Estimating County-Level Life Expectancy In China

Yong Cai
Department of Sociology
University of Washington

Latest Revision: February 28, 2006

Draft manuscript prepared for presentation at the annual meeting of the Population of Association of America, Los Angeles, March 30 – April 1, 2006. Email address correspondence to: caiyong@u.washington.edu .

Abstract

Using data collected in China's 2000 census, we construct sex-specific life tables at county-level. One challenge to this effort is to provide reliable and robust estimates of infant mortality rate, which are more prone to underreporting. In this paper, we construct life table for each of China's 2367 county-level units and compare three different methods of estimating infant mortality rate: conventional estimates based on reported births and deaths, Brass logit model adjustments based on mortality rates of other age groups and a standard mortality schedule, and empirical Bayes estimates based on infant mortality rates reported in the neighboring counties. The results show a good correspondence among these three estimates, particularly between the empirical Bayes estimates and the Brass model estimates. The study not only confirms the overall quality of Chinese mortality data, but also demonstrates the power of empirical Bayes method as a useful tool for small area population estimation.

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, CMLTC 1991, Li and Sun 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 2,367 county-level units.

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 influenced directly by population size. Eayres and Williams (2005) evaluated the methodologies for small area life expectancy and found that life expectancy tends to be overestimated with a small population size. They recommend a population years at risk of 5,000, above which life expectancy can be calculated with a reasonable confidence. The US National Center for Health Statistics (NCHS 1987, 1998) requires a total of 700 deaths (for each sex, in 3 years) for life table construction, implying a population size of 70,000 if we assume a death rate of 1 percent. Although there is no standard sample size requirement for life table construction, a population of 50,000-100,000 is desirable.

County is the lowest Chinese administrative unit that has reasonable population size 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 good population base for life table construction.

Life table at county level also provides important connection between mortality and local social condition. Counties in China are social as well as administrative units. 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 2,000 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.

We construct life tables for each of China's 2,367 county-level units. There are 2,870 county-level units in China's 2000 census. To incorporate the spatial dimension into the analysis, we use a modified version of the GIS basemap provided by University of Michigan China Data Center (CDC), in which urban districts have been merged into cities, with the exception of peripheral districts in four provincial municipalities (Beijing, Tianjin, Shanghai and Chongqing). The 2,870 county-level units are thus reduced to 2,367 counties and cities. To simply the narrative, we will refer all the 2,367 counties and cities as counties.

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. Others are because of data quality. For example, there is a rich literature on the underreporting of infant deaths in Chinese population data. We calculate the standard errors of estimated mortality rates and life expectancies to gauge the variation due to the sample size. We then 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. Finally, we use empirical Bayes method to cross-validate infant mortality estimates using information from neighboring units.

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 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*. 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+.

2.2 Estimation of ${}_nq_x$

A life table consists of a column of age interval $[x, x+n)$ [1] and various columns of age-specific mortality parameters, usually including but not limited to: the probability of dying in the interval (${}_nq_x$), the number alive at exact age x (l_x), total person-years lived in

[1] 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.

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).

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

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).

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

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.

		Males	Females
${}_1a_0$	if ${}_1m_0 \geq .107$	.330	.350
	if ${}_1m_0 < .107$	$.045+2.684 \cdot {}_1m_0$	$.053+2.800 \cdot {}_1m_0$
${}_4a_1$	if ${}_1m_0 \geq .107$	1.352	1.361
	if ${}_1m_0 < .107$	$1.651-2.816 \cdot {}_1m_0$	$1.522-1.518 \cdot {}_1m_0$

(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).

$$ {}_nq_x = \frac{n \cdot {}_nm_x}{1 + (n - {}_na_x) \cdot {}_nm_x} \tag{5} $$

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 (${}_nq_x$) 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 ${}_nq_x$ and e_x.

$$ S^2_{{}_nq_x} = \frac{n \cdot {}_nm_x(1 - {}_na_x \cdot {}_nm_x)}{{}_nP_x^m[1 + (n - {}_na_x)\,{}_nm_x]^3} \tag{6} $$

$$ S^2_{\hat{e}_x} = \frac{\sum_{x=y}^{80} l_x^2[\hat{e}_{x+n} + (n - {}_na_x)]^2 S^2_{{}_nq_x}}{l_x^2}, \quad y = 0,1,5,\ldots 80 \tag{7} $$

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.

Chiang method assumes that as the probability of survival in the final age interval is zero, and thus by definition, the associated variance is also zero. Silcocks et al. (1995)

argue that for the final age interval the life expectancy is dependent on the mean length of survival, and suggest including a term for the variance based upon this assumption. Since including such a term of variance for the final age interval has only minimal effect on the calculation of stand error for life expectancy at birth, we follow strictly with Chiang.

2.4 Steps in Life Table Construction

We follow the steps laid out by Chiang (1984) and Preston et al. (2001) to county life tables, with some minor modifications.

- All death and age structure data for each unit are assembled from the 2000 census;
- Observed age specific death rates ($_nm_x$) are calculated using Equations 2 and 4;
- $_nq_x$ values are converted from $_nm_x$ and $_na_x$ values using Chiang's method (Equation 5) and Coale and Demeny's estimation procedure (Equation 3). To avoid the small sample problem at old ages, the life tables are closed at age 85+ by setting $_\infty q_{85} = 1.00$;[2]
- 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 $l_{x+n} = l_x - {}_nd_x$;
- 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 $_\infty d_{85}/{}_\infty m_{85}$;[3]
- 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 $e_x = T_x/l_x$;
- The standard errors for the life expectancies are calculated following the procedure laid out in Chiang 1984:209-211 (Equations 6 and 7).

[2] 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.

[3] There are a few cases for which no deaths of 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 $l_{85}{}_\infty a_{85}$.

2.5 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).[4] 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."

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

We choose the Chinese provincial life tables constructed from the 2000 census as the standard life tables for subordinate county units. 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)[5]. Following Brass's argument that the model does not fit well at age one and at oldest 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^s 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 $L_{85}=l_{85}e_{85}$.

[4] $logit(l_x)$ is defined as $.5ln(l_x/(l_0-l_x))$. 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.

[5] One alternative often used is to use ordinary least square to fit a line between the observed mortality schedule and the standard mortality schedule.

2.6 Empirical Bayes Estimate

The empirical Bayes estimation procedures have long been used to estimate mortality and disease rate in epidemiology (Efron and Morris 1973, 1975; Clayton and Kaldor 1987; Marshall 1991). The idea is to pool information across areas to produce more stable and robust estimates using the empirical Bayes methods. Suppose the observed infant mortality in a give unit is m'. In the Bayes framework, the pooled infant mortality estimate (m'') from all the neighboring units are taken as the *priori* for the local estimate. The local estimate can be written as

$$_1\hat{q}_0 = m'' + (m'' - m')\,\mathrm{var}(m'')/\mathrm{var}(m') \qquad (10)$$

Generally speaking, there are two kinds of empirical Bayes mortality estimates, different by assumption on spatial variability. The "global" estimator adjusts the crude mortality rate observed in each unit by the "global" mean. In our case, which means the mortality rate for each county is shrunk towards the overall infant mortality of China, thus assuming spatial homogeneity of the mortality rate. Given there are large variations in mortality across China—in fact, that is the motivation to construct local level life tables, the "global" empirical Bayes estimate is not appropriate for this analysis. It is more sensible to assume that neighboring units are more like each other than units that are far away.

We will use the "local" version empirical Bayes estimator. It adjusts the curde mortality rates observed in each county by the mean value of its neighboring units, i.e. it shrinks the observed mortality in a unit towards the mean value of its neighboring units. Marshall (1991) proposed a moment-based estimation method to avoid computationally prohibitive iterative approach. We use the implementation of *EBlocal* function in **spdep** package (Bivand 2005).

Calculating the local empirical Bayes estimation of mortality requires the definition of each unit's neighborhood. There are two general approaches in defining the neighborhood with irregular lattice data: by adjacency and by distance. In the first case, neighborhoods are defined by units sharing boundaries. This is parallel to the order (lag) 1 in time series data. Similarly, the neighborhood definition can be expanded to from order 1, to order 2 and so on, by including neighbors' neighbor. In the second case, neighborhoods are defined by all the units within a specific distance from a single point in the unit, usually the geographic centroid, or administrative seat. Following the guidelines developed by Griffith (1996), we use a combination of first-order contiguous units and the closest five units to define the neighborhood of each unit in this analysis. We identify all the first-order contiguous neighbors. For those units who have less than five contiguous neighbors, we add the closest 5 units according to the distance definition. The motivation of adding 5 closest neighbors to the first order spatial contiguity neighbors is that there are some islands units with no or only few neighbors by the spatial contiguity standard. We also force the neighborhood to be symmetric: if A is B's

neighbor, B must be A's neighbor. As a result, each unit has on average 6.9 neighbors, 52 units have more 10 neighbors.

Based on this neighborhood definition and a binary weight structure, the spatial autocorrelation calculated for unadjusted male and female infant mortality rates are .60 and .56, both of them significant at .001 level.[6] This strong autocorrelation vindicates the usage of empirical Bayes method: the neighboring units are indeed having strong similarities. Similar results are found for mortality in other age groups. To save space, we will only discuss the empirical Bayes estimates for infant mortality.

3. RESULTS

The life expectancies derived from the life tables so constructed accord well with published estimates. The national level life expectancies, if using the same method, are 70.99 for male and 74.79 for female, which are only less than .03 years lower than those calculated by Li and Sun (2003) (male 71.01, female 74.77) using the same data 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. 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. The good concordance between these 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.

Table 1 presents the descriptive statistics of county-level life expectancy at birth and infant mortality using different methods: two methods for life expectancy estimates and three methods for infant mortality estimates.

There is a considerable inter-county variation in life expectancy and infant mortality using the raw data. Both of them are in part a tail phenomenon. Maximum and minimum county values represent extreme outliers in an otherwise modest range of variation. For male population, one county has a life expectancy higher than 85, and 7 counties with a life expectancy below 55. For female population, there are 21 county-level units with a life expectancy higher than 85, and 4 county-level units with a life expectancy below 55. Setting aside these extreme cases, about 90 percent of county-level units have a life expectancy at birth in a 12-year range. There are even more visible variation in infant mortality. Among 2,367 county level units, 106 have male ${}_1q_0$ below 5 per thousand and 37 reported infant mortality below 3 per thousand. This is a clear indication of infant death underreporting in the census because even countries with the

[6] These spatial correlations are significant at .001 level for assuming a normal distribution or using permutation test.

highest life expectancies such as Japan and Sweden still report infant mortality above 3 per thousand. Hong Kong, a highly relevant comparison to China, reports an infant mortality of 5 per thousand in 2000. At the same time, there are some very high values of ${}_1q_0$ at the other end of extreme, especially for females: 40 counties reported male ${}_1q_0$ higher than 100 per thousand, and 100 counties reported female ${}_1q_0$ higher than 100 per thousand.

Table 1. Descriptive Statistics of Life Expectancy at Birth (e_0) and Infant Morality Rate (${}_1q_0$) Based on Different Methods

			Percentile						
Variable	Mean	S.D.	Min	5	25	Median	75	95	Max
e_0 Male									
Raw	69.9	3.9	46.4	62.9	68.1	70.5	72.3	75.3	86.9
Brass	69.8	3.7	46.6	62.9	68.2	70.5	72.1	74.5	79.3
e_0 Female									
Raw	73.8	4.7	45.3	65.4	71.4	74.5	76.8	80.0	106.7
Brass	73.4	4.2	45.5	65.3	71.3	74.3	76.2	78.6	84.0
${}_1q_0$ Male (per thousand)									
Raw	25.7	23.4	0.0	4.6	12.2	22.9	42.2	91.1	376.7
Brass	26.0	22.1	0.9	6.1	13.6	23.8	42.3	93.0	355.2
Bayes	25.4	22.1	1.1	5.7	12.7	23.3	42.1	89.5	370.8
${}_1q_{0\ \text{Female}}$ (per thousand)									
Raw	32.8	32.0	0.0	4.9	11.4	18.9	31.6	68.6	307.3
Brass	33.6	31.3	1.2	6.6	12.7	19.2	31.8	66.9	267.7
Bayes	32.8	31.1	1.1	5.9	11.9	18.9	31.1	66.3	301.5

These extreme values of life expectancy and infant morality rates are not results of small population sizes. 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 larger visible effects on life table parameters. Underreporting, as has been noticed by other (e.g.

[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.

Zhang and Cui 2003, Banister and Hill 2004), presents a significant problem in the 2000 census. For example, 15 units reported no difference between the numbers of births in the year prior to the census, some of them having thousands of births, and the age 0 population, i.e. no deaths to the cohorts born in the census year.

Comparing the results from Brass smoothing with the raw data, we see that the Brass model smoothing corrects some of the irregularities and underreporting problems in mortality. As observed in Tables 1 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.

The relationships between the life expectancies at birth derived from the raw data and those derived from Brass smoothing are illustrated in Figure 1, separately for male and female life tables. The points in the figure are the life expectancies at birth calculated from raw data against their corresponding values calculated from the smoothed data. The dashed lines serve as a reference, representing an identity between the raw based and Brass smoothed values, i.e. assuming the Brass smoothed value is same as the raw data based life expectancy. There is a near perfect linear relationship between these two sets of values for male population. While there are more discrepancies between these two sets of estimates for female, the discrepancies seem to be limited only to a small number of counties. 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.

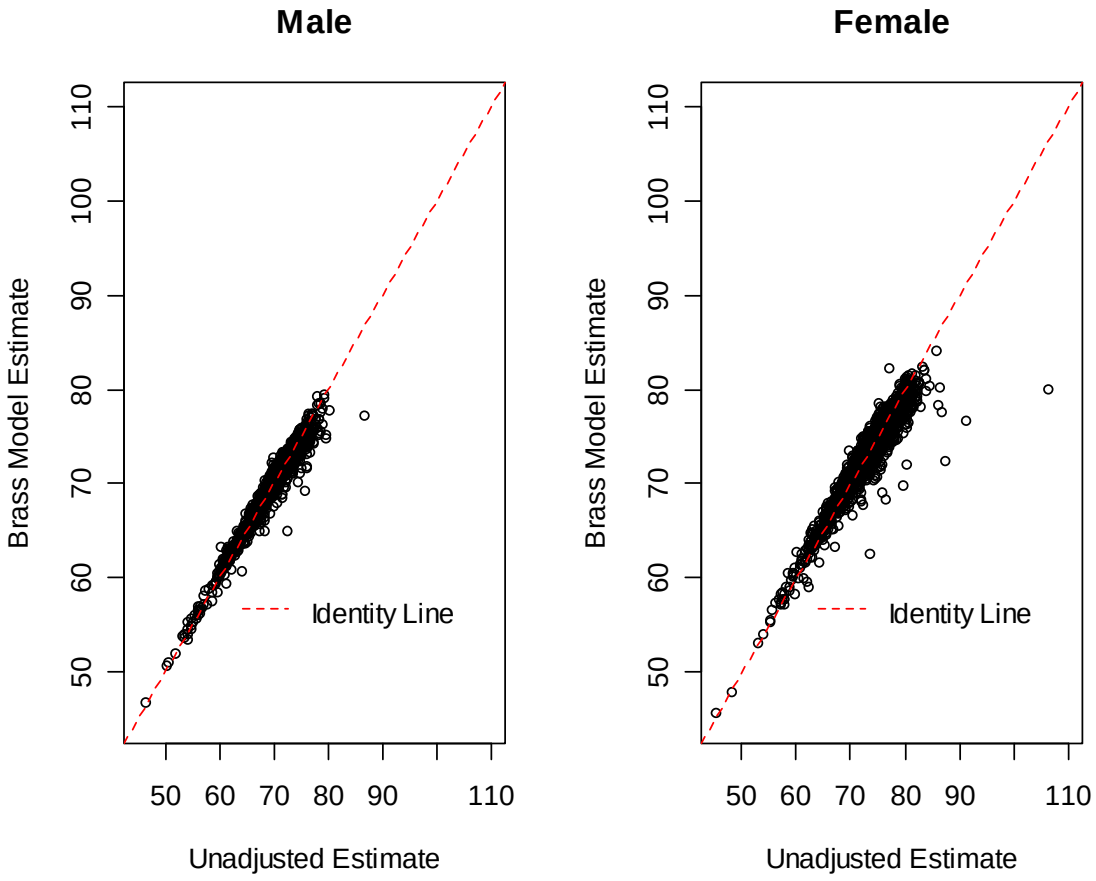

Figure 1. Life Expectancy at Birth: Brass Model Estimates vs. Raw Estimates

We observe a similar pattern between the raw and Brass model adjusted infant mortality, which is presented in Figure 2. The points in the figure are infant mortality rate calculated from raw data against their corresponding values calculated from the smoothed data. The dashed lines serve as a reference, representing an identity between the raw and smoothed values. Once again, the data suggests that the raw data is of good quality, or at least internally consistent.

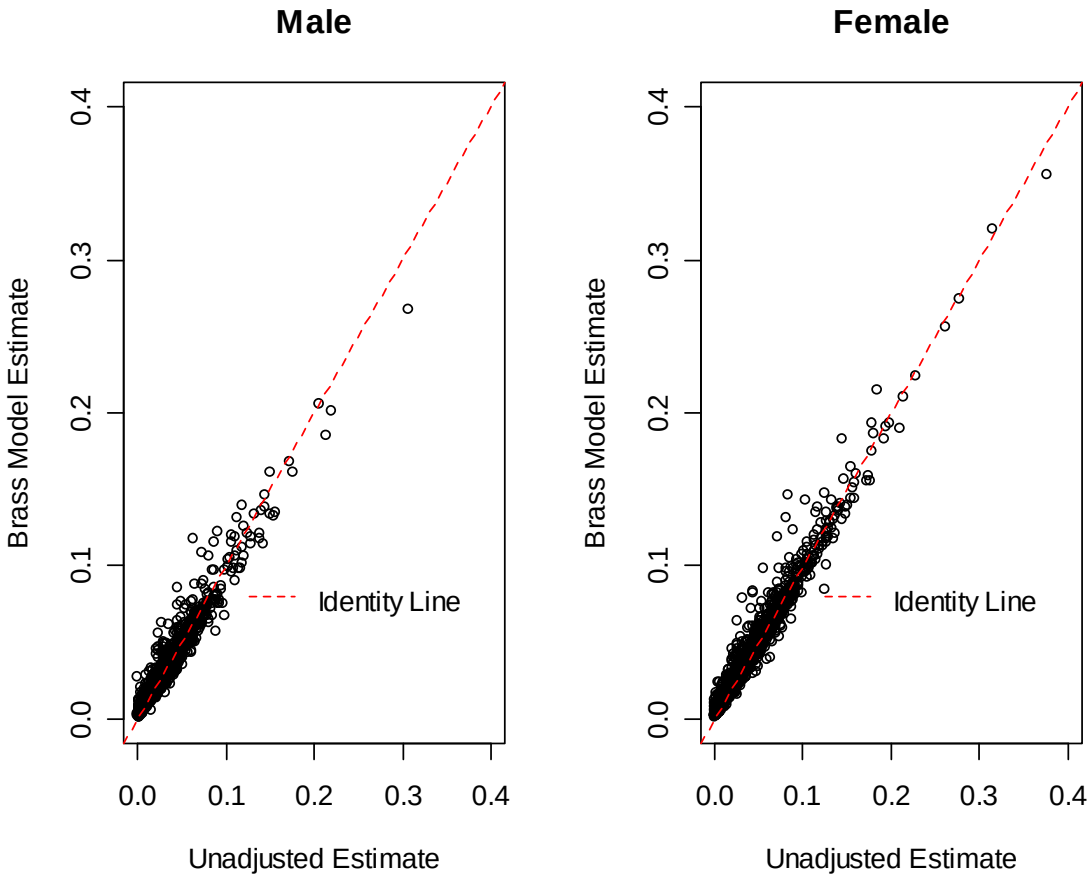

Figure 2. Infant Mortality Rate: Brass Model Estimates vs. Raw Estimates

The estimates of infant mortality using the empirical methods are very close to what we get from the Brass logit smoothing. Figure 3 displays the comparison of the infant mortality rates using empirical Bayes method and those based on Brass logit model smoothing. A regression line using the empirical Bayes estimate to predict the Brass logit model estimate is added to each plot, along with the identity line. It is rather obvious that the results from these two different methods are very close to each other. The regression line is almost identical to the identity line for female infant mortality, and off a small amount for male infant mortality. All the data points are clutter close to the regression line, with a correlation of .97 for male and .98 for female.

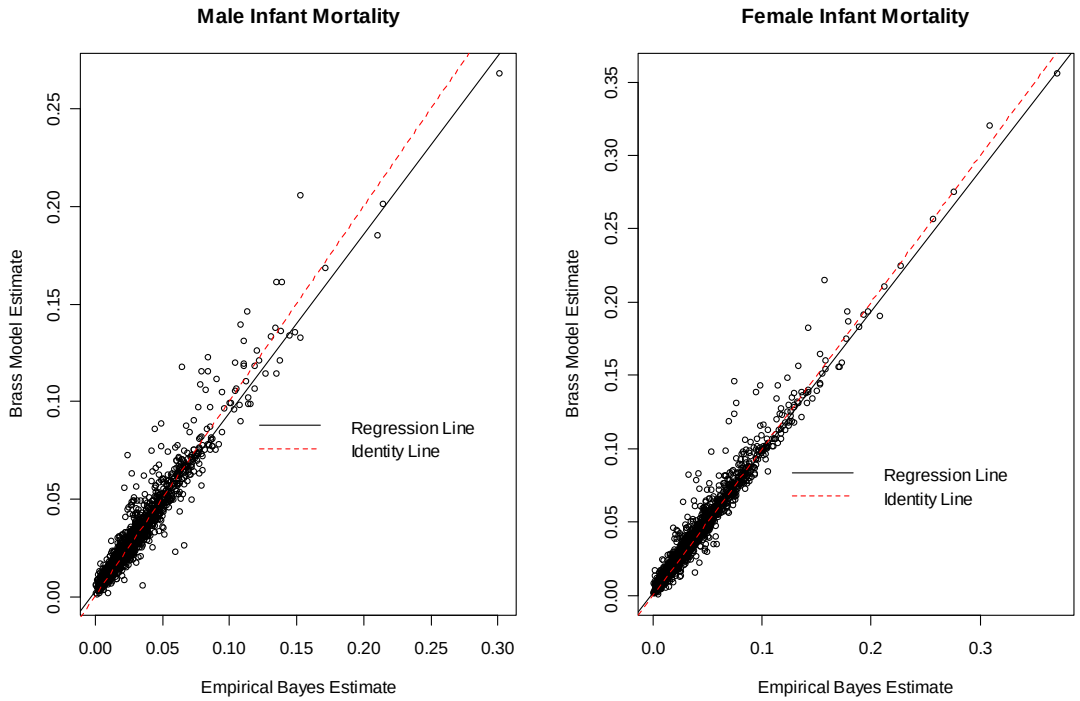

Figure 3. Infant Mortality Rates: Brass Model Estimates vs. Empirical Bayes Estimates

The good correspondence between the empirical Bayes based estimates of infant mortality and the Brass logit Model smoothed infant mortality rates indicates a good consistency in Chinese mortality data. The fact that these two sets of estimates are based on totally different aspects of mortality reporting—the Brass model is based on consistency across age, the Empirical Bayes method is based on spatial smoothness—but come out to have very close results alleviate some concerns of data quality. What left is the possibility that mortality is consistently underreported across all age groups as well as across administrative units, which is certainly possible, but less likely.

4. DISCUSSION

Using the population age structure and death data from the 2000 census, we have constructed sex-specific life tables for 2,367 county level administrative units of China. The results match well with official national and provincial life tables. Comparison of unadjusted life tables with the results of Brass Logit Model smoothing 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. Empirical Bayes method further confirms the overall data quality in Chinese mortality data.

The Brass Logit Model depends on the choice of the standard population. The life expectancies estimated from this exercise are likely to overestimate survival probabilities because we have used unadjusted provincial life tables as standard tables. The Coale-Demeny Model West tables are often considered as a good 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). One alternative is to use an iterative approach to derive an adjusted provincial standard life tables based on adjusted county level mortality estimates. Another alternative is to use the empirical Bayes method to estimate mortality for each specific mortality parameters and construct life tables based on those estimates. Both of those two alternative approaches are computationally expensive.

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 1,000 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, although it is possible there are small areas in China enjoy the low mortality level same as Japan's national average.

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. Constructing county level life tables addresses only the problem of population heterogeneity in high-level aggregates, but takes no account of population heterogeneity within counties. In particular, county level life tables 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.

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).