is divided into two parts. The first is an examination of the accuracy of age distribution, with a focus on possible age misreporting and heaping. The commonly used Whipple’s index and Myer’s index are employed here to examine the quality of age reporting. The second is an investigation of possible underreporting by analysing the intercensual survival rates from 1990 to 2000.

Table 1 presents the Whipple’s index and Myer’s index for all the five censuses in China. It is found that the Whipple’s index of the 2000 census is slightly higher than those of previous censuses, but it is still close to 100 for both males and females and is at an acceptable level according to international standards. This suggests that there is no obvious preference for the ages ending with numbers 0 and 5. Moreover, the Myer’s index is close to zero for both males and females, and is at a level comparable with those of the previous four censuses. In general, both Whipple’s index and
Myer’s index indicate that there is no obvious age preference and heaping in the 2000 census age reports.

Table 2 presents intercensal survival rates from 1990 to 2000 for five-year age cohorts. In the following discussion about survival rates, the age groups refer to those in 1990. Theoretically, for any age cohort, the survival rate from 1990 to 2000 should be below one. However, as is seen in Table 2, the survival rate is above one for several age cohorts, mostly children and young adults.

There are two possible explanations for this phenomenon. The first is that these age cohorts were undercounted in the 1990 census. The reason for the undercount of children may be attributable to the well-known underreporting of out-of-plan births in the 1980s. The reason for the undercounts of the young adults may be due to the underreporting of the rural migrants in urban areas who have no fixed place for household registration or who are separated from their household registration place. The second is the possible over-enumeration for the corresponding age groups in the 2000 census. It is quite possible that both explanations are at work here. However, we speculate that the first is more likely.

The sex differences in survival rates are mostly found in age groups 10–14 years old and 20–24 years old. That is, the survival rate is above one for females aged 10–14 and for males aged 20–24. It is possible that girls of 10–14 years old and male adults of 20–24 years old were underreported in the 1990 census.

As is well-known, the social and economic circumstances of the 2000 population census are much more complicated and challenging than those
of the 1990 census. These may affect the quality of the 2000 census more than the 1990 census, leading to a population undercount of the 2000 census higher than that of the 1990 census. Actually, the officially published undercount rate is 1.81% for the 2000 census, while it was only 0.06% in the 1990 census. It is quite possible that the characteristics of population underreporting observed in the 1990 census would have also occurred in the 2000 census. Therefore, it is reasonable to claim that population underreporting in the 2000 census is very likely to be concentrated in the age groups of children and young adults.

**Mortality Levels and Patterns of National Population**

Using the methods described above, and registered population and death data in the 2000 census, the reported life tables for both males and females of the national population, as well as of shi, zhen, and xian populations, could be computed. These figures provide information to describe the levels and patterns of reported mortality for different populations, and to further examine the dynamics of mortality levels and patterns in the 1990s.
Mortality Levels

Table 3 presents life expectancy at birth and infant mortality rates for males and females in the last three censuses. It should be noted that in this paper, while mortality indicators for the 1982 and the 2000 censuses are original and unadjusted, those for the 1990 census are adjusted. However, as the 1982 census is regarded as a high-quality census with reliable mortality indicators, in essence, only mortality indicators from the 2000 census are unadjusted from the original reported death data. This provides a basis on which to compare mortality indicators over censuses and hence to evaluate quality and reliability of reported mortality indicators in the 2000 census.

It can be seen that reported life expectancies at birth are probably too high in 2000. From 1989 to 2000, life expectancy at birth increased by 4.10 years and 4.78 years for males and females, respectively. Similarly, infant mortalities are too low in 2000. During the past eleven years, infant mortality rates declined very rapidly, to about 11.41‰ and 7.68‰ for males and females, respectively. In the 1990s, the enormous economic development of China became quite noticeable and has really improved the living standards of Chinese people. This factor may have contributed to the increase in longevity and decline in infant mortality. Still, the overall improvement is not sufficient to lead to such a large decline in mortality in the last decade. In fact, reported infant mortality is much less than that observed in the child health surveillance system organized by the Ministry of Public Health, which reports infant mortality of over 30‰. In addition, the regions of China with higher fertility are generally poor areas with a less developed economy, so the decline in infant mortality must be quite

Table 3: Mortality Levels Based on the 1982, 1990, and 2000 Censuses

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life expectancy at birth</td>
<td>66.43</td>
<td>66.91</td>
</tr>
<tr>
<td>Infant mortality rate (%)</td>
<td>36.47</td>
<td>32.19</td>
</tr>
</tbody>
</table>

limited. Moreover, research shows that when there is undercount of population deaths, infants and children usually suffer most. In general, it is very likely that mortality levels are seriously underreported in the 2000 census.

**Mortality Patterns**

Figures 2 and 3 present age-specific mortality rates in 2000 and in 1989 for males and females, respectively. It is true that the mortality suffers from serious underreporting in the 2000 census. However, if roughly the same underreporting rate is assumed across different age groups by sex, analysis of mortality patterns still provides us with some useful information about changes in mortality. It is apparent from Figures 2 and 3 that the mortality patterns in 2000 for both males and females exhibit the standard normal “J” shape, showing relatively high mortality among infants and in old age. Figure 4 presents age and sex-specific mortality ratios based on the 1990 and 2000 censuses. It can be seen that for both males and females, all mortality ratios are below one, indicating that the age-specific mortality rates in 2000 are generally lower than those in 1989. Of course, it is noted that the 2000 mortality suffers from serious underreporting, which contributes to an observed mortality decline over time. In addition, mortality declines more for children and young adults than for infants, older adults

**Figure 2: Age-specific Mortality Rates, Males, 1989 and 2000**

Source: Same as for Table 3.
Figure 3: Age-specific Mortality Rates, Females, 1989 and 2000

Source: Same as for Table 3.

Figure 4: Ratios of Mortality Rates of 2000 Census to 1990 Census, by Age and Sex

Source: Calculation based on Figures 2 and 3.
and the elderly. Also, female adults experience a greater decline in mortality than male adults. This is consistent with the theoretical expectations that during mortality decline, female mortality advantage is expected to increase. This is particularly true in transitional China, where male migrant labourers are at particular risk of mortality due to their hazardous occupations and poor working conditions. Of course, the possible contribution of the improvement in maternal health to the decline in female adult mortality cannot be omitted. Another possible reason might be the relatively greater undercount of female adult deaths than of male adult deaths.

Figure 5 presents the age-specific sex ratios of mortality for populations in 2000 and 1989. It is seen that except for infants and young children aged 1–4 years old, sex ratios of mortality are all above one across different age groups, indicating that in general mortality is higher for males than for females. A comparison of the years 1989 and 2000 reveals that the sex ratios of mortality increased for all age groups, indicating that males have a relatively high risk of dying. This situation is particularly true for adults, among whom female mortality declines more than male mortality. In addition to the possible reason mentioned above, that the undercount of female adult deaths is more than that of male adult deaths, it is also possible that the number of male adult deaths has been over-counted during the
Sex differences in infant and child mortality are particularly important for Chinese society, where a strong preference for sons and various forms of discrimination against girls are widespread and lead to a higher than expected mortality rate for female children, especially in rural areas.\textsuperscript{10} This is also evident in the 2000 census. Similar to the situation in 1989, infant and child mortality in 2000 was still higher for females than for males. Especially, as shown in Figure 5, while the sex ratio of child mortality aged 1–4 increased during 1989–2000, the sex ratio of infant mortality declined further. This suggests that discrimination against female infants was intensified, an observation consistent with other, micro-level, studies.\textsuperscript{11}

**Mortality Levels and Patterns for Shi, Zhen, and Xian Populations**

**Mortality Levels**

Since the definitions of *shi*, *zhen* and *xian* in the 2000 census are quite different from those of the previous censuses,\textsuperscript{12} in this paper no attempt is made to compare changes in mortality levels across the different censuses.

Table 4 presents life expectancy at birth and infant mortality rates for both males and females in *shi*, *zhen* and *xian* in 2000. It shows that for both males and females, while life expectancy at birth declines across *shi*, *zhen* and *xian*, the infant mortality rate increases across *shi*, *zhen* and *xian*. This indicates that mortality is highest in *xian* and lowest in *shi*, an observation consistent with the general socioeconomic and health conditions of these localities.

Comparing these reported mortality indicators for China’s *shi*, *zhen*...
and xian in 2000 with those observed in seven industrialized countries and some Asian countries provided by the United Nations, it is found that the life expectancy at birth for shi and zhen is very close to that of industrialized countries, and the life expectancy at birth for xian is quite high among Asian countries. Considering the real conditions of socioeconomic development, living standards and the health system in China, especially in rural areas, it is quite plausible to argue that mortality levels are seriously underreported across China’s shi, zhen and xian populations.

**Mortality Patterns**

Figures 6 and 7 present the age-specific mortality rates of males and females for shi, zhen and xian in 2000, respectively. In general, the mortality patterns of shi, zhen and xian all present the standard “J” shape, indicating that mortality is relatively high for infants and the elderly. Across all age groups, mortality is the highest for xian and the lowest for shi. The largest differences of mortality among shi, zhen and xian occur in the age groups of young children and the elderly.

Figure 8 presents sex ratios of mortality for shi, zhen and xian populations in 2000. The sex patterns of mortality for shi, zhen and xian are

![Figure 6: Age-specific Mortality Rates for Shi, Zhen and Xian, Males, 2000](image-url)
Figure 7: Age-specific Mortality Rates for Shi, Zhen and Xian, Females, 2000

Source: Calculation based on data from the 2000 censuses.

Figure 8: Sex Ratios of Age-specific Mortality Rates for Shi, Zhen and Xian, 2000

Source: Calculation based on data of Figures 6 and 7.
almost the same as those for the national population. That is, except for the age group under five years old, mortality is always higher for males than for females, and the largest sex difference of mortality appears to concentrate on young adults. Especially, the sex ratio of mortality for young adults is much higher for those in shi and zhen than for those in xian, which is again consistent with the relatively great occupational hazards for male rural migrants working in urban areas, as mentioned above.

Consistent with observation of the national population, across shi, zhen and xian, excess female mortality is found among children aged under five years old. It seems that discrimination against girls in contemporary China is universal across both urban and rural areas, with the situation worse in rural than in urban areas.

**Summary**

Using the reported tabulations of census data, this paper examines the quality of data and reported mortality indicators of the 2000 population census, as well as levels and patterns of reported mortality in the last decade in China. With respect to quality of age reports in the 2000 census, no obvious age preference or age heaping are found in the 2000 census. However, underreporting in population registration and population deaths in the 2000 census is found to be quite serious, especially among young children and young adults. This leads to reported mortality levels lower than real mortality levels, both for the national population and across populations in shi, zhen and xian. Furthermore, across all age groups, mortality is the highest in xian and the lowest in shi. Regarding the sex pattern of mortality in the 2000 census, it is found to be consistent with theoretical expectations, except for young children, across shi, zhen and xian, with females experiencing a more rapid decline in mortality than males, especially among young adults in urban areas. This might also be related to the more hazardous occupations of male migrant labourers in urban areas, as well as the possible more serious undercount of female adult deaths. For infants and children under five years old, females continue to suffer higher than expected mortality, and the situation for female infants has actually become worse in the last decade, especially in rural areas.

As mentioned at the beginning of this paper, this study is a first step in exploring mortality levels and patterns using the 2000 population census data. The results of this preliminary study suggest the direction of further
thorough analysis of mortality. First, age and death reports need to be adjusted based on the estimation of completeness of age and death reports. Based on this, real mortality can be estimated and its level and pattern can be correctly described and compared. In fact, when taking into account the underreporting issues of other sources of data used in the analyses, such as the population age distribution in the 1990 census, the undercount of the annually registered births, etc., these tasks become even more challenging. Second, analysis of the causes of deaths from other sources of data would help in the understanding of factors underlying the changes in the levels and patterns of mortality in the 1990s. Moreover, the excess mortality observed among female infants and children calls for more attention and research. Finally, the use of individual records in the micro data tape of the 2000 census would provide more valuable information about levels and patterns of mortality in the 1990s.

Notes

2. Ibid.
3. Ibid.
11. Li, Zhu and Feldman (Note 10).