Estimation of the Number of Missing Females in China: 1900-2000

Quanbao Jiang
School of Management
Xi'an Jiaotong University
Xi'an, Shaanxi Province, 710049
China

Marcus W. Feldman
Morrison Institute for Population and Resource Studies
Stanford University
Stanford, CA 94305, USA
Email: mfeldman@stanford.edu

Xiaoyi Jin
Institute for Population and Development Studies
School of Public Policy and Administration
Xi'an Jiaotong University
Xi'an, Shaanxi Province, 710049
China

June 15, 2005

ABSTRACT

Excess female mortality has existed throughout Chinese history. Since the 1980s the sex ratio at birth has risen significantly, resulting in a severely unbalanced sex ratio in the Chinese population. Data from the five censuses, with appropriate adjustments, permit estimates of the numbers of missing females from the period 1900–2000. The analysis is broken down by historically important periods and probes the reasons for the missing females in each period. Our estimates are compared with those of other investigators.

Keywords: missing females, excess female mortality, sex ratio at birth, China

BACKGROUND

Son preference and discrimination against female children in China has produced a consistently male-biased sex ratio at birth and excess girl child mortality. The result is a shortage of females and an unbalanced sex structure in population of China. Using 1990 census data, Coale and Banister (1994) estimated that with a normal sex ratio at birth and survival rate by sex, the sex ratio of China's total population would not exceed 1.02. Since the actual sex ratio reached 1.07; clearly there was a marked deficit of females. In order to measure the deficit of females caused by gender discrimination, Sen (1989, 1990) proposed the following estimate of 'missing females'. Compare the sex ratios by age in an actual population with those in a model population with a normal sex ratio at birth and non-gender-biased mortality. If the actual sex ratio exceeds that in the model population, the missing fraction of the female population who would have to be alive to equate the actual with the model sex ratio can be used to estimate the missing females (Klasen and Wink, 2002; Cai and Lavely, 2003).

There have always been missing females throughout Chinese history. For example, infanticide and abandonment of female children was prevalent before 1950 (Wolf and Huang, 1980), even in the royal families of the Qing Dynasty (Lee et al., 1994). In the first half of the 20th century, wars and famine compelled families to allocate their limited resources to sons, which resulted in higher than normal mortality of female children (Das Gupta and Li, 1999).

At present, the missing females comprise not only those missing before birth but also those after birth (Croll, 2001; Li et al., 2001; Banister, 2004). Missing before birth refers mainly to sex-selective induced abortion, which produces the high sex ratio at birth. Missing after birth is

the excess girl child mortality resulting from the biased treatment of girls' illnesses as well as infanticide and abandonment of female children (Banister, 1992; Croll, 2001; Li et al., 2004). Coale and Banister (1994) found that due to the prevalence of prenatal sex identification techniques, mainly the use of the ultrasound B, Chinese parents' approach to achieving their ideal number and sex composition of children has changed from infanticide and abandonment of female children to sex-selective induced abortion.

Cai and Lavely (2003) differentiated between *nominally* and *truly* missing females. Truly missing refers to females who have disappeared before or after birth mainly by the use of sex-selective induced abortion. The nominally missing includes not only the truly missing but also those who are not registered in the statistics, i.e. falsely missing. Coale and Banister (1994) analyzed the data from the first four censuses in China, and concluded that during the 50 years before 1990, almost all the missing was truly missing as a result of excess female mortality and sex-selective induced abortion. However, Cai and Lavely (2003) claimed that there was underreporting of female births in the census of 2000 and estimated the ratio of truly missing to the nominally missing.

The phenomenon of missing females reflects the ideology of son preference and discrimination against girls, which has deep social roots. The strict patrilineal family system that has existed through most of Chinese history has led to son preference. Patrilineage determines the dominant status of men in inheriting property, in living arrangements, in continuity of families, and in family power structure. The social structure is governed by men; only men can continue the family lineage, and a child can acquire his or her social status and enter the social

structure only through the father. A woman is evaluated by her fertility, especially the ability to bear male children. Patrilocality entails that all daughters marry into other families and then belong to the husbands' families. They keep only a sentimental relationship with their natal families and have no substantial rights and responsibilities. Most of the material resources are inherited by sons, and daughters can only acquire mobile articles in the form of dowries or inheritance. Patriarchy prohibits women from taking part in economic activities outside of the family and therefore from external communication. As a result, women have to depend on men, which results in women's low status (Skinner, 1997; Das Gupta and Li, 1999; Khan and Khanum, 2000; Das Gupta et al., 2002; Leone et al., 2003). The low status of women leads to various forms of discrimination against girls, including insufficient investment, excessive punishment, abandonment, and infanticide in extreme situations, and produces a relatively low survival rate of female babies and girls (Waldron, 1987).

Missing females indicates that females in China suffer from discrimination beginning at conception, which results in the disappearance of some females who are expected to be born and survive. The missing females may have helped to slow down China's population growth. But this has happened at the cost of deterioration of women's rights of survival, social participation and development. At the same time, missing females induces gender imbalance in the Chinese population and causes the male 'marriage squeeze' (Tuljapurkar et al., 1995; Das Gupta and Li, 1999). The result is not only marriage pressure on young males, but also a series of social problems, such as inferior physical and psychological health of the unmarried, instability of marriages and families, birth out of wedlock, old age support for those who never married,

increasing prostitution, and abduction of and trafficking in women. These problems and related social conflicts impair the welfare of the overall society and harm the long-term sustainable development of Chinese population and society.

Many scholars have estimated the number and percentage of missing females in China (Johansson and Nygren, 1991; Coale and Banister, 1994; Das Gupta and Li, 1999; Klasen and Wink, 2002; Cai and Lavely, 2003). Using published statistics from the census in 2000, Cai and Lavely (2003) estimated that there were about 12 million missing females born between 1980 and 2000 at the 2000 census. As this estimation did not adjust the published census statistics for unregistered individuals, it can be considered as an estimate of the nominally missing females. The authors assumed that the truly missing females comprised two thirds of the nominally missing. It was then estimated that the population of truly missing females born between 1980 and 2000 was 8.5 million, or 4.1 percent of the observed cohort in the 2000 census. Coale and Banister (1994) estimated the percentage of missing females born in 50 years between 1936 and 1985 in each census, but did not give the number of the missing females. Employing the method of Coale and Banister (1994), Das Gupta and Li (1999) estimated the percentage of missing females born in the 70 years between 1920 and 1990. From the census in 1990, Klasen and Wink (2002) estimated the number of missing females to be 34.6 million (corresponding to all living females at the time of 1990 census), the percentage of missing females being 6.3 percent. From the 2000 census, missing females were estimated to be 40.9 million corresponding to all the living females at the time of 2000 census, and the percentage of missing females reached 6.7 percent. These studies used mainly statistics for some specific years or years when a specific

Chinese census was conducted in the 20th century. In these analyses, census data are usually not adjusted.

From the historical dynamic of the number of missing females, we can explore the socio-demographic consequences of these missing females during the 20th century, in the hope that this knowledge will help to reduce gender discrimination and contribute to solving the problems of missing females. What is the number of missing females in the 20th century? What are the characteristics of missing females in various historical periods? What are the main social determinants of the gender bias in various historical periods? All of these are problems that need to be addressed. This paper focuses on the number and characteristics of missing females in China in the 20th century and comprises three parts. First, we introduce the data and methods, and conduct necessary evaluation and modification of the employed census statistics. Second, we report the results and offer explanations in terms of the relevant historical background. Finally, we discuss the estimates and what they mean for Chinese society.

DATA AND METHODS

Existing Estimation Methods

Methods to estimate the number and percentage of missing females can be generally divided into two kinds. One is to use the sex ratio of the total population (Sen, 1990; Klasen and Wink, 2002). Compared with the sex ratio of a chosen benchmark population and based on the benchmark sex ratio and the number of males in the investigated population, the expected number of females can be estimated in an observed population. The difference between the female expected and actual populations in the study is the number of missing females. Sen (1989,

1990) chose the population and sex ratio of Sub-Saharan Africa countries, as the benchmark. However, the high fertility, short life expectancy, and low sex ratio in African countries do not represent the situation in China (Coale, 1991; Klasen, 1994). Sen (1990) also chose 1.05, the sex ratio of some developed countries, as a benchmark. However, because of the influence of the Second World War and the large percentage of the elderly and especially female elderly, this sex ratio is even more problematic (Coale, 1991).

Klasen and Wink (2002) also employed the method based on the sex ratio of an overall population in estimating the number of missing females, but they modified the calculation of expected sex ratio used by Sen (1990). First, they defined the average life expectancy according to the particular situation of the target population. According to this life expectancy, they selected a stable population in the life table of the Coale-Demeny model as the benchmark and defined the sex ratio in this stable population as the expected sex ratio. Then the population of missing females reflected in the census of the target population was estimated based on the male population in this census and the expected sex ratio. It is necessary to point out that Klasen and Wink (2002) chose the east model life table as the benchmark because they considered that the gender bias in mortality of the east model life table was the smallest in Coale-Demeny models.

The other method to estimate the missing females is to study the missing females by cohort. First, assume that the sex ratio at birth and surviving status of males and females in each year by birth cohort are normal under a certain life expectancy, and estimate the expected sex ratio of each cohort in each census. Second, calculate the expected female population according to the surviving males in censuses and the expected sex ratio, and then estimate the number and

percentage of the missing females (Coale and Banister, 1994; Das Gupta and Li, 1999). Clearly the percentage of missing females in the same birth cohort differs at different censuses.

Methods

This paper uses the method of Coale and Banister (1994) and Das Gupta and Li (1999) to estimate the missing females from data of 1953, 1964, 1982, 1990 and 2000 censuses. However, the method is modified as follows.

We employ the surviving population of birth cohorts born between 1900 and 2000 at each census to estimate the number and percentage of missing females. As the surviving populations of these birth cohorts differ at different censuses, the sex ratio of each birth cohort is also different. To ensure the consistency of the estimated data, it is necessary to ascertain the observed sex ratio curve of each birth cohort at censuses in the 20^{th} century. Coale and Banister (1994) and Das Gupta and Li (1999) adopted the data of the census closest to each birth cohort and selected a short segment in the sex ratio curve of each census that when combined formed a sex ratio curve. Specifically, each birth cohort between 1920 and 1953 was based on the census of 1953, and those between 1954 and 1963 were based on the census in 1964 and so on. Although deficit of females mainly occurs at early stage of life (Coale and Banister, 1994), it also exists in the later part of life, as social events that place families under severe resource scarcity, such as famine, war, and declining fertility also strengthen gender discrimination and reduce the survival of female children (Das Gupta and Li, 1999). Therefore, rather than using the sex ratio curve of the census closest to each birth cohort, we use the following segments. The birth cohorts between 1900 and 1945 are based on the statistics of people aged from 7 to 52 years

old in the census of 1953. For the birth cohorts between 1946 and 1953, considering the increasing sex ratio and excess girl child mortality in the period of famine (Das Gupta and Li, 1999), we combine the birth cohorts between 1946 and 1953 with those between 1954 and 1963 and produce a sex ratio curve based on the statistics of the population aged from 0 to 17 years old in the census of 1964. China began its strict family planning policy in 1980, so the year 1980 is defined as a borderline. The sex ratio curve of birth cohorts between 1964 and 1980 is calculated from statistics of people aged from 1 to 17 years old in the census of 1982. For the birth cohorts between 1981 and 2000, the sex ratio curve is calculated based on statistics on population aged from 0 to 19 years old in the census of 2000.

After this segment partition is determined, the population of each birth cohort in the corresponding census in terms of different genders can be assessed and then the expected sex ratio of each birth cohort in the corresponding census calculated. We calculate the expected female population based on the expected sex ratio and estimate the number and percentage of missing females. P_x^m denotes the male population aged x in the census, P_x^f the female population aged x in the census, SRB the normal sex ratio at birth, SR_x^e the expected sex ratio in the census, N_x^{mf} the number of missing females, P_x^{mf} the percentage of missing females, and x age. Then we can get

$$SR_x^e = SRB \times \frac{L_x^m / L_0^m}{L_x^f / L_0^f} \qquad (1)$$

In formula (1), L_x^m represents the average surviving population years aged x between the

segment $(x, x+1)$ years old in the male life table, L_x^f the average surviving population years aged x between the segment $(x, x+1)$ years old in the female life table, and L_0^m and L_0^f have similar definition for age 0. Then, we get

$$N_x^{mf} = \frac{P_x^m}{SR_x^e} - P_x^f \qquad (2)$$

$$P_x^{mf} = \frac{N_x^{mf}}{N_x^{mf} + P_x^f} \qquad (3)$$

Note that although our goal is to study missing females, it is mainly the missing female children that are estimated in the census statistics of 1964, 1982, and 2000 from the above segment approach. In contrast to Cai and Lavely (2003), we correct the census statistics and adjust for underreporting as much as possible. Therefore, what this paper estimates is the actual population of missing females. Coale and Banister (1994) and Das Gupta and Li (1999) defined the percentage of missing females as the ratio between the missing female population and the observed surviving population, while here we define the percentage of missing females as the ratio between the missing female population and the expected female population.

In addition, we also briefly report results that use the method of Klasen (2002) to estimate the number and percentage of missing females.

Data Evaluation and Adjustment

Data employed here can be divided into two types according to the above mentioned methods: the first are the data of different ages and genders from the five censuses of 1953, 1964, 1982, 1990 and 2000 in China; the second constitute statistics used to calculate the expected sex ratio.

Evaluation and adjustment of census data

The accuracy of population statistics in terms of ages and gender exerts a rather strong influence on estimation of the missing female population. As the Chinese adopt Shengxiao (a lunar marking of years with twelve animals) to mark years, and almost everyone knows the animal corresponding to the year when he or she was born, it is possible to calculate one's age according to Shengxiao. Therefore, in each census there is rarely misreporting and heaping of age, and the record of ages is rather accurate (Coale, 1984; Coale and Banister, 1994) .

Two factors in the statistics of census deserve special consideration. One is the military population. In the 1953 and 1964 censuses, there is no record of soldiers by age. In the census of 1982, the statistics of the army in terms of different genders are provided in the form of lists of people in 5-year intervals. In the censuses of 1990 and 2000, the statistics of the army by gender and single age are provided. As there are a large number of soldiers and they are mostly males, if the population of soldiers is not corrected, the accuracy of missing females will be affected greatly. In order to adjust the army statistics in the census statistics, Coale and Banister (1994) designed a method employing the highest sex ratio among the cohorts aged between 16 and 34 ever recorded in the four censuses to adjust the statistics on soldiers, because it was believed that the former census gave more accurate numbers and sex ratio of these people because they were too young to join the army at that time. The next census would also be accurate because they became old and left the army. We adopt this method. The second factor is underreporting. In the census of 1953, there were altogether 10.4 million people whose gender and age were not clear. We assume the gender and age structure of these people is consistent with that recorded in the

census and allots them to various age groups. Different levels of underreporting existed in the censuses of 1982, 1990 and 2000 (Banister, 1987; Johansson and Arvidsson, 1994; Lavely, 2001), and the underreporting of young people is especially severe. This paper estimates the missing female population by birth cohorts between 1964 and 1980 and between 1981 and 2000 according to the census statistics of 1982 and 2000, respectively. In order to reduce the influence of underreporting of young people, we modify the results of the censuses of 1982 and 2000 in terms of different ages.

We adopt reverse survival analysis to adjust the statistics of 1982 with that of 1990. Survival rate determined by life table is used in reverse survival analysis. In order to choose the life table employed in reverse survival analysis, we first take the 1989 life table by Jiang et al. (1995) as the baseline and then transform it into a series life tables by logit transformation (Brass, 1977). It is assumed that there was little underreporting among the people aged 60 and over. In this way, people aged between 60 and 90 and above in the census of 1990 would be between 52 and 82 and above in the census of 1982. That life table with which the population estimated based on reversal survival analysis differs least from the observed population aged between 52 and 82^{+}, is selected. The calculated population by age and sex is considered as the modified one, as shown in Table 1. The specific modified population aged between 1 and 17 is shown in Appendix Table 1.

Table 1 here.

For the census statistics of 2000, the birth cohorts between 1991 and 2000 are 0 to 9 years old at the census of 2000. This population is modified using the survival analysis based on the

annual births provided by the National Statistics Bureau as well as the records of primary school enrollment. For the birth cohorts between 1981 and 1990, a survival analysis method similar to that of reversal survival analysis used for the census statistics of 1982 is adopted. In processing census statistics of 2000, survival analysis is adopted and the difference between the observed population in the census of 2000 and the estimated population aged 60 to 90 and above, which is estimated based on the population aged 50 to 80 and above in the census of 1990, is the smallest. The population aged 10 to 19 in the census of 2000 is modified according to this method. The specific modified population aged 0 to 19 is shown in Appendix Table 2. As the reference times of the 1990 and 2000 censuses are different, the latter census data were adjusted accordingly. After modification, it is found that in the birth cohorts between 1981 and 2000, the number of unregistered males is 16.68 million, 17.97 million for females, and 34.65 million for both genders in 2000 census.

Data to calculate the expected sex ratio

First, it is necessary to determine the baseline life table when calculating the expected sex ratio. Although Klasen and Wink (2002) argue that the Coale-Demeny west model life table underestimates the mortality by gender, many scholars still choose it as the basis for estimating the number of missing females (Coale and Banister, 1994; Das Gupta and Li, 1999; Cai and Lavely, 2003). We also adopt the west model life table. Second, for the normal sex ratio at birth, the universally accepted value in the international community is between 1.02 and 1.07 and many studies adopt 1.06 as a normal sex ratio at birth (Coale and Banister,1994; Cai and Lavely, 2003). This paper also adopts 1.06 as normal. Another factor that should be considered is the

female life expectancy at different ages. Different life expectancies determine the different expected sex ratios and hence the different estimated number of missing females. For life expectancy between 1900 and 1990, we use the data from Coale and Banister (1994) and Das Gupta and Li (1999). For life expectancy between 1990 and 2000, we use the estimated value from Jiang et al. (1995) and Li et al. (2004) as a reference. The specific data are shown in Scenario 1 of Table 2. However, life expectancy may not actually be constant at 25 years old for 40 years between 1900 and 1939. Therefore, this paper designs life expectancy Scenario 2, in which the life expectancy between 1900 and 1939 varies linearly. Generally speaking, as the life expectancy in Scenario 1 is lower, the estimated number of missing females is relatively higher. Therefore, we define the estimated number of missing females based on Scenario 1 as an upper limit and consider it as the main result to discuss, while taking the estimated results based on Scenario 2 as a lower limit.

Table 2 here.

In calculating the particular expected sex ratio curve, the segment method dealing with the birth cohorts between 1900 and 2000 is adopted. Based on the assumption of Coale and Banister (1994), the expected sex ratio of children aged from 0 to 4 is determined by the corresponding model life table, taking life expectancy in the year when the census was conducted; the expected sex ratio of people above 5 years old is determined by the corresponding model life table taking the average of life expectancy of the year when the child was born and that when the census was conducted.

With the above data and methods, we estimate, in different periods, the number and

percentage of missing females by birth cohorts between 1900 and 2000.

ESTIMATED NUMBER OF 'MISSING FEMALES'

Number and Trends of the Missing Females between 1900 and 2000

Our modified data and methods give an estimated number of missing females as 35.59 million corresponding to the Scenario 1 of Table 2 and the percentage is 4.65 percent. The number of missing females decreases to 31.78 million using Scenario 2 of Table 2 and the percentage also declines to 4.17 percent.

For comparison, we also used the method and assumptions of Klasen and Wink (2002) used to estimate the number of missing females; that is, east model life table and stable population, mortality level 22, and population growth rate of 12.5 per thousand. Sex ratio is defined as 1.06 instead of the value 1.05 used by Klasen and Wink (2002). Based on the 2000 census statistics modified by Li et al. (2004), i.e., 655.27 million males and 615.93 million females, the number of missing females is estimated to be 32.83 million and constitutes 5.06 percent of the expected female population, which can be considered as the median of missing females. Note that this estimated percentage of missing females is related to the surviving females at the 2000 census. This differs slightly from the percentage relative to the expected females reported in the previous paragraph. The specific estimated results are shown in Table 3.

Table 3 here.

Figure 1 represents the historical tendency of the estimated upper limit of missing females in birth cohorts between 1900 and 2000. It can be seen that the percentage of missing females is inconsistent, which reflects the quality of data. In order to analyze trends in the number and

percentage of missing females, this paper transforms the estimated percentage of missing females by a five year smoothing; that is, considering the mean of a certain year and the following four years as the percentage of missing females for that year.

Figure 1 here.

Figure 1 indicates that the percentage of missing females differs greatly in different historical periods. Before 1949, the overall percentage of missing females was relatively high. Since 1900, the percentage at first remained stable, and around 1910, that is the end of Qing dynasty, the percentage reached a local peak, after which it declined again. Since 1920, the percentage of missing females has generally tended to rise. In the middle and late 1930s, that is the beginning of the anti-Japanese War, it reached a peak value again, and then began to decline. After 1949, although the percentage of missing females fluctuated, it remained generally low until the middle of the 1970s. In this period, there are two local peaks: one is during the Great Famine in the late 1950s and the other is during the Cultural Revolution. Since the middle 1970s, the percentage of missing females has been continuously increasing, although the current value is still lower than the highest in the 20^{th} century.

Characteristics of Missing Females in Various Phases

Based on the important historical events and the trend indicated by Figure 1, we divide the period between 1900 and 2000 into five phases, namely 1900 to 1928, 1929 to 1949, 1950 to 1965, 1966 to 1979 and 1980 to 2000. The numbers and percentages of missing females in each phase are shown in Table 4.

Table 4 here.

Birth cohorts between 1900 and 1928

From 1900 the percentage of missing females remained stable at first and then began to rise. Around 1910, that is the end of Qing dynasty, it reached a peak and then declined again. After 1920, it began to rise again. Between 1900 and 1928, China experienced the end of Qing dynasty, the 1911 Republic Revolution; Yuan Shikai's assuming power and the subsequent Beiyang warlords, and the political situation was turbulent. For the birth cohorts around 1910, the percentage of missing females reached a local peak, which may be associated with the social turbulence at the end of Qing dynasty and during the 1911 Revolution. In the 1920s, the percentage of missing females continued to rise, possibly due to the large-scale fighting between warlords. Because of these conflicts, young males were forced to join the army, and the common people were required to supply operational rations, suffered harassment by the militias, and escaped from their hometowns when attacked (Chi, 1976; McCormack, 1977). Wars and famine resulted in severe resource scarcity and families would allot the limited resources to sons who were perceived as more useful, and consequently girls were more likely to suffer discrimination. Abandonment and infanticide of female children increased and their mortality rate was also higher than normal (Xin, 1989; Das Gupta and Li, 1999). The missing female population in the birth cohorts between 1900 and 1928 was 8.01 million and the percentage reached a relatively high level of 7.44 percent.

Birth cohorts between 1929 and 1949

The percentage of missing females in the birth cohorts between the end of the 1920s and the first part of the 1930s was high because of the continual and fierce conflicts among warlords and

between the Kuomintang and the Communist Party. After that, as Japan began to attack China on a large scale, the percentage of missing females rose quickly in 1937. During this war with Japan, abandonment and infanticide of female children was widespread (Tan, 1989). The sex ratio of yearly birth cohorts between 1937 and 1945 were respectively 118.88, 119.66, 118.26, 117.57, 116.61, 116.63, 115.42, and 113.84, with 112.86 at the census of 1953, much higher than normal. As the war between the Kuomintang and the Communist Party during 1946 to 1949 was less severe than the war with Japan, the percentage of missing females then decreased relative to the period between 1937 and 1945. Under the influence of conflicts between warlords, the war with Japan, and the continual wars between the Kuomintang and the Communist Party, the number of missing females in the birth cohorts between 1929 and 1949 reached 11.84 million or 9.56 percent, a rather high level.

Birth cohorts between 1950 and 1965

The percentage of missing females in the birth cohorts of the early 1950s is low. But during the period of Great Famine, there is a local peak after which the percentage of missing females decreases. After the founding of PRC in 1949, the long-term wars ended and there was a period of stability. With the establishment of communist rule, the idea of gender equality in the communist ideology reduced discrimination against girls (Das Gupta and Li, 1999). The ideas of "women can hold up half of the sky", equality between men and women, receiving the same pay for the same work, together with related social activities functioned to improve women's social status and reduce discrimination against females, although the tendency for son preference and discrimination against females was not completely eliminated. The government intervened

against female child abandonment and infanticide and other harmful practices, which helped to reduce the excess girl child mortality (Coale and Banister, 1994). The total number of missing females among the birth cohorts between 1950 and 1965 is 2.77 million and constituted a percentage of 2.42 percent, a low level.

Birth cohorts between 1966 and 1979

Compared with the general level between 1900 and 2000, the percentage of missing females in birth cohorts between 1966 and 1979 was low. After the Great Famine, a compensatory peak of fertility took place and the total population increased rapidly. Also, after the Cultural Revolution, there was a return to a high fertility situation. Although the Chinese government established a birth control policy characterized by 'later, more spacing, and fewer' in 1974, enforcement was not strict. In this period a couple may have had sons just by bearing more children, and the demographically detectable discrimination against girls decreased. The total number of missing females in the birth cohorts between 1966 and 1979 is 2.77 million, representing a percentage of 1.74 percent, a very low level.

Birth cohorts between 1980 and 2000

The percentage of missing females in the birth cohorts between 1980 and 2000 generally tended to rise. On the other hand, fertility was declining. The difference between the strict limitation on the number of children for each couple according to the birth control policy carried out in China since 1980s and the actual desired number of children for Chinese people is quite large, especially because the number and gender configuration of the children cannot be guaranteed at the low fertility. The popularization of prenatal sex identification techniques, such

as the ultrasound-B, has been the main approach used by Chinese parents to achieve their ideal number and gender configuration of children, a change from infanticide and abandonment of female children to sex-selective induced abortion in the previous half century (Coale and Banister, 1994). It has been proved that the recent rise in the sex ratio is the result of sex-selective induced abortion rather than underreporting of female children (Coale and Banister, 1994; Croll, 2001). The missing females of the birth cohorts between 1980 and 2000 reached 9.2 million and namely a percentage of 4.19 percent, a quite high level.

CONCLUSIONS

Demographic and social issues, such as the size and aging of the Chinese population, demand and supply of labor, and the marriage squeeze, have attracted much attention from academics, the public, and the government. The phenomenon of 'missing females' is a key component of these issues. For example, 'missing females' decrease the current population estimate. Although this decrease may have little influence on the current population size, its accumulated impact will exert a significant influence upon population growth and result in important social problems (Cai and Lavely, 2003). Our main findings are as follows.

The estimated total number of the missing females in the 20th century of the birth cohort between 1900 and 2000 is 35.59 million, representing 4.65 percent. However, adjustment of census data and choice of the key parameters strongly affect estimated number and percentage of the missing females. If life expectancy is assumed to increase linearly every ten years between 1900 and 1939 (that is scenario 2), the number of missing females decreases to 31.78 million and the percentage also declines to 4.17 percent.

From the 2000 census, Klasen and Wink (2002) estimated that there were 40.9 million missing females, or 6.7 percent, which differs greatly from our estimated number of the missing females between 1900 and 2000. This can be due to many factors. First we estimate the number of missing females in birth cohorts between 1900 and 2000 at each census, while Klasen and Wink (2002) estimated the number of missing females in the census statistics of 2000. Second, the methods employed to estimate the number of missing females are different. Klasen and Wink (2002) employed the sex ratio of the whole population, and adopted the Coale-Demeny east model life table as the stable population, while we used the Coale-Demeny west model life table and estimated the missing female population based on various birth cohorts. Third, our selection of some key parameters differs from theirs. For example, we adopt 1.06 as normal sex ratio at birth, while Klasen and Wink (2002) adopted 1.05. We employ different life expectancies in birth cohorts at different times, while Klasen and Wink (2002) used the same life expectancy. Finally, because of the underreporting in the 2000 census, we used the adjusted statistics of the 2000 census, while Klasen and Wink (2002) used the original reported data.

In order to conduct a comparative analysis, we also tried the method of Klasen and Wink (2002), but with the normal sex ratio at birth assumed as 1.06, our estimate for the total number of missing females is 32.84 million, almost 8 million fewer than estimated by Klasen and Wink (2002).

Our estimated percentage of missing females also differs from those of Coale and Banister (1994) and Das Gupta and Li (1999). Although the methods employed are generally similar, in order to estimate the number of missing females we used the surviving population of each birth

cohort at censuses and calculated the observed sex ratio according to the surviving population by single age group. Using a five-year smoothing, which is slightly different from that of Coale and Banister (1994) and Das Gupta and Li (1999), we follow trends in the missing females over the 20th century. At the same time, the census statistics selected for some birth cohorts are different. In addition, we adjusted the recorded surviving population in the censuses of 1982 and 2000. Moreover, our definition of the percentage of missing females is also different. These all lead to different estimates of percentage of missing females.

We estimate that there are 9.2 millions missing females in the birth cohorts between 1980 and 2000, namely 4.19 percent. This is 700 thousand more than the 8.5 million estimated by Cai and Lavely (2003) for the same period. Although both studies use the Coale-Demeny west model life table, we used the modified census data of 2000, while Cai and Lavely (2003) used the nominally missing female population in the original data to estimate the actual missing female population. Further, we use different life expectancies for different birth cohorts, while Cai and Lavely (2003) employed the same life expectancy for the different birth cohorts in the 2000 census. This also leads to different results.

It is proved that female child abandonment and infanticide existed widely in Chinese society before the 1950s (Wolf and Huang, 1980; Das Gupta and Li, 1999), and decreased after the 1950s. An issue deserving attention is the tendency for the proportion of missing females to increase after 1980. The missing female population in the birth cohorts between 1980 and 1989 was 2.94 million or 2.71 percent, while that between 1990 and 2000 was 6.26 million or 5.65 percent. This indicates that between 1980 and 2000, discrimination against females became

severe. Of course fertility also decreased markedly in this period. If powerful measures are not taken, it is possible that the sex ratio at birth and the percentage of missing females in the later birth cohorts will rise further.

Our estimation of missing females has some limitations. Because of data availability, this research could only use life expectancies reported in other studies, which may affect our estimates to some extent as the selection of life expectancy is critical for the estimation of the number and percentage of missing females. At the same time, this paper makes some modifications to the reported census statistics, but whether these adjustments are accurate needs further study. Further research might focus on the social consequences of missing females, especially in terms of the marriage squeeze historically during the 20th century and for the future prospects in the 21st century and the long-term effect of the missing females on Chinese population growth. It is important that policies for reducing the sex ratio at birth and excess female child mortality be developed.

References

1. Banister, J. 1987. *China's Changing Population*. Stanford: Stanford University Press.
2. Banister, J. 1992. China: Recent mortality levels and trends. Paper presented at the *Annual Meeting of the Population Association of America,* Denver, Colorado.
3. Banister, J. 1994. Implications and Quality of China's 1990 Census Data, in China State Council and National Bureau of Statistics (ed.), *1990 Population Census of China: Proceedings of International Seminar*, Beijing: China Statistics Press.
4. Banister, J. 2004. Shortage of girls in China today. *Journal of Population Research* 21(1): 19–45.
5. Brass, W. 1977. Notes on empirical mortality models. *Population Bulletin of the United Nations*, No. 9.
6. Cai, Y and W. Lavely. 2003. China's missing girls: Numerical estimates and effects on population growth. *China Review* 2(3): 13–29.
7. Chi, H. 1976. *Warlord Politics in China 1916–1928*, Stanford University Press.
8. Coale, A. J. 1984. *Rapid Population Change in China, 1952–1982*. National Academy Press.
9. Coale, A. J. 1991. Excess female mortality and the balance of the sexes in the population: An estimate of the number of 'missing females'. *Population and Development review* 17(3): 517–523.
10. Coale, A. J. and J. Banister. 1994. Five decades of missing females in China. *Demography* 31(3): 459–479.
11. Choe M., H. Hao, and F. Wang. 1995. Effects of Gender, Birth Order, and Other Correlation on Childhood Mortality in China. *Social Biology* 42: 50–64.
12. Croll, E. 2001. *Endangered Daughters: Discrimination and Development in Asia*. London:

Routledge.

13. Das Gupta, M. and P.N.M. Bhat. 1995. Fertility decline and increased manifestation of sex bias in India. *Harvard University Center for Population and Development Studies Working Paper* 95.02.
14. Das Gupta, M. and S. Li. 1999. Gender bias in China, South Korea and India 1920–1990: The effects of war, famine, and fertility decline. *Development and Change* 30(3): 619–652
15. Jiang, Z, N. Li, S. Li, F. Sun, H. Liu, Y. Du. 1995. *Mortality research on China's fourth census data*. In *Collection of Monograph on the 1990 census*. Beijing: China Statistics Press.
16. Johansson, S. and A. Arvidsson. 1994. Problems in Counting the Youngest Cohorts in China's Censuses and Surveys, in China State Council and National Bureau of Statistics (ed.), *1990 Population Census of China: Proceedings of International Seminar*, Beijing: China Statistics Press.
17. Johansson, S. and O. Nygren. 1991. The missing girls of China: A new demographic account. *Population and Development Review* 17(1): 35–51.
18. Khan, M. A. and P. A. Khanum. 2000. Influence of Son Preference on Contraceptive Use in Bangladesh. *Asia–Pacific Population Journal* 15(3): 43–56.
19. Klasen, S. 1994. 'Missing women' reconsidered. *World Development* 22:1061–1071.
20. Lavely, W. 2001. First Impressions from the 2000 Census of China. *Population and Development Review* 27(4): 755–769.
21. Lee, J., Wang F., and C. Cambell. 1994. Infant and child mortality among the Qing nobility: Implications for two types of positive check. *Population Studies* 48(3): 395–411.
22. Li S. and C. Zhu. 2001. *Research and practice of children's gender survival difference in China*. Beijing: China Population Publishing House.

23. Li S., F. Sun, Q. Jiang, X. Zou, R. Guan. 2005. Report of the study on China's mortality changes in the 1990s based on China's 2000 population census. [Unpublished research report]
24. Li S., F. Sun, Q. Jiang. 2005. Analysis of the total and structure of population in the 2000 census. Manuscript.
25. Li S., Zhu C., and M. Feldman. 2004. Gender differences in child survival in rural China: A county study. *Journal of Biosocial Science* 36(1): 83–109.
26. McCormack, G.1977. *Chang Tso-lin in northeast China, 1911–1928: China, Japan and the Manchurian idea*. Stanford University Press.
27. Sen, A. 1989. Women's survival as a development problem. *Bulletin of the American Academy of Arts and Sciences* 43:14–29.
28. Sen, A. 1990. More than 100 million women are missing *New York Review of Books*, December 20, 61–66.
29. Skinner, G. W. 1997. Family systems and demographic processes. In D. I. Kertzer and T. Fricke (eds.), *Anthropological Demography: Toward A New Synthesis*. Pp. 53–95. University of Chicago Press, Chicago, Illinois.
30. Stephan Klasen and Claudia Wink. 2002. A turning point in Gender bias in Mortality? An Update on the Number of Missing Women. *Population and Development Review* 28(2): 285–312.
31. Tan, A. 1989. The JoyLuck Club, New York: Putnam's.
32. Tuljapurkar, S., N. Li and M.W. Feldman. 1995. High sex ratios in China's future. *Science* 267: 874–876.
33. Waldron, I. 1987. Patterns and causes of excess female mortality among children in

developing countries. *World Health Statistics Quarterly* 40: 194–210.

34. Wolf, A. P. and C. Huang. 1980. *Marriage and Adoption in China, 1845-1945*. Stanford University Press, Stanford, Calif.

Table 1 Comparison of the reported and estimated population in the 1982 Census (Ten thousand)

Gender	EX_0	IMR(‰)	$52–82^+$ (years old)			1–17 (years old)		
			Estimated	Reported	Difference	Estimated	Reported	Difference
Male	66.3	36.15	6142.7	6135.9	6.8	20196.3	20047.0	149.3
Female	68.9	40.73	6498.9	6492.3	6.6	18914.9	18882.8	32.2

Table 2 Estimated female life expectancies at birth

Year	Scenario 1	Scenario 2
1900 – 1909	25.0	22.5
1910 – 1919	25.0	25.0
1920 – 1929	25.0	27.5
1930 – 1939	25.0	30.0
1940 – 1949	32.5	32.5
1950 – 1959	45.0	45.0
1960 – 1969	52.5	52.5
1970 – 1979	62.5	62.5
1980 – 1989	70.0	70.0
1990 – 2000	72.5	72.5

Sources: Coale and Banister (1994); Das Gupta and Li (1999); Jiang et al. (1995); Li et al. (2004).

Table 3 Estimates of missing females during the 20th century

	Number (Ten thousand)	Percentage (%)
Upper limit	3559	4.65
Median	3283	5.06
Lower limit	3178	4.17

Table 4 Number and percentage of missing females in China during the 20th century, by period

Year	1900–1928	1929–1949	1950–1965	1966–1979	1980–2000	1900–2000
Number (Ten thousand)	801	1184	377	277	920	3559
Percentage (%)	7.44	9.56	2.42	1.74	4.19	4.65

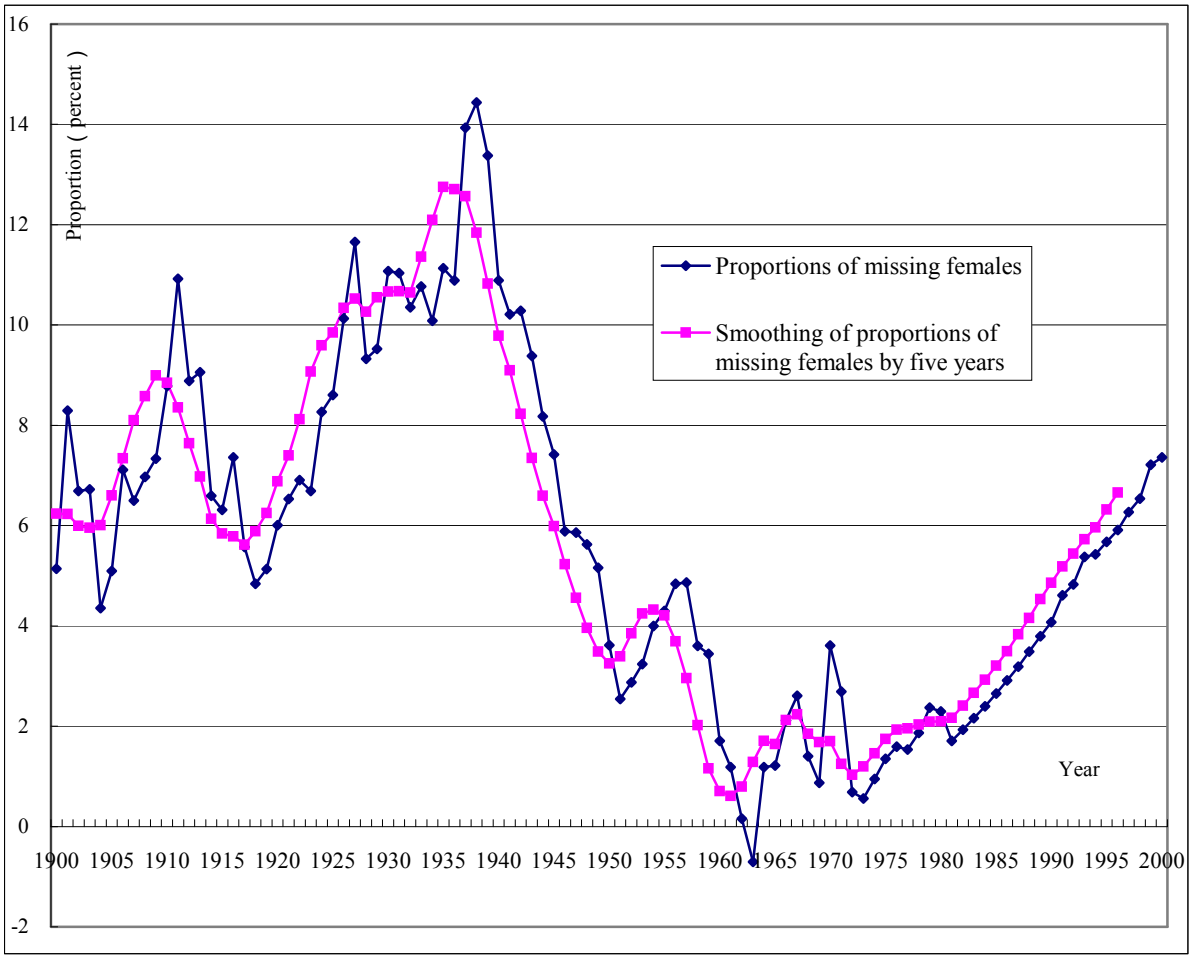

Figure 1 Percentages of missing females during 1900–2000

APPENDIX

Table 1 Adjusted numbers of population aged 1 to 17 in 1982 census (Ten thousand)

Birth year	Age	Adjusted			Reported			Difference		
		Male	Female	Total	Male	Female	Total	Male	Female	Total
1980	1	1283.0	1205.8	2488.8	901.5	836.1	1737.6	381.5	369.8	751.2
1979	2	1318.0	1238.0	2555.9	946.1	881.3	1827.4	371.9	356.7	728.6
1978	3	1208.1	1124.3	2332.5	1013.1	949.4	1962.6	195.0	174.9	369.9
1977	4	1304.4	1207.6	2512.1	959.0	903.0	1862.0	345.4	304.6	650.1
1976	5	1434.9	1344.8	2779.6	1000.6	941.5	1942.1	434.3	403.2	837.5
1975	6	1373.7	1294.3	2668.0	1052.9	990.4	2043.3	320.8	304.0	624.8
1974	7	1384.8	1274.9	2659.7	1121.6	1056.3	2177.9	263.2	218.6	481.7
1973	8	1291.6	1200.3	2491.9	1237.3	1166.0	2403.3	54.3	34.3	88.6
1972	9	1267.6	1202.0	2469.6	1290.2	1216.7	2507.0	-22.6	-14.7	-37.3
1971	10	1211.5	1150.0	2361.6	1299.0	1223.2	2522.3	-87.5	-73.2	-160.7
1970	11	1118.1	1056.9	2175.0	1407.2	1325.1	2732.3	-289.1	-268.2	-557.3
1969	12	1059.2	996.9	2056.1	1361.5	1287.3	2648.7	-302.3	-290.3	-592.6
1968	13	1004.2	942.6	1946.8	1452.2	1371.7	2824.0	-448.0	-429.2	-877.2
1967	14	978.6	918.8	1897.4	1263.9	1189.9	2453.8	-285.3	-271.1	-556.4
1966	15	1004.6	939.7	1944.3	1171.1	1104.0	2275.1	-166.5	-164.3	-330.8
1965	16	1004.9	934.8	1939.7	1319.0	1249.7	2568.7	-314.0	-314.9	-628.9
1964	17	949.1	883.1	1832.1	1250.8	1191.0	2441.8	-301.7	-308.0	-609.7
Total		20196.3	18914.9	39111.2	20047.0	18882.8	38929.7	149.3	32.2	181.5

Table 2 Adjusted numbers of population aged 0 to 19 in 2000 census (Ten thousand)

Birth year	Age	Adjusted			Reported			Difference		
		Male	Female	Total	Male	Female	Total	Male	Female	Total
2000	0	1013.7	887.7	1901.4	746.0	633.4	1379.4	267.7	254.3	522.0
1999	1	877.0	772.0	1649.0	633.2	516.3	1149.5	243.8	255.7	499.5
1998	2	971.3	859.1	1830.4	770.2	630.9	1401.1	201.1	228.2	429.3
1997	3	966.7	857.9	1824.6	789.7	655.7	1445.4	177.0	202.2	379.2
1996	4	991.3	883.4	1874.7	825.7	696.7	1522.4	165.6	186.7	352.3
1995	5	1018.8	910.5	1929.3	915.8	777.6	1693.4	103.0	132.9	235.9
1994	6	1039.2	931.9	1971.1	886.6	760.4	1647.0	152.6	171.5	324.1
1993	7	1075.1	964.9	2040.0	959.0	832.4	1791.4	116.1	132.5	248.6
1992	8	1156.5	1044.2	2200.7	1001.4	873.8	1875.2	155.1	170.4	325.5
1991	9	1278.9	1157.7	2436.6	1067.5	940.7	2008.2	211.4	217.0	428.4
1990	10	1302.0	1185.5	2487.5	1381.1	1239.9	2621.0	-79.1	-54.4	-133.5
1989	11	1321.5	1207.0	2528.5	1311.1	1202.7	2513.8	10.4	4.3	14.7
1988	12	1320.0	1209.7	2529.7	1278.0	1179.7	2457.7	42.0	30.0	72.0
1987	13	1253.8	1152.8	2406.6	1362.0	1266.3	2628.3	-108.2	-113.5	-221.7
1986	14	1160.3	1070.1	2230.4	1202.4	1116.6	2319.0	-42.1	-46.5	-88.6
1985	15	1090.3	1008.5	2098.8	1059.8	983.1	2042.9	30.5	25.4	55.9
1984	16	1065.9	988.8	2054.7	1047.3	984.5	2031.8	18.6	4.3	22.9
1983	17	1092.0	1015.8	2107.8	1029.4	979.0	2008.4	62.6	36.8	99.4
1982	18	1094.8	1021.3	2116.1	1217.6	1132.5	2350.1	-122.8	-111.2	-234.0
1981	19	1077.5	1008.0	2085.5	1014.7	938.0	1952.7	62.8	70.0	132.8
Total		22166.6	20136.8	42303.4	20498.5	18340.2	38838.7	1668.1	1796.6	3464.7