Center for Studies in Demography and Ecology

National, Provincial, Prefectural and County Life Tables for China Based on the 2000 Census

by

Yong Cai
Graduate Student
Department of Sociology
University of Washington

UNIVERSITY OF WASHINGTON CSDE Working Paper No. 05-03

National, Provincial, Prefectural and County Life Tables for China Based on the 2000 Census[*]

Yong Cai
Department of Sociology
University of Washington
Seattle, WA 98195-3340

caiyong@u.washington.edu

Latest Revision: April 16, 2005

Data available from the 2000 Chinese census make possible the construction of sex specific life tables for China at county level. This paper documents the construction of life tables for all 2,870 county-level units, as well as 345 prefectural-level units, 31 provincial-level units and the national totals of China. The life tables are first constructed with raw data published in the census volumes. The Brass relational model is then employed to assess data quality and to produce smoothed life tables. Resultant life tables are available by request.

[*] I am grateful for the advice and suggestions of William Lavely.

1. Introduction

Studies of Chinese mortality have been limited to major divisions such as urban/rural, Han/minority, or provincial administrations (e.g. Banister and Hill 2004, CCMLT 1991, Li and Sun 2003, Zhao 2003) and have thus necessarily assumed mortality homogeneity within major subdivisions of the population. China is a big country. Even Chinese provinces are large and diverse entities, several with populations exceeding that of many European states. For a better understanding of Chinese mortality and its variation, data at the sub-provincial level is clearly desirable. Yet systematic measures of Chinese mortality at the local level have been lacking. The situation has changed with the 2000 Census, for which aggregated data are available in unprecedented detail. The present exercise utilizes 2000 census data to construct sex-specific life tables for 2,870 county-level units and for all higher-level administrative units. The primary purpose is to provide a set of serviceable county-level life tables as a foundation for understanding local and regional variation in Chinese mortality.

The life table is one of the oldest and most important analytical devices in demography. A life table summarizes the mortality experience of a cohort or more commonly of a population in one short period, such as in the year prior to a census. It consists of age specific mortality parameters that characterize the mortality pattern. The most frequently used parameter is the life expectancy at birth (e_0)—the average number of years a hypothetical cohort would live if the mortality conditions embodied in the life table held throughout its life.

Life table construction requires a sizable population base. The table is based on the observed age-specific mortality rates. Because mortality may be a rare event in a given age group, the accuracy of its observation could be directly influenced by population size. Although there is no standard sample size requirement for life table construction, a population of 50,000-100,000 is often recommended. For example, the US National Center for Health Statistics (NCHS 1987, 1998) requires a total of 700 deaths (for each sex, in 3 years) for life tables construction, implying a population size of 70,000 if we assume a death rate of 1 percent[1]. The county is the lowest Chinese administrative unit that meets this population size criterion for period life table construction.

The average population size of Chinese counties enumerated in the 2000 census is about 400,000. Ninety-five percent of counties have a population size over 50,000 and 90% have a population size of over 100,000. As for deaths occurring in the year prior to the census, 95% of Chinese counties reported over 300 and 85% reported over 700. Thus, most Chinese counties provide a reasonable population base for life table construction.

---

[1] The required number of recorded deaths for life table construction in the United States has been on the decline. The requirement was 2000 deaths for the 1959-61 decennial life tables, and 1600 deaths for the 1969-71 decennial life tables.

The Chinese county is a social as well as political unit. Ever since the Qin unification of China in 221 B.C., a three-tier bureaucratic structure consisting of center, province (commandery), and county has been the backbone of China's state administration. Over the last 2000 years, China has experienced internal revolts, nomadic invasions, dynastic changes, territorial expansions and administrative reforms; however, these events never altered the county's position as the primary unit of administration. Over centuries the counties evolved to become local social and economic entities, in which people interact, exchange, form relationships, and take on a local identity.

To facilitate comparison we construct life tables for prefectural, provincial and national levels as well as for the county level. Under the Chinese constitution, the Chinese administrative system is a three-level hierarchy from the center to province and to county. County-level units include traditional counties (*xian* 县), county-level cities (*xianji shi* 县级市), urban districts (*qu* 区), and autonomous counties (*zizhixian* 自治县) and banners (*qi* 旗) in minority regions. Provincial-level units include provinces (*sheng* 省), autonomous regions (*zizhi qu* 自治区) and municipalities (*zhixia shi* 直辖市). The prefecture is a legacy administrative level that includes prefectures (*diqu* 地区), prefectural-level cities (*diji shi* 地级市), leagues (*meng* 盟), and autonomous prefectures (*zizhi zhou* 自治州). Most provincial-level units are divided into prefectural-level units; each prefectural-level unit contains about 10 or so county-level units. The hierarchical nature of these administrative units makes the life table at different levels complementary in the study of regional variation of mortality in China.

Given the nature of the data, some irregularities in the derived life table parameters are to be expected. Some are due to the relatively small population size of some units. For example, there are 8 units with a population smaller than 10,000, with a minimum of 6,384. Others are because of data quality. For example, there are counties that reported over 10,000 births but no infant deaths in the year prior to the census. We calculate the standard errors of estimated life expectancies to gauge the variation due to the sample size. We also use the Brass Logit Model (Brass 1971) to detect the irregularities of mortality reporting in the census data and to provide a smoothed version of life tables. The adjusted life tables parameters are compared with those calculated from raw data.

2. Data and Methods

2.1 Data

Data used are from the electronic version of the *Complete Collection of National and Provincial Population Census Data Assemblies* and the *Complete Collection of 2000 China County/District Population Census Data Assemblies*, distributed by the University of Michigan China Data Center. The data are derived from two sets of tables: Table 3-1 *Population by Age* *and Sex*, and Table 6-1 *Deaths by Sex, Age and Region* (1999.11.1-2000.10.31). The live population data are available by single years of age, with the exception of age 100 and above. The death data are aggregated in 5-year age groups, except for age 0, 1-4 and 100+. Data are available for 2,870 county-level units[2] (including 1,622 counties, 52 banners, 404 cities and 792 districts), as well as 345 prefectural-level units (including 38 prefectures, 258 prefectural level cities, 9 prefectural-level aggregates in 4 municipalities, 7 leagues, 30 autonomous prefectures, and 3 units directly administrated by provinces), 31 provincial-level units (22 provinces, 5 autonomous regions and 4 municipalities) and the national totals. For each unit, life tables for male, female and total population are constructed, yielding a total of 9,741 life tables.

2.2 Estimation of $_nq_x$

A life table consists of a column of age interval (x, $x+n$) and various columns of age-specific mortality parameters, usually including but not limited to: the probability of dying in the interval ($_nq_x$)[3], the number alive at exact age x (l_x), total person-years lived in the interval ($_nL_x$), total person-years lived beyond age x (T_x), and life expectation at age x (e_x). Each parameter and some of their combinations characterize the mortality pattern.

The key step of life table construction is the derivation of $_nq_x$ — the probability of dying between age x and $x+n$. Because the Chinese census data only report the live population at the time of the census and the deaths in the year prior to the census, the base population and the death population are not synchronized. For example, someone who died at age 10 in the year prior to census would have been age 11 at the time of the census had they survived. A common method in period life table construction is first to calculate the age-specific death rate ($_nm_x$) and then convert it to $_nq_x$. The observed death rate is defined as the number of deaths between age x and $x+n$ divided by the person-years lived ($_nP_x^m$) by the same age group in the interval. The census enumeration is the end-year population data for those who survived the risk of deaths in the year prior to the census. We approximate the person-years lived by taking the average of end-year population of two successive age groups plus the years lived by those who died in the interval (Equation 1).

$$_nP_x^m = \tfrac{1}{2}(_nP_x + {_nP_{x+1}}) + \tfrac{_na_x}{n} \cdot {_nD_x} \qquad (1)$$

---

[2] Three archipelagos in the south China sea, Xisha, Zhongsha and Nansha are listed as county-level units of Hainan province. Xisha reported a population of 517 and the other two reported none. These three units are excluded from this exercise. No data is available for Hong Kong, Macau and Taiwan.

[3] We follow the notation system commonly found in demographic literature: the right subscript x referring to the age at the beginning of the age interval, and the left subscript n referring to the length of the interval. Both are measured in exact number of years. The beginning of interval is included in the interval, but not the end of the interval.

where ${}_nP_x$ and ${}_nD_x$ are the population and deaths of age between x and $x+n$ reported in the census, ${}_nP_{x+1}$ is the population of age between $x+1$ and $x+1+n$, and ${}_na_x/n$ is the average number of years lived in the year prior to the census for those who died in the year prior to the census. Dividing ${}_nD_x$ by ${}_nP_x^m$, we obtain the observed ${}_nm_x$ (Equation 2).

$$
{}_nm_x = \frac{{}_nD_x}{\frac{1}{2}({}_nP_x + {}_nP_{x+1}) + \frac{{}_na_x}{n} \cdot {}_nD_x} \tag{2}
$$

Equation 2 requires ${}_na_x$, which will also be used for converting ${}_nm_x$ to ${}_nq_x$ and for calculating ${}_nL_x$. There are various approaches for choosing ${}_na_x$ values. All produce similar results (Namboodiri et al. 1987, Preston et al. 2001, Shryock et al. 1976). We use the rule of thumb ${}_na_x = n/2$, except for the two youngest age groups. For age 0 and age 1-4, we use Preston et al.'s (2001) adaptation of Coale and Demeny (1983) as shown in Equation 3.

$$
\begin{array}{llcc}
 & & \text{Males} & \text{Females} \\
{}_1a_0 & \text{if } {}_1m_0 \geq .107 & .330 & .350 \\
 & \text{if } {}_1m_0 < .107 & .045+2.684 \cdot {}_1m_0 & .053+2.800 \cdot {}_1m_0 \\
 & & & \\
{}_4a_1 & \text{if } {}_1m_0 \geq .107 & 1.352 & 1.361 \\
 & \text{if } {}_1m_0 < .107 & 1.651-2.816 \cdot {}_1m_0 & 1.522-1.518 \cdot {}_1m_0
\end{array} \tag{3}
$$

Note that the Coale and Demeny formulas require ${}_1m_0$, which itself is estimated from Equation 2 and requires ${}_1a_0$. This circularity problem can be resolved by combining Equation 2 and 3 into Equation 4, in which the Coale and Demeny coefficients are represented by α and β.

$$
{}_1m_0 = \frac{{}_1D_0}{\frac{1}{2}({}_1P_0 + {}_1P_1) + (\alpha + \beta \cdot {}_1m_0)\,{}_1D_0} \tag{4}
$$

Equation 4 can be solved numerically because there is only one unknown parameter (${}_1m_0$). When there are two solutions to Equation 4, we narrow it to one by applying the restriction given by Coale and Demeny, that ${}_1a_0$ is necessary between 0 and a ceiling value (.33 for male and .35 for female). ${}_4a_1$ and ${}_4m_1$ are calculated with the solution of Equation 4. The observed death rate (${}_nm_x$) calculated with Equation 2 is then converted to the probability of dying (${}_nq_x$) using Equation 5 (Chiang 1984).

$$ {}_{n}q_{x} = \frac{n \cdot {}_{n}m_{x}}{1 + (n - {}_{n}a_{x}) \cdot {}_{n}m_{x}} \tag{5} $$

2.3 Calculation of Standard Errors of Estimated Life Expectancies

The stochastic variation of life table functions can be substantial for units with a small sample size. To gauge the effect of stochastic variation, we calculate the sample variance and standard errors of life expectancies following Chiang (1984). Assuming that the age-specific death has a binominal distribution, the stochastic variation can be measured by standard errors of probabilities of dying (${}_{n}q_{x}$) and of the life expectancies (e_x). Equation 6 and 7 are the adaptations of formulas given in Chiang (1984) for the calculations of sample variance of ${}_{n}q_{x}$ and e_x.

$$ S^{2}_{{}_{n}q_{x}} = \frac{n \cdot {}_{n}m_{x}(1 - {}_{n}a_{x} \cdot {}_{n}m_{x})}{{}_{n}P_{x}^{m}[1 + (n - {}_{n}a_{x})\,{}_{n}m_{x}]^{3}} \tag{6} $$

$$ S^{2}_{\hat{e}_{x}} = \frac{\sum_{x=y}^{80} l_{x}^{2}[\hat{e}_{x+n} + (n - {}_{n}a_{x})]^{2} S^{2}_{{}_{n}q_{x}}}{l_{x}^{2}} \;, \; y = 0,1,5\ldots 80 \tag{7} $$

The cumulative nature of life expectancy means that the sample variance is also cumulated from the specific age to the end of life table. Thus, the standard error of life expectancy at birth is a good indicator of the overall effect of stochastic variation from sample size.

2.4. Steps in Constructing the Life Tables

To construct the life tables we follow the steps laid out by Chiang (1984) and Preston et al. (2001), with some minor modifications.

1) All death and age structure data for each unit are assembled from the 2000 census;

2) Observed age specific death rates (${}_{n}m_{x}$) are calculated using Equations 2 and 4;

3) ${}_{n}q_{x}$ values are converted from ${}_{n}m_{x}$ and ${}_{n}a_{x}$ values using Chiang's method (Equation 5) and Coale and Demeny's estimation procedure (Equation 3). To avoid the small sample problem at high ages, the life tables are closed at age 85+ by setting ${}_{\infty}q_{85} = 1.00$;[4]

---

[4] Given the high life expectancy in China, a large proportion of people live beyond age 85. For units with a large population size, it would be more appropriate to close the table at 100+ . A set of life tables at the national and provincial level closed at age 100+ are also constructed, but not discussed here.

4) The radix of the table (l_0) is set at 100,000. The populations at other ages are sequentially calculated with ${}_nd_x=l_x\,{}_nq_x$, and $l_{x+n} = l_x - {}_nd_x$;

5) The total number of person-year lived in the interval (${}_nL_x$) is computed as ${}_nL_x = {}_nd_x\,{}_na_x + l_{x+n}$, with the exception of ${}_\infty L_{85}$, which is set as ${}_\infty d_{85}/{}_\infty m_{85}$;[5]

6) The total number of person-years lived beyond age x and life expectancy at age x are calculated as $T_x= \sum_{a=x}^{\infty} {}_nL_a$ and $e_x=T_x/l_x$.

7) The standard errors for the life expectancies are calculated following the procedure laid out in Chiang 1984:209-211 (Equations 6 and 7).

3. Results

The life expectancies derived from the life tables so constructed accord well with published estimates. The national level life expectancies—male 70.99 and female 74.79 (see Appendix Tables A1 and Table A2)—are less than .03 years lower than those calculated by Li and Sun (2003) (male 71.01, female 74.77), which used the same data source but an iterative method to estimate mid-year population and death rates. NBS (2003:118) published a set of national and provincial life expectancies calculated from adjusted 2000 census mortality data—the adjustment was based on mortality rates observed in the annual population change surveys in the 1990s. Table 1 displays a comparison of our results with the NBS numbers. At the national level, NBS reports a life expectancy of 69.63 and 73.33, about 1.4 years lower than the results from this exercise. At the provincial level, these two sets of life expectancies are in good accord with one another, with the exception of Xinjiang, for which the NBS numbers are more than five years lower that unadjusted estimates. Although the NBS did not provide information on how the adjustments were made, the comparison indicates that adjustments were made for the undercounting of deaths in the census. Nevertheless, the good concordance between the NBS numbers and our results suggests that the life tables calculated from the raw census data, albeit underestimating the true mortality level in China, still provide a reasonable foundation for mortality analysis.

---

[5] There are a few cases for which no deaths for age 85 and above were reported (numbering, 3, 22, 11 for total, male and female population respectively). For these cases we take the provincial level ${}_\infty a_{85}$ to calculate ${}_\infty L_{85}$ as $l_{85}\,{}_\infty a_{85}$.

Table 1 Life Expectancies at Birth: A Comparison of Direct Estimates from Census Raw Data and NBS Adjustments, China 2000

	Derived from Raw Census Data			NBS Adjustments		
Region	Total	Male	Female	Total	Male	Female
National Total	72.82	70.99	74.79	71.40	69.63	73.33
Beijing	77.09	75.31	78.99	76.10	74.33	78.01
Tianjin	76.20	74.59	77.92	74.91	73.31	76.63
Hebei	72.86	70.98	74.93	72.54	70.68	74.57
Shanxi	72.38	70.67	74.33	71.65	69.96	73.57
Inner Mongolia	70.96	69.36	72.90	69.87	68.29	71.79
Liaoning	74.58	72.76	76.59	73.34	71.51	75.36
Jilin	73.51	71.75	75.53	73.10	71.38	75.04
Heilongjiang	74.53	72.55	76.83	72.37	70.39	74.66
Shanghai	79.29	77.33	81.20	78.14	76.22	80.04
Jiangsu	75.81	73.57	78.11	73.91	71.69	76.23
Zhejiang	75.32	73.12	77.83	74.70	72.50	77.21
Anhui	73.22	71.44	75.03	71.85	70.18	73.59
Fujian	74.50	72.18	77.02	72.55	70.30	75.07
Jiangxi	70.44	69.71	70.99	68.95	68.37	69.32
Shandong	74.25	72.00	76.62	73.92	71.70	76.26
Henan	73.25	71.30	75.19	71.54	69.67	73.41
Hubei	72.96	71.10	74.96	71.08	69.31	73.02
Hunan	72.87	71.17	74.75	70.66	69.05	72.47
Guangdong	75.20	72.65	77.85	73.27	70.79	75.93
Guangxi	73.84	71.50	76.38	71.29	69.07	73.75
Hainan	76.01	73.57	78.44	72.92	70.66	75.26
Chongqing	72.20	70.27	74.39	71.73	69.84	73.89
Sichuan	72.19	70.20	74.39	71.20	69.25	73.39
Guizhou	66.87	65.39	68.56	65.96	64.54	67.57
Yunnan	66.63	65.32	68.06	65.49	64.24	66.89
Tibet	66.08	64.44	67.62	64.37	62.52	66.15
Shaanxi	71.34	70.14	72.64	70.07	68.92	71.30
Gansu	69.07	68.30	69.91	67.47	66.77	68.26
Qinghai	69.04	67.56	70.69	66.03	64.55	67.70
Ningxia	72.20	70.76	73.84	70.17	68.71	71.84
Xinjiang	72.53	71.09	74.35	67.41	65.98	69.14

Table 1 also shows large regional variation in mortality across China at the provincial level. Shanghai's life expectancy at birth (e_0) is nearly 80, while Tibet's is in the mid-60s, with most provinces in the low 70s. The overall pattern of inter-provincial variation well reflects the general picture of economic development in China: coastal provinces have higher life expectancy than those in the interior.

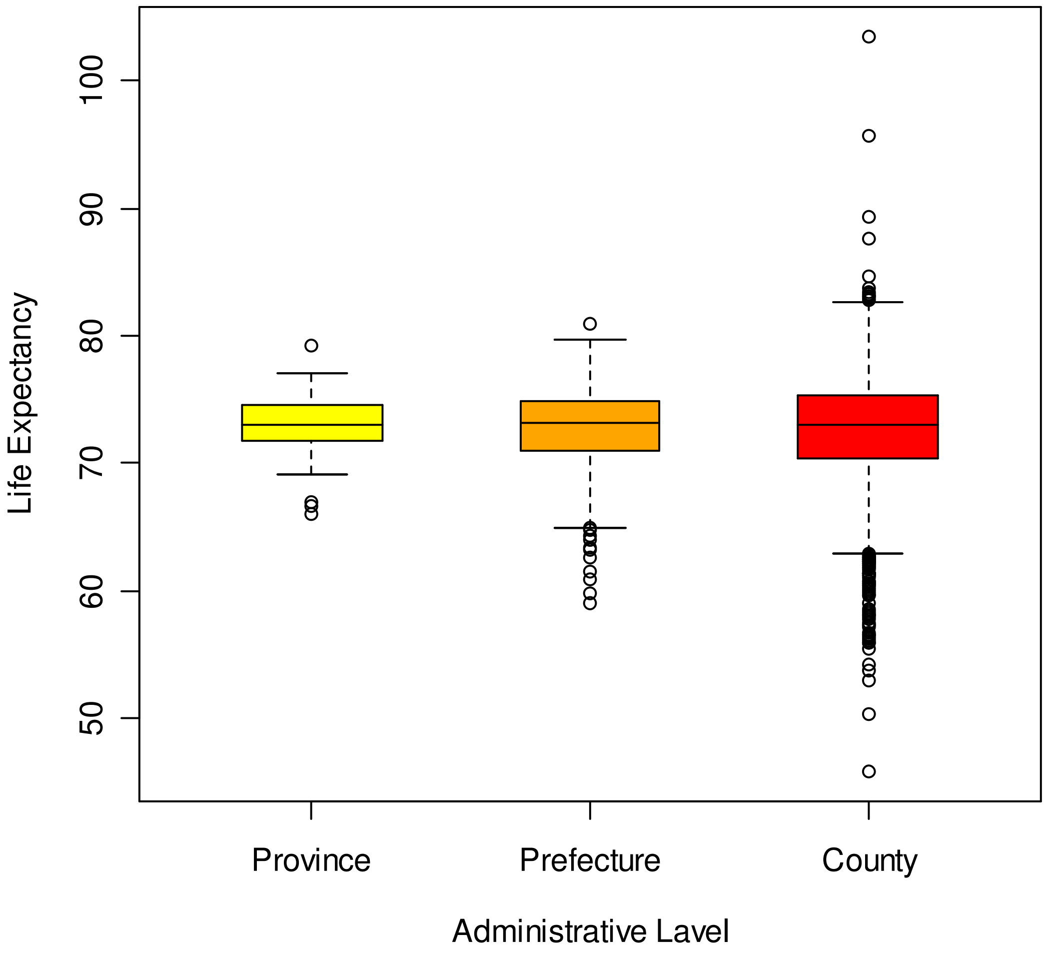
Chinese provinces are large, heterogeneous entities. It is not surprising that provincial tables conceal considerable mortality variation because aggregation tends to average out the extreme values in the lower level units. The top panel of Table 2 presents summary statistics on e_0 at provincial, prefectural and county levels. The distributions of e_0 for total population for the three administrative levels are shown as box plots in Figure 1. Box plots (Chambers 1983) are excellent tools for illustrating variations among different groups of data. The box represents the body of the data (middle 50%), with the central line indicating the median. The "whiskers" show the truncated range of the data, illustrated here as 1.5 times the inter-quartile range from the box. The dots outside of "whiskers" are "outliers."

Table 2 Descriptive Statistics of Life Expectancy at Birth (e_0) and Infant Morality Rate (q_0) for Provinces, Prefectures, and Counties, China 2000

				Percentile						
Variable	N	Mean	S.D.	Min	5	25	Median	75	95	Max
e_0 Provincial										
Total	31	72.8	3.0	66.1	66.4	71.3	73.0	74.6	78.0	79.3
Male	31	71.0	2.8	64.4	65.0	70.1	71.2	72.7	76.1	77.3
Female	31	74.7	3.3	67.6	67.9	72.9	75.0	77.0	79.9	81.2
e_0 Prefectural										
Total	345	72.6	3.4	59.0	65.9	70.9	73.2	74.9	77.0	81.0
Male	345	70.7	3.2	57.2	64.7	69.5	71.2	72.7	74.9	78.3
Female	345	74.6	3.8	60.6	66.7	72.9	75.4	77.2	79.6	83.5
e_0 County										
Total	2870	72.6	4.4	45.8	64.9	70.3	73.0	75.3	79.0	103.5
Male	2870	70.7	4.1	46.4	63.4	68.7	71.0	73.1	76.8	86.9
Female	2870	74.7	4.9	45.4	65.8	72.1	75.3	77.7	81.5	106.7
${}_1q_0$ Provincial (per thousand)										
Total	31	24.9	16.0	3.6	3.9	15.1	20.7	30.7	64.8	66.5
Male	31	21.9	14.0	3.6	3.8	11.7	17.7	28.4	56.9	58.5
Female	31	28.6	19.3	3.6	3.9	16.9	23.5	38.1	73.6	75.3
${}_1q_0$ Prefectural (per thousand)										
Total	345	25.8	20.8	2.6	6.1	11.9	20.0	31.9	62.7	138.2
Male	345	22.8	18.5	2.9	6.0	11.2	17.5	27.0	58.2	135.9
Female	345	29.6	25.1	2.4	5.8	11.6	22.4	39.7	83.9	162.7
${}_1q_0$ County (per thousand)										
Total	2870	25.7	25.2	0.0	3.4	9.7	18.3	32.9	73.4	342.2
Male	2870	23.0	22.5	0.0	3.2	9.4	16.7	28.8	64.0	307.3
Female	2870	29.2	31.2	0.0	2.8	9.4	19.1	37.5	88.7	376.7

Figure 1 reveals that the “core” variations of life expectancy in these three administrative levels are rather similar. In fact, the major difference between the provincial and prefectural level life expectancies is that the prefectural level life expectancy has a longer tail at the low end. In other words, except for some prefectures with very low life expectancies, the variation of mortality level at the provincial level is almost as large as it is at the prefectural level, suggesting that inter-provincial differences is the driving force of regional mortality inequality across China. An ANOVA analysis (not shown) confirms that inter-provincial variance explains about 60 percent of inter-prefectural variation and about 40 percent of inter-county variation.

Figure 1 Distribution of Estimated Life Expectancies for Both Sexes at Three Different Administrative Levels (Box Plots)



Inter-county variation is considerably larger than the inter-provincial and inter-prefectural variation. Again, this is in part a tail phenomenon. Maximum and minimum county values represent extreme outliers in an otherwise modest range of variation. For total population, there are four county-level units with a life expectancy higher than 85, and four county-level units with a life expectancy below 55. Setting aside the extreme cases, 90 percent of county-level units have a life expectancy at birth between 64.9 and 79.0, as compared with 66.4 and 78.0 for provinces and with 65.9 and 77.0 for prefectures.

Female life expectancies vary more than male life expectancies. Under normal circumstances, male and female life expectancies should move in parallel. The larger variation in female life expectancies begs for further investigation.

Variation of infant mortality is considerably larger than that of life expectancy. The bottom panel of Table 2 presents estimates of the infant mortality rate (${}_1q_0$). Among 2,870 county-level units, 271 have ${}_1q_0$ below 5 per thousand and 117 reported infant mortality below 3 per thousand. This is a clear evidence of infant death underreporting in the census. Even countries with the highest life expectancies such as Japan and Sweden still report infant mortality above 3 per thousand[6]. At the other extreme, there are some very high values of ${}_1q_0$, especially for females: 42 units reported male ${}_1q_0$ higher than 100 per thousand, and 106 units reported female ${}_1q_0$ higher than 100 per thousand. Most of those units are in Yunnan and Jiangxi provinces, where social and economic development has lagged.

These extreme values are not a result of small population size. Although some of the extreme values are observed in county-level units with small populations, most are not. The standard errors calculated for the life expectancies indicate that the stochastic errors could only contribute to these extreme values to a small degree. For example, the average standard errors of e_0 at county level for the female population is .19 (year), with a maximum of 1.4; the average standard errors of e_0 at county level for the total population is .14 (year), with a maximum of .96.[7]

Stochastic variation is not the only source of error in life table construction. Measurement errors, such as underreporting or age misstatement are likely to have more visible effects on life table parameters. Underreporting, as has been noticed by other, is present in the 2000 census. For example, there are 15 units that reported no difference between the numbers of births in the year prior to the census and the age 0 population, i.e. no deaths to the cohorts born in the census year. Table 3 lists the number of births for these 15 county-level units. Some of them have very large number of births, clearly signaling the undercount of infant deaths.[8]

---

[6] Hong Kong, a more relevant comparison to China, reports an infant mortality of 5 per thousand in 2000.

[7] The standard errors of the life expectancies reported here are based on the observed provincial average mortality aware of underreporting would underestimate the sample variance. We also calculated standard errors for the life expectancies based on the observed mortality. The difference between these two versions is relatively small.

[8] Some of the units reporting no infant deaths are clustered in space, suggesting that underreporting is a regional phenomenon that may be linked to local policy or variation in census administration.

Table 3 County-level Units Reporting No Deaths in the Cohort Born in the Year Prior to the 2000 Census

GB	English Name	Chinese Name	Number of Births
230712	Tangwanghe Qu	汤旺河区	259
230802	Yonghong Qu	永红区	1595
230804	Qianjin Qu	前进区	1236
230833	Fuyuan Xian	抚远县	543
320212	Mashan Qu	马山区	262
340103	Zhongshi Qu	中市区	1995
350202	Gulangyu Qu	鼓浪屿区	116
510302	Ziliujing Qu	自流井区	1404
510303	Gongjing Qu	贡井区	944
510304	Da'an Qu	大安区	3032
511602	Guang'an Qu	广安区	13074
511621	Yuechi Xian	岳池县	13277
511622	Wusheng Xian	武胜县	9190
622124	Subei Mongolian Zizhixian	肃北蒙古族自治县	115
650107	Nanquan Qu	南泉区	62

The obvious underreporting of infant morality in some units and the extreme values of life expectancy at birth raise important questions about the reliability of the life tables at the county level. There are other irregularities in the data as well, such as unusually high or low mortality in a particular age group that is inconsistent with the mortality levels of other age groups. To detect and smooth out such irregularities, we use Brass's logit relational model.

4. Smoothing with Brass's Relational Model

Brass (1971) found a simple linear relationship between the logit transformed life table parameter (l_x) of an observed population and of a standard population (Equation 8).[9] The Brass Relational Model is a powerful tool for assessing life tables, smoothing empirical data, completing a partial life table, and for population projection (Preston et al. 2001). The success of its application depends on the choice of the standard population, in particular, whether the standard population and study population belongs to the same "family." Brass (1971) suggests that the standard life table "must be some kind of average."

$$\text{logit}(\ l_x\ ) = \alpha + \beta\ \text{logit}(\ l_x^s\ ) \qquad (8)$$

---

[9] $logit(l_x)$ is defined as $.5ln(l_x/l_0/(1-l_x/l_0))$. In Brass's original notation, the life table starts with $l_0=1$, thus l_x is the same as the probability of surviving to age x.

We choose the Chinese provincial life tables constructed from the 2000 census as the standard life tables for subordinate prefectural and county units.[10] The provincial life tables presumably reflect the overall mortality pattern in each province and are obviously "some kind of average." We use the method outlined by Brass (1971:84) instead of ordinary least square to mitigate the influence of extreme values. Following Brass's argument that the model does not fit well at age one and at high ages because mortality in those age groups is subject to large reporting errors, we exclude l_1 and l_{85} from our calculation. l_x values are divided into two groups of equal number: one with l_x's from age 5 to 40, the other with l_x's from age 45 to 80. The mean values of $logit(l_x)$ are then calculated for each group. These values and their corresponding values from the standard populations define two points in space and determine a straight line, from which we derive the α and β values. The procedure is summarized in the Equation 9.

$$\begin{cases} \sum_{5}^{40} \text{logit}\,(l_x) = \alpha + \beta \sum_{5}^{40} \text{logit}(\,l_x^s\,) \\ \sum_{45}^{80} \text{logit}\,(l_x) = \alpha + \beta \sum_{45}^{80} \text{logit}(\,l_x^s\,) \end{cases} \tag{9}$$

Taking the estimated α and β values and l_x values from the standard population, we calculate a set of smoothed l_x values for each unit. Then a new set of life tables are constructed. We close the life table by taking the provincial level e_{85}'s to calculate $L_{85}=l_{85}e_{85}$. The life expectancies of prefectural- and county-level units after Brass-smoothing are summarized in Table 4.

Comparing the results for prefectural-level units in Table 2 and Table 4, we see that the Brass model produces little change in the life expectancies at the prefectural level. For over 90 percent of units the differences in life expectancies between derived from the raw and smoothed data are less than one year. The differences in ${}_1q_0$ values between derived from raw and smoothed data are more visible, but over 80 percent are still within 3 per thousand. Overall, the prefectural level life tables constructed from raw census data appear consistent and stable.

---

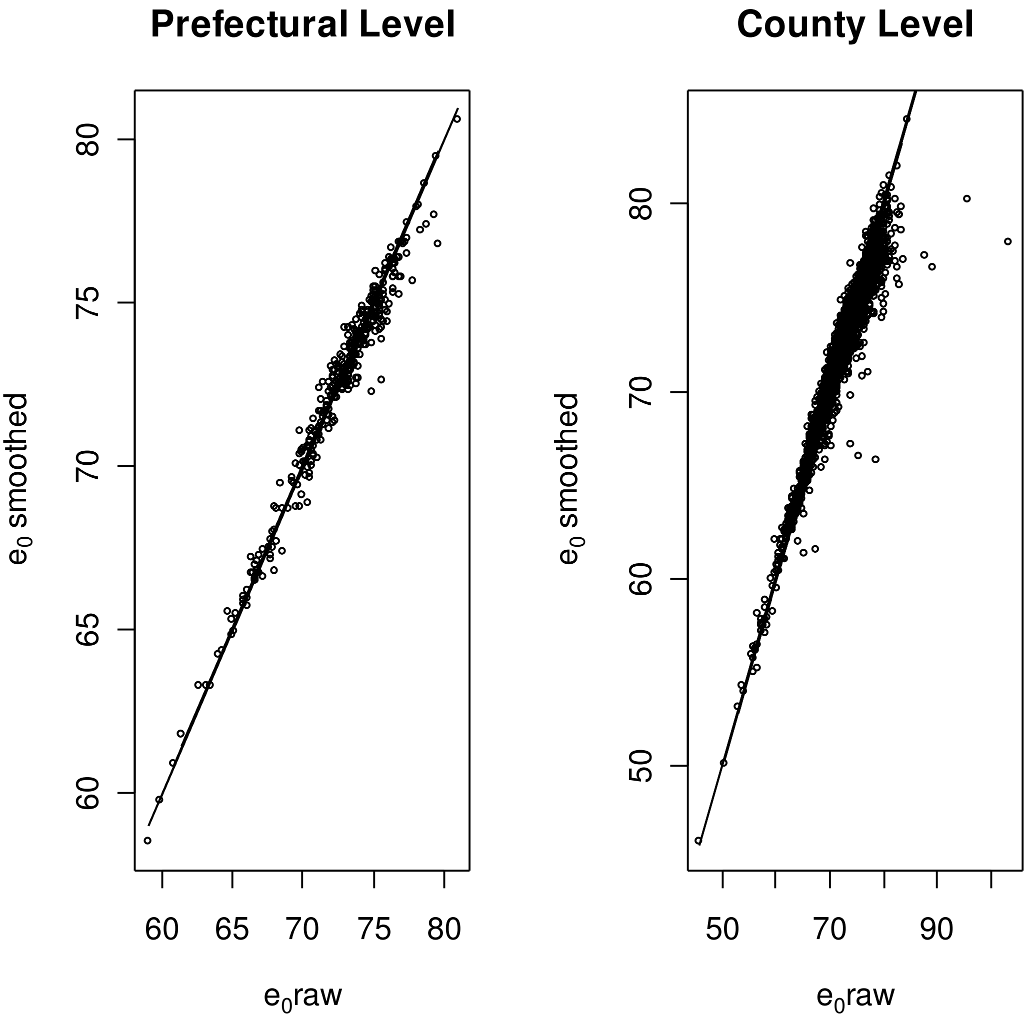
[10] The Coale-Demeny Model West tables were considered as an alternative standard. Using of Model West as a standard produces only small differences from tables that use provincial life tables as standards. The only exception is for the adjusted female ${}_1q_0$ and ${}_4q_1$ because of the unusually high and biased female infant and child mortality reported in the 2000 census (Cai and Lavely 2003).

Table 4 Descriptive Statistics of Life Expectancy at Birth (e_0) and Infant Mortality Rate (q_0) after Brass Logit Model Smoothing

				Percentile						
Variable	N	Mean	S.D.	Min	5	25	Median	75	95	Max
e_0 Prefectural										
Total	345	72.4	3.3	58.5	65.9	71.0	73.0	74.6	76.4	80.6
Male	345	70.6	3.1	57.5	64.6	69.5	71.2	72.6	74.6	78.5
Female	345	74.3	3.6	59.8	66.8	72.8	75.1	76.7	78.9	82.4
e_0 County										
Total	2870	72.2	3.9	46.0	64.8	70.3	72.8	74.7	77.4	84.5
Male	2870	70.5	3.8	46.6	63.4	68.7	70.9	72.7	75.6	82.8
Female	2870	74.0	4.3	45.6	65.9	72.0	74.8	76.9	79.6	86.0
${}_1q_0$ Prefectural (per thousand)										
Total	345	26.2	20.4	3.0	7.1	12.6	20.2	32.4	66.1	139.0
Male	345	23.1	18.0	3.0	7.5	12.2	17.5	26.9	59.1	136.9
Female	345	30.0	24.8	3.1	6.7	12.7	23.2	39.3	82.5	159.1
${}_1q_0$ County (per thousand)										
Total	2870	26.3	24.3	1.3	4.3	11.0	19.0	33.0	71.3	311.4
Male	2870	23.3	21.3	1.2	4.2	10.8	17.3	28.4	63.3	267.8
Female	2870	29.9	30.4	0.9	4.0	10.5	20.2	37.8	86.4	355.2

The Brass model smoothing plausibly corrects the irregularities and underreporting problems in mortality at the county-level. As observed in Tables 2 and 3, and as indicated by impossible e_0 values based on the raw data, deaths are seriously underreported in some county-level units. After smoothing, the highest e_0 values are reduced, while there are only minor changes at the low end, indicating that the smoothing is correcting for death underreporting.

Figure 2 Relationship between e_0 Derived from Raw Data and e_0 Derived from Brass Model Smoothed Life Tables



The relationships between the life expectancies at birth derived from the raw data and from Brass smoothing are illustrated in Figure 2, separately for prefectural- and county-level life tables. The points in the figure are the life expectancies at birth (for both sexes) calculated from raw data against their corresponding values calculated from the smoothed data. The solid lines serve as a reference, representing an identity between the raw and smoothed values. There is a near perfect linear relationship between these two sets of values. For most cases, the differences between the raw and smoothed values are small. This suggests that the raw data is of good quality, or at least internally consistent.

5. Discussion

Using the population age structure and death data from the 2000 census, we have constructed sex-specific life tables for 3,247 county-level and higher administrative units of China. The results match well with official national and provincial life tables. Comparison of prefectural- and county-level tables with the results of Brass Logit Model between indicates that the raw data life tables are of reasonable quality. The Brass smoothing corrects county life tables for the most egregious irregularities, whether due to stochastic variation or to underreporting of deaths.

One note of caution: the Brass Logit Model cannot correct for underreporting problems that exist in the standard population or that occur consistently across all ages in a specific unit. As shown in the comparison of the NBS publication and the results derived from the raw census data, the life expectancies estimated from this exercise are likely to overestimate survival probabilities. Even after Brass Logit smoothing, there are cases with remarkably high life expectancy and low infant mortality by international standards. According to the most recent United Nations Human Development Report (2003), Japan reported infant mortality as low as 3 per 1000 and life expectancy as high as 81.6. Considering China's current socioeconomic development level, units that match or exceed these Japanese levels probably reflect death underreporting.

Census and survey data from China were once praised for their remarkable accuracy (Coale 1984). The social and political conditions that favored a high-quality enumeration have faded with the increase in migration and relaxation of bureaucratic control (Lavely 2001). Underenumeration problems undoubtedly affect the 2000 census, as we have seen in this exercise. However, available evidence suggests that the data quality issues of the 2000 census should not be exaggerated. Using the general growth balance method, Banister and Hill (2004) studied mortality levels and trends from the 1960s to 2000 and concluded that the quality of data from the Chinese censuses is "quite high" and that census coverage has improved gradually over time.

Mortality risk relates to many factors. County-level life tables address the problem of population heterogeneity in higher level aggregates, but take no account of population heterogeneity within counties. In particular, life table are no substitute for micro-level mortality data, as yet unavailable for the 2000 Chinese census. Nonetheless, the present county-level tables provide an important foundation for the study of regional mortality variation in China.

Appendix Abridged National Life Tables, China 2000

Table A1 Abridged Life Table for Male Population

age x	1000mx	1000qx	lx	dx	Lx	Tx	ex
0	22.5	22.1	100000.0	2206.0	97906.6	7099442.5	70.99
1	1.5	5.9	97794.0	574.9	389653.7	7001536.0	71.59
5	0.6	3.2	97219.1	313.0	485313.1	6611882.5	68.01
10	0.5	2.5	96906.1	244.2	483920.2	6126569.5	63.22
15	0.8	3.8	96662.0	371.6	482380.8	5642649.5	58.38
20	1.2	6.0	96290.4	580.9	479999.6	5160268.5	53.59
25	1.4	6.8	95709.5	646.7	476930.5	4680269.0	48.90
30	1.7	8.3	95062.8	783.9	473354.1	4203338.5	44.22
35	2.2	10.7	94278.9	1006.9	468877.2	3729984.3	39.56
40	3.0	15.1	93272.0	1409.3	462836.6	3261107.0	34.96
45	4.3	21.4	91862.7	1962.0	454408.2	2798270.5	30.46
50	6.7	32.9	89900.6	2956.7	442111.5	2343862.3	26.07
55	10.5	51.2	86944.0	4454.7	423583.1	1901750.6	21.87
60	17.8	85.0	82489.3	7012.4	394915.3	1478167.5	17.92
65	29.2	135.9	75476.9	10257.9	351739.6	1083252.1	14.35
70	49.8	221.3	65219.0	14430.6	290018.2	731512.6	11.22
75	76.8	322.2	50788.3	16365.1	213028.8	441494.4	8.69
80	125.0	476.1	34423.2	16387.1	131148.3	228465.6	6.64
85+	185.3	1000.0	18036.1	18036.1	97317.3	97317.3	5.40

Table A2 Abridged Life Table for Female Population

age x	1000mx	1000qx	lx	dx	Lx	Tx	ex
0	32.0	31.1	100000.0	3110.2	97082.5	7478655.0	74.79
1	1.5	5.9	96889.8	574.8	386042.3	7381572.5	76.19
5	0.4	2.2	96315.0	212.1	481044.9	6995530.0	72.63
10	0.3	1.7	96103.0	159.7	480115.6	6514485.0	67.79
15	0.5	2.4	95943.3	227.4	479147.9	6034369.5	62.90
20	0.7	3.6	95715.9	346.0	477714.5	5555221.5	58.04
25	0.8	4.2	95369.9	399.7	475850.3	5077507.0	53.24
30	1.0	4.9	94970.2	464.6	473689.4	4601657.0	48.45
35	1.2	5.9	94505.6	557.9	471133.1	4127967.5	43.68
40	1.7	8.5	93947.7	794.1	467753.2	3656834.5	38.92
45	2.6	12.8	93153.6	1187.7	462798.7	3189081.3	34.23
50	4.2	20.7	91965.9	1901.3	455076.2	2726282.5	29.64
55	6.6	32.5	90064.6	2928.1	443002.6	2271206.3	25.22
60	11.4	55.2	87136.5	4813.3	423649.0	1828203.8	20.98
65	18.9	90.1	82323.2	7420.9	393063.6	1404554.8	17.06
70	33.5	154.7	74902.3	11589.6	345537.4	1011491.1	13.50
75	54.2	238.6	63312.7	15109.1	278790.7	665953.6	10.52
80	93.0	377.1	48203.6	18178.9	195570.7	387162.9	8.03
85+	156.7	1000.0	30024.7	30024.7	191592.2	191592.2	6.38

References

Anderson, Barbara. 2004. "Undercount in China's 2000 Census in Comparative Perspective." PSC Research Report 04-565. September 2004. Population Study Center, University of Michigan.

Banister, Judith and Kennth Hill. 2004. “Mortality in China 1964-2000.” *Population Studies* 58:55-75.

Brass, William. 1971. “On The Scale of Mortality” pp.66-110 in W. Brass, ed., *Biological Aspects of Demography*. London: Taylor and Francis Ltd; New York: Barns & Noble Inc.

Cai, Yong and William Lavely. 2003. "China's Missing Girls: Numerical Estimates and Effects on Population Growth." *China Review* 3:13-29.

Chiang, Chin Long. 1984. *The Life Table and its Applications*. Malabar, Fla.: R.E. Krieger Pub Co.

Coale, Ansley. 1984. *Rapid Population Changes in China, 1952-1982.* Washington, D.C., National Academy Press.

Coale, Ansley J. and Paul Demeny with Barbara Vaughan. 1983. *Regional Model Life Tables and Stable Populations.*

Chambers, John, William Cleveland, Beat Kleiner, and Paul Tukey. 1983. *Graphical Methods for Data Analysis.* Wadsworth and Brooks.

Compilation Committee of China Classified (Regional) Model Life Tables (CCMLT). 1991. *China Classified (Regional) Model Life Tables.* China City Publishing House, Beijing.

Lavely, William. 2001. “First Impressions from the 2000 Census of China”. *Population and Development Review* 24:755-69.

Li, Shuzhuo and Fubin Sun. 2003. "Mortality Analysis of China's 2000 Population Census Data: A Preliminary Examination." *China Review* 3:31-48.

Namboodiri, N. Krishnan and C. M. Suchindran. 1987. *Life Table Techniques and Their Applications*. Orlando Fla.: Academic Press.

National Bureau of Statistics of China (NBS). 2003. *The Complete Collection of National and Provincial Population Census Data Assembly*. Electronic version. University of Michigan China Data Center, CDC-S-2003-201.

National Center for Health Statistics and T.N.E. Greville (NCHS). 1967. Methodology of the National, Regional, and State Life Tables for the United States: 1959-1961. *Life Tables: 1959-61.* Vol. 1, Public Health Service Publication No. 1252, Vol. 1. No. 4.

National Center for Health Statistics and T.N.E. Greville (NCHS). 1975. Methodology of the National, Regional, and State Life Tables for the United States: 1959-1961. *Life Tables: 1969-71.* Vol. 1, No. 3. DHEW pub No. (HRA) 75-1150.

National Center for Health Statistics, R. J. Armstrong and L. R. Curtin (NCHS). 1987. Methodology of the National and State Life Tables. *U.S. Decennial Life Tables for 1979-81.* Vol. 1, No. 3. DHHS pub. No. (PHS) 87-1150-3.

National Center for Health Statistics and R. J. Armstrong. (NCHS). 1998. Methodology of the National and State Life Tables. *U.S. Decennial Life Tables for 1989-91.* Vol. 1, No. 2. DHHS pub. No. (PHS) 98-1150-2.

Preston, Samuel H., Patrick Heuveline, and Michel Guillot. 2001. *Demography: Measuring and Modeling Population Processes*. Malden, MA: Blackwell Publishers.

Shryock, Henry S., Jacob S. Siegel and Associates. 1976. *The Methods and Materials of Demography*. Academic Press, Inc.

Tu, Ping and Zhiwu Liang. 1993. "An Evaluation of the Quality of Enumeration of Infant Deaths and Births in China's 1990 Census." *Paper Presented in International Seminar on China's 1990 Census* (in Chinese).

United Nations Development Programme. 2003. *Human Development Report*. United Nations.

Zhao, Zhongwei. 2003. “On the Far Eastern Pattern of Mortality.” *Population Studies* 57:131-147.

Zhang, Weimin and Hongyan Cui. 2003. “Estimates of the Completeness of the China’s 2000 Census”. *Population Research* 27 (4):25-35 (in Chinese).