RESEARCH REPORT SERIES
(Statistics #2006-3)

Summary of Accuracy and Coverage Evaluation for Census 2000

Mary Mulry

Statistical Research Division
U.S. Census Bureau
Washington, DC

Report Issued: February 28, 2006

Summary of Accuracy and Coverage Evaluation for Census 2000

Final Report
June 8, 2005

Mary H. Mulry
Statistical Research Division

Table of Contents

1.	Introduction	1
2.	U.S. Census	2
2.1	Constitutional purpose of the census and other uses	2
2.2	Need for coverage measurement	3
3.	Census 2000	4
3.1	Universe for Census 2000	4
3.2	Phases in taking Census 2000	5
3.2.1	Address list construction	5
3.2.2	Questionnaire mail strategy	5
3.2.3	Nonresponse Followup	6
3.3	Multiple ways to respond	7
3.4	Short and long Census forms	7
3.5	Unclassified units and missing data	8
4.	Estimates of census coverage error	8
4.1	Types of census coverage error	8
4.2	Methods to measure census coverage error	9
4.3	Coverage error in Census 2000	12
5.	Coverage evaluation survey for Census 2000	17
5.1	Design of Accuracy and Coverage Evaluation Survey	17
5.2	Data collection for A.C.E.	18
5.3	Matching operation and followup for A.C.E.	18
5.4	Revision necessitated by undetected duplication and other measurement error	19
5.5	Research for A.C.E. Revision II	20
5.6	Description of dual system estimator used for A.C.E. Revision II estimates	22
5.7	Small area estimates	24
6.	Assessment of quality of A.C.E. Revision II estimates	25
6.1	Comparison with Demographic Analysis	25
6.2	Evaluation of A.C.E. Revision II estimates	29
6.3	Concerns about estimates from A.C.E. Revision II	31
7.	Historical perspective	32
8.	Summary	34
8.1	Influence of A.C.E. Revision II on planning of the 2010 Census	34
8.2	Coverage measurement for the 2010 Census	35
References		36
Appendix A		43

Summary of Accuracy and Coverage Evaluation for Census 2000

1. Introduction

The U.S. Census Bureau evaluated how well Census 2000 counted the population by conducting a coverage measurement survey known as the Accuracy and Coverage Evaluation survey (A.C.E.). The Census Bureau considered adjusting the Census 2000 population total of 281,421,906 to correct for coverage error on three occasions, but each time decided not to adjust.

As part of the preparations for Census 2000, the Census Bureau planned to decide in March 2001 whether to adjust Census 2000 numbers legally required to be produced by April 1, 2001 for the purpose of redistricting of the U.S. Congress (Prewitt 2000). Initially the A.C.E. measured a 1.2 percent net undercount, but after much debate, the Census Bureau decided not to adjust Census 2000 for redistricting. The reason for not adjusting was that discrepancies between the A.C.E. estimate and another method of measuring census coverage, Demographic Analysis, could not be explained by the deadline for the decision. Demographic Analysis(DA) initially produced a base estimate of a -0.7 percent net undercount (an overcount). Demographic Analysis subsequently produced an alternative estimate 0. 3 percent net undercount with an alternative assumption that net undocumented migration was twice as much as included in the base estimate (Robinson 2001). For the 1980 and 1990 censuses, the estimates of the percent net undercount from the two programs had been within 0.2 percentage points. The discrepancies in the estimates of Census 2000 coverage error caused the Census Bureau to not be sure that an adjustment would improve the census numbers (U.S. Census Bureau 2001b).

Next, the Census Bureau planned to consider adjusting Census 2000 for purposes other than redistricting when the results of evaluations planned for the summer of 2001 became available (U.S. Census Bureau 2001c). Also during that summer, an extensive Demographic Analysis research program addressed the discrepancy between its estimate and the A.C.E. estimate. The research produced a Revised Demographic Analysis estimate of 0.1 percent net undercount. The decision in October 2001 was to use the unadjusted census for purposes other than redistricting because the evaluations of the A.C.E. found a large number of erroneous enumerations that the A.C.E. failed to detect (U.S. Census Bureau 2001a). At that time the Census Bureau issued an A.C.E. Preliminary Revised estimate of a 0.06 percent net undercount (Thompson, Waite, and Fay 2001). The Census Bureau also decided to conduct further research on census duplication and other measurement errors and planned to consider whether to incorporate a revised net undercount estimate into the census base used in the intercensal estimates program.

The revision, known as A.C.E. Revision II, estimated the percent net undercount to be -0.5 percent, (an overcount). Although the A.C.E. Revision II estimates were believed to be an improvement over the A.C.E. estimates, the Census Bureau also believed the revised estimates contained troubling anomalies and unexplained results. The Census Bureau was not confident an adjustment would improve the census population numbers for the level of geography for which the intercensal estimates are produced, namely counties and places. The decision was not to adjust the census base for the intercensal estimates (U.S. Census Bureau 2003c).

This document gives an overview of the methodology used to estimate the error in the coverage of Census 2000 and presents the results. Also, there is information about the methodology used in taking Census 2000 as well as a discussion of the quality of the A.C.E. Revision II estimates. A report issued by the U.S. Census Bureau (2004)contains detailed explanations of the design and methodology for A.C.E. and A.C.E. Revision II. Another source for discussion and critique of Census 2000, the A.C.E., and A.C.E. Revision II is a report by the National Research Council (2004).

2. U.S. Census

Understanding the purpose of the census and its uses sheds light on the concerns about the quality of the coverage of the census.

2.1 Constitutional purpose of the census and other uses

A census of the population is required by the U.S. Constitution every 10 years for the purpose of apportionment of the U.S. House of Representatives. The number of representatives a state receives depends on the size of a state's population relative to the size of the population of other states. In contrast, the Senate contains two Senators from each state, regardless of the size of the state. The framers of the Constitution constructed the Congress with the House of Representatives and the Senate to balance the interests of the large and small states.

The House of Representatives selects the algorithm that determines the number of representatives from each state. Currently the number of seats in the House of Representatives is 435. An algorithm is necessary since only whole seats can be assigned to states. If portions of seats could be assigned, then the number of seats would be the percentage of the U.S. population in a state multiplied by 435. Over the years, different methods have been used to assign the number of representatives to states. In fact, a new apportionment was not done after the 1920 Census because the House of Representatives could not agree on a method. At that time, a large migration from rural states to urban areas was underway causing the more rural states to lose enough population to reduce their number of representatives. For the 1930 Census the House of Representatives chose a method known as Hill's algorithm, and this algorithm is still in use (Balinski and Young 2001) and described in Appendix A. The Supreme Court ruled that Hill's algorithm was fair as recently as 1992 (Department of Commerce et al 1992).

Hill's algorithm in effect minimizes the quantity

$$\sum_i a_i \left(\frac{P_i}{P_+} - \frac{a_i}{435} \right)^2$$

where P_i is the population for state I, P_+ is the total census count for the 50 states, and a_i be the number of representatives a state receives (Balinski and Young 1982, Spencer 1985).

This perspective on the algorithm emphasizes the importance of the accuracy of a state's proportion of the population, called its population share, as well as its number of people. Since a state's population share affects the number of seats it receives in the House of Representatives, the share also affects the state's representation in the Electoral College where the votes for President are cast. The number of votes a state receives in the Electoral College is equal to the number of representatives it has in Congress, which is comprised of the number of seats in the House of Representatives plus the two seats that every state receives in the Senate. The District of Columbia receives three votes in the Electoral College.

The census numbers also are used by states in forming the districts for electing their Members of the House of Representatives, called redistricting. In addition, states use census numbers in forming districts for their legislatures, and local governments draw districts for their elected officials using census numbers. The states could do their own censuses and use those numbers for constructing districts. However, the expense of a census is very large and hard to justify when the U.S. Census numbers are available. In addition, if there are any challenges to the fairness of a redistricting plan, the courts rely on census data in making decisions.

Over the years, the census has evolved to be the definitive source of data for federal fund allocations to states and small areas such as counties, places, and school districts. In the fund allocation formulas, often there is a fixed amount of money being allocated among the areas. These are known as 'fixed pie' allocations. For this type of allocation, the accuracy of an area's share of the population is important in determining the amount of funds it receives (Citro and Cohen 1985, Spencer 1980).

States, local governments, and businesses use census data in planning and decision-making. No other source of data for small areas is available on such a wide basis.

2.2 Need for census coverage measurement

The Census Bureau conducted its first coverage measurement survey, a post-enumeration survey, after the 1950 Census (U.S. Census Bureau 1960, Marks, Mauldin, and Nisselson 1953). An evaluation of the 1940 Census demonstrated the need for an assessment of coverage of a census. A match of draft registration records to the 1940 Census evaluated the coverage of adult males of an age eligible for the draft. Surprisingly, there were more males registered for the draft than enumerated in the census. (Price 1947)

The Census Bureau also conducted a post-enumeration survey after the 1960, 1980, 1990, and 2000 Censuses and in test censuses in the 1970s, 1980s, and 1990s in preparation for censuses. The design of the post-enumeration survey has evolved over the years. The implementation and estimation methodology in the 1950 and 1960 post-enumeration surveys had some basic differences with the designs used in 1980, 1990, and 2000 (Mulry 2000).

The 1950 and 1960 post-enumeration surveys did not use dual system estimation to estimate census coverage error (U.S. Census Bureau 1960 , Marks, Mauldin, and Nisselson 1953 , Marks and Waksberg 1966). Rather the idea was to form a much higher quality census interview with highly trained interviewers. The results of the post-enumeration survey interview were assumed to be the truth. However, the results were disappointing. Neither the 1950 nor 1960 post-enumeration surveys measured as large of an undercount as found by Demographic Analysis (Mulry 2000).

The Census Bureau did not evaluate the 1970 Census using a post-enumeration survey. Improvements in the methodology came from development sponsored by the United Nations (Marks, Seltzer, and Krotki 1974) where dual system estimation (Chandrasekar and Deming 1949) was applied in other countries. Dual system estimation required assuming the post-enumeration survey was only a second independent enumeration, not that it was a perfect enumeration (Mulry2000). The 1980, 1990, and 2000 implementations did use dual system estimation (Fay, Passel, Robinson, and Cowan 1988, Hogan 1992, 1993 , U.S. Census Bureau 2001b). Dual system estimation is discussed further in Sections 4.2 and 5.5. Section 7 contains historical estimates of census coverage error.

3. Census 2000

An overview of the methodology for taking the census gives context to the description of the program for evaluating the coverage of Census 2000.

3.1 Universe for Census 2000

The universe of the census is the resident population of the U.S. Census 2000 included members of the military and their dependents stationed outside the U.S. Also, U.S. citizens who were employees of the U.S. government and their dependents stationed outside the U.S. were included in the census numbers used in apportionment, but excluded for other purposes. Other U.S. citizens living outside the U.S. were not included in the census. The resident population includes citizens of other countries who are part of an established household while working or studying in the U.S. and members of their families with them (U.S. Census Bureau 1999).

The Census Bureau attempted to enumerate each person at his or her “usual residence”, defined as the place where the person lives and sleeps most of the time surrounding April 1, 2000, Census Day. This concept of usual residence was used in the first U.S. decennial census in 1790 and has been the main principle in determining where a person should be counted in every subsequent census. Using the concept of usual residence means that people may not have been counted where they were on April 1, 2000. For example, people temporarily away from their home, such as on a business trip or on vacation, on April 1, 2000 were counted at their usual residence rather than at their hotel. People with more than one residence were counted at the one where they stayed most of the time which may not have been the one where they stayed on April 1, 2000. People who did not have a usual residence on April 1, 2000 were counted where they were staying on that day (U.S. Census Bureau 1999).

3.2 Phases in taking Census 2000

The census in the U.S. is a massive undertaking. The methodology for taking Census 2000 involved several steps. Most of the census questionnaires were delivered by mail although in some of the rural areas, Census Bureau employees delivered the questionnaires. The questionnaire asked for a list of those who lived at the address on April 1, 2000 in addition to a few other questions. Approximately 16 percent of the population received a long form which asked additional questions. A member of the household filled out the form and returned it by mail. When a questionnaire was not received, an enumerator visited the address and conducted an interview in person. All the forms were scanned by optical character and mark recognition technologies and converted to electronic format for processing (Kline 2004).

The Census Bureau spent approximately $100 million on an advertising campaign to inform people about the census and encourage them to respond. In addition, a Promotion and Outreach Program established the largest number of partnerships ever with a wide range of organizations to implement promotional activities to educate the public about the importance of participating in the census (Edwards and Wilson 2003) .

3.2.1 Address list construction

Distributing the census by mail required forming a list of addresses. For Census 2000, the construction of the address list involved merging the address list from the 1990 Census with the list of addresses where the U.S. Postal Service delivers mail, called the Delivery Sequence File, in addition to other operations (Vitrano, Pennington, Treat 2004). The address list development occurred over several years, and resulted in a final census count of 115,904,641 addresses (Vitrano, Pennington, Treat 2004).

One major change from previous censuses was that Census 2000 included the Local Update of Census Addresses program during which the Census Bureau solicited help of local and tribal governments who worked with the census address list to improve its quality (Vitrano, Pennington, Treat 2004). In addition, the New Construction Program allowed local governments to submit addresses for housing units that had been built subsequent to the completion of the address list in January 2000. The address list development process included several quality assurance programs (Vitrano, Pennington, Treat 2004).

3.2.2 Questionnaire mail strategy

The strategy for mailing the questionnaires included three steps. First the Census Bureau mailed advance letters delivered between March 6 and March 8, 2000 that said the questionnaires were coming. Then the Census Bureau mailed the questionnaires so that they were delivered between March 13 and March 15, 2000. The Census Bureau also mailed post cards that reminded the members of the household to fill out their questionnaires if they had not and thanked them if they had. Approximately 100 million pieces of mail were sent first class for each mailing. Census

Day was April 1, 2000 (Stackhouse and Brady 2003b).

In some areas where the Census Bureau anticipated problems with mailing the questionnaires, census workers updated the address list as they covered the area and left census questionnaires at each address. About 22 million questionnaires were delivered by census workers (Stackhouse and Brady 2003a) In the more rural areas, census workers created a list of addresses and conducted an interview with the residents to collect the information on the census questionnaire (Vitrano, Pennington, Treat 2004).

A member of the responding household filled out the census questionnaire, answering the questions for all the members of the household, and mailed the questionnaire back to the Census Bureau. The U.S. Postal Service returned questionnaires when the dwelling at the address was vacant or the address did not exist.

The Census Bureau received questionnaires for 64.3 percent of the addresses on the address list by April 18, 2000, called the mail response rate (Stackhouse and Brady 2003a). The mail response rate for the 1990 Census was 65 percent (Stackhouse and Brady 2003a). The percentage of occupied units that returned their mail questionnaire by April 18, 2000 was 74.1 percent (Stackhouse and Brady 2003b), called the mail return rate. The return rate for the 1990 Census was 74 percent. Although the comparison between 1990 and 2000 is not exact because of changes during the intervening decade, the Census Bureau considered maintaining the return rate a success in light of the downward trend in response rates for censuses and surveys in general (Stackhouse and Brady 2003b).

3.2.3 Nonresponse Followup

When the Census Bureau did not receive a questionnaire from an address, and the U.S. Postal Service did not identify it as vacant or nonexistent, an interviewer visited the household and conducted an interview. This operation was known as Nonresponse Followup (NRFU). In Census 2000, approximately 42 million households were enumerated in the NRFU (Borsa and Hough 2004).

For most local census offices, NRFU was completed as scheduled in a nine-week period between April 27, 2000 and June 26, 2000 (Borsa and Hough 2004). One reason the Census Bureau was able to complete NRFU on schedule was the creation of a restructured pay scale that established competitive local pay rates to attract and hire sufficient numbers of “temporary” quality field workers to conduct the census (Borsa and Hough 2004).

The Coverage Improvement Followup (CIFU) checked on 6.5 million addresses that were determined during the NRFU operation to be vacant or nonexistent. This operation also was used for addresses that required followup but were identified too late to be included in the NRFU. CIFU contacted 8.9 million housing units and enumerated 5.3 million people from June 26, 2000 until August 23, 2000. (Clark and Moul 2004)

3.3 Multiple ways to respond

People could be counted in Census 2000 in ways other than a mail questionnaire or by a NRFU or CIFU interviewer. The alternatives included responding by telephone, over the Internet, or through a “Be Counted” form. In addition, questionnaires were available upon request in five languages and language assistance guides were available in 49 languages.(Smith and Jones 2002)

The Census Bureau’s Telephone Questionnaire Assistance Program had a toll-free telephone number listed on the census questionnaires. People could call the number for information or help with their questionnaire. In addition, the operators at the Telephone Questionnaire Assistance center could take the person’s census information over the telephone if the person desired to respond over the telephone (Chestnut 2002).

People also were able to be counted in the census through the Census Bureau’s website. This capability was not widely advertised so only169,257 people in 63,053 households were counted in Census 2000 through the Internet (Whitworth 2002).

Persons who did not believe they had been counted in Census 2000 could obtain a “Be Counted” form, list themselves at their April 1, 2000 address, and mail the form to the Census Bureau. The Census Bureau placed these forms at convenient locations such as Department of Motor Vehicles, libraries, and subway stations. Approximately 600,000 “Be Counted” forms were processed and about 240,000 of the forms added people not counted on other forms(Carter 2002).

3.4 Short and long census forms

There were two types of census forms, a short form and a long form. The short form requested the basic information about the people and the housing unit at the address. For the people, the questionnaire asked name, birth date, race, and Hispanic ethnicity. For the housing unit, the questionnaire asked whether it was owned or rented and whether it was a single-family house, in a multi-unit structure, or a mobile home. The questions on the short form are called the “100 percent data” because these questions were asked of everyone.

The long form had questions in addition to the questions on the short form. These additional questions are known as the “sample data” since only a sample of the population was asked to answer them. The 20 additional questions asked of the household members included things such as their highest level of education and where they were born. Examples of the information requested in 26 additional questions about the housing unit included the number of rooms and the type of heating fuel. The sample that received the long form was selected from the census address list. The long form data were needed for other government programs that required the estimates for such small areas that the census was the most efficient collection method. Also, the long form data provide a wealth of information to businesses and local governments as well as federal agencies. Tabulations of the short and long form data can be found on the Census Bureau's website www.census.gov.

Long forms were mailed to about 16 percent of the addresses overall. The percentage of addresses that received a long form varied by the size of the place or county. For large areas, the percentage was about 12.5 percent while in the medium-sized areas 16 to 25 percent of the addresses received the long form. In the very small places, about half of the addresses received a long form. (Hefter 1999)

3.5 Unclassified units and missing data

Despite all its efforts, the Census Bureau was not able to obtain all the information for some housing units on the address list. The Census Bureau used a statistical process known as imputation to attribute all characteristics to people when none were collected. Also, imputation was used to assign whether a housing unit was occupied or vacant when the occupancy status of a housing unit was unknown. In a few cases, whether the housing unit existed had to be imputed before the occupancy status. The imputation process used characteristics of other household members, when available, and characteristics of the neighborhood. For some housing units, the Census Bureau obtained the characteristics for some of the individuals but not others. There were 2.3 million people in this situation, and the Census Bureau imputed their characteristics. For another 2.2 million people, the interviewers obtained the number of people that lived in a housing unit but could not collect their characteristics, and the imputation process provided their characteristics (Zajac 2003, U.S. Census Bureau 2003b). There were some housing units where interviewers were unable to determine whether the units were occupied or vacant. For other housing units, the Census Bureau interviewers determined they were occupied but could not obtain information about the number of individuals living there. The imputation process added 1.2 million people to the census in housing units where either the occupancy status was unknown and both the occupancy status and number of occupants were imputed or the number of people living in the occupied unit was unknown and imputed (U.S. Census Bureau 2003b).

When questionnaires were returned but some of the information was missing, the Census Bureau imputed the missing items. For the items known as the 100 percent items, the imputation rates for people where at least some of their characteristics had self-responses ranged from 1.1 percent for sex to 4.1 percent for tenure (Norris 2003).

4. Estimates of census coverage error

4.1. Types of census coverage error

There are two basic types of census enumeration errors:

- omissions from the census
- erroneous enumerations in the census.

The types of census enumerations considered to be errors include:

- enumerations for people who died before Census Day
- enumerations for people born after Census Day
- enumerations for people from another country on vacation in the U.S.
- enumerations for people enumerated at homes other than their usual residence
- duplicates of other enumerations.

A census omission occurs when a person who should be enumerated is not. For example, a boarder may be missed because the person filling out the form only lists family members. For another example, if a house has a basement apartment and the census address list does not have a separate record for the apartment, then the apartment would not receive a mail questionnaire and the residents could be missed.

The Census Bureau used measurements of omissions and erroneous enumerations to estimate the net coverage error, the difference in the omissions and erroneous inclusions. The net coverage error evaluates the overall coverage of the census. Studying the omissions and erroneous inclusions enables the Census Bureau staff to gain insight about the causes of census coverage error and improve census-taking methods in future censuses (Griffin and Moriarity 1992). The measurements of omissions and erroneous enumerations used in the formation of the dual system estimates are not unbiased estimates because of the implications of the treatment of enumerations with incomplete data and the limitation on the geographic distance feasible to search for enumerations. However, those implications are designed to offset in the dual system estimator so that it produces an unbiased estimate of that population when there is no model error.

The percent net undercount is defined as

[(True population size - Census count)/True population size] x 100.

The percent net undercount is estimated by substituting an estimate of the true population size from a coverage measurement survey or alternative method, such as Demographic Analysis, in the formula.

4.2 Methods to measure census coverage error

The Census Bureau has used two methods to measure the net coverage error in censuses, the post-enumeration survey and Demographic Analysis. A post-enumeration survey is a type of coverage measurement survey conducted after the census. Another type of coverage measurement survey is a pre-enumeration survey that is conducted before rather than after the census. A test of a pre- enumeration survey in conjunction with the 1986 Census Test showed the undesirable potential for affecting census results (Wolfgang 1987) so the Census Bureau has continued to use the post-enumeration survey for coverage measurement. The 2000 Accuracy and Coverage Evaluation Survey (A.C.E.) was a post-enumeration survey that used dual system estimation.

Demographic Analysis estimates of coverage error in Census 2000 also were developed and used to evaluate the A.C.E. and A.C.E. Revision II as well as Census 2000. The Demographic

Analysis methodology and estimates used in these evaluations are discussed in Section 6.1, and some estimates for previous censuses are presented in Section 7. Demographic Analysis estimates are formed using historical vital records and therefore are available only at the national level and only for the racial groups Black and non-Black because older records do not contain other racial and ethnic groups. A major concern about Demographic Analysis estimates is the low quality of vital records on migration in and out of the country.

An advantage of a post-enumeration survey is that it can provide estimates below the national level and for racial groups other than Black and non-Black. The post-enumeration survey questionnaire may include questions concerning the same racial groups as measured in the census it evaluates. However, the estimates from a post-enumeration survey definitely contain sampling error and may contain other errors, called nonsampling errors, arising from violations of underlying assumptions during the implementation. A description of the implementation of the 2000 A.C.E. is given in Section 5. The next few paragraphs mention some issues present in the design of post-enumeration surveys in general to set context for the estimates from A.C.E. Revision II and the 1990 Post-Enumeration Survey(PES) in Section 4.3.

A post-enumeration survey is composed of two samples, the enumeration sample (E-Sample) and the population sample (P-Sample). The E-Sample is a sample of census enumerations and designed to measure erroneous enumerations. The P-Sample is a sample of the population selected independently of the census and designed to measure census omissions. The members of households interviewed in the P-Sample are matched to the census on a case-by-case basis to determine whether they were enumerated in the census (Mulry 2000). Both the 2000 A.C.E. and the 1990 Post-Enumeration Survey (Hogan 1992 1993) used dual system estimation to produce estimates of the population size. This estimate of population size was used in calculating the percent net undercount. The basic dual system estimator used by the Census Bureau for a post-enumeration survey is as follows:

$$DSE = Cen \times r_{DD} \times \frac{r_{CE}}{r_M} \tag{1}$$

where:

Cen is the census count.

r_{DD} is the census data-defined rate, which is the percentage of enumerations with at least two characteristics reported. A name counts as one characteristic.

r_{CE} is the correct enumeration rate estimated by the percentage of the enumerations in the E-Sample that are correct.

r_M is the census inclusion rate estimated by the percentage of individuals in the P-Sample that match a census enumeration, called the match rate.

The designers of every post-enumeration survey have to address two implementation issues. One issue is whether a person was considered correctly enumerated if he or she is counted at any address or if there is only one address where the person had to be counted. The other issue is how to handle people who move between Census Day and the survey interview (Mulry 2000).

The decision for the A.C.E. for Census 2000 was that a person was correctly enumerated only if he or she was enumerated at his or her usual residence. The 1990 Post-Enumeration Survey used the same definition.

When deciding how to treat movers, the designers of a post-enumeration survey have three options, PES-A, PES-B, and PES-C. In PES-A, the members of the P-Sample are the residents of a sampled housing unit on Census Day. In PES-B, the members of the P-Sample are the residents of the housing unit when the P-Sample interview is conducted. A hybrid known as PES-C uses the number of movers into the sampled housing units (PES-B) to estimate the number of movers and the match rate for the people who have moved out (PES-A) to estimate the match rate for movers. The match rate for movers is important to measure because it tends to be lower than the match rate for those who do not move. All three variations in the treatment of movers have advantages and disadvantages. With PES-A, the people who have moved have to be traced so that they can be interviewed in person, or the interviewers have to rely on proxy responses from landlords, neighbors, or others. However, the matching is much easier because the enumeration is at the sampled housing unit if the movers were enumerated. With PES-B, the people are easy to find because they now live at the sampled housing unit, but they must remember their former address well enough for census staff to find it in the census records. PES-C attempts to incorporate the best of both, but that means the interviewers have to collect much more information than they do with PES-A or PES-B (Mulry 2000).

The design of the A.C.E. for Census 2000 used the PES-C treatment of movers. The Census Bureau chose PES-C design because it was the best choice to implement when conducting the census Nonresponse Followup on a sample basis rather than visiting all the nonresponding units(Thompson and Fay 1998). However, the plans for sampling for Nonresponse Followup were canceled when the Supreme Court ruled that the law covering the census, Title XIII, did not permit the sampling for apportionment of the House of Representatives (Department of Commerce 1999). In Census 2000, interviewers visited all nonresponding units. (Borsa and Hough 2004)

The implementation of PES-C relied on proxy response for those who moved out of the sample block between Census Day and the P-sample interview. The Census Bureau did not attempt tracing outmovers because an evaluation in the Census 2000 Dress Rehearsal found that the additional cost and time for tracing was not warranted. The evaluation compared match rates from data for outmovers obtained from proxy interviews and from self-interviews after tracing and found little difference (Raglin and Bean 1999). The 1990 Post-Enumeration Survey used the PES-B treatment of movers.

A technical concern about post-enumeration surveys that use dual system estimation is that they may be subject to a bias, called correlation bias, that arises because of a violation of the assumption of independence between the census and the P-Sample or because of a violation of the assumption that the enumeration probabilities are equal. The Census Bureau attempts to preserve the independence of the census and P-Sample by keeping the A.C.E. data collection and processing operations completely separate from the census data collection and processing. Poststratification of the respondents by geography, sex, age, racial and ethnic groups, and population density reduces the bias by grouping together people with similar chances of being counted, as estimated by the match rate (Chandrasekar and Deming 1949). However, the poststratification may not describe all the heterogeneity of enumeration probabilities and thereby may not eliminate all correlation bias (Mulry 2000). Correlation bias is most likely a source of downward bias in dual system estimates. A correction for correlation bias in dual system estimates for adult males has been developed using Demographic Analysis estimates of the ratio of males to females under the assumption that correlation bias is not present for adult females (Bell 1993). The 2000 A.C.E. Revision II estimates contain a correction for correlation bias, but the 1990 Post-Enumeration Survey estimates do not.

4.3 Coverage error in Census 2000

The estimates from A.C.E. Revision II are believed to be the best estimates of coverage error in Census 2000. They provide valuable information about the quality of the census. However, there are some technical concerns about the estimates that are discussed in Section 6.2.

Table 1 shows the estimated percent net undercount for the whole U.S. for Census 2000 and the 1990 Census and the estimated percent net undercount and percent differential undercount for major groups in the population. The percent net undercount estimates for the 1990 Census provide some context for viewing the percent net undercount estimates for Census 2000. The 1990 Census was evaluated using the 1990 Post-Enumeration Survey (PES) which used the same basic methodology as the A.C.E. However, differences in methodology have to be weighed when viewing the two sets of estimates. The 1990 Post-Enumeration Survey included the population living in noninstitutional group quarters along with the population living in housing units while the A.C.E. Revision II estimates are only for the population living in housing units. In addition, the A.C.E. Revision II estimates for adult males included a correction for correlation bias while the 1990 Post-Enumeration Survey estimates did not. If the 1990 estimates had contained a correction for correlation bias, the net undercount rates would have been larger, particularly for Black males (Bell 1993). Estimates of bias for individual error components for the 1990 Post-Enumeration Survey from an evaluation program that included reinterviews and expert recoding are contained in a paper by Mulry and Spencer (1993). A computerized search of the 1990 Census for duplicates was not one of the evaluations because the names on the census questionnaires were not converted to electronic format. Table 2 shows the estimated net undercounts for the U.S. and the major groups for Census 2000 and the 1990 Census. Table 3 displays the estimated percent net undercount for owners and renters within the major groups for the last two censuses.

A.C.E. Revision II estimates a negative percent net undercount of the Census 2000 household population. The estimated percent net undercount of -0.49 with a standard error of 0.20 is significantly different from zero at the 10-percent significance level. In contrast, the 1990 PES estimated a 1.61 percent net undercount (standard error of 0.20) in the 1990 census (U.S. Census Bureau 2003b).

Among the A.C.E. Revision II coverage estimates by race and Hispanic ethnicity domains, only those for the Non-Hispanic White and Non-Hispanic Black domains show estimated percent net undercounts that differ significantly from zero. The Non-Hispanic White domain has a negative estimated percent net undercount of -1.13 percent, reflecting an overcount, while the Non-Hispanic Black domain has an estimated net undercount of 1.84 percent (U.S. Census Bureau 2003b).

The 1990 Post-Enumeration Survey estimated very similar net undercount rates for the Non-Hispanic Blacks and Hispanics. The A.C.E. Revision II estimate for the Hispanic domain is a net undercount of 0.71 percent, which is not as similar to the Non-Hispanic Black estimate as it was in 1990. This may partly be due to sampling variation. However, a comparison is difficult because the estimates for Hispanics adult males contained the correlation bias correction based on all non-Blacks while Black adult males had correlation bias correction based only on Blacks. Regardless, the A.C.E. Revision II net undercount estimates for the Non-Hispanic Black and Hispanic domains are not significantly different from one another (U.S. Census Bureau 2003b).

A comparison between the A.C.E. Revision II and the 1990 PES net undercount estimates for the American Indian and Alaska Native On Reservation population is difficult because of the large standard errors for the estimates. A.C.E. Revision II estimated a percent net undercount of -0.88 percent for the American Indian and Alaska Native On Reservation in Census 2000 with a standard error of 1.53. The 1990 PES estimated a percent net undercount of 12.22 for the American Indian and Alaska Native On Reservation population with a standard error of 5.29. (U.S. Census Bureau 2003b)

Table 3 shows differences in coverage error estimates with respect to tenure. Nationally, A.C.E. Revision II estimates owners to have a net undercount of -1.25 percent and non-owners a net undercount of 1.14 percent. These estimated net undercount rates are statistically different from zero, and their difference is also statistically significant. The 1990 PES estimated an even more dramatic difference in coverage between owners and non-owners, though in the same direction (higher estimated undercount for non-owners) (U.S. Census Bureau 2003b).

The A.C.E. Revision II estimates show coverage differences by age and sex. In particular, statistically significant net overcounts were estimated for children age 10-17 and for adult females 18-29, 30-49, and 50 and over, as well as for males 50 and over. In contrast, statistically significant percent net undercounts were estimated for males 18-29 and 30-49, and the percent net undercount estimate for children 0-9 was not significantly different from zero. The coverage differences by sex are affected by the correlation bias adjustments that increase the undercount estimates for adult males (U.S. Census Bureau 2003b).

Table 1: Estimated Percent Net Undercount for Major Groups

Characteristic	A.C.E. Revision II Undercount		1990 PES Undercount	
	Est. (%)	S.E. (%)	Est. (%)	S.E. (%)
Total	-0.49	0.20	1.61	0.20
Race/Hispanic Origin Domain				
Non-Hispanic White*	-1.13	0.20	0.68	0.22
Non-Hispanic Black	1.84	0.43	4.57	0.55
Hispanic	0.71	0.44	4.99	0.82
Non-Hispanic Asian**	-0.75	0.68	2.36	1.39
Hawaiian or Pacific Isl**	2.12	2.73	2.36	1.39
AI on Reservation***	-0.88	1.53	12.22	5.29
AI off Reservation***	0.62	1.35	0.68	0.22
Tenure				
Owner	-1.25	0.20	0.04	0.21
Non-Owner	1.14	0.36	4.51	0.43
Age/Sex				
0 - 9****	-0.46	0.33	3.18	0.29
10 - 17****	-1.32	0.41	3.18	0.29
18 - 29 Male	1.12	0.63	3.30	0.54
18 - 29 Female	-1.39	0.52	2.83	0.47
30 - 49 Male	2.01	0.25	1.89	0.32
30 - 49 Female	-0.60	0.25	0.88	0.25
50+ Male	-0.80	0.27	-0.59	0.34
50+ Female	-2.53	0.27	-1.24	0.20

The A.C.E. Revision II is for the household population.
The 1990 net undercount is for the PES universe which included noninstitutional, nonmilitary group quarters in addition to the household population.
*For 1990, AI off Reservation was included in the Non-Hispanic White Race/Hispanic Origin Domain. Therefore, the net undercount and standard error for these domains are identical.
**For 1990, Asian or Pacific Isl. was a single Race/Hispanic Origin Domain. Therefore, for Non-Hispanic Asian and for Hawaiian or Pacific Isl, the net undercount and standard error are repeated.
***For the A.C.E. Revised Preliminary estimates, American Indian and Alaskan Native was a single Race/Hispanic Origin Domain. Therefore, for AI on Reservation and for AI off Reservation, the net undercount and standard error are identical.
****For the 1990 PES, the “0 - 17” Age/Sex group was a single group. Therefore, the net undercount and standard error for children “0 - 9” and “10 - 17” are identical.
A negative net undercount denotes a net overcount.

Table 2: Net Undercount Estimates for Major Groups (in thousands)

Characteristic	Census 2000	A.C.E. Revision II		1990 PES	
		Est.	S.E.	Est.	S.E.
Total	273,587	-1,332	542	3,994	488
Race/Hispanic Origin Domain					
Non-Hispanic White*	192,924	-2,151	382	1,277	417
Non-Hispanic Black	33,470	628	146	1,389	168
Hispanic	34,538	248	152	1,102	181
Non-Hispanic Asian**	9,960	-74	67	174	103
Hawaiian or Pacific Isl**	590	13	16		
AI on Reservation	540	-5	8	52	22
AI off Reservation*	1,565	10	21		
Tenure					
Owner	187,925	-2,320	372	71	334
Non-Owner	85,662	988	310	3,871	368
Age/Sex					
0 - 9***	39,642	-180	130	2,084	191
10 - 17***	32,307	-422	129		
18 - 29 Male	21,594	245	138	792	130
18 - 29 Female	21,576	-295	111	687	113
30 - 49 Male	41,297	848	104	685	114
30 - 49 Female	42,783	-257	105	326	95
50+ Male	33,798	-270	90	-160	93
50+ Female	40,590	-1,001	107	-419	98

The Census count is for the household population.
The A.C.E. Revision II net undercount is for the household population.
The 1990 net undercount is for the PES universe which included noninstitutional, nonmilitary group quarters in addition to the household population.

*For 1990, AI off Reservation was included in the Non-Hispanic White Race/Hispanic Origin Domain.
**For 1990, Asian or Pacific Isl. was a single Race/Hispanic Origin Domain. Therefore, the net undercount and standard error displayed is for the Asian or Pacific Isl Domain.
***For the 1990 PES, the “0 - 17” Age/Sex group was a single group. Therefore, the net undercount and standard error displayed are for the “0 - 17” Age/Sex group.

Table 3: Estimated Percent Net Undercount: Race/Hispanic Origin Domain by Tenure

A.C.E. Revision II			1990 PES		
Characteristic	Est. (%)	S.E. (%)	Est. (%)	S.E. (%)	Characteristic
Total	-0.49	0.20	1.59	0.19	Total*
Owner	-1.25	0.20	0.04	0.21	Owner
Non-Owner	1.14	0.36	4.51	0.41	Non-Owner
Non-Hispanic White	-1.13	0.20	0.68	0.22	Non-Hispanic White
Owner	-1.46	0.20	-0.26	0.23	Owner
Non-Owner	-0.07	0.41	3.06	0.50	Non-Owner
American Indian off Reservation	0.62	1.35			
Owner	-1.53	1.77			
Non-Owner	3.54	2.18			
Non-Hispanic Black	1.84	0.43	4.57	0.53	Black
Owner	0.56	0.49	2.26	0.56	Owner
Non-Owner	3.06	0.60	6.48	0.83	Non-Owner
Hispanic	0.71	0.44	4.99	0.78	Hispanic
Owner	-1.08	0.50	1.82	0.68	Owner
Non-Owner	2.35	0.62	7.43	1.18	Non-Owner
Non-Hispanic Asian	-0.75	0.68	2.36	1.36	Asian or Pacific Isl.
Owner	-1.71	0.85	-1.45	1.47	Owner
Non-Owner	0.68	0.98	6.96	2.50	Non-Owner
Hawaiian or Pacific	2.12	2.73			
Owner	0.67	3.87			
Non-Owner	3.64	3.60			
American Indian on Reservation	-0.88	1.53			
Owner	-0.74	1.74			
Non-Owner	-1.17	1.71			

* Excludes American Indians on Reservations.
A negative net undercount denotes a net overcount.
The 1990 Hispanic domain excludes Blacks, Asian or Pacific Islanders, and American Indians on Reservation. The1990 net undercount is for the PES universe which included noninstitutional, nonmilitary group quarters in addition to the household population.

5. Coverage evaluation survey for Census 2000

The A.C.E., and thereby A.C.E. Revision II, followed the basic methodology of post-enumeration surveys. A.C.E. Revision II used data collected for A.C.E. along with additional data collected and processed in subsequent evaluations and research. Both A.C.E. and A.C.E. Revision II used dual system estimation, but A.C.E. Revision II incorporated additional data as explained in this section.

5.1 Design of Accuracy and Coverage Evaluation Survey

The A.C.E. evaluated the coverage of Census 2000 only for individuals living in housing units. The A.C.E. did not assess the coverage of individuals living in group quarters such as college dormitories, nursing homes, and prisons.

The design of the A.C.E. applied the basic concepts of a post-enumeration survey discussed in Section 4.2 (Hogan 2003, Childers 2001). The E- and P-Samples for the A.C.E. used the same sample of block clusters, which may be one block or several smaller blocks grouped together. A sample of 11,303 block clusters was selected from all block clusters in the 50 states and the District of Columbia. In the very large block clusters, a subsample was selected. The E-Sample had 311,029 housing units, and the P-Sample had 310,913 housing units (U.S. Census Bureau 2004). All the enumerations geocoded to the sample block clusters, or subsampled portion, were in the E-Sample. An independent listing of housing units that did not rely on any of the Census 2000 addresses was constructed in the sample block clusters. All people living in the housing units in the sample block clusters, or subsampled portion, were in the P-Sample.

The A.C.E. considered a person to be correctly enumerated only if he or she was enumerated at his or her usual residence. This decision meant that the matching operation searched for an enumeration only in the vicinity of a person's usual residence. Also, the interviewers asked questions designed to determine a person's usual residence.

As discussed in Section 4.2, the design of the A.C.E. used the PES-C treatment of movers, and the implementation specified that interviewers rely on proxy interviews with landlords, neighbors, or others to gather information about those who moved out.

The A.C.E. design had several basic steps, which are explained in more detail in the following sections. First, interviewers created a list of all the housing units in the sample block clusters. Next, P-Sample interviews were conducted at all the housing units. With the data collected, cases were matched to the census on a case-by-case basis. When there was uncertainty about whether a P-Sample person was enumerated or an E-Sample enumeration should have been in the sample blocks, a followup interview was conducted. With the additional information, a final matching operation was conducted.

5.2 Data collection for A.C.E.

Prior to the A.C.E. interviewing, interviewers listed all the addresses in the sample blocks. The operation was designed to create a list independent from the census. No information from the census address list was used. The interviewers could not work on the A.C.E. address list in any area where they had worked on the census. Then the A.C.E. list of addresses was matched to the census list of addresses for the sample blocks. When there were discrepancies or uncertainty as to whether an address on the A.C.E. list was the same as an address on the census list, these cases were sent to the field for interviewers to collect additional information.

After the resolution of discrepancies in the address lists, a merged list of addresses for each sample block cluster was loaded into the laptop computer for the interviewer assigned to the area. The A.C.E. interviewers used laptop computers to collect data from respondents. The laptop had the advantage of containing the entire questionnaire and could automatically do the branching when different answers required different subsequent questions. After the interviews were completed, the interviewers transmitted the data electronically to the A.C.E. processing center. Having the data in electronic format enabled the processing to proceed faster than it would have if the interview had been conducted with a paper questionnaire that required a keying operation to convert the answers to electronic format.

Because the A.C.E. used the PES-C treatment for movers, the interviewers collected information about the residents on the day of the interview and the residents on Census Day, who may or may not have been the same people. The interviewing operation obtained interviews about the residents of 97 percent of the 261,060 housing units in the P-Sample that were occupied on Census Day and 98.9 percent of the 267,155 P-Sample housing units that were occupied on the day of the interview (Cantwell and Ikeda 2003). The estimation included an adjustment for the housing units not interviewed (Cantwell and Ikeda 2003).

5.3 Matching operation and followup for A.C.E.

After the data collected in the A.C.E. interviews were transmitted to a processing center, a computer-assisted clerical operation matched the P-Sample people to the census enumerations in the sample blocks. The easier cases were matched by computer before the matchers received the data. The matchers used an automated system to view the census data and the P-Sample data. For the E-Sample enumerations, the matchers determined whether they were correct, erroneous, or unresolved. For the P-Sample people, the matchers determined whether they were residents of the sample block and whether they matched a census enumeration, did not match, or were unresolved. The matchers used the P-Sample person's name and other characteristics to decide if the person matched a census enumeration. Other information collected on the questionnaires aided the matchers in determining if the P-Sample people were residents of the sample block cluster on Census Day. The matchers entered the codes that they assigned to cases reflecting their conclusions into the automated system.

When the matchers could not resolve whether an E-Sample enumeration was correct or erroneous or whether a P-Sample person was a resident of the sample block on Census Day or matched a census enumeration, the case was sent for further data collection, called a followup. In the field the interviewers conducted interviews with paper questionnaires that sometimes had special questions tailored to resolve the ambiguity. With the additional data, the matchers made final determinations for the cases sent to followup.

Sometimes the ambiguity was not clarified. These cases retained a final status of unresolved, and an imputation of their final status was entered. In the E-Sample, 3 percent of the enumerations were unresolved and received imputed probabilities of being a correct enumeration. For 2 percent of the P-Sample people, the operation could not determine whether they were a resident of the sample block on Census Day and a probability of being a resident was imputed. The match status was unresolved for 1 percent of the P-Sample people, and a probability of being included in the census was imputed. Those whose match status was unresolved also had an unresolved residence status so no one received an imputed match status without receiving an imputed residence status (Cantwell and Ikeda 2003).

5.4 Revision necessitated by undetected duplication and other measurement error

The Census Bureau followed a process with pre-defined criteria for its decision in March 2001 regarding the adjustment of Census 2000 for the purpose of redistricting. After much debate regarding the quality of the A.C.E. estimate of 1.18 percent net undercount based on the criteria, the Census Bureau decided not to adjust the census numbers for redistricting because of differences between the A.C.E. estimates and Demographic Analysis and other anomalies that could not be explained by the deadline of April 1, 2001 (U.S. Census Bureau 2001b). The Census Bureau continued evaluating the A.C.E. estimates throughout the summer of 2001 in preparation for a decision in October 2001 on whether to adjust other Census 2000 data products.

Through two separate evaluation studies, the Census Bureau discovered the A.C.E. failed to detect a large number of erroneous census enumerations. The Measurement Error Reinterview (Krejsa and Raglin 2001) gave the initial indication that the A.C.E. processing was not identifying all the erroneous enumerations. The Measurement Error Reinterview evaluated the A.C.E. through independent reinterview in January and February 2001 of a subsample of E-Sample and P-Sample cases selected based on their matching codes from a subsample of the A.C.E. block clusters. The subsample was called the Evaluation Sample and contained 2,259 block clusters. The reinterview was administered in person using the Evaluation Followup questionnaire, which differed from the A.C.E. followup questionnaire by including more specific questions to identify people who were not residents of the sample blocks (Krejsa and Raglin 2001). When the Measurement Error Reinterview found a large number of erroneous enumerations that the A.C.E. did not detect, a more careful processing of a subsample of 17,522 E-Sample cases in the Measurement Error Reinterview confirmed the first results and found the estimate of correct enumerations was at least 1.45 million too high (Adams and Krejsa 2001).

In parallel, a computerized search of the census also provided evidence of a large number of duplicate census enumerations that the A.C.E. processing had not detected and estimated the number of correct enumerations was 2.9 million too high (Mule 2001, Fay 2002b). The innovation that allowed discovery of the duplicate enumerations was that names from the census questionnaires were available in electronic format. The nationwide search methodology could not have been used in previous censuses because names on enumerations were not converted to an electronic format that permitted using computers. The large number of undetected duplicate enumerations of people was very surprising because 3.6 million enumerations of people were removed earlier during the census data collection operations when the Census Bureau discovered and removed 1.4 million duplicate listings of housing units on the census address list (Nash 2000, Miskura 2000).

Since undetected duplicate enumerations in the census were a major source of error in the A.C.E. estimates, the Census Bureau produced the A.C.E. Revision Preliminary estimates in October 2001, which indicated the net undercount was 0.06 percent (Thompson, Waite, Fay 2001). The estimates included adjustments to account for duplicate census enumerations and other E-Sample measurement errors detected by the Measurement Error Reinterview (Krejsa and Raglin 2001) and the Matching Error Study (Bean 2001). The Matching Error Study evaluated the error in the E-Sample processing by using expert matching team to reprocess all the data in the block clusters in the Evaluation Sample and found negligible error in the E-Sample correct enumeration rate. The Matching Error Study also evaluated the error in the P-Sample matching and measured a significantly higher P-Sample match rate than the A.C.E. did, although the P-Sample results were not included in the A.C.E. Revision Preliminary estimates (Bean 2001).

Since the A.C.E. Revision Preliminary estimates were only at the national level, the Census Bureau pursued a complete revision of the A.C.E. estimates, the A.C.E. Revision II, to provide census coverage estimates for geographic areas and possible use in the Intercensal Population Estimates Program. No new data were collected for the revision. The revision conducted research with the census data, developed improved methods for processing data collected for evaluations of the A.C.E., and developed an estimator that incorporated the results (U.S. Census Bureau 2003b).

5.5 Research for A.C.E. Revision II

The research for A.C.E. Revision II found duplicate enumerations of E-Sample enumerations both within their block cluster and outside their block cluster. The Further Study of Person Duplication (FSPD)(Mule 2002, Fay 2002a) estimated there were 5.8 million duplicate enumerations in the census, which is considered to be a lower bound on the number of duplicates in the census. Within the A.C.E. block clusters, the FSPD found about 1.2 million duplicates in a computerized search of enumerations in the E-Sample but was not as effective as the A.C.E. clerical person matchers who found about 1.9 million duplicates for this group (Mule 2002). However, when clerical matchers looked for additional duplicates in households where FSPD linked a member to someone whose housing unit was outside the search area, they found very few (Byrne, Beaghen, Mulry 2003). Interestingly, the duplicates occurred disproportionately among the population under 30 years of age. The group under 30 years of age is 42.2 percent of the census population while they are 53.6 percent of the estimated 5.8 million duplicates. In particular, the group aged 18 to 29 is 16.5 percent of the census population but 24.0 percent of the duplicates (Mule 2002). The causes of duplicate enumerations within a block cluster appear to arise from operational errors, such as a dwelling having two different addresses on the address list or a Nonresponse Followup interviewer going to the wrong address. Early indications are that the causes of duplicate enumerations in different states or counties include moving situations, people visiting family/friends, people with vacation/seasonal homes, college students, and children in shared custody situations (Smith 2004).

In a similar fashion, FSPD performed a computerized search of the census to look for links between the P-Sample people and enumerations. For P-Sample people who were not movers and who said their usual residence was in the A.C.E. sample blocks on Census, FSPD estimated that 6.1 million linked to census enumerations outside the A.C.E. search area surrounding the sample blocks. Of the 6.1 million, 3.2 million also matched to enumerations within the A.C.E. search area and 2.8 million did not match (Mule 2002). These links raised doubts about whether the P-Sample persons actually were Census Day residents at their given address (U.S. Census Bureau 2003b). Reviews of the questionnaires found unreliable identification of moves, second homes, and stays in group quarters residences. These reviews concluded that there were some fundamental problems with how the questionnaires asked about the usual residence of a person on Census Day and improving the quality of data collected would require intensive questionnaire design and testing. (Martin, Fay, and Krejsa 2002).

Since the evaluations of the A.C.E. found errors in the assignment of enumeration and residence status codes for the E- and P-Samples, respectively, the A.C.E. Revision II methodology included recoding the subsample of the A.C.E. sample that were interviewed in the Evaluation Followup and used the results in the estimation. The A.C.E. Revision II Measurement Coding Operation had to re-process 69,318 E-Sample people and 52,671 P-Sample people. The operation assigned some of the E- and P-Sample codes by a computer algorithm, with the rest assigned clerically by the Census Bureau's elite matching team. The clerical matchers assigned codes to 23,988 people (Adams and Krejsa 2002b). Although the strategy of combining automated and clerical coding permitted recoding of a larger sample in the time available, most likely reducing the variance of the A.C.E. Revision II, there was concern that the automated assignment of enumeration and residence status for some of the cases increased the possibility of error in the A.C.E. Revision II dual system estimates. An evaluation based on a subsample of cases coded both ways showed that the potential error from the automated coding was very small (Adams and Krejsa 2002b).

The A.C.E. Revision II Measurement Coding Operation did not include a further evaluation of the coding of match status, but used the coding of match status for the P-Sample from the Measurement Error Review (Krejsa and Raglin 2001). The additional matches in the surrounding block discovered by the Matching Error Study (Bean 2001) also were incorporated for the A.C.E. Revision II match status. Drawing from the results of these evaluations appeared to provide the best coding available, and any further recoding probably would not provide better codes (U.S. Census Bureau 2003b).

The A.C.E. Revision II dual system estimates included an adjustment for correlation bias, discussed in Section 4.2. Evidence of correlation bias in national estimates comes from sex ratios (number of males divided by number of females) for A.C.E. Revision II estimates that were low relative to ratios derived from Demographic Analysis (Robinson and Adlakha 2002).

In previous post-enumeration surveys, the erroneous inclusions appeared to be much smaller than omissions. In this setting not adjusting estimates for correlation bias had the effect of understating the net undercount, which resulted in corrections to the census that were in the right direction but not large enough. In the presence of overcounts, it is possible that corrections without correlation bias might not even be in the right direction, and could actually increase errors relative to no adjustment (U.S. Census Bureau 2003b).

Estimates of correlation bias were calculated using the sex ratios from Demographic Analysis estimates. The model used for the correlation bias adjustment in the A.C.E. Revision II estimates was about the simplest possible, and assumed that relative correlation bias was constant for adult males within the age-race groups. The correlation bias estimates are made only for adult males under the assumption of no correlation bias for adult females. Also, correlation bias was not estimated for children. The correlation bias adjustments were done separately for Blacks and non-Blacks within three age categories: 18-29, 30-49, and 50 and over, with the exception of non-Black males 18 to 29 years of age, a group for which the data would not support estimation of correlation bias for males. (U.S. Census Bureau 2003b) For Blacks, the correlation bias adjustment rates were 8 percent for 18-29, 10 percent for 30-49, and 5 percent for 50+. For non-Blacks, the correlation bias adjustment rates were 2 percent for 30-49 and 1 percent for 50+ (Shores 2002). The assumptions and model underlying the measurement of correlation bias are discussed in detail in a paper by Bell (2001) and further discussion of alternative models can be found in Bell (1993).

5.6 Description of dual system estimator used for A.C.E. Revision II estimates

The A.C.E. Revision II estimation methodology uses dual system estimates (DSEs) that incorporate corrections for measurement errors obtained from two sources: the recoded cases from the A.C.E. evaluation data by the A.C.E. Revision II Measurement Coding Operation (Adams and Krejsa 2002a) and census duplicates identified by the Further Study of Person Duplication (Mule 2002). Both of these corrections affect the estimated E-Sample correct enumeration rates and the P-Sample match rates. The DSEs for adult males are also inflated by correlation bias adjustment factors estimated using Demographic Analysis sex ratios for the adult age groups (18-29, 30-49, 50+) at the national level by Black versus non-Black race groups. Also for the first time, the A.C.E. Revision II poststratification reflected one set of factors related to erroneous inclusions and a different set of factors related to omissions. Previous estimates of census coverage error used the same factors for poststratification of the E- and P-Samples. (U.S. Census Bureau 2003b)

The specific form of the A.C.E. Revision II DSE is given in equation (2) and discussed below (U.S. Census Bureau 2003). For a detailed discussion of the estimator, see U.S. Census Bureau (2004) or Kostanich (2003a) .

$$DSE_{ij} = Cen_{ij} \times r_{DD,ij} \times \frac{r_{CE,i}}{r_{M,j}} \times \phi \qquad (2)$$

where:

i and j denote the E- and P- Sample poststrata used to estimate the correct enumeration and match rates, respectively.

Cen_{ij} is the census count of the household population for the cross-classification of poststrata *I* and *j*. Includes the cases removed during the census because they potentially were duplicates but subsequently were reinstated.

$r_{DD,ij}$ is the census data-defined rate for the cross-classification of poststrata *I* and *j*. The reinstated cases are included in the denominator but not in the numerator.

$r_{CE,i}$ is the estimated correct enumeration rate for E-Sample poststratum *I*.

$r_{M,j}$ is the estimated match rate for P-Sample poststratum *j*.

ϕ is the correlation bias adjustment factor (for adult males, distinct for a given age-race group)

The numerator of the data-defined rate, $r_{DD,ij}$, is the count of census data-defined persons, which is the census count excluding whole-person imputations and all "reinstated" persons (those who were removed from the census because they potentially were duplicates but then reinstated as part of the Housing Unit Duplication Operation) (Miskura 2000, Nash 2000). The denominator of $r_{DD,ij}$, is the census count, so that the product, $Cen_{ij} \times r_{DD,ij}$, at the level of the *ij* poststrata is the count of data-defined persons that were eligible for A.C.E. matching. The correct enumeration rate, $r_{CE,i}$, is the ratio of the estimated E-Sample correct enumerations to the weighted estimate of data-defined persons for E-Sample poststratum *I*. The product, $Cen_{ij} \times r_{DD,ij} \times r_{CE,i}$, effectively estimates correct enumerations for the detailed *ij* poststratum under the synthetic assumption that correct enumeration rates are constant over persons within E-Sample poststratum *I*. (U.S. Census Bureau 2003b)

The results of the A.C.E. Revision II Measurement Coding Operation and Further Study of Person Duplication affect the estimates of correct enumerations that are the numerators of the correct enumeration rates, $r_{CE,i}$. The denominators of the correct enumeration rates are not affected. For example, E-Sample cases with duplicates that were originally coded as correct enumerations are given reduced correct enumeration probabilities, which reduces tabulated estimates of correct enumerations. The basic principle followed was that each detected duplicate pair would contribute one correct enumeration in the A.C.E. Revision II estimation and on the average, the probability of being a correct enumeration for the member of the pair in the E-Sample was ½. This principle was used in the construction of the A.C.E. Preliminary Revised Estimates (Fay 2002b). Otherwise, for E-Sample people without duplicates, an adjustment based on the results of the A.C.E. Revision II Coding Operation was applied to their total number of correct enumerations estimated from the original A.C.E. For details and further discussion, see U.S. Census Bureau (2004) or Kostanich (2003a).

The match rate, $r_{M,j}$, is the ratio of estimated matches to estimated Census Day residents for P-Sample poststratum *j*. The match rate estimates the enumeration rate for the P-Sample people. The A.C.E. Revision II Measurement Coding Operation and Further Study of Person Duplication also affect both the estimates of matches and the estimates of P-Sample residents that are the numerators and denominators of the match rates, $r_{M,j}$. (U.S. Census Bureau 2003). For the P-Sample people that linked to an enumeration outside the A.C.E. search area surrounding the sample block, there was no symmetry argument for estimating the probability of being a resident of the sample block that was analogous to that used for estimating correct enumeration probabilities for E-Sample people with duplicates. Therefore, the P-Sample matches and nonmatches linking to enumerations outside the search area were assigned probabilities of being a resident equal to the probabilities of being a correct enumeration derived for corresponding E-Sample matches and nonmatches with duplicates. Otherwise, for P-Sample people without a link to an enumeration outside the search area, adjustments based on the results of the A.C.E. Revision II Coding Operation were applied separately to their population total and number of matches estimated from the original A.C.E. The specifics are somewhat complicated. For details and discussion, see U.S. Census Bureau (2004) or Kostanich (2003a).

The correction for correlation bias is expressed as multiplicative factors ϕ which correct the adult male DSEs for this estimated correlation bias. For children and adult females the factors ϕ are 1 (U.S. Census Bureau 2003b). Although the adjustment to correct for correlation bias removed a significant source of error, the demographic analysis results used in the adjustment are limited to the two race categories – Black and non-Black – and are available only at the national level by age and sex. No data were available to detect differential coverage by race and Hispanic ethnicity with the non-Black group. In addition, there were no data to use in estimating how correlation bias in the estimates varied between geographic areas below the national level. (U.S. Census Bureau 2003b)

5.7 Small area estimation

Synthetic assumptions were used in the construction of the A.C.E. Revision II estimates. One type of synthetic assumption involved correcting the individual poststratum estimates for errors estimated at more aggregate levels, such as the corrections for correlation bias, duplicates, and measurement coding error. All involve estimates at a very aggregate level with little or no information available about how the effects being estimated truly affect correct enumeration rates and census inclusion probabilities for individual poststrata. For example, the correction for correlation bias relies on a synthetic assumption that correlation bias for adult males is constant over persons within the age-race groups. (U.S. Census Bureau 2003b)

Another synthetic assumption is used in the application of poststratum coverage correction factors to specific areas within the poststratum. The assumption is that the net coverage error rate is constant within the poststratum. Equation (2) shows how the A.C.E. Revision II estimates are constructed for the cross-classified *ij* poststratum. To produce estimates for specific areas or population subgroups we first define coverage correction factors (CCFs) by dividing the dual system estimates from equation (2) by the corresponding census counts, i.e.,

$$CCF_{ij} = DSE_{ij} / Cen_{ij} = r_{DD,ij} \times \frac{r_{CE,i}}{r_{M,j}} \times \phi \qquad (3)$$

To produce the estimate for any area or population subgroup *a*, the CCFs from equation (3) are applied synthetically:

$$\sum_{ij} Cen_{a,ij} \times CCF_{ij} = \sum_{ij} Cen_{a,ij} \times r_{DD,ij} \times \frac{r_{CE,i}}{r_{M,j}} \times \phi \qquad (4)$$

where the summation is over all the cross-classified *ij* poststrata and $Cen_{a,ij}$ is the census count in poststratum *ij* for area or subgroup *a* (U.S. Census Bureau 2003b). A.C.E. Revision II estimates for states, counties and places were estimated using Equation (4) and may be found in Kostanich (2003b).

6. Assessment of quality of A.C.E. Revision II estimates

Evaluation studies examined the quality of A.C.E. Revision II estimates of census coverage error (U.S. Census Bureau 2003). The A.C.E. Revision II estimates are subject to both sampling error as well as nonsampling errors due to violations of assumptions underlying the methodology and estimation. The evaluations of the A.C.E. Revision II estimates included evaluations that focused on individual error components as well as comparisons of the relative error between the census and the A.C.E. Revision II. This section discusses a comparison of net undercount estimates from Demographic Analysis and A.C.E. Revision II, the evaluations of components of error in A.C.E. Revision II, and the concerns the Census Bureau had about using the A.C.E. Revision II estimates.

6.1 Comparison with Demographic Analysis

Demographic Analysis is a tool for examining the relative error between the census and A.C.E. Revision II (Robinson and Adlakha 2002). Robinson (2001) describes the construction of the Demographic Analysis estimates. Demographic Analysis uses vital records to form an estimate of the size of the population. The basic approach is to start with births, subtract deaths, and add the net migration, the difference between immigrants and emigrants. Demographic Analysis uses administrative statistics on births, deaths, authorized international migration, and Medicare enrollments, as well as estimates of legal emigration and net unauthorized immigration. (Robinson and Adlakha 2002)

Table 4 shows that the Demographic Analysis estimate of the percent net undercount in the Census 2000 count of 281.4 million is 0.12 percent. Overall, the Demographic Analysis estimates indicate that the net census undercount is at substantially low levels except for the two groups, Black adult men with a 5.15 percent net undercount rate and young children ages 0-9 with a 2.56 net undercount rate. For these two groups, the percent net undercount is disproportionately high in that their percent net undercount rate was 2 percentage points or more higher than the percent net undercount rate for the total population. Table 5 shows that the disproportionately high percent net undercount rates hold when the adult Black men are divided into three age groups of 18-29 years, 30-49 years, and 50 and over years of age. Also, the disproportionately high undercount rates also hold when the children ages 0-9 are divided into Black and non-Black and by males and female (U.S. Census Bureau 2003b).

The A.C.E. Revision II estimate of 280.1 million is 1.7 million below the Demographic Analysis estimate. The A.C.E. Revision II estimate implies a net overcount of 1.3 million, or -0.48 percent[1], compared to the Demographic Analysis estimated net undercount of 0.12 percent. The consistency between Demographic Analysis and A.C.E. Revision II displayed in Table 5 is basically a consequence of using the Demographic Analysis sex ratios to remove the correlation bias. The A.C.E. Revision II estimates for females (especially non-Black females) show patterns similar to the Demographic Analysis estimates for ages 10 and over, even thought they did not receive an adjustment for correlation bias (U.S. Census Bureau 2003b).

The A.C.E. Revision II and the Demographic Analysis estimates remain inconsistent with regard to coverage rates for children aged 0-9. In contrast to Demographic Analysis results which show a relative large undercount of children (both Black and non-Black), the A.C.E. Revision II results estimates show a net overcount (not statistically significant) of non-Black children and small net undercount of Black children (Table 5). The causes of the inconsistency of the DA and A.C.E. Revision II results for young children is not known at this time.(U.S. Census Bureau 2003b)

For ages 50 and over, a smaller but systematic gap is observed between the Demographic Analysis estimate and A.C.E. Revision II estimate for each race-sex group (Table 5). For Black males, the Demographic Analysis percent net undercount is higher than the corresponding A.C.E. Revision II estimate; for non-Black males Demographic Analysis measures a small net undercount and the A.C.E. Revision II estimates a small net overcount; for Black females and non-Black females both Demographic Analysis and the A.C.E. Revision II measure a net overcount but the DA estimate is smaller.   (U.S. Census Bureau 2003b)

[1]This estimated net undercount from A.C.E. Revision II is slightly different from the -0.49 estimate cited earlier because it is relative to the entire resident population including persons in group quarters.

Table 4. Census Count, Demographic Analysis (DA) Estimate and A.C.E. Revision II Estimate for the U.S. Resident Population: April 1, 2000. Standard errors are in parentheses.

	Count or Estimate
1. Census Count	281,421,906
2. DA Estimate	281,759,858
3. A.C.E. Revision II Estimate	280,090,250
	(541,631)
Net Census Undercount (Amount)	
4. DA Estimate (=2-1)	337,952
5. A.C.E. Revision II Estimate (=3-1)	-1,331,656
	(541,631)
Net Census Undercount (Percent)	
6. DA Estimate (=4/2*100)	0.12
7. A.C.E. Revision II Estimate (=5/3*100)	-0.48
	(0.20)

Note: 1) A.C.E. Revision II estimate includes an adjustment for correlation bias, based on the DA sex ratios for adult males.
2) DA reflects revised estimate published in Robinson (2001)

Table 5. Estimates of Percent Net Undercount for Census 2000 by Race, Sex, and Age based on Demographic Analysis (DA) and A.C.E. Revision II			
Category	**DA**	**A.C.E. Revision II**	**A.C.E. Revision II Standard Error**
BLACK MALE			
All ages	5.15	4.19	0.43
0-9	3.26	0.72	0.73
10-17	-1.88	-0.59	0.68
18-29	5.71	6.14	0.94
30-49	9.87	8.29	0.53
50+	3.87	2.43	0.62
BLACK FEMALE			
All Ages	0.52	-0.61	0.43
0-9	3.60	0.70	0.73
10-17	-1.20	-0.55	0.68
18-29	-0.66	0.00	0.94
30-49	1.28	-0.40	0.53
50+	-1.03	-2.51	0.62
NON-BLACK MALE			
All Ages	0.21	-0.19	0.22
0-9	2.18	-0.68	0.33
10-17	-2.01	-1.46	0.42
18-29	-0.63	0.19	0.70
30-49	0.63	1.05	0.23
50+	0.14	-1.10	0.27
NON-BLACK FEMALE			
All Ages	-0.78	-1.41	0.20
0-9	2.59	-0.68	0.33
10-17	-1.55	-1.44	0.42
18-29	-1.94	-1.54	0.53
30-49	-1.01	-0.63	0.25
50+	-1.18	-2.42	0.27
Notes: See Table 4.			

6.2 Evaluations of A.C.E. Revision II estimates

The evaluations of bias were relatively limited because data that previously were used to estimate biases in the A.C.E. estimates were used in the production of the A.C.E. Revision II estimates to correct for the major biases. The limited data available for evaluation of bias does not itself reflect negatively on the A.C.E. Revision II estimates. Rather, the Census Bureau believes that because of the corrections for these major errors, the A.C.E. Revision II estimates are of much higher quality than the March 2001 A.C.E. estimates (U.S. Census Bureau 2003b).

The evaluations of the A.C.E. Revision II estimates of census coverage error confirmed the presence of the large number of census duplicates and gave insight to the quality of the A.C.E. Revision II estimates themselves (Mulry and Petroni 2003). By the time a computerized search of the census provided evidence of a large number of duplicate census enumerations in October 2001(Thompson, Waite, Fay 2001), field tests for validation were not practical. Studies of duplication as part of the A.C.E. Revision II program estimated 5.8 million duplicate enumerations in Census 2000 (Mule 2002). The estimate of the number of duplicates in the census was validated in two evaluation studies. One found 6.7 million duplicates using only the census duplicates identified by administrative records (Mulry, Bean et al 2003). In the other study, an expert matching team clerically examined the duplicate enumerations and agreed that about 95 percent were duplicates (Byrne, Beaghen, Mulry 2003). The expert matching team found very few additional duplicates in the households they reviewed (Byrne, Beaghen, Mulry 2003).

Other evaluations examined the potential for bias in the match rate due to inconsistent reporting in the census and the P-Sample of the variables used in the poststratification as well as the violation of assumptions about movers and missing data in the estimation. These evaluations found these potential sources of error had very little effect on the A.C.E. Revision II estimates or their estimated variances (U.S. Census Bureau 2003b).

Another evaluation examined the relative error in the census and the A.C.E. Revision II estimates through using the data from the other evaluations to estimate bias (systematic error) and variance (random error) to construct bias-corrected confidence intervals and in a loss function analysis (Mulry, ZuWallack, and Spencer 2003). The loss function analysis examined the relative accuracy by using the estimates of sampling variance and nonsampling bias and variance to estimate the aggregate expected loss for the census and the A.C.E. Revision II for population levels and shares for counties and places across the nation and within state. The measure of accuracy used by the loss functions was the weighted mean squared error with the weights set to the reciprocal of the census count for levels and the reciprocal of census share for shares.

The evaluations of bias were relatively limited because data that previously were used to estimate bias were incorporated into the A.C.E. Revision II estimates in order to correct for major errors discovered in the March 2001 A.C.E. estimates. Although the evaluations did account for the variance arising from the corrections for bias, the corrections for bias in the A.C.E. Revision II estimates may themselves have been subject to bias, the magnitude of which was not quantified. This was particularly true for the corrections for correlation bias and for P-Sample cases that linked to census enumerations outside the A.C.E. search area since several models fit the data, and no information was available to enable selecting among the models on the basis of bias. Other criteria, such as variance, had to be used (U.S. Census Bureau 2003b) .

The loss function analysis accounted for some but not all error components that could be identified in the A.C.E. Revision II estimates. More specifically, the bias estimate included error components for inconsistency of poststratification assignments based on census versus A.C.E. data, for error from estimating the numbers of outmovers by the numbers of inmovers, and most importantly, for error in the estimates of census duplicates although evaluations indicate that this error may have been misestimated. The variance estimate included sampling error components from both phases of sampling in A.C.E. Revision II estimates, and also random nonsampling error components from choice of imputation models and for models used to account for P-Sample cases that matched census enumerations outside the search area. On the other hand, the loss function analysis did not account for the following errors: synthetic estimation error; bias from response error and coding error in P-Sample residency status, match status, and mover status; bias from response error and coding error in E-Sample correct enumeration status; bias in correlation bias adjustments to the estimates due to error in the Demographic Analysis sex ratios and to the choice of model used to implement the adjustments; and bias due to the choice of model used to adjust the DSE for E-Sample cases with duplicate links (Mulry, ZuWallack, and Spencer 2003). Though not included in the loss functions, a partial evaluation of synthetic error is described earlier in this section.

The synthesis of the available evaluation data detected a small amount of bias in the A.C.E. Revision II estimate of the net undercount rate at the national level, only - 0.16 percent. Based on the bias-corrected 95-percent confidence intervals, both the census and the A.C.E. Revision II estimates were too low for Non-Hispanic Blacks and both Non-Hispanic Black Owners and Renters. The intervals showed the census was too high for Non-Hispanic Whites, Owners, White Owners, and Hispanic Owners. All other census and A.C.E. Revision II estimates were covered by their bias-corrected 95-percent confidence intervals. The source of most of the bias estimate was based on the evaluation of the identification of duplicates using administrative records (Mulry, ZuWallack, and Spencer 2003).

The loss function analyses indicated that the A.C.E. Revision II was more accurate than the census for every loss function considered with the exception of levels for places with population of at least 100,000. The bulk of the error in the A.C.E. Revision II for places with population of at least 100,000 appeared to lie in the nine (9) places with population of at least 1 million. More research is needed to understand the one exceptional result. The validity of the loss function analysis depended on the quality of the estimates of components of error in the A.C.E. Revision II, and some of those components were not accurately quantified (Mulry, ZuWallack, and Spencer 2003).

In general, one cannot be certain whether omitted biases tend to make any given loss function analysis overstate or understate the comparative accuracy of the A.C.E. Revision II estimates relative to the census. Whether omitted biases caused the loss function to favor the census or the A.C.E. Revision II depended on the signs of the correlations between the omitted biases and the expected undercount rate for the areas considered (Mulry and Spencer 2001). While the loss function evaluations suggested the superiority of the A.C.E. Revision II estimates, concerns did remain about whether the bias estimates used in the loss function analysis were of sufficient quality to assure the correctness of the results (Mulry, ZuWallack, and Spencer 2003).

An evaluation focused on the synthetic estimation bias that arises when the census from different areas but in the same poststratum have different coverage error rates but have the same census coverage correction factors. The evaluation attempted to estimate synthetic estimation bias in undercount estimates from analysis of "artificial populations" or "surrogate" variables whose geographic distributions are known. These surrogate variables are constructed as best as possible to have patterns similar to coverage error. Sensitivity analyses of four artificial populations generally did not change the overall loss function findings (Griffin 2002), though the analyses were not considered conclusive. Any limitations of the loss functions regarding synthetic error are expected to be more important when comparing small places or counties than for large places or counties. The use of different poststrata by the E-and P-Samples may have reduced synthetic error for the correct enumeration rate that is present when the E- and P-Samples use the same poststrata as in the Census Bureau's previous applications of dual system estimation. However, the use of different E- and P-Sample poststrata has increased concern for synthetic error in the small areas affected by extreme coverage error estimates.

6.3 Concerns about estimates from A.C.E. Revision II

The Census Bureau considered using the A.C.E. Revision II estimates in an adjustment of the census base for the intercensal population estimates program. In the decade between censuses, this program produces updated population estimates for the nation, states, and counties by age, sex, race, and Hispanic ethnicity, as well as estimates of total population for small places. The intercensal estimates are used in fund allocation formulas in many federal programs.

After a review of the A.C.E. Revision II estimates, the Census Bureau decided not to use them in the intercensal population estimates. Troubling anomalies and unexplained results meant that the Census Bureau could not be confident of improvements in accuracy at the levels of geography for which estimates are produced. This lack of confidence is especially important in the context of the intercensal estimates program, because of the challenge process available to local officials (U.S. Census Bureau 2003c).

The concern about the A.C.E. Revision II estimates focused on the correlation bias adjustment, the estimation methodology for small areas, and inconsistencies with demographic analysis. (U.S. Census Bureau 2003c)

- The adjustment to correct for correlation bias was limited to the two race categories – Black/non-Black – and was available only at the national level by age and sex. There were no data to use in estimating how correlation bias in the estimates varied by race/Hispanic origin within the non-Black group or between geographic areas below the national level (U.S. Census Bureau 2003c).
- The concern about the estimation methodology for small areas rested in the first use of different poststrata for the E- and P-Samples and the particular poststrata selected. The estimates of extreme overcounts for some small places that could not be validated or explained led to the concern that a systematic problem existed in the estimates for all small areas (U.S. Census Bureau 2003c).
- While the A.C.E. Revision II estimates and the Demographic Analysis estimates for the whole U.S. were reasonably consistent, the inconsistency in estimates for children aged 0-9 raised concerns. The consistency at the national level was enhanced somewhat by the use of demographic analysis sex ratios in the adjustment for correlation bias.

However, for children aged 0-9, A.C.E. Revision II estimated a small net overcount (not statistically significant) while demographic analysis estimated a net undercount of 2.56 percent. Demographic analysis estimates for this group are believed to be highly accurate because they rely primarily on recent birth registration data (U.S. Census Bureau 2003c).

A thorough discussion of the technical limitations of the estimates is contained in "Technical Assessment of A.C.E. Revision II" (U.S. Census Bureau 2003b). Many of the limitations are of greater concern for subnational than for national estimates.

7. Historical perspective

The estimates of percent net undercount shown in Figure 1 show that at the national level the estimates from the 1980, 1990, and 2000 implementations of post-enumeration surveys are comparable to those from Demographic Analysis. An advantage of the post-enumeration survey is that it provides estimates for levels of geography below the national level and for racial and ethnic groups although the estimates have sampling error and a vulnerability to violations of underlying assumptions. The advantage of Demographic Analysis is that it uses vital records, which may be less error prone, but there are problems with estimation of net migration into the U.S. In addition, demographic analysis estimates are not available below the national level or for racial and ethnic groups other than Black and non-Black because the historical records contain only those groups.

A measure useful in combination with Demographic Analysis in assessing historically the relative coverage of Blacks and non-Blacks is the differential undercount. The percent differential undercount for a group within the population provides a measure of the percent net undercount for the group relative to the percent net undercount for the whole population. The percent differential undercount for a group within the population (U.S. Census Bureau 2001b) is defined as

Percent net undercount for the group - Percent net undercount for the U.S.

The differential undercount is important because many uses of the data depend on population shares as discussed in Section 2.1. The differential undercount indicates whether there is error in the population share for a group or area (Spencer 1980). If a group's or area's differential undercount is negative, then its share is too large. Conversely, a positive differential undercount for a group or area means its share is too small.

Estimates of differential undercount for Blacks and non-Blacks are available from demographic analysis back to the 1940 Census and shown in Figure 2 (Long, Robinson, Gibson 2003). Note that the percent differential undercount for Blacks stayed at about 3.5 percent fairly consistently from 1940 through 1990. Census 2000 was able to lower the percent differential undercount for Blacks substantially to 2.7 percent.

Figure 1.

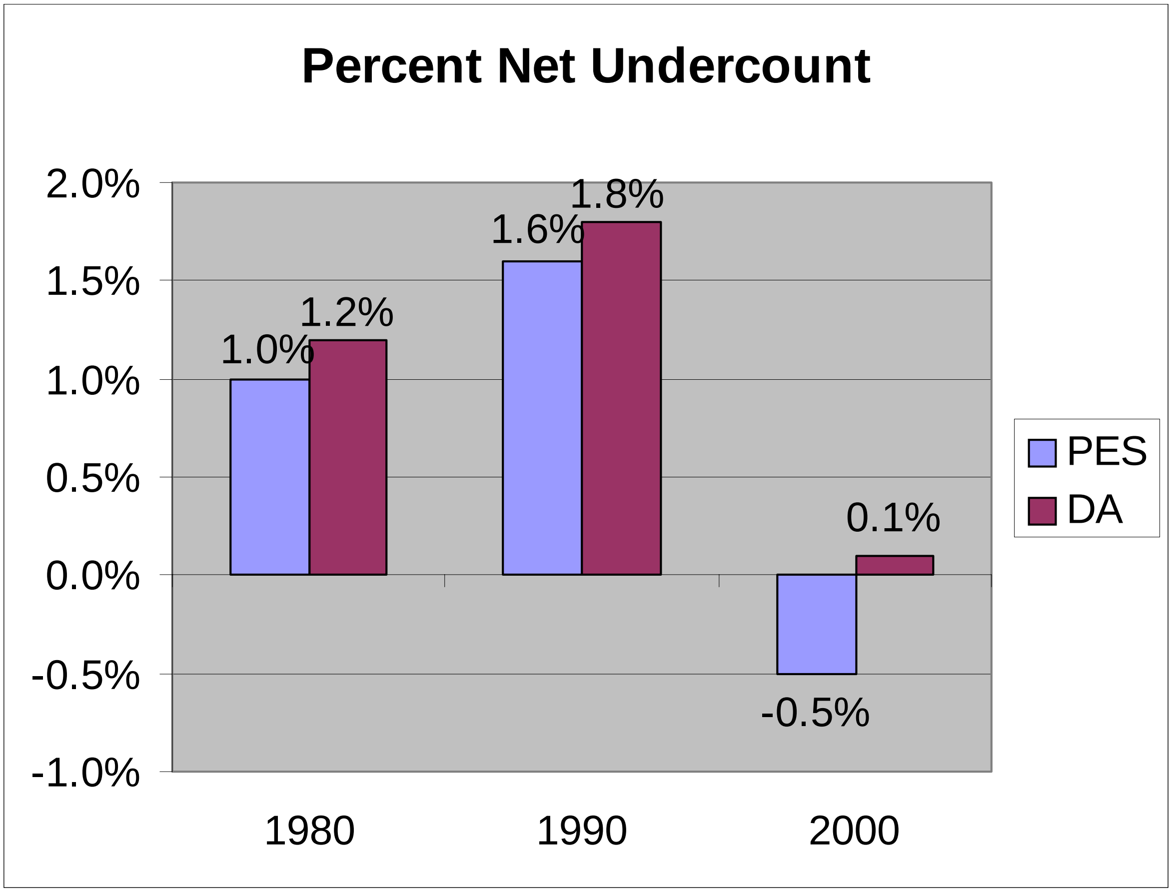
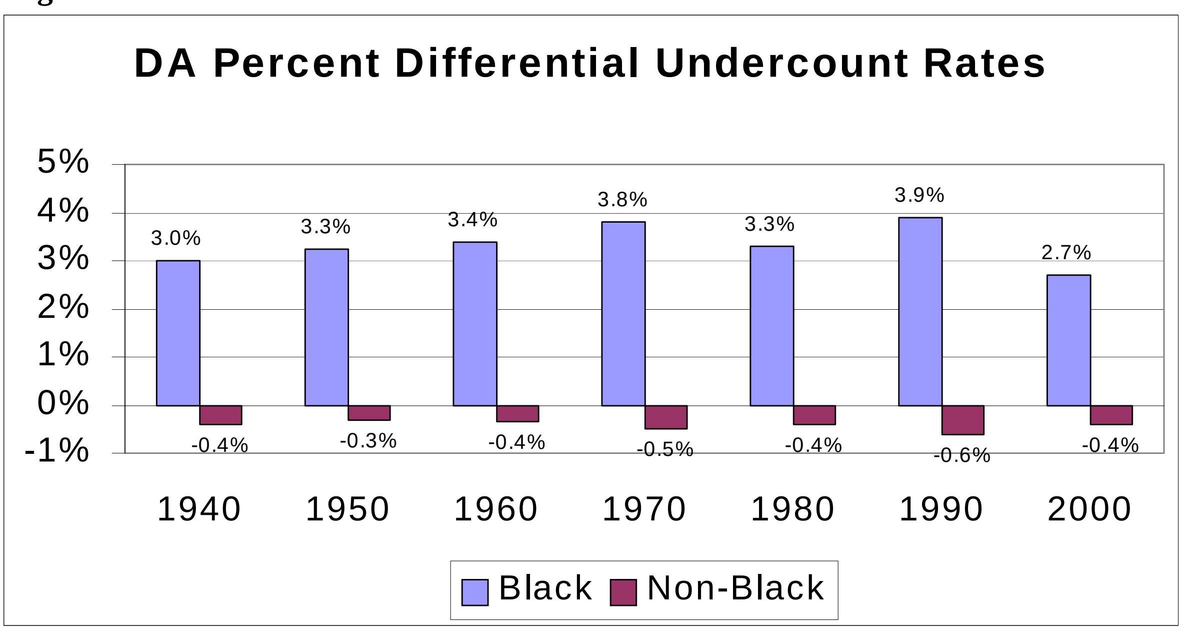


Figure 2.



Evidence of the persistent differential undercount for Blacks has prompted interest in adjustment of the census using the estimates from post-enumeration surveys. The issue of adjustment of the census for coverage error first arose when Newark, NJ filed a lawsuit challenging the Department of Treasury over biases in fund allocation arising from the undercount in the 1970 Census. The court ruled in favor of the Department of Treasury, and unadjusted 1970 Census numbers continued to be used in fund allocation (City of Newark, NJ et al. 1978)

The controversy continued when New York City led a lawsuit asking for the results of the 1980 Post-Enumeration Program to be used in an adjustment of the 1980 Census. The Census Bureau took the position that the results of the 1980 Post-Enumeration Program were not of high enough quality for census adjustment, and the courts did not order an adjustment. Again in 1990, the Census Bureau considered adjusting the 1990 Census for undercount using the results of the 1990 Post-Enumeration Survey. However, the Census Bureau decided in July 1991 not to adjust the 1990 Census. An extensive evaluation of the 1990 PES estimates was conducted in preparation for the decision. In 1992 the Census Bureau revised the estimates and considered using the revised estimates to adjust the census base in the Postcensal Estimates Program. However, the Census Bureau decided not to adjust the 1990 Census with the 1992 estimates of coverage error (U.S. Census Bureau 1992).

For Census 2000, the Census Bureau approached the census adjustment issue by constructing a plan for a decision on correcting the census numbers for redistricting. The plan for the adjustment decision including pre-specifying the A.C.E. data collection, processing, and estimation operations. In addition the Census Bureau announced in advance the criteria it would use in the decision on whether the A.C.E. estimates were of sufficient quality to use in an adjustment of the census. The Census Bureau considered the operational data to validate the conduct of A.C.E. was successful, assessed whether the A.C.E. measurements of undercount were consistent with historical patterns of undercount and independent Demographic Analysis benchmarks, and reviewed measures of quality that were available in March 2001 (Prewitt 2000). In the end, the inconsistency between the A.C.E. and Demographic Analysis estimates caused the Census Bureau to decide not to adjust the census numbers for redistricting (U.S. Census Bureau 2001b). Subsequently, the Census Bureau decided in October 2001 not to adjust Census 2000 for other purposes because newly available evaluation data indicated problems with the A.C.E. detection of erroneous enumerations (U.S. Census Bureau 2001a). In March 2003 after producing the A.C.E. Revision II estimates, the Census Bureau decided not to adjust the census base for the Intercensal Population Estimates Program for reasons discussed in Section 6.3.

8. Summary

Although concerns about some of the technical limitations of the A.C.E. Revision II estimates led to the Census Bureau's decision not to adjust the population base for the intercensal estimates, the results are informing ongoing research for the 2010 Census and the evaluation of the coverage of the 2010 Census. Research efforts aimed at formulating corrections for measurement errors found in the A.C.E. have provided ground-breaking information on the source and magnitude of coverage error components. These results have prompted an expansion of the coverage measurement goals for 2010 beyond estimating net error to include estimating individual components of error.

8.1 Influence of A.C.E. Revision II on planning of the 2010 Census

The current planning for the 2010 Census is drawing on the wealth of information about the census and coverage measurement evaluation that A.C.E., A.C.E. Revision II, and their evaluations provided. The 2004 and 2006 Census Tests (U.S. Census Bureau 2003a, Kostanich, Whitford, and Bell 2004) include examining methodology for detecting and removing duplicate enumerations before the census is completed in addition to methodology designed to reduce the number of people missed by the census. These results have prompted an expansion of the coverage measurement goals for 2010 beyond estimating net error to include estimating individual components of coverage error.

8.2 Coverage measurement for the 2010 Census

The research and preparations for the 2010 Census coverage measurement are focusing on providing information useful for designing improvements in census-taking methodology. The new emphasis on improving the quality of future censuses re-orders the goals for coverage measurement in the three previous censuses. This direction for the program requires evaluating the sources and magnitude of coverage error rather than only the net effect of all the sources combined.

The most important objective for the 2010 Census Coverage Measurement(CCM) program is to obtain separate estimates of erroneous census inclusions and census omissions. The A.C.E. Revision II results showed that a very small net overcount was a result of large numbers of erroneous census inclusions that would have had to be offset by an almost equally large number of census omissions. The 2010 Census Coverage Measurement program has a greater emphasis on understanding the relationship between the census methodology and the measurements of erroneous inclusions and omissions (Kostanich, Whitford, and Bell 2004).

Producing estimates of net error for the 2010 census continues to be an important objective. The plans include estimating the net undercount and the differential net undercount for demographic groups and by geography. However, estimates of net error for small geographic areas are not important because the emphasis is not on adjusting census counts (Kostanich, Whitford, and Bell 2004).

Since the Census Bureau will not attempt to adjust the states' total census numbers that must be produced by December 31, 2010 for apportionment or the census numbers for redistricting that have a deadline of April 1, 2011, the schedule for the evaluation of the coverage of the 2010 Census is not constrained by these dates. Testing of methodology and operations for the Census Coverage Measurement program will be included in the 2006 Census Test (Kostanich, Whitford, and Bell 2004). The design of the Census Coverage Measurement program is drawing on the methodological and operational lessons learned from the A.C.E. (Childers and Petroni 2004) and the A.C.E. Revision II (Mulry 2004).

In addition, plans include an evaluation of the estimates of census coverage error. Depending on assessments and evaluations of the accuracy of the 2010 Census Coverage Measurement estimates, consideration may be given to using these results to adjust the base for the intercensal population estimates (Kostanich,Whitford, and Bell 2004). Evaluations of the estimates of the census coverage error have provided valuable information about the quality of the census as well as the undercount estimates themselves. The evaluations of the original A.C.E. estimates detected that duplicate enumerations were a much larger problem in Census 2000 than observed or suspected in previous censuses (Thompson, Waite, Fay 2001). The evaluations of the A.C.E. Revision II estimates of census coverage error confirmed the presence of the large number of census duplicates and provided insight into the quality of the A.C.E. Revision II estimates themselves.

References

Adams, T. and Krejsa, E.(2002a) "A.C.E. Revision II Measurement Subgroup Documentation". DSSD A.C.E. Revision II Memorandum Series #PP- 6. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/pp-06r.pdf

Adams, T. and Krejsa, E. (2002b) "Evaluation of At-Risk Codes". DSSD A.C.E. Revision II Memorandum Series #PP- 45. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/pp-45r.pdf

Adams, T., and Krejsa, E. A. (2001) "ESCAP II: Results of the Person Followup and the Evaluation Followup Forms Review." Executive Steering Committee For A.C.E. Policy II, Report No. 24. October 12, 2001. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/report24.pdf

Balinski, M. L. and Young, H. P. (2001) *Fair Representation*. Second Edition. Brookings Institution Press. Washington, DC.

Bean, S. L. (2001) "ESCAP II: Accuracy and Coverage Evaluation Matching Error." Executive Steering Committee For A.C.E. Policy II, Report No. 7. October 12, 2001. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/report7.pdf

Bell, W. R. (2003) "Alternative Options for Tabulating Estimates of Census Correct Enumerations Allowing for Duplicate Links". DSSD A.C.E. Revision II Memorandum Series #PP- 3. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/pp-03.pdf

Bell, William (2001) "ESCAP II: Estimation of Correlation Bias in 2000 A.C.E. Using Revised Demographic Analysis Results" Executive Steering Committee For A.C.E. Policy II, Report No. 10. dated October 13, 2001. Census Bureau. http://www.census.gov/dmd/www/pdf/report10.pdf

Bell, W. R. (1993) "Using Information from Demographic Analysis in Post-Enumeration Survey Estimation". *Journal of the American Statistical Association*, 88, 1106-1118.

Borsa, F. R. and Hough, C. L. (2004) "Data Collection in Census 2000". Census 2000 Testing, Experimentation, and Evaluation Program. Topic Report No. 13, TR-13. U.S. Census Bureau. Washington, DC. http://www.census.gov/pred/www/rpts/TR13.pdf

Byrne, R., Beaghen, M., and Mulry, M. H. (2003) "Clerical Review of Census Duplicates". *2003 ASA Proceedings.* American Statistical Association. Alexandria, VA. CD-ROM.

Cantwell, P. and Ikeda, M. (2003) "Handling Missing Data in the 2000 Accuracy and Coverage Evaluation Survey". *Survey Methodology*, 29, 2, 139 - 145 .

Carter, N. (2002) "Be Counted Campaign for Census 2000". Census 2000 Evaluation A.3. Decennial Statistical Studies Division. U.S. Census Bureau, Washington, DC. http://www.census.gov/pred/www/rpts/A.3.pdf

Chandrasekar, C. and Deming, W. E. (1949) "On a Method of Estimating Birth and Death Rates and the Extent of Registration". *Journal of the American Statistical Association*, 44, 101-115.

Chestnut, John (2002) "Telephone Questionnaire Assistance". Census 2000 Evaluation A.1.a. Decennial Statistical Studies Division. U.S. Census Bureau, Washington, DC. DC. http://www.census.gov/pred/www/rpts/B.1.b.pdf

Childers, D. R. (2001) "Accuracy and Coverage Evaluation: The Design Document". DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1, Revised. Decennial Statistical Studies Division. January 26, 2001. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/S-DT-1R.pdf

Childers, D. R. and Petroni, R. J.(2004) "Coverage Measurement from the Perspective of March 2001 A.C.E." Census Testing, Experimentation, and Evaluation Program, Topic Report Series, No. 4, TR-4. U.S. Census Bureau, Washington, DC. http://www.census.gov/pred/www/rpts/TR4.pdf

Citro, C. F. and Cohen, M. L. (1985) *The Bicentennial Census: New Directions for Methodology in 1990.* National Academy of Science Press. Washington, DC.

City of Newark, NJ et al v. W. Michael Blumenthal, Secretary of Treasury et al, No. 74-548 (D.D.C.) January 17, 1978.

Clark, J. R. and Moul, D. (2004) "Coverage Improvement in Census 2000 Enumeration". Census 2000 Testing, Experimentation, and Evaluation Program. Topic Report 10, TR-10. U.S. Census Bureau. Washington, DC. http://www.census.gov/pred/www/rpts/TR10.pdf

Department of Commerce v. House of Representatives, 525 U.S. 316 (1999).

Department of Commerce v. Montana (91-860), 503 U.S. 442 (1992).

Edwards, W. S. and Wilson, M. J. (2004) "Evaluations of Census 2000 Partnership and Marketing Program". Census 2000 Testing, Experimentation, and Evaluation Program. Topic Report No. 6, TR-6. U.S. Census Bureau. Washington, DC. http://www.census.gov/pred/www/rpts/TR6.pdf

Fay, R. E. (2002a) “Probabilistic Models for Detecting Census Person Duplication”. *2002 ASA Proceedings.* American Statistical Association. Alexandria, VA. CD-ROM.

Fay, R. E. (2002b) “Evidence of Additional Erroneous Enumerations from the Person Duplication Study”. Executive Steering Committee For A.C.E. Policy II, Report No. 9. Revised March 2002. Census Bureau, Washington, DC.
http://www.census.gov/dmd/www/pdf/report9revised.pdf

Fay, R. E., Passel, J. S., and Robinson, J. G., with assistance from C. D. Cowan. (1988) *The Coverage of Population in the 1980 Census*, 1980 Census Of Population and Housing, Evaluation and Research Reports PHC80-E4. Washington, D. C.: U.S. Department of Commerce.

Griffin, D. H. and Moriarity, C. L. (1992) “Characteristics of Census Errors”. *1992 Proceedings of the Section on Survey Research Methods.* American Statistical Association. Alexandria, VA. 512 - 517.

Griffin, R.A. (2002) “Analysis of Synthetic Assumption”. DSSD A.C.E. Revision II Memorandum Series #PP-49. U.S. Census Bureau, Washington, DC.
http://www.census.gov/dmd/www/pdf/pp-49r.pdf

Hefter, S. P. (1999) “Projected Long Form Sample Size for Census 2000". DSSD Census 2000 Procedures and Operations Memorandum Series, # LL-2. U.S. Census Bureau. Washington, DC.
http://www.census.gov/dmd/www/pdf/LL-2.pdf

Hogan, H. (2003) “The Accuracy and Coverage Evaluation: Theory and Design”. *Survey Methodology*, 29, 2, 129-138.

Hogan, H. (1993) “The 1990 Post-Enumeration Survey: Operations and Results”, *Journal of the American Statistical Association*, 88, 1047-1060.

Hogan, H. (1992) “The 1990 Post-Enumeration Survey: An Overview”, *The American Statistician*, American Statistical Association, 261-269.

Kline, D. (2004) “Census 2000 Data Capture”. Census 2000 Testing, Experimentation, and Evaluation Program. Topic Report No. 3, TR-3. U.S. Census Bureau. Washington, DC.
http://www.census.gov/pred/www/rpts/TR3.pdf

Kostanich, D. L.,Whitford, C., and Bell, W. R. (2004) “Plans for Measuring Coverage of the 2010 Census”. *2004 ASA Proceedings.* American Statistical Association. Alexandria, VA. CD-ROM.

Kostanich, D. (2003a) “A.C.E. Revision II: Summary of Methodology”. DSSD A.C.E. Revision II Memorandum Series #PP-35, U.S. Bureau of the Census, Washington, DC.
http://www.census.gov/dmd/www/pdf/pp-35r.pdf

Kostanich, D. (2003b) "A.C.E. Revision II: Adjusted Data for States, Counties, and Places." DSSD A.C.E. Revision II Memorandum Series #PP-60, U.S. Bureau of the Census, Washington, DC. http://www.census.gov/dmd/www/pdf/pp-60r.pdf

Krejsa, E. A. and Raglin, D. A. (2001) "ESCAP II: Evaluation Results for Changes in A.C.E. Enumeration Status." Executive Steering Committee For A.C.E. Policy II, Report No. 16. October 15, 2001. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/Report3.pdf

Long, J. F., Robinson, J. G. and Gibson, C. (2003). "Setting the Standard for Comparison: Census Accuracy from 1940 to 2000". *2003 ASA Proceedings.* American Statistical Association. Alexandria, VA. CD-ROM.

Marks, E. S., Seltzer, W., and Krotki, K. (1974) *Population Growth Estimation*, The Population Council, New York.

Marks, E. S., and Waksberg, J. (1966) "Evaluation of Coverage in the 1960 Census Through Case-by-Case Checking." *1966 Proceedings of the Social Statistics Section*. American Statistical Association.

Marks, E. S., Mauldin, W. P., and Nisselson, H. (1953) "The Post-Enumeration Survey of the 1950 Census: A Case History in Survey Design. " *Journal of the American Statistical Association,* 48, 220 – 243.

Martin, E. A, Fay, R. E., and Krejsa, E. A. (2002) " Analysis of Questionnaire Errors in Survey Measurements of Census Coverage".*2002 ASA Proceedings.* American Statistical Association. Alexandria, VA. CD-ROM.

Miskura, S. M. (2000) "Results of Reinstatement Rules for the Housing Unit Duplication Operations". Memorandum for P. J. Waite, Decennial Management Division, U.S. Census Bureau, November 21, 2000

Mule, T. (2002) "Further Study of Person Duplication in Census 2000". DSSD A.C.E. Revision II Memorandum Series #PP- 51. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/pp-51r.pdf

Mule, T. (2001) "Person Duplication in Census 2000." Executive Steering Committee For A.C.E. Policy II, Report No. 20. October 11, 2001. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/Report20.pdf

Mulry, M. H. (2004) "Methodological Lessons from Census 2000 Coverage Error Measurement". *2004 ASA Proceedings.* American Statistical Association. Alexandria, VA. CD-ROM.

Mulry, M. H. (2000) "Post-Enumeration Survey." in Margo Anderson (ed.). The *Encyclopedia of the U.S. Census*. Congressional Quarterly, Inc. Washington, DC, 303-306.

Mulry, M. H. and Petroni, R. J. (2003) “Evaluation of the A.C.E. Revision II Estimates of Census 2000 Coverage Error”. *2003 ASA Proceedings.* American Statistical Association. Alexandria, VA. CD-ROM.

Mulry, M. H., ZuWallack, R. S., and Spencer, B. D. (2003) “Loss Function Analysis for A.C.E. Revision II Estimates of Census 2000 Coverage Error”. *2003 ASA Proceedings.* American Statistical Association. Alexandria, VA. CD-ROM.

Mulry, M. H., Bean, S. L., Bauder, D. M., Wagner, D., Mule, T., Petroni, R. J. (2003) “Census and Administrative Records Duplication Study”. *Proceedings of the Federal Committee on Statistical Methods Research Conference*. Office of Management and Budget. Washington, DC, 36-43. http://www.fcsm.gov/03papers/MulryBean.pdf

Mulry, M. H. and Spencer, B. D. (2001) “Overview of Total Error Modeling and Loss Function Analysis.” DSSD Census 2000 Procedures and Operations Memorandum Series, Number B-19*. Decennial Statistical Studies Division. February 28, 2001. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/Fr19.pdf

Mulry, M. H. and Spencer, B. D. (1993) “Accuracy of the 1990 Census and Undercount Adjustments”. *Journal of the American Statistical Association*, 88, 1080-1091.

Nash, F. (2000) “Overview of the Duplicate Housing Unit Operations,” Census 2000 Informational Memorandum Number 78, Decennial Management Division, U.S. Census Bureau, November 7, 2000.

National Research Council (2004) *The 2000 Census: Counting Under Adversity*. Panel to Review the 2000 Census. C. F. Citro, D. L. Cork, and J. L. Norwood, editors. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. The National Academies Press. Washington, DC.

Norris, S. (2003) “Analysis of Item Nonresponse Rates for 100 Percent Housing and Population Items from Census 2000". Census 2000 Evaluation B.1.b. U.S. Census Bureau. Washington, DC. http://www.census.gov/pred/www/rpts/B.1.b.pdf

Prewitt, K. (2000) “Statement on the Feasibility of Using Statistical Method to Improve the Accuracy of Census 2000".   June, 2000. U.S. Census Bureau, Washington, DC. http://www.census.gov/Press-Release/www/feasibility.pdf

Price, D. O. (1947) “A Check on Underenumeration in the 1940 Census,” *American Sociological Review*, 12(1), 44-49.

Raglin, D. A. and Bean, S. L. (1999) “Outmover Tracing for the Census 2000 Dress Rehearsal”. *1999 Proceedings of the Section on Survey Research Methods.* American Statistical Association. Alexandria, VA. 456 - 461.

Robinson, J. G. and Adlakha, A. (2002) "Comparison of A.C.E. Revision II Results with Demographic Analysis". DSSD A.C.E. Revision II Memorandum Series #PP- 41. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/pp-41r.pdf

Robinson, J. G. (2001) "Demographic Analysis." Executive Steering Committee For A.C.E. Policy II, Report No. 1. October 13, 2001. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/Report1.pdf

Shores, R. (2002) "Adjustment for Correlation Bias". DSSD A.C.E. Revision II Memorandum Series #PP- 53. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/pp-53r.pdf

Smith, D. R. (2004) "Long Distance Duplicate Telephone Followup Operation". DSSD 2003 Memorandum Series Chapter #A-03. July 8, 2004. U.S. Census Bureau, Washington, DC.

Smith, D. R., and Jones, J. (2002) "Use of Non-English Questionnaires and Guides in Census 2000 Language Program". Census 2000 Evaluation A.4. U.S. Census Bureau. Washington, DC. http://www.census.gov/pred/www/rpts/A.4.pdf

Spencer, B. D. (1980) *Benefit-Cost Analysis of Data Used to Allocate Funds*. Springer. New York.

Spencer, B. D. (1985) "Statistical Aspects of Equitable Apportionment". *Journal of the American Statistical Association*, 80, 815-822.

Stackhouse, H. F. and Brady, S. (2003a) "Census 2000 Mail Response Rates". Census 2000 Evaluation A.7.a. U.S. Census Bureau. Washington, DC. http://www.census.gov/pred/www/rpts/A.7.a.pdf

Stackhouse, H. F. and Brady, S. (2003b) "Census 2000 Mail Return Rates". Census 2000 Evaluation A.7.b. U.S. Census Bureau. Washington, DC. http://www.census.gov/pred/www/rpts/A.7.b.pdf

Thompson, J., Waite, P., Fay, R., (2001), "Basis of 'Revised Early Approximations' of Undercounts Released Oct. 17, 2001." Executive Steering Committee for A.C.E. Policy II, Report 9a. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/report9a.pdf

Thompson, J. H. and Fay, R. E. (1998) "Census 2000: The Statistical Issues". *1998 Proceedings of the Survey Research Methods Section.* American Statistical Association. Alexandria, VA, 101 - 110.

U.S. Census Bureau (2004) "Accuracy and Coverage Evaluation of Census 2000: Design and Methodology". DSSD/03-DM. Issued September 2004. U.S. Census Bureau, Washington, DC.

U.S. Census Bureau (2003a) "2004 Census Test Operational Plan". Dated September 29, 2003. Decennial Management Division. U.S. Census Bureau, Washington, DC.

U.S. Census Bureau (2003b) “Technical Assessment of A.C.E. Revision II” March 12,2003. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/ACETechAssess.pdf

U.S. Census Bureau (2003c) “Decision on Intercensal Population Estimates”. March 12, 2003. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/dipe.html

U.S. Census Bureau (2001a) “Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy on Adjustment for Non-Redistricting Uses”. October 17, 2001. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/Recommend2.pdf

U.S. Census Bureau (2001b) “Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy”. March 1, 2001. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/ESCAP2.pdf

U.S. Census Bureau (2001c) “Analysis Plan for Further ESCAP Deliberations Regarding the Adjustment of Census 2000 Data for Future Uses”. July 26, 2001. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/ACEplan726.pdf

U.S. Census Bureau (1999) “Plans and Rules for Taking the Census”. Dated August 9, 1999. Revised April 25, 2003. Population & Housing Branch, Population Division. U.S. Census Bureau, Washington, DC. www.census.gov/population/www/censusdata/resid_rules.html

U.S. Census Bureau (1992) “Assessment of Accuracy of Adjusted Versus Unadjusted 1990 Census Base for Use In Intercensal Estimates”. Report of the Committee on Adjustment of Postcensal Estimates. August 7, 1992. U.S. Census Bureau, Washington, DC. http://www.census.gov/dmd/www/pdf/Adjresearch.pdf.

U.S. Bureau of the Census (1960) *The Post-Enumeration Survey: 1950*. Bureau of the Census Technical Paper No. 4, Suitland, Maryland: U.S. Bureau of the Census.

Vitrano, F. A., Pennington, R. A, Treat, J. B. (2004) “Address List Development in Census 2000". Census 2000 Testing, Experimentation, and Evaluation Program. Topic Report No. 8, TR-8. U.S. Census Bureau. Washington, DC. http://www.census.gov/pred/www/rpts/TR8.pdf

Whitworth, Erin (2002) “Internet Data Collection”. Census 2000 Evaluation A.2.b. U.S. Census Bureau, Washington, DC. http://www.census.gov/pred/www/rpts/A.2.b.pdf

Wolfgang, G. (1987) “The Pre-enumeration Survey of the 1986 Census of Central Los Angeles County”. *1987 Proceedings of the Survey Research Methods Section.* American Statistical Association. Alexandria, VA, 475 - 479.

Zajac, K. (2003) “Analysis of Imputation Rates for the 100 Percent Person and Housing Unit Data Items from Census 2000". Census 2000 Evaluation B.1.a. U.S. Census Bureau. Washington, DC. http://www.census.gov/pred/www/rpts/B.1.a.pdf

Appendix A
Hill's Algorithm for Apportionment of U.S. Congress

To describe the Hill algorithm, let P_i be the population for state I and a_i be the number of representatives a state receives. The Hill algorithm first gives every state one representative, which is a requirement in the U.S. Constitution. Then the algorithm assigns seats one at a time to the states until all 435 seats have been assigned. The steps in the algorithm are listed below (Balinski and Young 2001).

Step 1. Initially set $a_i = 1$ for every state i.

Step 2. For each state i, calculate $\frac{P_i}{\sqrt{a_i(a_i+1)}}$

Step 3. Assign the next seat to the state with the maximum value and increase that state's number of seats by 1.

Step 4. Repeat Steps 2 and 3 until all 435 seats are assigned which occurs when the sum $\sum_i a_i = 435$.