The Role of Population Estimates and Projections in the Evaluation of Censuses

Dr. Peter D. Johnson
Dr. Peter O. Way
International Programs Center
Population Division
U.S. Census Bureau

Paper presented at the 2003 Population Census Conference of the Association of National Census and Statistics Directors of America, Asia, and the Pacific in Kyoto, November 19-21, 2003

The Role of Population Estimates and Projections in the Evaluation of Censuses

Introduction

When the results of a new census become available, there may be a tendency to immediately move to the analysis and then use the results as the basis for new yearly estimates as well as longer-term projections of the population. This paper will discuss several reasons why it is a good idea to pause and reflect before abandoning estimates and projections based on an earlier census to be sure that all the lessons learned from the comparison to the new census can be appreciated. We will explore several processes where the comparison of census and population estimates or projections can be useful and also will look at examples of these processes and pose some questions about further research. Our goal is to increase discussion of lessons learned by comparing census results to population estimates and projections.

We often compare our jobs as demographers to that of detectives. We can think of the true demographic situation in a country as an object inside a box with small windows on the outside. Each demographic observation can be thought of as a little window that lets us look inside from a slightly different perspective, providing another clue about what is inside. However, through each window we only see part of the object, and the windows may have defective glass that distorts the picture slightly. The demographer's job is to reconcile all the views of the population, taking into consideration the quality of each view.

Figure A shows some of the interrelationships between population censuses and population estimates or projections. Estimates and projections (referred to hereafter as estimates) are obviously important in preparing for a census in order to gauge the resources needed for the census (number of questionnaires to print, enumerators to hire, etc.). They should also be used (especially at the subnational level) as results begin to come in as a check on the results and to identify areas where extra resources may be needed.

(Figure A about here)

The focus of this paper, however, is on the use of the results of a new census, and will be on the following processes:

A. Post Enumeration Survey (PES) evaluation of census results
B. Evaluation of census results based on estimates or demographic analysis
C. Evaluation of intercensal components of growth
D. Adjustment of postcensal components of growth
E. Evaluation of prior census results
F. Intercensal estimates (mathematical formula)
G. Intercensal estimates (based on adjusted censuses and/or components)

Once the census has been processed, the comparisons of the census results to estimates or other demographic analysis procedures (process B) can provide an assessment of the quality of the census, including estimates of the net census error by age and sex. Ideally, this assessment can be compared to the results of a post enumeration survey (PES, process A) to obtain a clearer picture of the population at the time of the census.

Very little in demography is completely accurate or certain, so one can also turn the census/estimate comparison around and use the census to evaluate the estimates. At the International Programs Center of the U.S. Census Bureau (IPC), this often takes the form of using the residual method to estimate net migration during the intercensal period (process C). These estimates are often used directly in the postcensal estimates or used to adjust other data on components (process D).

The new census can also be used to evaluate the previous census (process E). The underenumeration of young children, a frequent problem in censuses, can often be checked using this method.

The final processes we will discuss are the creation of intercensal estimates (processes F and G). Here we distinguish between intercensal estimates (those that benefit from census data for earlier and later dates) and postcensal estimates (those produced after a census, but before any data are available from a later census). Ideally, adjusted censuses and components of growth resulting from processes A, B, C and E can form the basis for creating the intercensal estimates in a demographically-based way (process G). However, this is usually not done. Rather, a more simplistic mathematical formula (process F) is used to derive the intercensal estimates (either totals or by age and sex).

There exist multiple alternative explanations (i.e., fertility, mortality, migration, under- or overenumenumeration) for observed differences between E/P and a census that theoretically could explain any of the differences. The challenge is to use the information we have to narrow down the alternatives with the "final" analysis being the solution we think is most plausible – i.e., most plausibly reflects demographic reality – taking into account the available data, and recognizing the strengths and weaknesses of the alternative data sources. There exists a single "right" answer to this demographic puzzle. But we should resign ourselves to the reality that it will never be known. However, for many reasons, it is important to continue to strive to improve our estimates and our census processes, and to reduce the sources of error that undermine both. In our experience, the full exploitation of the two-way street of censuses and estimates will enhance both.

We will now briefly consider how estimate/census comparisons have been used in the demographic analysis of different countries. These analyses are a combination of work done by the U.S. Census Bureau in evaluating its own census, a review of publicly available materials (printed or online), and analysis done at the International Programs Center of the U.S. Census Bureau.

United States Census 2000

When the Census 2000 total population came in at 281.4 million, it was almost 7 million higher than the estimates based on the 1990 census. There were several explanations for the difference:

a. Census 2000 had better population coverage than the 1990 census
b. Census 2000 was overenumerated
c. The postcensal estimates were too low

In order to find which explanation (or combination) was correct, the U.S. Census Bureau conducted both a PES and demographic analysis (DA). The U.S. Census Bureau has a long history of census evaluation, with PES and DA evaluations going back to 1950. For Census 2000, the PES was called the Accuracy and Coverage Evaluation survey (A.C.E.). The DA estimates and postcensal estimates at the Census Bureau are independently prepared using different methodologies. The postcensal estimates move forward from the previous census (without adjustment for undercount) using components of growth, while the DA estimates are constructed independent (for the most part) of census data. DA estimates are based on administrative records of births and deaths (with adjustments as needed), recorded immigration (from the U.S. Citizenship and Immigration Services formerly known as the U.S. Immigration and Naturalization Service (INS)), estimates of other migration, and data from Medicare for the population 65 years and over. The DA estimates implicitly correct for net undercount in the census. A difficulty with the DA estimates is the lack of accurate data on emigration and undocumented immigration.

The Executive Steering Committee for the A.C.E. Policy (ESCAP – not the one in Bangkok) directed an extensive evaluation of both the A.C.E. and DA estimates . The initial estimates of the net census errors were quite different (see table 1 and Figures B-1 and B-2): a net undercount of 3.3 million (1.15 percent) from the ACE and a net overcount of 1.8 million (0.65 percent) from DA. Concern over the underestimation of undocumented migration resulted in an Alternative DA which produced an estimated net undercount of 0.9 million (0.32 percent). Had only one estimate of coverage been available, the decisions regarding adjustment of the census may have been misguided.

Due to the discrepancies between DA and A.C.E., further work was done to validate the DA results. The revised DA estimate implied a net undercount of 0.3 million (0.12 percent). Preliminary review of the A.C.E. resulted in an estimated net undercount of 0.06 percent. This prompted further review of the A.C.E. results, which resulted in the A.C.E. II estimate of a net overcount of 1.3 million (0.48 percent).

(Table 1 and Figures B-1 and B-2 about here)

In the end, all three explanations of the discrepancies between the estimates and Census 2000 were partially correct: Census 2000 did show higher levels of population coverage than earlier censuses, but that was partly due to overcounts balancing undercounts. In

addition, the population estimates were found to be too low, due primarily to underestimating international migration.

As a result of this research, work is already underway to plan for the evaluation of the 2010 census to correct some of the errors found in Census 2000. In addition, the detailed DA may help in finding ways to improve postcensal components of growth (process D) and the postcensal estimates.

The 1990-2000 intercensal estimates (process F) were calculated using a mathematical procedure that redistributes the difference between the census and estimates over the period (U.S. Census Bureau, 2002a). The procedure is referred to at the Census Bureau as the “Das Gupta method 6” or the "constant exponential adjustment method."

The mathematical expression for this method is:

$$P_t = Q_t \left(\frac{P_{10}}{Q_{10}} \right)^{(t/120)}$$

Where

t = time (in months) elapsed since April 1, 1990
P_t = Population estimate at time t
Q_t = Postcensal estimate at time t
P_{10} = April 1, 2000 Census Count
Q_{10} = April 1, 2000 Postcensal Estimate

Alternatively it can be expressed as:

$$P_t = Q_t e^{rt}$$

Where

$$r = \frac{1}{10} \ln \left(\frac{P_{10}}{Q_{10}} \right)$$

and time, t, is expressed in years.

This is a very convenient way to construct yearly population estimates that are consistent with two census results (either enumerated or adjusted). The problem with this method is that it does not try to assign the differences to either a component (e.g. mortality, fertility, or migration; process C) or to differential coverage (processes A, B, and E). Process C (evaluation of the components of growth in the intercensal period) could be beneficial to improving the postcensal estimates by trying to find out where the errors were located.

Japan Intercensal Estimates

We did some preliminay research on how other developed countries in the region deal with intercensal estimates after a new census. Japan's approach to the 1995-2000 intercensal estimates (process F) is very straightforward (Japan Statistics Bureau, no date a): simply divide the difference between the census and the estimate by the number of years (or months) in between. Japan has taken censuses every five years since before 1950 and this same approach has been used for the intercensal periods since 1970 (Japan Statistics Bureau, 2003). For the period 1950 to 1970 the intercensal adjustments are close to constant for each year, but not quite. Our limited review of the available sources in English did not reveal any further research into the sources of the discrepancies (e.g. differential coverage, unmeasured migration) or any attempt to correct the intercensal time series of individual components for the errors discovered. We would welcome discussion of the causes of the discrepancies and any analysis that may have been done on this.

Figure C shows the pattern of intercensal adjustments for the years 1950 to 2000. The pattern does not show any discernable trend. In fact, the differences between successive intercensal adjustments always seem to change direction: if it goes down in one period it goes up in the following period. It might be interesting to compare these figures to the intercensal natural increase and international migration figures to see if there is any correlation.

(Figure C about here)

IPC Estimates of Intercensal Migration in the Northern Mariana Islands , 1995 to 2000

Some of the Pacific island areas (e.g. Northern Mariana Islands, American Samoa, and Guam) vary in the quantity and quality of data. As with many countries, the component least well measured is migration. For several of these countries IPC used the difference between the estimates and the census as an estimate of intercensal net migration (process C) and used those results to inform the projections beyond the recent census (process D). Figures D-1 and D-2 show for Northern Mariana Islands the comparison of the Census 2000 results to estimates without and with the residual migration included. In this case we fit the Rogers and Castro migration model to the residual migrant numbers by age to smooth out the irregularities. You can see that there is close agreement between the estimates including migration and the census.

(Figure D-1 and D-2 about here)

Comparison of IPC Estimates to China 2000 Census Results

Comparisons of IPC estimates for China to the 2000 census results (process B) implied significant levels of undercount in the population under age 10 (Figure E). We based our estimates on official adjusted crude birth rates that include adjustment for underreporting of births and mortality based on data from the 1990 census. Although we include net international migration by age and sex, we do not have detailed data to estimate this. If the 2000 census results for children are accurate, it would imply that the births were overreported, which is not likely. Results of the 2000 census may also be used to estimate undercounts of young children in the 1990 census (process E) that may be the cause of implied overcounts at ages 10-19 in the 2000 census. The implied undercounts of young children in the 1990 census are far lower than those in the 2000 census.

(Figure E about here)

Comparison of IPC Estimates to the Philippines 1995 and 2000 Censuses

The comparison of IPC estimates to the 1995 and 2000 censuses of the Philippines (Figures F-1 and F-2) generally implied underenumeration of children under age 5, men 15 to 29, and women 15 to 79. Both males and females ages 80 years and over appear to be overenumerated, as are men 35-49 or 35-54. Given the frequent underenumeration of young children, we would tend to adjust the census in that age group. The implied underenumeration of young adults (ages about 15-29) may point to an underestimate of net out migration in this age group, although this group can also be difficult to enumerate due to changing circumstances and/or migration related to higher education, starting a new job, or getting married and starting a family. The differential in the older adult ages: implied underenumeration for women 35 to79 but overenumeration for men 35 to 49 or 54 is more difficult to explain. One explanation could be an underestimate of return migration of men, and an underestimate of permanent out migration of women.

(Figures F-1 and F-2 about here)

Conclusion

In this brief presentation, we hope that we have been able to show enough results from various national settings to convince you of the value in comparing the "old" postcensal estimates to a new census. These estimates can tell us a lot:

- About the new census
- About errors in the estimates (including the previous census)
- About "correcting" our demographic picture of our own or of other countries to better understand where we have been, where we are, and where we are going, demographically speaking.

References

Japan Statistics Bureau. 2003. Japan Statistical Yearbook 2003. Tables 2-1B and 2-3.

Japan Statistics Bureau. no date a. “Intercensal Adjustment of Current Population Estimates (1995 - 2000).” <http://www.stat.go.jp/english/data/jinsui/2-1.htm>. (Accessed November 4, 2003).

Japan Statistics Bureau. no date b. “Population by Sex for Japan.” <http://www.stat.go.jp/english/data/jinsui/zuhyou/15k4-1.xls>. (Accessed November 4, 2003).

U.S. Census Bureau. 2001a. "Monthly Estimates of the United States Population: April 1, 1980 to July 1, 1999, with Short-Term Projections to November 1, 2000." <http://eire.census.gov/popest/archives/national/nation1/intfile1-1.txt>

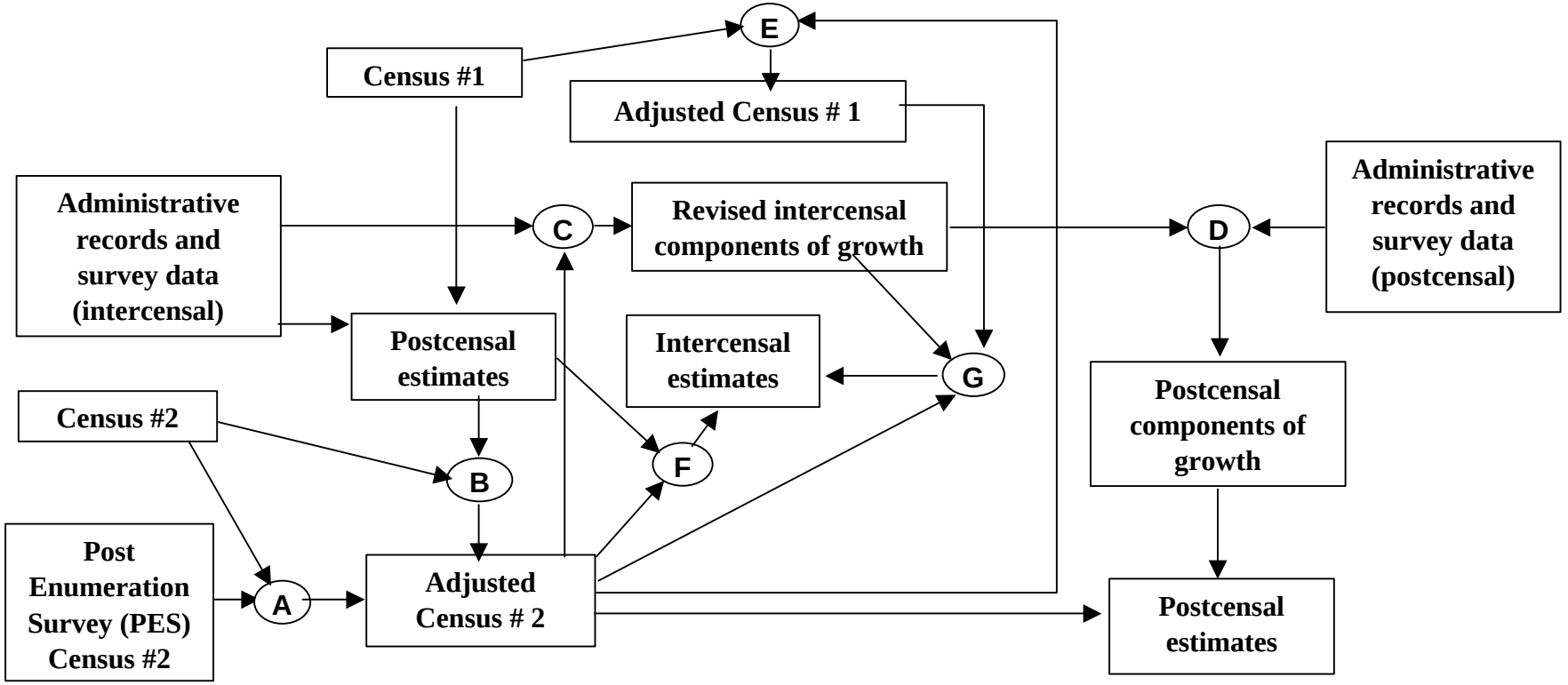
U.S. Census Bureau. 2001b. "Basis of 'Revised Early Approximation' of Undercounts Released October 17, 2001." by John H. Thompson, Preston J. Waite, and Robert E. Fay. <http://www.census.gov/dmd/www/pdf/report9a.pdf>

U.S. Census Bureau. 2002a. “National Intercensal Estimates.” <http://eire.census.gov/popest/data/national/intercensal.php>. (Last updated: September 13, 2002).

U.S. Census Bureau. 2002b. "Comparison of A.C.E. Revision II Results with Demographic Analysis." by J. Gregory Robinson and Arjun Adlakha. DSSD A.C.E. Revision II Estimates Memorandum Series #PP-41. <http://www.census.gov/dmd/www/pdf/pp-41r.pdf>

U.S. Census Bureau. 2003. "Coverage Measurement From the Perspective of March 2001 A.C.E." by Rita Petroni and Danny R. Childers. Census 2000 Testing, Experimentation, and Evaluation Program. Topic Report Series, no. 4. Table 2, p. 12 <http://www.census.gov/pred/www/rpts/Cov_Meas_TR_Final.PDF>

Figure A. Interrelationships between Population Censuses and Estimates



Boxes are results
Circles are processes:

- **A. Post Enumeration Survey (PES) evaluation of census results**
- **B. Evaluation of census results based on estimates or demographic analysis**
- **C. Evaluation of intercensal components of growth**
- **D. Adjustment of post-censal components of growth**
- **E. Evaluation of prior census**
- **F. Intercensal estimates (mathematical formula)**
- **G. Intercensal estimates (based on adjusted censuses and/or component)**

Table 1. U.S. Census Population and Alternate Population Estimates: April 1, 2000

Source (release date)	Population	Difference	
		Count	Percent
Enumerated (December 2000)	281,421,906		
Estimates (January 2001)	274,520,000	6,901,906	2.51
Base DA (March 2001)	279,598,121	1,823,785	0.65
Alternative DA (March 2001)	282,335,711	-913,805	-0.32
Revised DA (October 2001)	281,759,858	-337,952	-0.12
A.C.E. (March 2001)	284,683,782	-3,261,876	-1.15
A.C.E. revised (October 2001)			-0.06
A.C.E. II (Mar 2003)	280,090,250	1,331,656	0.48

Notes:
The difference count is defined here as enumerated census minus alternate estimate.
The difference percent is defined here as the census difference divided by the alternate estimate.
A.C.E. difference figures are for the household population.
The "A.C.E. revised (October 2001)" percent difference was calculated relative to the household population

Sources:

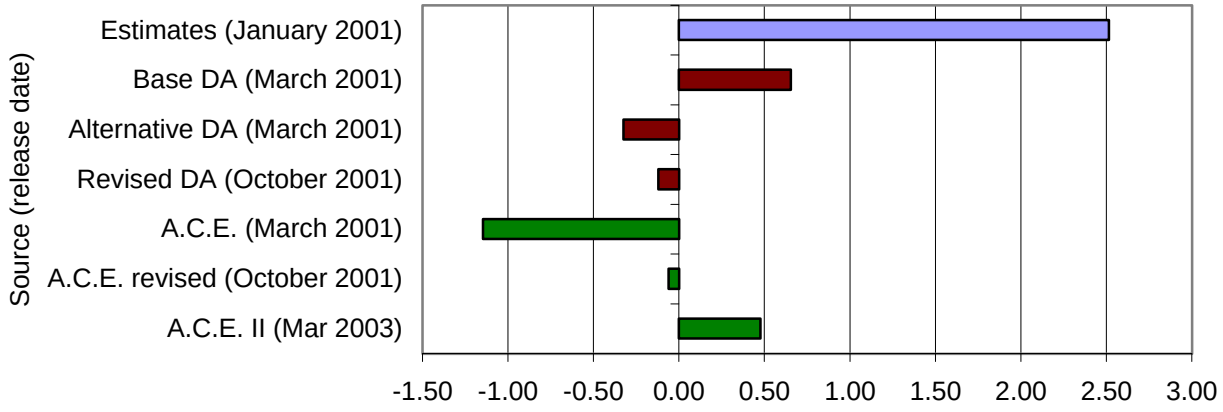
U.S. Census Bureau. 2001. "Monthly Estimates of the United States Population: April 1, 1980 to July 1, 1999, with Short-Term Projections to November 1, 2000." <http://eire.census.gov/popest/archives/national/nation1/intfile1-1.txt>

U.S. Census Bureau. 2001. "Basis of 'Revised Early Approximation' of Undercounts Released October 17, 2001." by John H. Thompson, Preston J. Waite, and Robert E. Fay. <http://www.census.gov/dmd/www/pdf/report9a.pdf>

U.S. Census Bureau. 2002. "Comparison of A.C.E. Revision II Results with Demographic Analysis." by J. Gregory Robinson and Arjun Adlakha. DSSD A.C.E. Revision II Estimates Memorandum Series #PP-41. <http://www.census.gov/dmd/www/pdf/pp-41r.pdf>

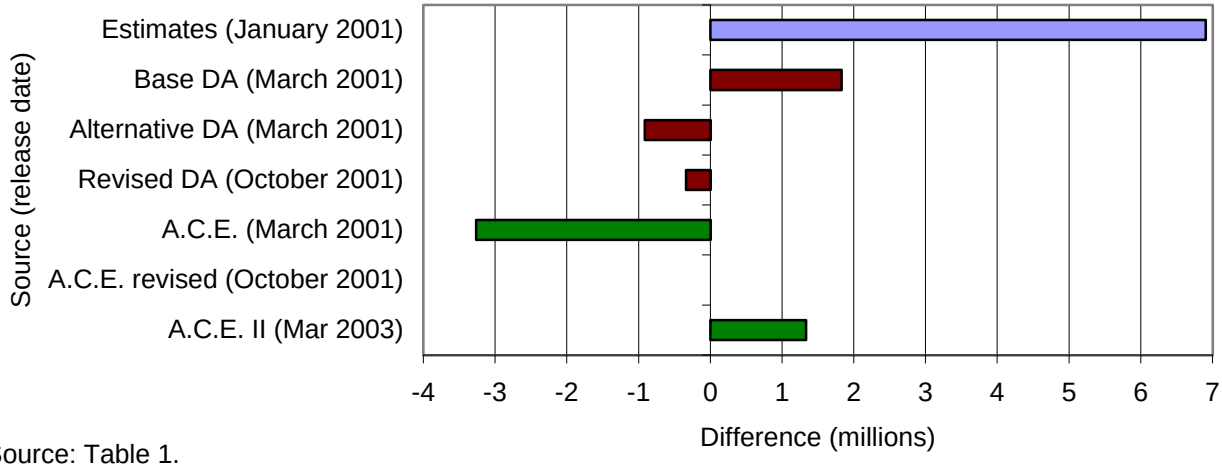
U.S. Census Bureau. 2003. "Coverage Measurement From the Perspective of March 2001 A.C.E." by Rita Petroni and Danny R. Childers. Census 2000 Testing, Experimentation, and Evaluation Program. Topic Report Series, no. 4. Table 2, p. 12 <http://www.census.gov/pred/www/rpts/Cov_Meas_TR_Final.PDF>

Figure B-1. Percent Difference Between Census Count and Alternative Estimates for the United States: 2000



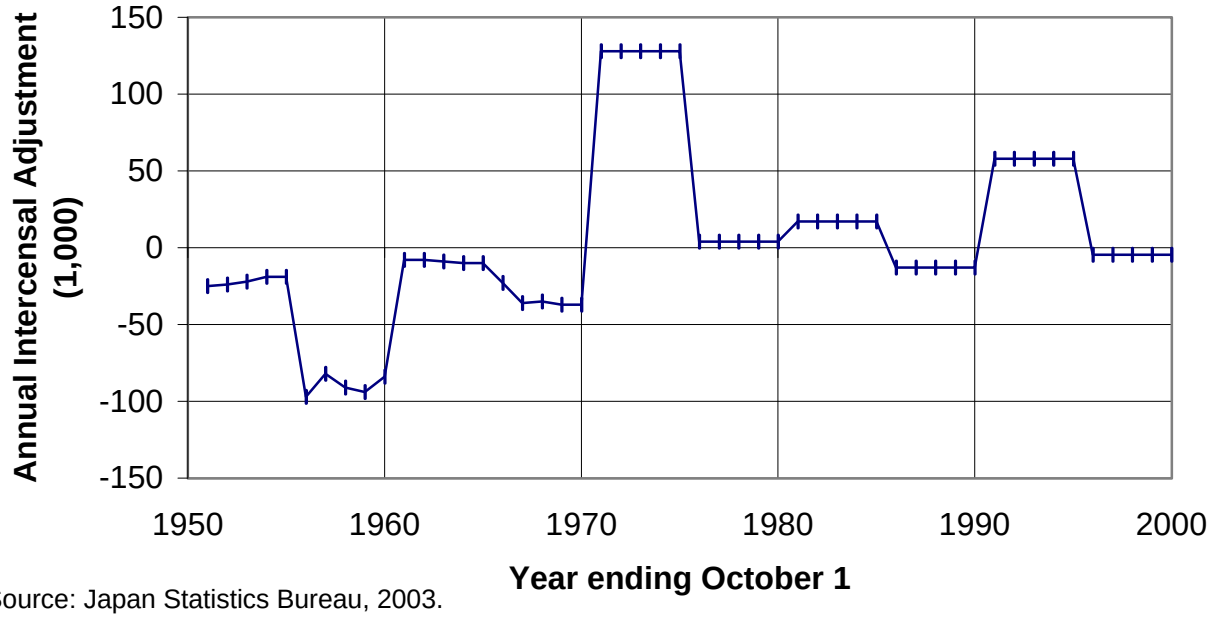
Source: Table 1.

Figure B-2. Difference Between Census Count and Alternative Estimates for the United States: 2000



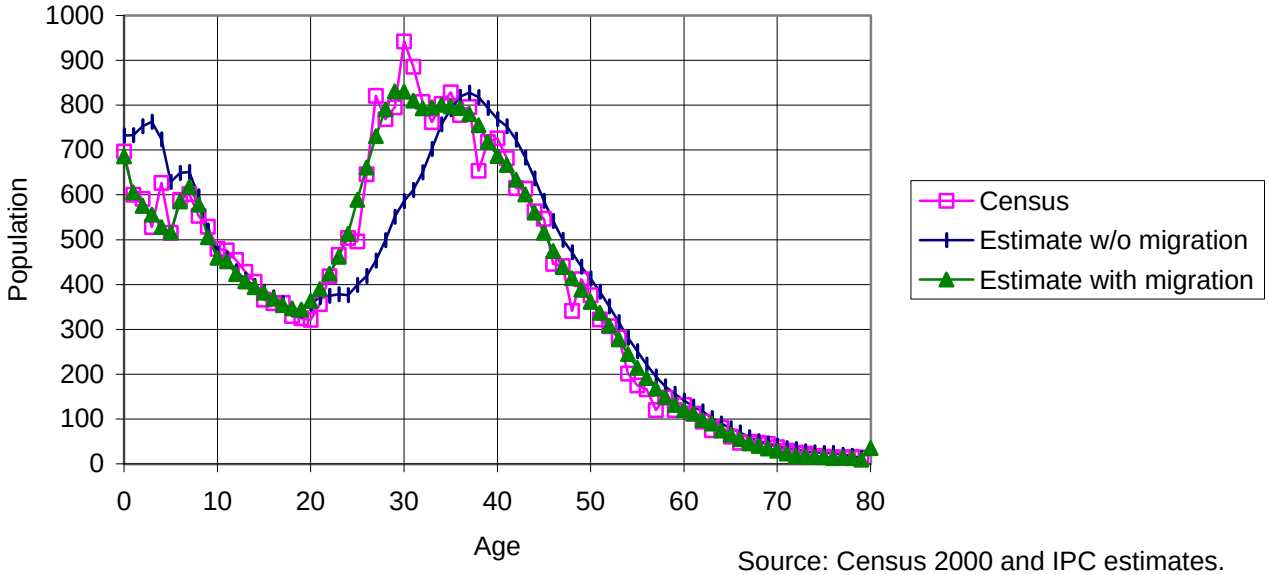
Source: Table 1.

Figure C. Annual Intercensal Population Adjustment for Japan: 1951-2000



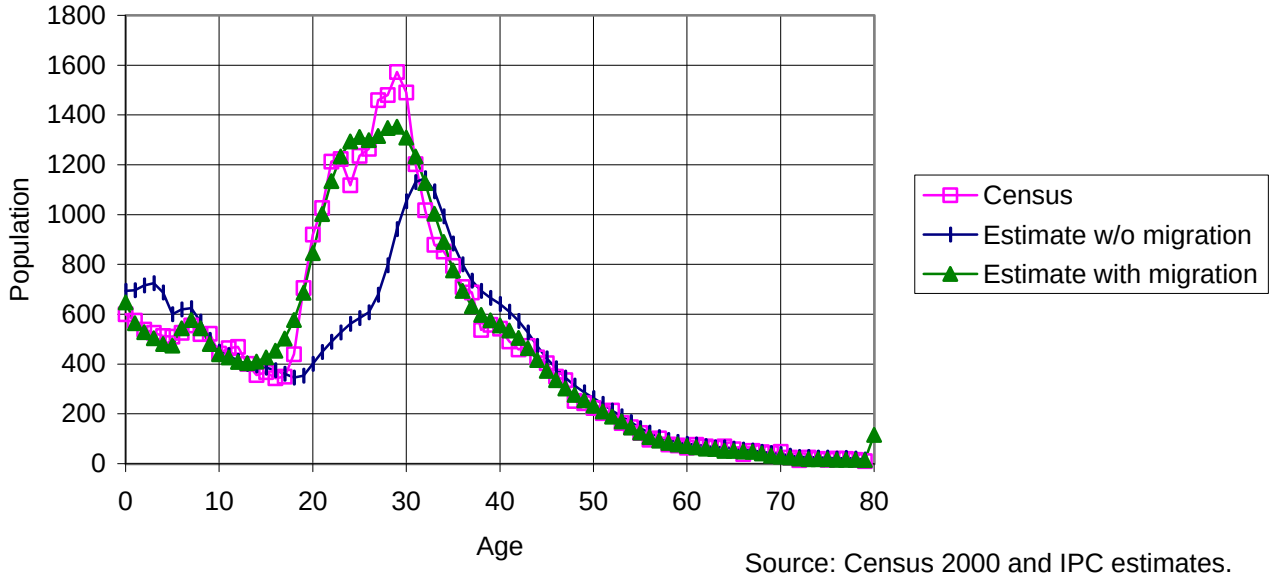
Source: Japan Statistics Bureau, 2003.

Figure D-1. Comparison of Census 2000 to Estimates of the Male Population of the Northern Mariana Islands



Source: Census 2000 and IPC estimates.

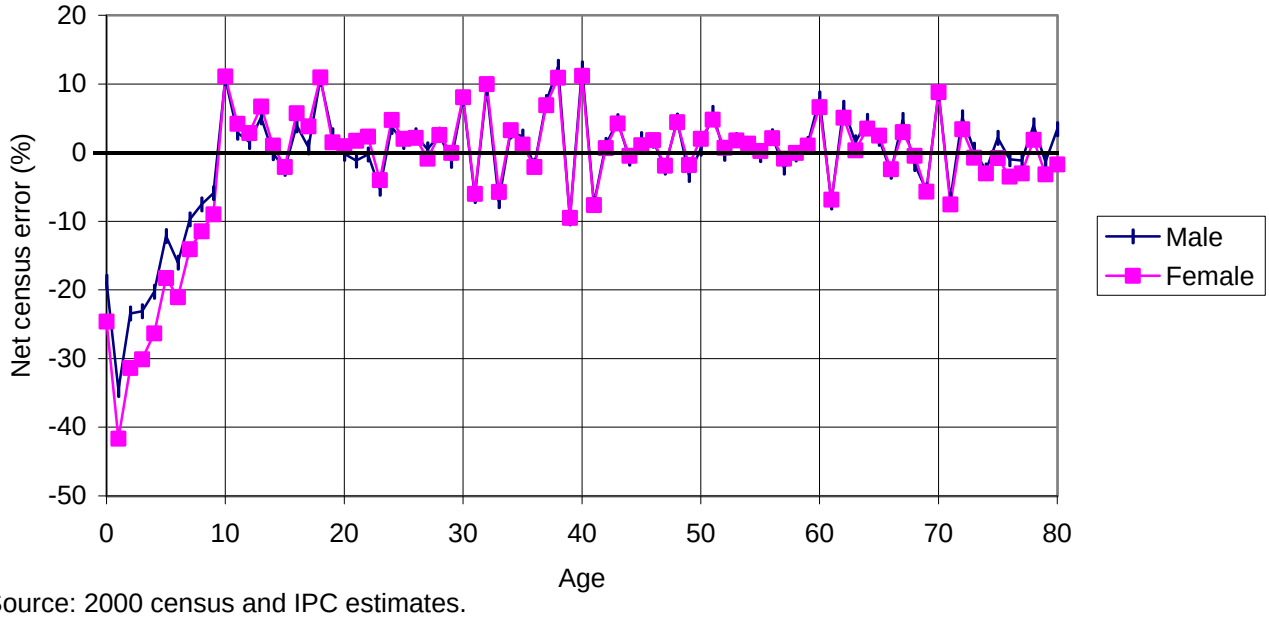
Figure D-2. Comparison of Census 2000 to Estimates of the Female Population of the Northern Mariana Islands



Source: Census 2000 and IPC estimates.

Figure E. Implied Net Census Error (Percent) For China 2000 Census

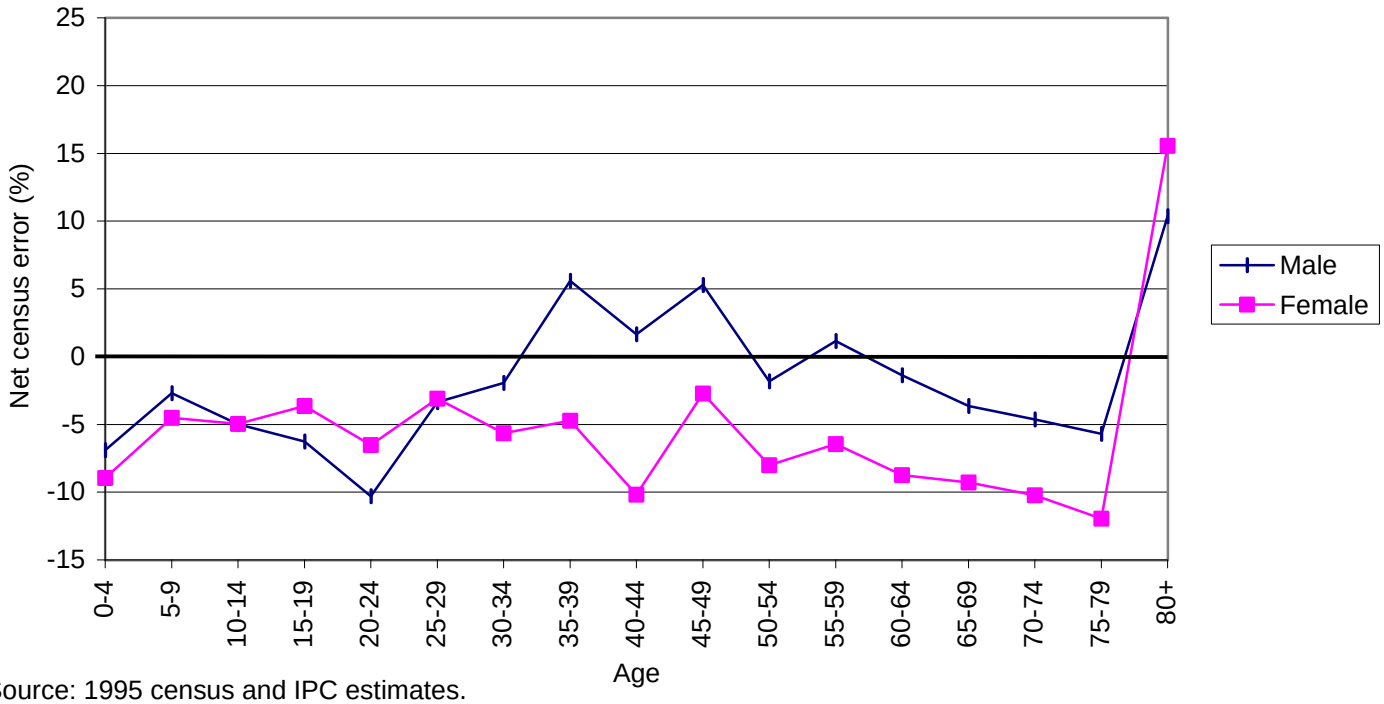
(A minus sign denotes a net census undercount.)



Source: 2000 census and IPC estimates.

Figure F-1. Implied Net Census Error (Percent) for the Philippines 1995 Census

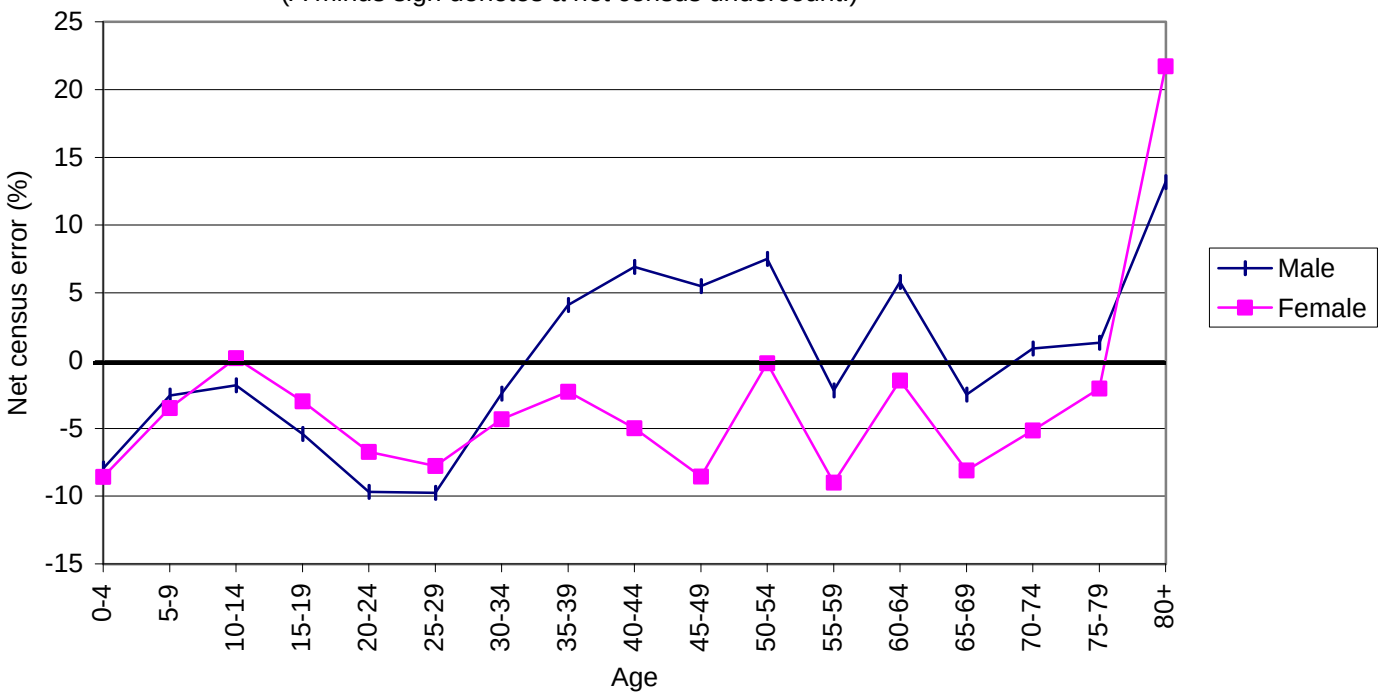
(A minus sign denotes a net census undercount.)



Source: 1995 census and IPC estimates.

Figure F-2. Implied Net Census Error (Percent) for the Philippines 2000 Census

(A minus sign denotes a net census undercount.)



Source: 2000 census and IPC estimates.