Original Article

Evaluation and Estimation of the Provincial Infant Mortality Rate in China's Sixth Census*

HU Song Bo[1], WANG Fang[2], and YU Chuan Hua[1,3,#]

1. Department of Epidemiology and Biostatistics, School of Public Health, Wuhan University, Wuhan 430071, Hubei, China; 2. School of Science, Jiujiang College, Jiujiang 332005, Jiangxi, China; 3. Global Health Institute, Wuhan University, Wuhan 430072, Hubei, China

Abstract

Objective To assess the data quality and estimate the provincial infant mortality rate ($_1q_0$) from China's sixth census.

Methods A log-quadratic model is applied to under-fifteen data. We analyze and compare the average relative errors (AREs) for $_1q_0$ between the estimated and reported values using the leave-one-out cross-validation method.

Results For the sixth census, the AREs are more than 100% for almost all provinces. The estimated average $_1q_0$ level for 31 provinces is 12.3‰ for males and 10.7‰ for females.

Conclusion The data for the provincial $_1q_0$ from China's sixth census have a serious data quality problem. The actual levels of $_1q_0$ for each province are significantly higher than the reported values.

Key words: Infant mortality rate; The sixth census; AREs

Biomed Environ Sci, 2015; 28(6): 410-420 *doi: 10.3967/bes2015.058* *ISSN: 0895-3988*

www.besjournal.com (full text) *CN: 11-2816/Q*

INTRODUCTION

The infant mortality rate (IMR, denoted $_1q_0$), defined as the ratio of the number of newborns less than one year of age who die divided by the number of live births in the same year, has an important place in mortality research. The IMR, which has previously been regarded a highly sensitive measure of population health[1], is also an important indicator for measuring population health[2] and reflects a country or region's economic and social development. However, IMRs widely vary due to under-reporting or misreporting of newborn births and deaths[3-4], especially in China and other underdeveloped and developing countries[3,5].

The sixth national census in China occurred between November 1, 2009 and October 30, 2010. According to the Chinese 2010 census data published by Chinese Statistics Press, China's $_1q_0$ is approximately 4‰ for both sexes combined. This rate is much less than the 13.1‰ IMR that was announced by the national health and family planning commission based on the 2010 cause of death surveillance system. However, the regions in the surveillance system do not span the entire country. Additionally, the surveillance data have problems with quality that result from a lack of statistical standardization, low management efficiency and difficult follow-up. Recent research on the sixth census data evaluation suggests that infant mortality is underreported in the sixth census and in the entire county, where the underreporting rate is

*This research is supported by a grant from the National Science Foundation of China: A Study on the Mortality Pattern of Chinese Population and Related Statistical Models (81273179); China's sixth census excluds the data of Hong Kong SAR, Macao SAR, and Taiwan.

#Correspondence should be adreessed to Professor YU Chuan Hua, PhD, Tel/Fax: 86-27-68759299, E-mails: yuchua@163.com; yuchua@whu.edu.cn

Biographical note of the first author: HU Song Bo, male, born in 1981, PhD, majors in methods and applications of biostatistics.

greater than 60%[5-6].

Census infant mortality statistics are retrospective surveys, which are prone to missing reports in China for three main reasons. First, respondents avoid reporting their child's death due to their sadness. Second, respondents subjectively conceal newborn's births and deaths because of China's one-child policy. The third reason is that statistical work is sometimes negligent because China has a large population and grass-roots investigators differ in professionalism. Additionally, respondents may have forgotten the infant's death due to the long reporting period.

China has a total of 6 censuses. The previous census in China also had missing death reports; however, the overall quality was acceptable. The missing deaths reporting problem was more serious in the sixth census than the fifth census. According to research on missing death rates, approximately 10% of deaths were underreported in the 2000 fifth census (8% by Zhang[7]; 10%-15% by Zhai[8]), which increased to 20% in the 2010 sixth census[9].

In this paper, we assess the quality of each province's infant mortality rate from China's sixth census and estimate the actual mortality rates. The rest of this paper is organized into the following sections. Section 2 provides an overview of the data used in this paper. Section 3 presents the log-quadratic model and the probability of dying between ages 1 and 15 (denoted ${}_{14}q_1$) to model the relationship with ${}_1q_0$. Section 4 assesses the quality of each province's ${}_1q_0$ from the sixth census by comparing the errors from the model that was described in Section 4. In Section 5, we estimate the actual levels of each province's ${}_1q_0$ from the sixth census. In Section 6, we discuss the results and provide conclusions.

In this paper, we present mortality indicators by sex, rather than for both sexes combined. This approach was chosen because most model life table systems that are used to adjust the ${}_{14}q_1$ value rely on mortality information by sex.

All calculations were completed using the R statistical programming. The average number of years lived for each age interval (x, x+n) by those dying during the age interval are obtained from the Coale and Demeny West region relationships[10].

METHOD

Data

The data used in this paper come from two primary sources. First, we collected 872 (for each sex) period life tables from the Human Mortality Database (HMD) that covered (mostly) 5-year time intervals[11]. The HMD is a publicly available dataset that is maintained by the University of California in Berkley, California, USA and the Max Planck Institute for Demography Research in Rostock, Germany. The HMD dataset has many advantages, which include high data quality as well as standardized age and time intervals. This dataset contains mortality profiles from 37, mostly developed, world regions, with the earliest tables dating back to the mid-18th century and the most recent from 2009, which represents more than 98 billion person-years of risk exposure. In this analysis, we used all under-fifteen mortality rates. The data are divided into 40 countries or areas and contains an aggregate under-fifteen population of approximately 7.9 billion person-years. All under-fifteen mortality rates in this collection were computed directly from observed deaths and population counts without adjustments. Table 1 summarizes the four sets of life tables that were used for this study. Table 2 provides the simple statistical result of ${}_1q_0$ that was used for this study, where the minimum value is 2.27‰ for males in Luxembourg and 1.62‰ for females in Iceland across the 2005-2009 time period.

The second and third data sources are the Chinese 2010 (the sixth) and 2000 census data (the fifth), which are published by Chinese Statistics Press. The ${}_1q_0$ by sex from 31 provinces in the two censuses is presented in Table 3, which shows that the infant mortality rate for the entire country decreased from 20.47‰ for males and 28.4‰ for females to approximately 4‰ for each sex during the decade. In the sixth census, most provinces have incredibly low values. For example, there are 7 areas where the ${}_1q_0$ for males is less than 2.5‰. In the Henan province, the ${}_1q_0$ is only 1.16‰ for males and 1.14‰ for females, which is much less than the minimum rate in the HMD dataset.

Methods

There are many statistical models that can be used to estimate infant mortality rates, including the Heligman-Pollard and Weibull models[12] or the Lee-Carte and Renshaw-Haberman models[13-14]. However, there is no accepted, effective and reliable method that is used to estimate ${}_1q_0$ in China because there is a lack of high quality data. The most commonly used method in China is implemented by the sampling strategy, in which the ${}_1q_0$ is adjusted,

Table 1. A Summary of the Life Tables Used in this Study

Country or Area	Year(s)	Number of Tables	Under-fifteen Population (Millions of Person-Years) and its Proportion (%)
Australia	1921-2009	18	89.75 (8.3)
Austria	1947-2009	13	31.06 (6.5)
Belgium	1841-2009	34	105.8 (8.5)
Bulgaria	1947-2009	13	36.24 (7.0)
Belarus	1960-2009	10	34.85 (7.3)
Canada	1921-2009	18	147.00 (8.3)
Switzerland	1876-2011	28	52.14 (7.8)
Chile	1992-2004	3	16.75 (8.9)
Czech Republic	1950-2011	13	42.54 (6.9)
Germany	1956-2009	26	329.84 (5.6)
Denmark	1835-2011	36	52.42 (8.9)
Spain	1908-2009	21	264.87 (8.4)
Estonia	1960-2009	10	4.72 (6.7)
Finland	2005-2009	27	45.93 (9.2)
France	1816-2009	78	1410.53 (8.7)
Northern Ireland	1922-2009	18	11.57 (9.0)
United Kingdom	1922-2009	18	338.88 (7.3)
Scotland	1855-2009	31	67.72 (9.5)
England and Wales	1841-2009	68	1109.44 (8.9)
Hungary	1950-2009	12	41.39 (6.8)
Ireland	1950-2009	12	18.35 (9.1)
Iceland	1838-2009	35	2.39 (10.4)
Israel	1983-2009	6	16.01 (10.4)
Italy	1872-2009	28	534.95 (8.9)
Japan	1947-2009	13	495.02 (7.2)
Lithuania	1960-2009	10	12.30 (7.3)
Luxembourg	1960-2009	10	1.23 (6.4)
Latvia	1960-2009	10	7.84 (6.5)
Netherlands	1850-2009	32	128.51 (9.3)
Norway	1846-2009	33	43.69 (9.2)
New Zealand	1901-2008	48	39.85 (9.2)
Poland	1958-2009	11	141.58 (7.7)
Portugal	1940-2009	14	52.96 (8.1)
Russia	1960-2009	10	485.79 (7.0)
Slovakia	1950-2009	12	23.80 (8.3)
Slovenia	1983-2009	6	2.92 (5.4)
Sweden	1751-2011	53	117.99 (9.3)
Taiwan,China	1970-2009	8	65.93 (8.4)
Ukraine	1960-2009	10	162.77 (6.7)
USA	1933-2009	16	1310.44 (8.2)
Total		872	7897.76 (8.1)

Note. (1) Life tables by sex are only counted once, and under-age five populations contain both sexes. Throughout table 1, there is a maximum of one life table per country-period. (2) Data for Germany contain east Germany, west Germany and the total population after 1990. Data for France, England and Wales contain the civilian and total populations. Data for New Zealand refer to the non-Maori population and the full national population and Maori population after 1950. (3) All under-five mortality rates have been computed from directly observed deaths and population counts without adjustment. Source: Human Mortality Database, www.mortality.org (accessed: March 14, 2014).

Table 2. A Summary of $_{1}q_{0}$ in the HMD Dataset (‰)

Sex	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Male	2.27	11.98	30.19	65.54	112.20	360.60
Female	1.62	9.38	23.66	54.67	94.41	314.40

Table 3. The Reported $_{1}q_{0}$ Value from the Fifth and Sixth National Censuses in China (‰)

Area or Province	Male		Female	
	Fifth	Sixth	Fifth	Sixth
Total	20.47	4.27	28.4	4.44
Beijing	3.64	1.34	3.66	1.26
Tianjin	4.03	1.81	4.03	1.17
Hebei	15.01	2.87	20.65	2.63
Shanxi	15.19	4.25	18.05	4.58
Inner Mongolia	26.77	4.86	30.08	3.72
Liaoning	9.37	3.07	9.81	2.34
Jilin	15.59	1.76	15.98	1.48
Heilongjiang	9.53	1.75	8.5	1.31
Shanghai	4.07	3.7	4.42	2.62
Jiangsu	10.89	2.33	12.88	2.04
Zhejiang	10.09	4.29	11.75	3.98
Anhui	21.77	4.8	34.05	4.91
Fujian	14.49	3.45	21.38	3.81
Jiangxi	25.92	3.9	60.93	4.8
Shandong	12.7	2.07	15.71	2.1
Henan	15.72	1.16	24.89	1.14
Hubei	14.79	3.57	19.28	3.67
Hunan	19.89	2.58	27.21	2.61
Guangdong	11.11	2.9	18.59	3.15
Guangxi	17.81	3.89	31.91	4.07
Hainan	12.74	5.48	23.71	7.31
Chongqing	20.64	4.03	20.78	3.57
Sichuan	19.38	3.86	20.29	3.59
Guizhou	51.01	14.61	64.51	17.62
Yunnan	52.61	14.33	67.55	15.82
Tibet	37.82	13.67	36.7	13.7
Shanxi	23.25	2.14	35.34	2.05
Gansu	35.3	7.47	48.2	9.25
Qinghai	37.67	12.72	40.12	11.55
Ningxia	22.87	11.51	22.09	10.41
Xinjiang	28.2	9.04	24.98	7.61

including the under-reporting rates, capture-recapture methods and so on. These methods are disadvantaged due to implementation challenges and result in the final estimates being strongly affected by the sampling strategy design.

This paper used the log-quadratic model to estimate the infant mortality rate, which is as follows:

$$y=a_0+a_1x+a_2x^2+\varepsilon \quad (1)$$

where y=log($_{1}q_{0}$), x=log($_{14}q_{1}$), and ε is an error term. This log-quadratic model was also used by Wilmoth and colleagues to generate a model life table system[15]. We use this model to fit different levels from the Human Mortality Database and estimate the infant mortality rate from the fifth and sixth census data that were obtained from 31 provinces in China. By comparing the error terms from the two censuses, we assess the infant mortality data quality from the sixth national census in China and estimate its actual levels based on the adjusted $_{14}q_{1}$ value.

We use $_{14}q_{1}$ to model the relationship with $_{1}q_{0}$ for two reasons. First, $_{14}q_{1}$ is a predictor that is used to estimate $_{1}q_{0}$ because they are strongly correlated. Table 4 shows the Pearson correlation coefficients between $_{1}q_{0}$ and the probability of dying age intervals (1,x) (x=5, 10, …, 60) in a log-log scale using the life tables at different levels of $_{1}q_{0}$. We also highlight the largest correlation coefficients for each level (i.e. row). We found that the maximum correlation coefficient for each sex is always between log($_{1}q_{0}$) and log($_{14}q_{1}$) at all levels of $_{1}q_{0}$ except for level '$_{1}q_{0}$<1', where the maximum value, which is between log($_{1}q_{0}$) and log($_{4}q_{1}$), is very close to the value between log($_{1}q_{0}$) and log($_{14}q_{1}$) for each sex. Hence, the $_{1}q_{0}$ can be interpreted well by $_{14}q_{1}$. Second, we used $_{14}q_{1}$ rather than $_{4}q_{1}$ or $_{9}q_{1}$, which also interprets $_{1}q_{0}$ well, because the $_{14}q_{1}$ is more relatively robust and reliable. In general, the census data quality problems are mainly due to data inaccuracies for child and elderly populations.

RESULTS

Evaluation of the Quality of the Sixth Census

The model is used for assessing the under-fifteen mortality data quality in a given area by comparing and analyzing the average relative errors

(AREs) of the estimated ${}_1q_0$ values with the AREs of the HMD dataset, which contains several different under-fifteen mortality patterns.

The AREs between the predicted and observed values were calculated for every area included in the HMD dataset using the leave-one-out cross-validation method. That is, each country or area is deleted from the HMD dataset in turn, and the log-quadratic model that is fitted to the remaining dataset is used to predict the ${}_1q_0$ value for the deleted area. Table 2 shows that the median of ${}_1q_0$ is 30.19‰ for males and 23.66‰ for females, which means that half of the tables have ${}_1q_0$ values that are greater than approximately 30‰, which is inconsistent with the social development and economies in most developed and developing countries in 2010. Hence, we also considered the results from the different ${}_1q_0$ levels. The resulting AREs for ${}_1q_0$ are presented in Table 5 by sex, where the ${}_1q_0$<1 means that all life tables from the HMD data ware used, while the ${}_1q_0$≤0.03 means that only those life tables that satisfied ${}_1q_0$≤0.03 were used instead of all the life tables.

First, we found that Taiwan, China, whose ARE values for both sexes are outliers according to Grubb's criteria (*P*-value<0.001 for both sexes), has poor infant mortality information quality, which has been validated by Chen LM et al.[16]. Second, Hungary, Japan and Poland have relatively larger ARE values for both cases. Norway has a maximum ARE for case ${}_1q_0$<1 but only approximately 20% for case ${}_1q_0$≤0.03. Excluding Taiwan, China, the maximum AREs are approximately 54% for both sexes when ${}_1q_0$≤1, but approximately 31% for males and 37% for females when ${}_1q_0$≤0.03. In addition, the average AREs for all countries, excluding Taiwan China, are 21.6% for males and 20.9% for females when ${}_1q_0$<1; and 19.1% for males and 19.7% for females when ${}_1q_0$≤0.03.

We now use a log-quadratic model to fit the under-fifteen mortality data from the HMD and estimate the infant mortality rate from the fifth and sixth national censuses in China. The life tables from two different levels from the HMD are selected to train the log-quadratic model. Tables 6 for males and 7 for females presents the resulting AREs, where level 1 reflects ${}_1q_0$<1, i.e., all data are included. Level 2 is more complex, as each province in China was estimated independently. In level 2, for each province, life tables from the HMD were selected when their ${}_1q_0$s were less than this province's reported ${}_1q_0$+0.01 and more than its ${}_1q_0$-0.01 to fit the model and estimate the provincial ${}_1q_0$.

For fifth census, we find that the AREs for the entire country, especially for level 2, are consistent with Table 5 for both levels and both sexes. Three provinces, Beijing, Tianjin and Shanghai, have apparently larger AREs, which appears to be abnormal. Moreover, there are more abnormal areas for males compared to females. It is concluded

Table 4. Correlation Coefficients Between ${}_1q_0$ and the Probabilities of Dying Age Intervals (1, x) in a Log-log Scale Calculated for Different Levels of ${}_1q_0$

Levels	Age 1-5	Age 1-10	Age 1-15	Age 1-20	Age 1-25	Age 1-30
Male						
${}_1q_0$<1	**0.9715**	0.9716	0.9708	0.9611	0.9496	0.9411
${}_1q_0$≤0.010	0.7674	0.7891	**0.8041**	0.7796	0.7247	0.6854
${}_1q_0$≤0.015	0.8456	0.8582	**0.8689**	0.8439	0.7865	0.7322
${}_1q_0$≤0.020	0.8857	0.8956	**0.9009**	0.8783	0.8197	0.7616
${}_1q_0$≤0.025	0.9054	0.9131	**0.9165**	0.8908	0.8319	0.7710
${}_1q_0$≤0.030	0.9133	0.9191	**0.9212**	0.8923	0.8317	0.7697
Female						
${}_1q_0$<1	**0.9748**	0.9735	0.9720	0.9672	0.9646	0.9637
${}_1q_0$≤0.010	0.8165	0.8257	**0.8364**	0.8182	0.7857	0.7706
${}_1q_0$≤0.015	0.8851	0.8934	**0.8977**	0.8857	0.8632	0.8479
${}_1q_0$≤0.020	0.9005	0.9065	**0.9099**	0.8989	0.8788	0.8653
${}_1q_0$≤0.025	0.9124	0.9166	**0.9182**	0.9051	0.8862	0.8741
${}_1q_0$≤0.030	0.9185	0.9211	**0.9216**	0.9068	0.8876	0.8761

Note. (1) The results are calculated from the HMD dataset. (2) The '${}_1q_0$<1' means that all life tables from the HMD were used to calculate the correlation coefficients. Similarly, the '${}_1q_0$≤0.010' means that the life tables that satisfied the condition '${}_1q_0$≤0.010' were used.

Table 5. The AREs for the Predicted and Observed Values Using the Leave-one-out Cross-validation Method Across $_1q_0$ Levels (%)

Country or Area	$_1q_0<1$		$_1q_0\le 0.03$	
	Male	Female	Male	Female
Australia	14.5	15.0	7.9	4.6
Austria	31.4	30.2	23.7	22.8
Belgium	16.8	16.1	16.4	16.6
Bulgaria	19.3	18.0	13.6	16.9
Belarus	21.4	16.7	15.9	18.2
Canada	17.6	14.5	16.1	18.2
Switzerland	21.2	20.2	14.6	13.2
Chile	3.8	10.5	6.5	9.4
Czech Republic	20.1	23.9	22.8	21.2
Germany	18.8	20.9	14.3	16.3
Denmark	16.7	14.4	18.1	19.9
Spain	10.6	6.7	19.1	24.1
Estonia	30.6	22.8	22.8	22.9
Finland	18.1	17.9	16.6	12.3
France	10.7	9.5	17.5	20.0
Northern Ireland	16.6	21.0	20.3	29.7
U.K.	15.8	22.0	25.3	20.8
Scotland	20.5	22.2	19.7	22.2
England and Wales	9.7	11.0	26.2	20.6
Hungary	44.0	44.3	30.8	37.1
Ireland	23.6	19.2	17.7	21.6
Iceland	31.0	34.8	19.9	18.1
Israel	13.2	9.6	14.9	8.5
Italy	6.1	9.3	29.3	31.6
Japan	38.7	35.6	29.0	28.6
Lithuania	30.7	20.7	22.1	21.2
Luxembourg	15.7	10.3	16.7	10.9
Latvia	44.8	28.3	29.9	26.3
Netherlands	21.6	23.1	15.2	21.6
Norway	54.0	54.6	19.6	22.1
New Zealand	15.0	19.8	12.1	15.6
Poland	42.1	43.3	31.4	33.0
Portugal	10.3	10.0	27.8	24.0
Russia	21.2	14.0	9.8	11.3
Slovakia	25.8	27.0	23.5	19.3
Slovenia	11.4	12.6	13.8	11.8
Sweden	13.2	18.6	4.9	5.3
Taiwan, China	117.2	99.3	98.4	96.3
Ukraine	35.7	35.9	21.1	33.7
USA	10.8	11.2	16.1	16.7
Average	24.0	22.9	21.0	21.6

Note. Results are calculated from the HMD dataset. The '$_1q_0$<1' and '$_1q_0$≤0.03' means are explained above the table.

that, first, the fifth census has a problem with missing death reports; however, the overall data are relatively reliable, which is consistent with the related study. Second, China's mortality rates in the 2000 fifth census, specifically at most provincial levels confirm the HMD data. The inconsistent areas, Beijing, Tianjin and Shanghai, may not confirm the HMD experience, but we ascribe this to the problems with data quality, which may be correlated with China's one-child policy. In addition, the AREs for level 2 are significantly lower than level 1, which indicates that the pattern of under-fifteen mortality differs across mortality levels.

In the sixth census, for the entire country and each province, males and females, and level 1 and level 2, almost all the AREs are more than 100% and only the AREs for Guizhou and Hainan provinces appear normal for females. For each province and the entire country, the MRE values for both sexes and levels are significantly higher in the sixth census compared to the fifth census, even for the Guizhou and Hainan provinces. We believe that this large performance gap over the past decade cannot solely be explained by changes in under-fifteen mortality patterns or different HMD experiences; thus, it is important to acknowledge poor data quality problems. In addition, we can see from Tables 6 and 7 that the average ARE for level 1 is always more than that for level 2 for each sex, which indicates that the under- fifteen mortality rate patterns vary with the mortality levels.

The Actual Level Estimations for Each Provincial $_1q_0$ in the Sixth Census

In section 4, we found that almost all provincial under-fifteen data from China's sixth census, compared to the fifth census, have problematic data quality. In this section, we use the log-quadratic model to estimate actual levels in the sixth census for each provincial $_1q_0$, except for the Guizhou and Hainan provinces for females. The results will be affected by the quality of $_{14}q_1$; however, completely accurate population and mortality data are difficult to acquire in China. In general, the census data quality problems are due to inaccuracies in child and elderly populations, especially for infants, while the data are more robust for those aged 5-60. Hence, the estimations for $_1q_0$ in this section are based on the following two assumptions from the sixth census: (1) the quality of $_{14}q_1$ is relatively accurate, and only requires a small adjustment or no adjustment rather than substantial changes; and (2) data quality for those aged 15-60 is more accurate and reliable.

Table 6. ARE Comparisons Between Predicted and Reported Values from the Fifth and Sixth National Censuses in China for Males (%)

Area or Province	Fifth Census		Sixth Census	
	level 1	level 2	level 1	level 2
Total	29.1	11.0	229.8	154.8
Beijing	207.3	157.3	259.6	275.4
Tianjin	161.5	125.6	189.9	198.1
Hebei	7.3	2.4	243.8	197.6
Shanxi	16.2	4.1	127.1	102.4
Inner Mongolia	43.6	22.3	77.7	65.6
Liaoning	88.9	52.9	201.8	168.5
Jilin	13.5	16.5	262.0	256.0
Heilongjiang	18.1	8.6	271.0	263.0
Shanghai	151.3	119.1	97.6	89.8
Jiangsu	65.0	37.8	325.5	265.6
Zhejiang	109.1	62.0	204.6	143.1
Anhui	6.9	0.2	145.5	107.4
Fujian	38.2	18.4	176.9	145.5
Jiangxi	17.1	1.4	327.9	198.8
Shandong	34.7	16.1	282.3	254.8
Henan	26	10.9	657.2	562.2
Hubei	68.9	36.7	244.1	176.9
Hunan	51.2	20.0	513.7	319.0
Guangdong	87.1	50.3	235.3	191.9
Guangxi	44.8	18.4	257.8	173.8
Hainan	77.0	42.5	161.7	107.8
Chongqing	78.7	24.3	348.8	203.1
Sichuan	108.4	32.4	451.2	233.1
Guizhou	6.2	2.4	87.7	47.9
Yunnan	8.0	2.8	129.2	70.5
Tibet	67.9	15.9	203.2	107.4
Shaanxi	3.3	5.9	439.5	330.4
Gansu	25.0	8.2	138.8	82.3
Qinghai	4.1	0.5	311.1	140.6
Ningxia	35.8	10.4	122.5	68.8
Xinjiang	43.3	11.1	190.1	96.4
Average	54.5	29.6	247.3	181.2

Note. Levels 1 and 2 represent selecting two different under-fifteen dataset levels from the HMD to fit the log-quadratic model and estimate the ${}_1q_0$ for China's censuses. (1) Level 1 represents the condition '${}_1q_0<1$'; (2) in level 2, we fit the model independently for each province, where the life tables are selected to fit the model when the ${}_1q_0$s are less than the province's reported ${}_1q_0+0.01$, and more than its ${}_1q_0-0.01$.

Table 7. ARE Comparisons Between Predicted and Reported Values from the Fifth and Sixth National Censuses in China for Females (%)

Area or Province	Fifth Census		Sixth Census	
	level 1	level 2	level 1	level 2
Total	18.8	8.4	141.0	114.6
Beijing	133	116.1	160.1	157.2
Tianjin	83.9	75.0	212.0	209.5
Hebei	48.3	34.3	158.1	147.3
Shanxi	17.8	14.8	58.5	51.7
Inner Mongolia	57.6	26.3	83.8	76.7
Liaoning	50.3	38.5	170.8	161.4
Jilin	34.0	29.7	248.2	241.0
Heilongjiang	8.4	5.3	306.9	296.6
Shanghai	56.9	51.0	130.3	124.1
Jiangsu	12.5	7.6	235.1	216.3
Zhejiang	29.3	21.6	149.9	124.6
Anhui	43.9	14.2	56.0	49.4
Fujian	22.6	16.4	101.7	90.6
Jiangxi	52.8	2.1	165.5	129.6
Shandong	12.8	10.3	164.6	157.4
Henan	36.9	20.5	383.1	369.0
Hubei	5.0	0.6	144.4	124.3
Hunan	4.9	2.6	337.5	267.9
Guangdong	7.6	7.6	142.4	128.3
Guangxi	27.6	10.4	158.1	129.0
Hainan	16.7	10.6	38.4	31.9
Chongqing	47.7	22.6	249.8	191.4
Sichuan	65.0	31.5	345.4	236.2
Guizhou	17.3	0.3	18.0	10.1
Yunnan	37.8	3.1	65.3	38.6
Tibet	61.4	8.0	193.8	135.3
Shaanxi	46.3	13.1	393.6	321.1
Gansu	46.0	12.1	85.6	63.4
Qinghai	11.8	4.9	285.3	155.7
Ningxia	17.2	4.3	83.6	61.7
Xinjiang	37.8	15.2	164.9	110.2
Average	36.6	20.0	176.0	147.6

Note. The level 1 and level 2 means are explained above the table.

The Adjustment of $_{14}q_1$

There are two questions that must be answered about the provincial $_{14}q_1$ adjustment. First, which province needs to be adjusted? Second, what methods can be used to adjust? We answer these two questions in this subsection.

The mortality rate usually reflects social and economic development as well as the health services. When an area is socioeconomically developed, people's incomes are higher and the mortality level is lower. In contrast, public finance government expenditures on medical health rarely affects health and mortality risk, even with high investments. Thus, whether or not $_{14}q_1$ is adjusted is based on the assumption that the $_{14}q_1$ value is negatively interrelated to per capita GDP and total health expenditures. Table 8 lists the types of per capita GDP, total health expenditures (THE) and the $_{14}q_1$ values for 31 provinces in China in 2010, where the $_{14}q_1$ is ranked from low to high and the other two indices are ranked from high to low. There should be few differences between each $_{14}q_1$ ranking and the other two rankings. In this study, $_{14}q_1$ will not be adjusted if its ranking falls between the rankings of per capita GDP and THE. We are cautious to adjust based on this assumption (1). The data that are reported for $_{14}q_1$ are also available for the difference between two rankings in the corresponding range; thus, we will not adjust the $_{14}q_1$s if the rankings satisfy the following condition:

min[rank (PGDP), rank (THE)]-2≤rank ($_{14}q_1$) ≤max [rank (PGDP), rank (THE)]+2 (2)

The results for adjustments by sex are presented in Table 9. There are 9 provinces that need to be adjusted: Shanxi, Jilin, Heilongjiang, Jiangsu, Zhejiang, Hubei, Guangdong, Hainan, and Gansu for males. For females, Gansu and Hainan do not need adjustments because the $_1q_0$ does not need to be estimated. These results answer the first question.

The second question is how to adjust the mortality rates. The Clark model life table[17], which is the latest life table system, will be used to adjust the $_{14}q_1$ values based on the previous assumption (2). The Clark system outperforms existing model life table systems in estimating errors[17]. Its calculation procedure can be implemented with the 'LifeTables' package in R, which implements a HMD-calibrated model in the R statistical software package[18]. We do not use the Coale-Demeny [10] or UN model life tables for developing countries[19], which were the most widely used model life table systems for the past three decades because the Clark system can reflect contemporary mortality experiences, including extremely low mortality $_1q_0$ levels. The Coale-Demeny and UN model life table systems, which were constructed using early data, do not contain the lower mortality levels' model life tables or cover the entire range of epidemiological situations and trajectories between $_1q_0$ and $_{14}q_1$[20].

There are five mortality patterns in the Clark system and each pattern contains some model life tables with different mortality levels. The detailed adjustment method for each province is as follows, with the Jilin province as an example. For each Clark system pattern, we find a model life table for the Jilin province life table, such that it has the minimum modeled and reported ARE values for the mortality rate among those aged 15-60. So there are 5 different pattern's model life tables we found for Jilin, but we do not have an accurate child mortality rate to discriminate the Jilin pattern[17]. The $_{14}q_1$, which is closest to the reported value, is selected from these 5 different model life table patterns as the adjusted value for the Jilin province. The results are presented in Table 9 and include the reported and adjusted values as needed.

Estimated $_1q_0$ Results

A log-quadratic model is applied to fit the HMD data, excluding Taiwan China, and estimates the 31 provincial infant mortality rates by sex in China's sixth national census. To acquire good estimations for each province, the method is similar to the previous level 2 that was used in Tables 6 and 7. For each province, those life tables whose $_1q_0$ was less than the reported $_1q_0$+0.015 and more than its $_1q_0$-0.005 are selected from the HMD as training data to fit the model and estimate the $_1q_0$ according to the adjusted value of $_{14}q_1$. This asymmetric choice of training data is due to the underestimated reporting levels of $_1q_0$.

The results are presented in Table 10. There are significant increases for each provincial estimated $_1q_0$ value compared to the reported value, especially for the Henan province, where the estimated value increases to almost 7 times for males and 5 times for females. The estimated $_1q_0$ minimum is 5.0‰ for males and 3.2‰ for females in Beijing. The maximum is 27.1‰ for males and 29.4‰ for females in Tibet. The estimated average $_1q_0$ value for the 31 provinces is 12.3‰ for males and 10.7‰ for females, which is more than 2 times larger than the reported average values for both sexes and is slightly below the levels for the entire country that were announced by the national health and family planning commission in 2010.

DISCUSSIONS

This paper used log-quadratic models to assess the under-fifteen mortality rate data quality from China's 2010 sixth census. The ${}_{14}q_1$ is selected to model the relation to ${}_{1}q_0$ by comparing correlation coefficients from the Human Mortality Database. Different HMD infant mortality levels are used to train the log-quadratic models and estimate the ${}_{1}q_0$ for China's fifth and sixth censuses. We found that the AREs for ${}_{1}q_0$ in the 2000 fifth census performs well, which is consistent with the overall reliable data. We also found that the 2010 sixth census performs worse compared to the fifth census for almost all provinces; this finding suggests that there is a serious data quality problem. In addition, we confirmed that Taiwan, China has poor infant mortality information available in the HMD by using the leave-one-out cross-validation method at different ${}_{1}q_0$ levels and comparing their AREs.

Table 8. The Index Rankings and Results for Adjusting the ${}_{14}q_1$

Area or Province	**PGDP**	**THE**	**Male**		**Female**	
			${}_{14}q_1$	Adjust	${}_{14}q_1$	Adjust
Beijing	2	8	1	F	1	F
Tianjin	3	24	2	F	2	F
Hebei	12	7	13	F	9	F
Shanxi	18	19	**11**	T	**12**	T
Inner Mongolia	6	22	7	F	10	F
Liaoning	8	11	9	F	8	F
Jilin	11	18	**3**	T	**3**	T
Heilongjiang	16	14	**4**	T	**4**	T
Shanghai	1	10	5	F	7	F
Jiangsu	4	3	**14**	T	**11**	T
Zhejiang	5	4	**18**	T	**17**	T
Anhui	26	13	16	F	14	F
Fujian	10	17	10	F	15	F
Jiangxi	24	21	22	F	23	F
Shandong	9	2	6	F	6	F
Henan	21	5	8	F	5	F
Hubei	13	12	**17**	T	**16**	T
Hunan	20	9	21	F	21	F
Guangdong	7	1	**12**	T	**13**	T
Guangxi	27	15	19	F	20	F
Hainan	23	28	**20**	T	**19**	F
Chongqing	14	23	24	F	22	F
Sichuan	25	6	25	F	24	F
Guizhou	31	25	28	F	28	F
Yunnan	28	16	29	F	29	F
Tibet	29	31	30	F	30	F
Shaanxi	15	20	15	F	18	F
Gansu	30	27	**23**	T	**25**	T
Qinghai	22	30	31	F	31	F
Ningxia	17	29	26	F	26	F
Xinjiang	19	26	27	F	27	F

Note. Source: Chinese Statistical Yearbook 2011; Chinese Health Statistics Yearbook 2011; Research Report of Chinese total health expenses 2011. T indicates a need for adjustment; F indicates no adjustments are needed.

Table 9. Comparisons for Reported and Adjusted $_{14}q_1$ by Sex

Area or Province	Male		Female	
	Reported	Adjusted	Reported	Adjusted
Beijing	0.00253	0.00253	0.00162	0.00162
Tianjin	0.00271	0.00271	0.00177	0.00177
Hebei	0.00463	0.00463	0.00302	0.00302
Shanxi	0.00454	0.00408	0.00321	0.00304
Inner Mongolia	0.00412	0.00412	0.00304	0.00304
Liaoning	0.00438	0.00438	0.00284	0.00284
Jilin	0.00318	0.00322	0.00237	0.00229
Heilongjiang	0.00323	0.00399	0.00244	0.00223
Shanghai	0.00357	0.00357	0.00272	0.00272
Jiangsu	0.00465	0.00380	0.00304	0.00313
Zhejiang	0.00597	0.00350	0.00430	0.0029
Anhui	0.00543	0.00543	0.00337	0.00337
Fujian	0.00450	0.00450	0.00338	0.00338
Jiangxi	0.00752	0.00752	0.00547	0.00547
Shandong	0.00382	0.00382	0.00253	0.00253
Henan	0.00418	0.00418	0.00251	0.00251
Hubei	0.00564	0.00443	0.00390	0.0037
Hunan	0.00715	0.00715	0.00491	0.00491
Guangdong	0.00457	0.00418	0.00336	0.00279
Guangxi	0.00633	0.00633	0.00453	0.00453
Hainan	0.00651	0.00451	0.00437	0.00437
Chongqing	0.00813	0.00813	0.00536	0.00536
Sichuan	0.00955	0.00955	0.00688	0.00688
Guizhou	0.01241	0.01241	0.00909	0.00909
Yunnan	0.01508	0.01508	0.01174	0.01174
Tibet	0.01962	0.01962	0.01977	0.01977
Shaanxi	0.00533	0.00533	0.00437	0.00437
Gansu	0.00802	0.00695	0.00741	0.00677
Qinghai	0.02594	0.02594	0.02252	0.02252
Ningxia	0.01155	0.01155	0.00830	0.00830
Xinjiang	0.01184	0.01184	0.00879	0.00879

Table 10. Comparisons for Reported and Estimated $_1q_0$ by Sex (‰)

Area or Province	Male		Female	
	Reported	Estimated	Reported	Estimated
Beijing	1.3	5.0	1.3	3.2
Tianjin	1.8	5.4	1.2	3.6
Hebei	2.9	9.2	2.6	6.7
Shanxi	4.3	8.2	4.6	6.8
Inner Mongolia	4.9	8.4	3.7	6.8
Liaoning	3.1	8.7	2.3	6.3
Jilin	1.8	6.4	1.5	4.9
Heilongjiang	1.8	8.0	1.3	4.7
Shanghai	3.7	7.2	2.6	6.0
Jiangsu	2.3	7.6	2.0	7.0
Zhejiang	4.3	7.1	4.0	6.4
Anhui	4.8	10.9	4.9	7.6
Fujian	3.5	8.9	3.8	7.6
Jiangxi	3.9	13.9	4.8	12.3
Shandong	2.1	7.6	2.1	5.5
Henan	1.2	8.3	1.1	5.4
Hubei	3.6	8.8	3.7	8.3
Hunan	2.6	13.1	2.6	10.8
Guangdong	2.9	8.3	3.2	6.2
Guangxi	3.9	12.1	4.1	10.2
Hainan	5.5	9.2	7.3	-
Chongqing	4.0	14.8	3.6	11.8
Sichuan	3.9	16.5	3.6	14.5
Guizhou	14.6	23.2	17.6	-
Yunnan	14.3	25.3	15.8	23.4
Tibet	13.7	27.1	13.7	29.4
Shaanxi	2.1	10.3	2.1	9.7
Gansu	7.5	13.8	9.3	15.6
Qinghai	12.7	24.4	11.6	28.3
Ningxia	11.5	21.4	10.4	18.5
Xinjiang	9.0	21.6	7.6	19.0
Average	5.1	12.3	5.2	10.7

To estimate each province's actual infant mortality rates in China's 2010 sixth census, we make simple adjustments to the ${}_{14}q_1$. By comparing 31 provincial rankings on three indices: ${}_{14}q_1$, per capita GDP and total health expenditures, we determine which provinces need adjustments. Moreover, we use the Clark model life table system to adjust ${}_{14}q_1$. Finally, the actual infant mortality levels for each province were estimated using the log-quadratic model and adjusted ${}_{14}q_1$. The results show that the actual infant mortality level for each province is significantly increased compared to the reported levels. We found that the average ${}_{1}q_0$ for 31 provinces is 12.3‰ for males and 10.7‰ for females.

However, the actual level estimates are based on the assumptions that the Chinese sixth census data for individuals aged 1-60 is relatively accurate and reliable. In contrast, we use the HMD data to train the log-quadratic model and estimate the infant mortality rate in China's census. The HMD dataset mainly reflects OECD historical and contemporary experiences, even though the AREs perform well when applied to the fifth census. There is a need for a systematic evaluation and the application of other methods for the ${}_{1}q_0$ in China's sixth census.

Note. The fifth and sixth national censuses in China mentioned above exclude the data of Hong Kong SAR, Macao SAR, and Taiwan.

Received: January 23, 2015;
Accepted: May 14, 2015

REFERENCES

1. Blaxter M. The health of children: a review of research on the place of health in cycles of disadvantage. Heinemann Educational: London, 1981; p272.
2. Reidpath DD, Allotey P. Infant mortality rate as an indicator of population health. J Epidemiol Community Health, 2003; 57, 344-6.
3. Huy TQ, Johansson A, Long NH. Reasons for not reporting deaths: a qualitative study in rural Vietnam. World Health Popul, 2007; 9, 14-23.
4. Aleshina N, Redmond G. How high is infant mortality in central and eastern Europe and the commonwealth of independent states? Popul Stud (Camb), 2005; 59, 39-54.
5. Rongqing H, Xianxin Z. Infant Mortality Reported in the 2010 Census: Bias and Adjustment. Population Research, 2013; 37, 3-16. (In Chinese)
6. Menghan Z, Fan Y. Infant Mortality and Under-five Mortality Rate in China's 2010 Census: An Assessment. Population Research, 2013; 37, 68-80. (In Chinese)
7. Weiming Z, Hongyan C. An Evaluation of Data Accuracy of the 2000 Population Census of China. Population Research, 2003; 27, 25-35. (In Chinese)
8. Dehua Z. Indirect Estimation on Population Mortality Based on the Fifth Census in China. Population & Economics, 2003; 140, 31-65. (In Chinese)
9. Hongyan C, Lan X, Rui L. An Evaluation of Data Accuracy of the 2010 Population Census of China. Population Research, 2013; 37, 10-21. (In Chinese)
10. Coale AJ, Demeny P, Vaughan B. Regional model life tables and stable populations. 2nd ed.ed.; Academic Press: New York, 1983; p 496.
11. University of California, Berkeley and Max Planck Institute for Demographic Research. Human Mortality Database. URL http://www.mortality.org. (accessed in 2014).
12. Heligman L, Pollard JH. The age pattern of mortality. J of the Institute of Actuaries, 1980; 107, 49-80.
13. Carter LR, Lee RD. Modeling and forecasting U.S. sex differentials in mortality. Int J Forecast, 1992; 8, 393-411.
14. Renshaw AE, Haberman S. A cohort-based extension to the Lee-Carter model for mortality reduction factors. Insurance: Mathematics and Economics, 2006; 38, 556-70.
15. Wilmoth J, Zureick S, Canudas-Romo V, et al. A flexible two-dimensional mortality model for use in indirect estimation. Popul Stud (Camb), 2012; 66, 1-28.
16. LM C, CA S, DM W, et al. Under registration of neonatal deaths: an empirical study of the accuracy of infantile vital statistics in Taiwan. J Epidemiol Community Health, 1998; 52, 289-92.
17. Clark SJ, Sharrow DJ. Contemporary Model Life Tables for Developed Countries: An Application of Model-based Clustering. In Center for Statistics and the Social Sciences, University of Washington: Seattle, 2012.
18. Sharrow DJ. A package to implement HMD model life table system: Package 'LifeTables'. http://cran.r-project.org/ web/ packages/LifeTables/LifeTables.pdf (accessed in. 2014)
19. United Nations Model life tables for developing countries. United Nations: New York, 1982; p 351.
20. Guillot M, Gerland P, Pelletier F, et al. Child mortality estimation: a global overview of infant and child mortality age patterns in light of new empirical data. PLoS Med, 2012; 9, e1001299.