Underreporting of Births: the Cause of the Chaos of the National Population Statistics in China

Xiaochun Qiao, Ph.D.
Professor, Institute of Population Research
Renmin University of China, Beijing, P. R. China

Introduction

Every one knows that China is the largest country in the world. It is large because of its number of the total population. It must be right if you say that the number of the total population in China is ranked first in the world. But I believe that you cannot confirm how many people are in current mainland China, can you? As a researcher, you liked to accept the number published by the national government, I believe, utilized the data for conducting scientific research, and finally elaborated some fancy and delicate results. I did that like this too. Once you found that the data you carefully selected and used were carelessly changed by the data managers without any notice and evidence just after your research was finished, you were so embarrassed. In matter of fact, data, especially relatively correct data, are crucial to scientific research. However, we find that the quality of population data in China is not as good as what some researchers assessed before (Banister 1984; Coale 1984). The objective of the paper is to reveal what kinds of data problems exists, why the numbers of births and total population in directly enumerated data are so lower, how the published data are always being changed, and why such problems cannot be settled in China .

We will review some assessments to the data of China's 1982 census first, and then introduce the national system of Chinese data collection and publication. Following the time series, started in 1982, the time of the third population census, we will discuss how the published data were obtained in 1982-1989, and how they were adjusted subsequently. The data published in 1990s will be highlighted and some explanations of the chaos will

be given. Since the original causes for underreporting of births still existed, the 2000 census encountered the same problem as in 1990s. This led to a series of data problems in the 2000 population census even though such problems were tried to be covered.

Brief Review

China has conducted five population censuses since it was founded in 1949, that is, the years of 1953, 1964, 1982, 1990, and 2000. It was first time that people in the world knew the total number of Chinese population, that is, 544 million, in 1953. The data of 1964 census were not published afterward so that no one knew the total number of Chinese population in the duration from 1953 to 1982 even though some national data were used by scholars (Liu 1980; Aird 1980; Zhu 1980; Coale 1981). Since the family planning program was started in 1973 and the economic reform and opening started in 1978, Chinese government decided to conduct the third population census in 1982 (Qiao 1995). The third census was the first census using computer to process the national data (Li 1984, Banister 1984, Qiao 1995). Before the third census, Chinese government worried about the quality of the population census because they had no experiences in conducting such census with the largest number of population and without computer professionals at that time. With extreme effort, the third population census achieved a great success. After assessment in detail, Judith Banister (1984) mentioned that "The reporting of age in all three censuses (1953, 1964, and 1982) conducted by the People's Republic of China has been extraordinarily accurate." Ansley Coale (1984), using survival rates between 1953,1964, and 1982 censuses and fertility rates derived from the independently conducted 1982 fertility survey, constructed a population for 1982; after compared this population at each age to population enumerated by the 1982 census, he found that the agreement between the two independent sources of data was "extraordinary."

Is the quality of the current population data as good as evaluated above? No, because it happened in early 1980s. With China entering 1990s, one found that the official data published were less reliable. Attane and Sun (1999) found that fertility rates directly derived from1992 survey conducted by National Family Planning Commission (NFPC)

and from 1995 one-percent survey conducted by National Bureau of Statistics (NBS) were all adjusted upward. Zeng (1995) even adjusted the fertility rates of 1992 survey further to higher fertility rates. At same time, M. Giovanna Merli and Adrian E. Raftery, based on 1992 survey of four counties in the northeast of China, found that the number of births was underreported in rural China in response to China's population policies. In matter of fact, the birth underreporting has existed for long time since the beginning of family planning program. There is a word in China that the more the datum is focused by the government in it work, the less the accuracy of the datum in statistics. In order to clarify the quality of population data in China, we have to know the national system of population data collection.

Sources of National Population Data

There are three main sources for national population data, that is, the National Bureau of Statistics (NBS), the National Family Planning Commission (NFPC)[1], and the Ministry of Public Security (MPS). In correspondence with their working features, the three institutions focus on different aspect of population statistics.

NBS is the main institution on collection and publication of population data. It is regulated by the central government that NBS is the only institution to publish the official data on behalf the central government. All the governmental agencies are required to provide relevant data to NBS. In addition, NBS itself organizes some surveys. In population field, the main task for NBS is to conduct the population census by each ten years regulated by China Statistics Law issued in 1986, one-percent intercensal sampling survey (or called small census), and the sampling survey on population changes in each year. China has conducted five population censuses in 1953, 1964, 1982, 1990, and 2000; two intercensal surveys in 1987 and 1995. The first national sampling survey on population changes started at the end of 1982, just after the 1982 census. The sample scale was 500 thousand people from 1982 to 1988. The sample expanded to 1.8 million from 1989 to 1992 in order to make the representation in provincial level. Since 1994, the sample decreased to 1.2 million (Hu 2005).

NFPC takes charge of the statistics dealing with family planning and reproductive health. It has its own institution on yearly report from the grassroots family planning commission to the NFPC. However, due to the inaccuracy of the reported data, such data have not been used to evaluate the achievement of the work for long time. In addition, there are regular sampling surveys dealing with family planning and reproductive health for almost each five years beginning from 1982. Such sampling surveys are significant in family planning studies. The 1982 and 1988 surveys were retrospective surveys which have provided valuable data for reproductive and contraceptive history. The survey in 1992 was a cross-sectional survey in general, but it recorded the birth year of the last four children as well. The survey in 1997 was called the population and reproductive health survey. It is first time that researchers can get the reproductive health information in national level. The purpose of the 2001 family planning and reproductive health survey intended to follow up the same villages sampled in 1997 survey (Financial Department 2003). Unfortunately, it was not a typical longitudinal survey because the individuals were not followed up.

MPS takes the responsibility for dynamic statistics through vital registration system or household registration system. It takes charge of the registration of four dynamic enumerations such as birth, death, in-migration, and out-migration. Due to such work nested in MPS, the data relevant to public security, such as abnormal deaths, would be emphasized and cared, and the accuracy of other data else may not be quite focused. That is why the data from MPS are always inaccurate.

Chinese government regularly publishes an annually national statistical report at the early of next year. In the official report, there are only four sorts of population data, that is, total number of national population at the end of the year, number of new births and crude birth rate (CBR) of the year, number of deaths and crude death rate (CDR), and the number of the net population increase and the natural increase rate (NIR). Before 1982, all the published population data at the official report were adopted from the statistics of

---

[1] The National Family Planning Commission has changed its name to National Population and Family Planning

MPS. However, the data, especially birth rate and death rate, from MPS and NFPC were relatively lower than those, especially the data from censuses, from NBS (Attane and Sun 1999). Since 1982, the data from NBS have been utilized in the official report.

How Was the Total Population Estimated?

One can easily obtain CBR, CDR, as well as NIR from the annual sampling surveys on population changes for each year of the survey. Using the total number of the population from, for example, the 1982 census as the baseline population, and the formula:

$$P(t) = P(t-1) \, e^{\, r\,(t)}$$

Where P (t) stands for the national population at the end of the t year; r (t) expresses the NIR, which can be derived from the annual survey by subtracting the CDR from CBR. This formula is appropriate if we assume that the net international migration is equal to zero.

Because the reference time of the 1982 census was on July 1st, in reality, one had to estimate the number of population on December 31, 1982, that is, P (1982). As we had obtained the r (1983) from the sampling survey, based on the formula above, we can calculate the total number of population at the end of 1983 by:

$$P(1983) = P\,(1982) \, e^{\, r\,(1983)}$$

Following the same step, one can calculate the subsequent numbers of population in following years, that is, P (1984), P (1985), P (1986), etc. The estimated total population, CBR, CDR, and NIR then were published at the official report and the annual published Statistical Yearbook of China.

Adjustment to the Already Published Data in 1980s

The total population, CBR, CDR, and NIR are regularly published at the National Statistical Yearbook just at the following years. However by subsequent inference, started in 1982 census, until July 1, 1990, the time of the fourth census, the published

---

Commission in 2003. If some one used NPFPC, it would be same unit named NFPC before.

total population derived by inference may or may not be consistent with the population derived from the census. In reality, the total population from the census was 1134 million on July 1st 1990 and, the published population based on the subsequent inference should be 1119 million at the same time of the census. The number of national population obtained from the census was 15 million more than the already published population. In fact, such result implied that the quality of the census data was higher. The reason was that the higher risk of the census enumeration error was underreporting the births, due to the family planning policy, rather than overreporting. If this assumption were correct, the more the enumerated population, the higher the quality of the data. One of the simple ways for evaluation of the data quality is to check its fertility rate or birth rate. In 1990 census, the total fertility rates (TFR) directly calculated was 2.31 in 1989 and 2.25 in 1990. The fertility and birth rate seemed appropriate at that time even though underreporting was still happening.

Once we admitted that the census data were better than the inferred data, the inferred and published data had to be corrected. In order to harmonize the data, the National Bureau of Statistics had to change the previously published data after the third census, which was shown at table 1.

Table 1. The adjustment for the Total Population, CBR, and CDR, 1982-1989

Year	Total Population(million)			Birth Rate (per thousand)			Death Rate (per thousand)		
	Former	Adjusted	Gap	Former	Adjusted	Gap	Former	Adjusted	Gap
1982	1015.90	1016.54	0.64	21.09	22.28	1.19	6.60	6.60	0
1983	1027.64	1030.08	2.44	18.62	20.19	1.57	7.08	6.90	-.18
1984	1038.67	1043.57	4.90	17.50	19.90	2.4	6.69	6.82	.13
1985	1050.44	1058.51	8.07	17.80	21.04	3.24	6.57	6.78	.21
1986	1065.29	1075.07	9.78	20.77	22.43	1.66	6.69	6.86	.17
1987	1080.73	1093.00	12.27	21.04	23.33	2.29	6.65	6.72	.07
1988	1096.14	1110.26	14.12	20.78	22.37	1.59	6.58	6.64	.06
1989	1111.91	1127.04	15.13	20.83	21.58	0.75	6.50	6.54	.04

Note: "Former" reflects the data published before the 1990 census which is cited from 1990 China Statistical Yearbook. "Adjusted" is from 1991 China Statistical Yearbook. "Gap" = "Adjusted" – "Former".

The aim for the adjustment was to add the 15 million in to the populations from year 1982 to 1989. However, there is something hard to be explained: 1) the underreporting mostly occurred in births rather than in death, but the death rates were also adjusted, and even decreased in 1983; 2) it is hard to find the law backing the adjustment in both birth rate and death rate. It seems that the CBRs in the middle of the duration might have been gained more than the two extreme parts or that the lower CBRs might have been gained more extent, but there were still some exceptions. Unfortunately, there was no any explanation or clue on how the adjustments were made by NSB but just noted at the bottom of the table 3-2 dealing with CBR, CDR, and NIR from 1949-1989 in the 1991 Statistical Yearbook of China that the data in 1982-1989 were predicted based on 1982 and 1990 censuses, the data in 1990 was from 1990 sampling survey of population change, and the others were from MPS. However, this is still a mystery.

Underreporting of Births in 1990s

After 1990 census, NBS has learnt the experiences that the number of new births enumerated at the sampling survey is always less the reality, and the number of births had to be adjusted upward afterward. In stead of adjusting the number of birth at the time of next census, NBS started to adjust the number of birth before the data were published at the annual Statistical Yearbook. Even though, NBS has never officially published the enumerated TFR and CBR and the methods for the adjustment of the TFR and CBR, we can find, from some other sampling surveys and personal publications, that the TFR or CBR directly calculated from sampling surveys in 1990s were extremely lower. Making some adjustment seems appropriate. Now, we provide some examples below.

1) Fertility sampling survey conducted by NFPC in 1992

This survey conducted in October 1992 covered 385 thousand samples, with 80 thousand women in reproductive ages, which was nationally representative. However, the directly calculated TFRs were 1.65 in 1991 and 1.52 in 1992. NFPC invited some scholars to assess the results, and finally they made adjustment. The published TFRs were 1.87 in 1991 and 1.72 in 1992 (Attane and Sun 1999). However, such results were still relatively lower because the TFRs derived from 1990 census were 2.31 in 1989 and 2.25 in 1990,

and it seemed impossible that the fertility rate could be dropped so fast at that time. Zeng (1995) made his own estimation on the TFR, that is, 2.2 in 1991 and 2.1 in 1992.

2) Sampling survey of population changes conducted by NBS in 1993 and 1994

In general, fertility survey conducted by family planning commission can easily lead to birth underreporting because respondents, especially women in reproductive ages, are afraid of being charged if they have unplanned births. Under such assumption, fertility or CBR derived from surveys by NBS should be more accurate than by NFPC. However, the CBR directly calculated from sampling surveys of population changes conducted by NBS was also lower. For instance, the directly calculated CBR were 15.58 ‰ in 1993 and 15.32 ‰ in 1994 (Jia 1995), and officially published CBR at the annual Statistical Yearbook were 18.09‰ in 1993 and 17.70‰ in 1994. There were 2.51 per thousand points, about 2.9 million births, in 1993 and 2.38 per thousand points, about 2.8 million births in 1994 added into the published CBR and total number of population. Such amount of births added means that there would be over 20% of births unreported if the published data were correct. Unfortunately, like the adjustment in 1980s, why and how the CBR were adjusted have never been explained by NBS.

3) Surveys for checking the quality of family planning statistics

Due to lower fertility directly calculated from 1992 sampling survey by NFPC and from sampling surveys of population changes by NSB, one found that it would be difficult for any official survey to obtain accurate results such as fertility rate and CBR. In order to clarify the reality of the fertility level in China, NFPC carried out a provincial survey to check the quality of family planning statistics. The survey started in 1993, and was continued subsequently. First survey went to Hebei and Hubei provinces in October 1993. The nature of the survey was that the survey was not informed to the sampled village until the interview began. There were 32 villages in the two provinces suddenly surrounded by investigators, who were selected from else provinces and trained before the survey, at the same time of a mid-night. No one was allowed to get out of the village until the survey finished. The new births in 1992 and within nine months of 1993 were enumerated. The underreporting rate was 37.3%, with 937 reported births and 1494

checked-out births in years of 1992 and 1993, for the two provinces. Such result astonished the central leaders and all the leaders in NFPC, as well as scholars. However, this result could not be inferred to whole nation, and it was still hard to estimate the extent of underreporting rate in China. Such surveys were continuously carried out in Henan in April 1994 with 27.3% underreporting rate, Shandong in October 1994 with 10.9% underreporting rate, Gansu in May 1995 with 30.6% underreporting rate, etc. Because such checking for data quality was no longer effective since the grassroots personnel knew how to deal with it, the survey could not be used for data checking now.

4) One-percent sampling survey conducted by NBS in 1995

One-percent sampling survey belongs to the intercensal survey between 1990 and 2000 censuses. There were 12 million people sampled and enumerated. Unfortunately, the directed calculated TFR from the survey was 1.46 in 1995, still too low. The department of population and employment of NBS adjusted the TFR to 1.85 (Zhang, al. et. 1997).

5) Population and reproductive health survey conducted by NFPC in 1997

It is the first national survey dealing with reproductive health in national level. The survey divided into two stages. At first stage, there were 180 thousand samples selected, and then 15 thousand women in reproductive ages were selected from the samples of the first stage. The samples are nationally representative. As the women responded to their delivery history, we could calculate the TFR for previous years. The result of directly derived TFR is given at Figure 1.

Figure 1. Total Fertility Rate in China, 1990-1997

	1990	1991	1992	1993	1994	1995	1996	1997
TFR	2.35	1.75	1.57	1.55	1.35	1.36	1.38	1.35
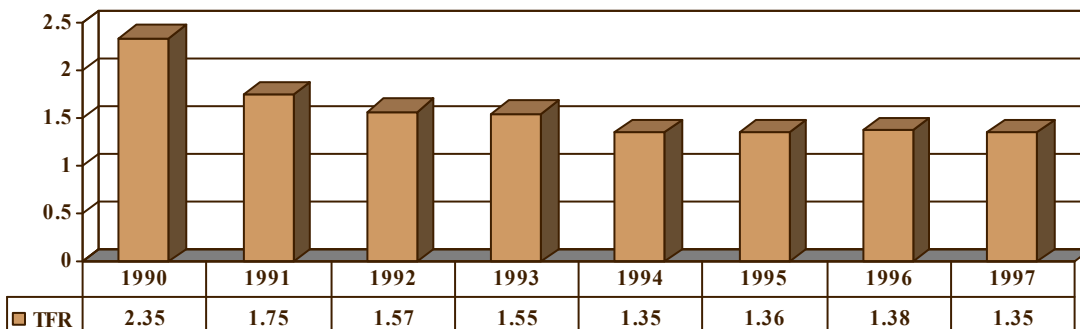
Source: The results were calculated directly from the raw data of the 1997 national population and reproductive health survey.

The TFRs directly calculated from the survey in late 1990s were too low to be accepted. However, something is curious that the TFR in 1990 seems appropriate, and TFRs in 1991 and 1992 were higher than those directly derived from 1992 survey. The TFR in 1995 was lower than that directly derived from 1995 survey.

From the example given above, we can see that almost all the fertility rates from the surveys in 1990s were quite low, no matter where, who, and how the surveys were conducted. How was such phenomenon produced in China? In brief, there are three reasons: 1) Couples did not report their unapproved children in order to avoid the financial charge. 2) The local leaders and the grassroots family planning personnel intended to underreport the births in order to increase their official achievement. 3) It was hard to collect the information of new births from floating population. There were 144 million floating peoples whose living place was separated from their household registration areas where their new babies should be counted or registered.

Data Adjustment in 1990s

As TFRs and CBRs directly derived from the surveys in 1990s were too low, NSB changed their strategy in data publication. In 1980s, the adjustment was made after 1990 census; we called the post adjustment. In 1990s the adjustment was made right after the time of survey so that the adjusted data can be published at the time just after the survey year; we called pre adjustment. For post adjustment, the correction can be accounted by the difference of the published data and the census data, which can be used as a standard of the adjustment. However, for the pre adjustment, it seems no direct clue and excuse to change the results of the original data, but one just noted that the published data were adjusted based on the sampling error and enumeration error. Unfortunately, researchers have never been officially informed on how big the sampling error and/or enumeration error in any sampling survey are.

We use the CBR as the indicator to estimate the sampling error and enumeration error. Based on the extent of upward adjustment of CBR in 1993 and 1994, we can see that there were about 2.45 per thousand points in CBR or over 2.9 million new births or 13.7% of underreporting rate in births per year added into the published CBR. Let me guess how the adjustment was made reasonably: "as knowing that underreporting of births was overwhelming due to the family planning program, they put all effort into adding more births to published data by taking the first upper bound of sampling error, nearly 1 per-thousand point, taking the second upper bound of enumeration error, nearly 2 per thousand point. All together less than 3 per thousand points of CBR, nearly 3 million births, were added into the total population each year." In fact, such a great amount of errors was not easily being added rationally and reasonably into the CBR. Unfortunately, the sampling error could be very small because the sample size in the survey was so huge and the samples were stratified which can decrease the variance. In addition, the inference to population from sample value is an interval estimate with a given confidence; point estimate is the estimate with highest probability in the interval; it would be absurd if one of extreme values were used as the estimate to the value of the population. In regard to the enumeration error, even though post enumeration has been used since 1992, the effect of the post enumeration can be only achieved under the condition that the error produced in the enumeration is random; if a conventional idea or behavior, which may lead to systematic error, such as subjective underreporting of the new birth, existed, the enumeration and the post enumeration would be affected to same extent unless the conventional idea or behavior could be changed before the post numeration. In fact, finding out an adequate evidence to adjust the CBR upward would be difficult in current China even though such adjustment were coincidently correct.

Such upward adjustment in CBR and total population produced some chaos in national data: 1) the summation of the provincial populations was no longer equal to the national population; 2) the CBRs in provinces were no longer consistent with the CBR in national level; 3) the micro data were not consistent with the aggregate data, which caused that the results derived from the micro data were different from the aggregate results.

How were the published data of 2000 census processed?

In 1990s, almost all the sampling surveys dealing with fertility showed lower estimates in CBR or TFR. All the published CBR, TFR, and total population provided by NBS were adjusted upward without any explanation on how the results were adjusted. Under such phenomenon, as time approaching to the 2000 census, NFPC and NSB held different attitude and hope to the result of the census.

Before the census, what NFPC worried was a more number of national populations produced than their hope or the number already published. There were some reasons: 1) severe underreporting of the births had lasted for many years, and there might be many unreported people out there who would be enumerated this time; 2) family planning is the national policy which highly emphasized by the central government, if the number of national population produced by the census were beyond what they knew, all the leaders in charge of the family planning program in all levels would be dismissed; 3) some researchers and people predicted that the number of population in China would reach 1.3 billion in 2000; such result might be possible. In order to avoid the unpleasant consequence happened, NFPC initiated a national work called "clarifying and checking" (qingli qingcha) in the spring of 1998. The purpose of the work was to recheck the new births in whole China in order to confirm if the unreported births were within their expected.

What the NBS worried was just opposite as NFPC worried. They worried about a less number of population produced by the 2000 census (Board of Population Research, 1999). The reason was that they had already artificially added almost 3 million new births each year, and the results had already been published by the annual Statistical Yearbook of China. All together, almost 30 million were added into the total population from 1990 to 1999 without any clear evidence if such amount of population really existed. If such amount of people added could not be shown up in the 2000 census or the total number of the population in 2000 was less than the total number published before, it would prove that the amount of the births added in the ten years were wrong, so that the leaders in NBS would be dismissed.

Who will win the game? The NFPC or the NBS?

NBS had an advantage over NFPC in the competition, because NBS was taking charge of the census and they could do whatever they wanted to do. One of NBS strategy was to encourage the central government issuing a document requiring the local government not to charge some one who had unplanned births once they reported the births. Unfortunately, such effort could not success because NFPC insisted that the population policy is the principle national policy (ji ben guo ce), and the document would be harmful to the implementation of the national policy.

What would the number of national population in November 2000 be mostly accepted by all parties? The mostly accepted number should be 1270 million, which would be consistent to the published data. If the added births from 1990 to 1999 could not exist, based on the directly enumerated population, the total number of national population would be 1246 million. If the reality were as the prediction of some researchers, the total number of the national population would be 1300 million. In brief, the acceptable interval of the number should be between 1246 to 1300 million, and the best one should be 1270 million. Due to the effect of family planning policy, the underreporting rate must be greater than the overreporting rate. Under such circumstance, what we, as scientific researchers, expect is that the total population should be greater than 1270 million, the more the better; less than 1270 million means that there would be some extent of underreporting, the less the worse.

As planned, the publication of major data of the 2000 census should be on January 1, 2001. However, the publication had been delayed for almost two months due to some reasons, and was finally published on March 28, 2001. In the first Data Report, it firstly mentioned that the total population of mainland China was 1265.83 million. Just based on this data, compared with what we expected above, this result seems close to the 1270 million but with over 4 million differences, which would mean that the data could be acceptable, but there were some extent of underreporting. Unfortunately, a note at the end

of this report broke our sweet dream, which mentioned: "After the enumeration, post enumeration was conducted at 602 sampled areas. The result is that the underreporting rate is 1.81%. The total number of the mainland population published has already included the population calculated through the underreporting rate." The note tells us three things: 1) the underreporting rate; 2) the rate is coming from the post enumeration; 3) the underreporting population has been added to the national population.

I wandered why NBS did not publish both the directly enumerated population and the error like the ways of previous census publication and publications of other countries but the population including the error, the way which has never been used before. The reason might be that they wanted to avoid the embarrassment by hiding the number of the directly counted population in the census. The directly enumerated population should exclude the underreporting population, 22.46 million (Zhang and Xu 2002), from the total population, 1265.83 million. Once we did this, we found that the directly enumerated population was only 1243.37 million, which is lower than the lower pound of our expected number, 1246 million, and less than the published national population at end of 1998, 1248.1 million. This means that the nearly 30 million people added to the total population published in 1990s by the NBS could not show up in the count of the census. It seems that the NBS has lost the game. However, the thing would not be as simple as this.

I wonder how the 1.81% underreporting rate came from. I doubt that the rate were not from the post enumeration. There are three reasons: 1) The Bylaw of Quality Control and Checking of the Census had regulated before the census that the net error of the total population should be less than 4‰, and required that if the error were greater than 4‰, the questionnaire should be refilled until it reached the requirement. Based on this regulation, the underreporting error should be unable to pass the 4‰. It is unimaginable that the overall error can be as high as 18.1‰, much higher than the census requirement. 2) In population census, usually, there are a set of errors such as underreporting rate, overreporting rate, and net underreporting rate (= underreporting rate – overreporting rate). For previous four censuses in China, all these error were published after the

censuses. However, in the Data Report of the 2000 census, it gave only one error, the underreporting rate. We did not know if the 1.81% underreporting rate indicated the underreporting rate or the net underreporting rate. If it indicated the underreporting rate, what was the overreporting rate? If it indicated the net underreporting rate, what was the underreporting rate? 3) China's previous censuses had very low net error. The net underreporting rates were 1.16‰ in 1953, 0.01‰ in 1964, 0.71‰ in 1982, and 0.6‰ in 1990. However, the underreporting rate suddenly increased to 18.1‰, 30 times higher than that in 1990. It seems impossible. 4) All most all the provincial underreporting errors were much lower than that of the national error, 4‰. For example, Beijing recounted 18 census areas (4 of them were the areas for national post enumeration), and the underreporting rate was 6‰, and the rate assigned to Beijing by the national census office were 18.1 ‰. In fact, NBS assigned most of the "underreported population", called Matching (xian jie) population, to provinces. Table 2 shows the assignment. You can see that the finally published provincial populations quite differed from the enumerated populations. In order to share the burden of the 2246 million national-artificially-drafted underreported populations, provinces had to accept the assignment given by the NBS. However, it is hard to find the reasons why they assigned the number of matching population quite different among the provinces.

Table 2. Enumerated Population, Matching Population, and Underreporting Rate by Provinces

Provinces	Published population[1] （Million）	Enumerated Population[2] （Million）	"Matching"Population[3] （Million）	"Underreporting Rate"[4] （%）
Beijing	13.82	13.57	0.25	1.81
Tianjin	10.01	9.85	0.16	1.60
Hebei	67.44	66.68	0.76	1.13
Shanxi	32.97	32.47	0.50	1.52
Neimeng	23.76	23.32	0.44	1.85
Liaoning	42.38	41.82	0.56	1.32
Jilin	27.28	26.80	0.48	1.76
Heilongjiang	36.89	36.24	0.65	1.76
Shanghai	16.74	16.41	0.33	1.97
Jiangsu	74.38	73.04	1.34	1.80
Zhejiang	46.77	45.93	0.84	1.80
Anhui	59.86	59.00	0.86	1.44
Fujian	34.71	34.10	0.61	1.76

Jiangxi	41.40	40.40	1.00	2.42
Shandong	90.79	89.97	0.82	0.90
Henan	92.56	91.24	1.34	1.43
Hubei	60.28	59.51	0.77	1.28
Hunan	64.40	63.27	1.13	1.75
Guangdong	86.42	85.23	1.19	1.38
Guangxi	44.89	43.85	1.04	2.32
Hainan	7.87	7.56	0.31	3.94
Chongqing	30.90	30.51	0.39	1.26
Sichuan	83.29	82.35	0.94	1.13
Guizhou	35.25	35.25	0	0.00
Yunnan	42.88	42.36	0.52	1.21
Tibet	2.62	2.62	0	0.00
Shannxi	36.05	35.37	0.68	1.89
Gansu	25.62	25.12	0.50	1.95
Qinghai	5.18	4.82	0.36	6.95
Ningxia	5.62	5.49	0.13	2.31
Xinjiang	19.25	18.46	0.79	4.10
Province total	1262.28	1242.61	19.67	1.56
Army	2.50	2.50	0	0.00
Prov. and Army Total	1264.78	1245.11[5]	19.67	1.56
National Matching			1.05[6]	0.08[7]
General Total	1265.83		20.72	1.64

Note: 1. Data from "Data Report 2". 2. Data from computer processing. 3. Equals to the published number minus the enumerated number. 4. Equals to the published population being divided by the matching population. 5. The data are not same as the enumerated because this one is from artificial processing, and the other is from computer processing. 6. This is the difference between published total and province and army total. 7. Obtained by 126583/105

If the underreporting rate were not exactly drawn from post enumeration, I wonder why they could not provide higher underreporting rate in order to make the total population close to 1270, rather than 1265.83, so that another 4 million added could be soaked up. The explanation of the president of NBS on the quality of 2000 census was that, based on the internationally admitted standard, the quality of the census data would be credible when the underreporting rate was under 2%, so the data of the 2000 census was credible (Zhu 2000). This might be the reason why the census office could not let the underreported error be greater than 2% by balancing the total number of the population.

The Adjustment for population in 1990s

Even though the underreporting rate introduced could make up the most of the losses of the total population in order to make the census result consistent to the published total, the lower TFR, 1.22, still revealed the problem of the census. However, as the published number could not reach 1270 million, the NBS has to adjust the published data in 1990s again. The original published national populations in 1990s can be found at 2001 China Population Statistical Yearbook edited by the Department of Population, Social, Scientific, and Technology Statistics, National Bureau of Statistics, but the numbers were changed in 2002 Yearbook. The difference from the adjustment of 1980s data is that the published totals have to be decreased, rather than to be increased (see Table 3). Interestingly, unlike the adjustment of the 1980s', they only adjusted CBRs in 1998 and 1999, and CBRs from 1991 to 1997 and CDRs in all 1990s could not be adjusted. This is understandable because only a few couple of million populations should be eliminated from the published number, so they do not need to touch all the data in 1990s.

Table 3. The Adjustment of Total Population and CBR, 1996-2000

Year	2001 Yearbook		2002 Yearbook	
	Total Population (Million)	CBR (‰)	Total Population (Million)	CBR (‰)
1996	1223.89	16.98	1223.89	16.98
1997	1236.26	16.57	1236.26	16.57
1998	1248.10	16.03	1247.61	15.64
1999	1259.09	15.23	1257.86	14.64
2000	1265.83	-	1267.43	14.03

Source: 1. Department of Social, Scientific, and Technology Statistics, NBS: China Population Statistical Yearbook 2001 and 2002. China Statistical Press, 2001 and 2002.

Conclusion

Every one knows that TFR = 1.22 in 2000 was too low to be accepted, and no one knows what the TFR should be since 1990. Without knowing the fertility, we would be probably not aware of the number of national population in China even though the census has given a number. In matter of fact, we cannot say that the adjusted population by NBS is far beyond the reality. It may be coincidentally perfect. But what we worried about is that the evidence for the correction is not sufficient, and the methods used for the adjustment

would be incorrect. It is normal that data from statistics produce some extent of errors. We are not afraid of the errors, but we are afraid of the inappropriate ways to deal with the errors. What we need to know is the reality, not something meshed in haze.

It is curious that it has been over 15 years that we do not know a relatively accurate or a commonly acceptable number of fertility rates in China since 1990 census. Official documents started using TFR as 1.8 to represent the national fertility level in the early 1990s. Ironically, this number, 1.8, has been still used in official documents now even though one has perceived that the fertility rate is decreasing. Current Chinese fertility policy is to "stabilize the low fertility". However, today no one knows how low the fertility is. Now it becomes a crucial issue to make clear of the numbers in fertility level and total population. Without knowing the numbers, population research in China seems impossible.

Reference:

Aird, John S. (1980). Reconstruction of an official data model of the population China, paper presented at Conference on Population of China, East-West Population Institute, East-West Center, Honolulu, May 1980.

Attane, Isabelle and Sun Minglei (1999). Birth rates and fertility in China, How credible are recent data? Population: An English Selection, Vol. 11: 151-260

Banister, Judith (1984). An analysis of recent data on the population of China. Population and Development Review 10(2): 241-271

Board of Population Research (1999). The Fifth population census: how to make the number correct. (Chinese) 1999 (6)

Coale, Ansley J. (1981). Population tends, population policy, and population studies in China. Population and Development Review 7(1): 85-97

Coale, Ansley J. (1984). Rapid population change in China, 1952-1982. Committee on Population and Demography, Report No. 27, Washington, D.C.: National Academy Press

Financial Department of NFPC (2003). Proceedings of documents of national family planning surveys (Chinese).

Hu, Ying (2005). China's annual population survey: a review. Journal of Population Research (Chinese), 29(1):37-42

Jia, Tongjin (1995). Analysis on the current situation and problems of the sampling survey of population changes. Journal of Population Research (Chinese), 1995(5)

Liu, Zheng (1980). The recent selection and development of China's population, paper presented at Beijing International round Table Conference on Demography, October 1980

Merli, M. Giovanna and Adrian E. Raftery (2000). Are births underreported in rural China? Manipulation of statistical records in response to China's population policies. Demography, 37(1): 109-126

Qiao, Xiaochun (1995). Research on Population Censuses in China. Population Publishing House (Chinese)

Qiao, Xiaochun (2002). Some Problems of the Fifth Population Census in Reference to the "Major Figures" (Chinese). Chinese Population Science, 2002(5))

Zhang, Weimin, Yu Hongwen, and Chui Hongyan (1997). Current changes of China's population. Symposium on Demography of China, China Population Association (Chinese), Beijng October 1997: 36-71

Zhang, Weimin and Xu Gang (2002). Effect and challenge of the implementation of the 2000 census, paper presented at the 8th Conference of the National Population Science (Chinese), June, Beijing

Zhu, Zhengzhi (1980). China's population problem at present and its developing trend. Jingji Kexue (Chinese), March 1980.

Zeng, Yi (1995). Is fertility far below replacement in 1991 and 1992 in China? Journal of Population Research (Chinese), 19(3): 7-14