UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

February 28, 2001

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES B-2*

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: Donna Kostanich DK
Assistant Division Chief, Sampling and Estimation
Decennial Statistical Studies Division

Prepared by: James Farber JEF
Sample Design Team

Subject: Quality Indicators of Census 2000 and the Accuracy and Coverage Evaluation

The attached document was prepared, per your request, to assist the Executive Steering Committee on A.C.E. Policy in assessing the data with and without statistical correction.

This report summarizes some overall quality indicators of Census 2000 and the Accuracy and Coverage Evaluation.

DSSD Census 2000 Procedures and Operations Memorandum Series B-2*
February 28, 2001

Quality Indicators of Census 2000 and the Accuracy and Coverage Evaluation

James Farber
U.S. Census Bureau

Table of Contents

Tables

Figures

Quality Indicators of Census 2000 and the Accuracy and Coverage Evaluation

prepared by James Farber

Introduction

The Accuracy and Coverage Evaluation (A.C.E.) survey checks the quality of Census 2000 population coverage. The population of the United States is too dynamic to capture with certainty, so the census will always be subject to some kind of coverage error. Using a sample of the population, the A.C.E. estimates the net proportions of people missed by the census (undercount) or erroneously included in the census (overcount), and produces population estimates that are corrected for these census coverage errors. Details of the A.C.E. design and operations can be found in Childers and Fenstermaker (2000).

The results of the A.C.E. indicate that the net undercount rate of the total population residing in housing units in Census 2000 is 1.18 percent, with a standard error of 0.13 percent. There is a differential undercount in Census 2000, although it appears to be smaller than in the 1990 Census. The net undercount rates for major race/Hispanic origin groups ranges from 0.67 percent (0.14 percent standard error) for the Non-Hispanic White or Other Race domain to 4.74 percent (1.20 percent standard error) for the American Indian on reservation domain.

Both the census and the A.C.E. can potentially produce unreliable results due to statistical or operational errors. This document describes high level indicators of Census 2000 and A.C.E. quality to inform executive assessment of the A.C.E. population estimates as compared to the Census 2000 counts. The specific indicators in this document are actually performance measures for various parts of the census or A.C.E. They are used here to get an idea about the quality of Census 2000 and the A.C.E. because the comprehensive Census 2000 and A.C.E. evaluation program will not be finished for some time. The timing of the recommendation on whether to correct Census 2000 with A.C.E. results prohibits waiting for the evaluations, and thus performance measures included here and in the other memoranda in the DSSD Census Procedures and Operations Memorandum Series B must be considered instead.

The assessment of the Executive Steering Committee for A.C.E. Policy (ESCAP) on whether to correct Census 2000 depends on a complex and interrelated set of quality indicators that exceeds the capacity of a single document. This report is not intended as a comprehensive evaluation of Census 2000 or A.C.E. quality, but rather as a summary of those indicators of quality from the other B-series memoranda that the ESCAP may want to consider in making its determination on census correction using the A.C.E. We have included references at the end of this report to other documents with further results or more detailed information beyond what is given here.

Census 2000 Quality Indicators

Census 2000 counted people by:

- determining where people lived
- delivering forms and receiving responses
- following up nonrespondent addresses and verifying coverage in the field
- processing the data

Errors in the census can arise due to issues in any of those phases. This section gives some quality indicators for those parts of the census.

Address List Development

A foundation of the decennial census process is the list of housing units that need to receive census forms. To minimize the chance that housing units are omitted from the address list, known as the Decennial Master Address File (DMAF), many sources of information are combined and unduplicated, including the 1990 Census address list, United States Postal Service address files, and the results of census field operations to canvass areas and list housing units.

When were housing units added to the census address list?

The DMAF is a dynamic list, with updates occurring at several phases throughout the census. One measure of its quality is the timing at which housing units were added. If census enumerators went into the field and found few new addresses, that could indicate the DMAF was highly accurate before field operations began. Table 1 shows how many housing units were added to the DMAF in each phase of Census 2000, overall and for each region of the United States. This table is a classification of all housing unit records that were ever delivered to the DMAF throughout Census 2000.

Table 1. Distribution of Housing Unit Additions to the Decennial Master Address File

	Total Number of Housing Units	Pre-Questionnaire Delivery Operations	Questionnaire Delivery	Post-Questionnaire Delivery Operations	Operation Undetermined
United States	128,691,771	96.7%	1.8%	1.3%	0.3%
Northeast	24,545,009	96.7%	1.7%	1.4%	0.3%
South	47,344,579	96.3%	2.1%	1.4%	0.3%
Midwest	29,750,345	97.6%	1.0%	1.2%	0.3%
West	27,051,838	96.4%	1.9%	1.3%	0.4%

(Source: Baumgardner et al, 2001)

Table 1 shows that the DMAF was nearly 97 percent complete overall and in each region of the nation before the census forms were mailed out or delivered. These addresses received preprinted census forms that might have increased the probability of initial response to the census. The two fastest growing regions, the South and West, not surprisingly had slightly lower percentages of housing unit coverage before the census and higher rates of added housing units during questionnaire delivery.

How many housing units were removed from the census address list?

Housing units were removed from the DMAF when they were determined not to exist or to be duplicates of other housing units. Table 2 gives information on the number and percent of housing units removed from the DMAF for the nation and each region. The number retained on the DMAF is the final number of housing units in Census 2000. Similar data for the 1990 Census are not available.

Table 2. Housing Units Removed from the Decennial Master Address File

	Total DMAF Housing Unit Records	Retained on the DMAF	Removed from the DMAF	
			Number	Percent
United States	126,276,807	115,904,641	10,372,166	8.2%
Northeast	24,260,015	22,180,440	2,079,575	8.6%
South	46,216,140	42,382,546	3,833,594	8.3%
Midwest	29,305,631	26,963,635	2,341,996	8.0%
West	26,495,021	24,378,020	2,117,001	8.0%

(Source: Baumgardner et al, 2001)

Census 2000 Enumeration

People initially received census forms either through the mail or directly from census enumerators. Either type of enumeration methodology relied on respondent cooperation to maximize the initial response to the census, which reduces cost and also increases the amount of respondent-provided data in the census.

What are the mail return rates in Census 2000 and the 1990 Census?

Table 3 shows the percentages of forms that were returned by mail out of the total that were mailed to occupied housing units for the nation and for each region. The 1990 Census mail return rate is available only for the national level.

Table 3. Mail Return Rates for Census 2000 and the 1990 Census

	Census 2000	1990 Census
United States	72.0%	74%
Northeast	71.8%	
South	69.6%	
Midwest	76.3%	
West	71.7%	

(Source: Baumgardner et al, 2001)

Note that differences in the assignment of housing units to an enumeration methodology between the 1990 Census and Census 2000 have likely accounted for some of the slight reduction in mail return rates. The 1990 Census had more than 10 times the number of housing units in list/enumerate areas, where enumerators compiled the address list and simultaneously collected data from respondents. Thus, list/enumerate housing units were excluded from the 1990 mail return rate calculation. In Census 2000, many of these housing units were in mailback enumeration areas and included in the mail return rate. These housing units are typically more difficult to locate and enumerate, which could contribute to the slight reduction in mail return rates. The lower Census 2000 mail return rate does not necessarily imply less respondent cooperation.

Followup

Two major followup operations took place in Census 2000: nonresponse followup (NRFU) and coverage improvement followup (CIFU). The NRFU operation involved field followup of housing units that did not return a census form within about two weeks after Census Day. The CIFU operation was a followup of certain types of housing units to maximize census coverage.

What was the size and outcome of nonresponse followup?

There were a total of about 42 million addresses in the NRFU universe, or about 35.4 percent of eligible addresses on the version of the DMAF used for census form delivery. Certain types of addresses, such as those in list/enumerate areas, are not eligible for NRFU because enumeration continues in those areas until response attains 100 percent. The distribution of outcomes of the NRFU operation is shown in Table 4.

Table 4. Distribution of Nonresponse Followup Outcomes

Outcome	Number of Addresses	Percent of NRFU Workload
Occupied	25,988,521	62.3
Vacant	9,754,928	23.4
Delete	5,979,600	14.3
Unresolved	5,344	0 0
Total	41,728,393	100 0

(Source: Baumgardner et al, 2001)

About 38 percent of the NRFU addresses were found to be vacant or delete. A primary factor contributing to initial nonresponse was that no one was available to return the form.

What were the sources and outcomes of coverage improvement followup?

The CIFU operation followed NRFU as a final large-scale field check of census coverage. Addresses were eligible for CIFU if they were found to be vacant or delete in NRFU, were added through some census operation, if their census forms were lost or blank, or for some other reason. Table 5 shows the distribution of addresses by the reason they were in CIFU and their outcomes.

Table 5. CIFU Workload Source of Followup by Outcome

Source of Followup	CIFU Outcome (Percent)				
	Occupied	Vacant	Delete	Unresolved	Total
Vacant or Delete in NRFU	1,521,059 (23.5)	3,088,989 (47.8)	1,856,459 (28.7)	487 (0.0)	6,466,994 (74.6)
New Construction	100,632 (27.1)	74,304 (20.0)	196,783 (52.9)	11 (0.0)	371,730 (4.3)
Adds from Update/Leave	319,877 (44.6)	278,946 (38.9)	118,265 (16.5)	12 (0.0)	717,100 (8.3)
Lost or Blank Return	190,586 (35.4)	251,430 (46.7)	96,876 (18.0)	22 (0.0)	538,914 (6.2)
Other	182,847 (32.1)	54,801 (9.6)	332,132 (58.3)	10 (0.0)	569,781 (6.6)
Total	2,315,001 (26.7)	3,748,470 (43.3)	2,600.506 (30.0)	542 (0.0)	8,664,519 (100)

(Source: Baumgardner et al, 2001)

More than 23 percent of the housing units that NRFU found vacant or delete were in fact occupied, with similar or greater percentages of occupancy in newly constructed housing units, adds, and other types of CIFU housing units. The CIFU operation successfully enabled these potential census omissions to count in Census 2000, thus improving census coverage.

Data Processing

The large amount of address information used to compile the DMAF along with an increased number of response opportunities increased the chance of duplicate returns at the same DMAF housing unit and also at different DMAF housing units. That is, a single household might return a mailback form late and also give information to a NRFU enumerator. Or a single household might return two forms sent to their single housing unit because that unit is included under two different addresses in the DMAF. Census 2000 included several data processing steps to handle multiple census returns for a single housing unit and returns to duplicate housing units.

How many forms were returned to each housing unit?

The distribution of Census 2000 housing units by the number of returns is in Table 6.

Table 6. Census Returns Per Housing Unit

Number of Returns	Number of Housing Units	Percent of Housing Units
1	107,305,027	90.5
2	10,740,311	9.1
3+	473,635	0.4
Total	118,518,973	100.0

(Source: Baumgardner et al, 2001)

More than 90 percent of Census 2000 housing units had only 1 census return. Of those returning 2 or more, Baumgardner et al (2001) shows that more than 78 percent of the duplicate returns actually represented the household of people, making unduplication a simple process. Only about two percent of the census housing units had multiple returns representing different households. Census 2000 data processing included rules to determine which household should be counted at the housing unit.

How many duplicate housing units were there?

Early comparisons of the DMAF to independent benchmarks suggested that the DMAF had overcoverage of housing units (Miskura, 2000). While the NRFU and CIFU operations included components to delete housing units from the DMAF, Census 2000 also had a data processing

phase to determine which housing units were duplicates and should be removed from the DMAF. This process relied on sophisticated matching of address information and matching of person data collected in the census. After housing units were identified as potential duplicates, a research operation looked in more detail and identified some housing units for reinstatement. The results of the housing unit unduplication operation are presented in Table 7.

Table 7. Number of Census 2000 Duplicate Housing Units with Population Counts[1]

	Number of Housing Units	Population Count
Potential Deletes	2,411,743	6,010,110
Reinstated	1,019,057	2,366,140
Deleted	1,392,686	3,643,970

[1] All data include the United States and Puerto Rico.
(Source: Miskura, 2000)

The housing unit unduplication operation attempted to balance between erroneously excluding people who lived in valid, unique housing units and erroneously including people who lived in true duplicates of other housing units. The percentage of housing units and people deleted represents about one percent of the total housing units and people included in Census 2000.

How much data were missing in Census 2000?

The quality of the census data is affected by the amount of respondent-provided information: lower rates of missing data generally mean higher data quality. Table 8 shows the missing data rates for the 100 percent data characteristics for Census 2000. Note these missing data rates are not comparable to the E-sample missing data rates presented later due to different definitions. The census missing data rates include edits, allocations, and substitutions, while the E-sample rates include only allocations and substitutions.

Table 8. Missing Data Rates for Census 2000 100 Percent Characteristics

Characteristic	Percent with Missing Data
Race	5.0%
Hispanic Origin	5.4%
Age	7.2%
Sex	3.0%
Tenure	5.3%

(Source: Baumgardner et al, 2001)

Accuracy and Coverage Evaluation Quality Indicators

This section provides some quality indicators for each component of the A.C.E., their relationship to census or A.C.E. quality, and the benchmark or statistics used to measure quality. While some tables are included in the text below, there are many detailed tables and figures in appendices or separate documents as indicated.

A.C.E. Estimation and Variance

The A.C.E. estimates are derived from two samples. A sample of census people, called the Enumeration sample or E sample, is compared to the independent A.C.E. sample of people, called the Population sample or P sample, to estimate net census coverage error. Dual system estimation (DSE) methodology is used for the comparison and to calculate the A.C.E. estimates. The DSEs are computed within post-strata, classifications of the P-sample and E-sample people based on geographic characteristics, housing unit characteristics, and person characteristics. The goal of post-stratification is to maximize the homogeneity of the samples with respect to the probability that a person was correctly enumerated in the census. Because the P sample and E sample are independent, the net proportions of matches, missed census people, and correctly enumerated E-sample people within each post-stratum can be generalized to the entire census using synthetic estimation.

Like all statistical estimates based on samples, the A.C.E. estimates are subject to sampling variation. The coefficient of variation (CV) summarizes the variance for the A.C.E. estimates. In addition, several types of bias may affect the estimates, including correlation bias (Bell, 2001) and synthetic bias (Malec and Griffin, 2001). The estimates are also subject to nonsampling error, such as noninterview and item nonresponse, which are discussed in later sections.

What are the undercount rates?

Table 9 summarizes the Census 2000 coverage rates with standard errors (SE) as measured by the Accuracy and Coverage Evaluation (A.C.E.) for the total population and seven race/Hispanic origin domains. See Davis (2001) for more details on A.C.E. results and Appendix 1 for details on the race/origin domain definitions. Note that the A.C.E. net undercount percent measures the undercount only in regular housing units, which comprised the A.C.E. sampling universe. Housing units in remote areas of Alaska and group quarters were ineligible for the A.C.E.

Table 9. A.C.E. 2000 Estimated Undercounts by Race/Origin Domain

Demographic Domain	A.C.E. Net Undercount Percent[1] (SE)
Total Population	1.18 (0.13)
Domain 1: American Indian on reservations	4.74 (1.20)
Domain 2: American Indian off reservations	3.28 (1.33)
Domain 3: Hispanic	2.85 (0.38)
Domain 4: Non-Hispanic Black	2.17 (0.35)
Domain 5: Native Hawaiian/Pacific Islander	4.60 (2.77)
Domain 6: Non-Hispanic Asian	0.96 (0.64)
Domain 7: Non-Hispanic White/Other Race	0.67 (0.14)

[1] Undercounts are computed for the population residing in housing units, excluding remote Alaska housing units and group quarters.
(Source: Davis, 2001)

The national net undercount of the household population for Census 2000 as measured by the A.C.E. is 1.18 percent (standard error, 0.13 percent). The net undercount for the 1990 Census as measured by the 1990 Post-Enumeration Survey (PES) was 1.61 percent (standard error, 0.20 percent). The A.C.E. results show differential undercoverage among the race/Hispanic origin domains, tenure, and age/sex groups. Davis (2001) contains more detailed estimation results for major demographic groups and for each individual post-stratum.

The standard errors in general are lower than expected. The A.C.E. has about twice the sample size of the 1990 PES, which should make the A.C.E. variances lower. But the A.C.E. variance reductions are larger than the expected reductions due to the sample size increase. This indicates not only high quality in the A.C.E. but also in Census 2000. The factors that contribute to variance, such as weight variation, A.C.E. interview and match rates, and census enumeration rates, combined to further reduce the A.C.E. variances.

How do the A.C.E. results compare to demographic analysis? What are the historical patterns in demographic analysis?

Demographic analysis (DA) uses administrative information, such as birth and death statistics, immigration data, and Medicare enrollments, to estimate the population by age, sex, and race (Black and Non-Black) at the national level. DA thus provides an independent assessment of the coverage of the census and the quality of the A.C.E., and has done so over the course of several decennial censuses. The DA methodology also produces sex ratios, the proportion of males per 100 females, to further check the demographic consistency of the A.C.E. estimates. Due to multiple race reporting in Census 2000, two different DA models are used to assess Black and

Non-Black Census 2000 counts. From 1940 to 1990, one DA model was used for the Black and Non-Black DA estimates. Note the population estimates used to derive the DA net undercount percentages include the total population residing in housing units and group quarters, while all A.C.E. undercounts in this document apply only to the population in housing units. Also note that measures of uncertainty are unavailable at this point for the DA estimates. See Robinson (2001) for more details on the two DA models, limitations of the DA estimates, and further DA results and comparisons.

Table 10A. Demographic Analysis Estimates of Net Undercount Percent
(A negative value denotes a net overcount.)

Group	DA Model[1]	DA Net Undercount Percent					A.C.E. Net Uct. % (SE)[2]
		1960	1970	1980	1990	2000	2000
Total	-	3.08	2.71	1.22	1.85	-0.65	1.18 (0.13)
Black	1	6.57	6.48	4.50	5.68	4.67	2.17 (0.35)
	2					0.93	
Non-Black	1	2.65	2.21	0.77	1.29	-1.48	1.04 (0.14)
	2					-0.90	

[1] Under DA Model 1, the Black group includes only people who mark "Black" as their single race on the census form, and all remaining people are in the Non-Black group. Under DA Model 2, the Black group includes people who mark "Black" alone or with any other races, and all remaining people are in the Non-Black group.
[2] In the A.C.E., the Black group includes all people in the Non-Hispanic Black domain, and people in all other race/Hispanic origin domains are in the Non-Black group.
(Source: Robinson, 2001, Davis, 2001, and DSE output files)

Table 10B. Demographic Analysis Estimates of Sex Ratios
(Number of males per 100 females.)

Group	1990			2000			
	DA	PES	Census	DA	A.C.E.	Census	
						DA Model 1	DA Model 2
Black	95.2	90.4	89.6	94.9	91.1	90.6	90.7
Non-Black	97.2	96.5	95.9	97.7	97.9	97.2	97.2

(Source: Robinson, 2001)

As Table 10A shows, DA estimates a net overcount in Census 2000. However, the DA estimate may be revised based on additional research, particularly on undocumented immigration. As

Robinson (2001) notes, an alternative DA methodology, based on an assumption that doubles the estimated rate of undocumented immigration during the 1990s, produces a net undercount of 0.32 percent. This is still lower than the A.C.E. estimate but not to the extent of the base DA overcount of 0.65 percent. Research on the DA sources and methodology will continue to further refine the DA estimates.

The undercount for the Black population is lower than in 1990 according to DA. However, the Black undercount estimate varies greatly for the two DA models, with the A.C.E. net undercount between the two. Table 10B indicates the A.C.E. may have correlation bias, particularly for the Black population. Compared to DA, the A.C.E. sex ratios for the Black population are lower, meaning the A.C.E. potentially understates the undercount rates of Black males. The low Census 2000 sex ratios also indicate a higher undercount for Black males relative to Black females.

How do the A.C.E. results compare to 1990 Post-Enumeration Survey results?

The Census Bureau measured the coverage of the 1990 Census using the PES. The basic methodology of the 1990 PES was similar to the A.C.E. survey, but there are some notable differences. Both surveys relied on a population sample independent from the census to determine census omissions and an enumeration sample of census records to determine erroneous inclusions. But the A.C.E. sample of more than 300,000 housing units is about twice as large as the 1990 PES. The sample for American Indians on reservations is also significantly higher in the A.C.E. than in the PES. More detailed differences between the A.C.E. and the 1990 PES are noted in sections below as appropriate.

Table 11 shows the net undercount percentages from the 1990 PES for the total population and five major race/Hispanic origin groups. Note that these major race/origin groups are directly comparable to the seven A.C.E. race/origin domains due to differences in the post-stratification designs. In particular, the A.C.E. design allows for multiple race reporting, which was not available in the 1990 Census. In addition, the larger sample size of A.C.E. enables the addition of post-stratification variables such as return rate. Haines (2000) details other features of the A.C.E. post-stratification design. Also note that the 1990 PES results in Table 11 include the group quarters population, unlike A.C.E. results given in this report. See Thompson (1992) and Bureau of the Census (1992) for additional 1990 PES results.

Table 11. 1990 Census Undercount Rates by Race/Origin Group

Race/Origin Group	PES Net Undercount Percent[1] (SE)
Total Population	1.61 (0.20)
Non-Hispanic White/Other	0.68 (0.22)
Black	4.57 (0.55)
Asian or Pacific Islander	2.36 (1.39)
American Indian on Reservations	12.22 (5.29)
Hispanic[2]	4.99 (0.82)

[1] Undercounts are computed for the population residing in housing units and certain types of group quarters eligible for the 1990 PES.
[2] The Hispanic group excludes Blacks, Asians and Pacific Islanders, and American Indians on reservations.
(Source: Davis, 2001)

Comparing Table 9 and Table 11, the total undercount and the differential undercount have decreased from the 1990 Census to Census 2000. Even though the PES groups and A.C.E. domains are not defined identically, they can be reasonably compared. The undercounts for major race/Hispanic origin demographic groups are lower in Census 2000 than in the 1990 Census. The Black and Hispanic undercounts dropped about 50 percent.

Also, the demographic estimates are more reliable under the A.C.E. design than the 1990 PES design. The A.C.E. standard errors for the total population and for the race/Hispanic origin domains are generally smaller than expected, after accounting for the sample size differences between the A.C.E. and 1990 PES. As discussed above, this indicates high quality both in the A.C.E. and in Census 2000.

How does each estimation step contribute to the final A.C.E. estimates?

Mule (2001) shows how each step of the estimation process contributes to the final A.C.E. estimates for the total population and for each of the 416 post-strata used in DSE after collapsing due to sample size and variance considerations. Table 12 summarizes the effects of these estimation steps on the total population correct enumeration rate, match rate, and the correction ratio, which is the correct enumeration rate divided by the match rate. The estimation steps considered include noninterview adjustment, characteristic imputation, imputation of unresolved status, and targeted extended search.

Table 12. Effects of the A.C.E. Estimation Steps on the National Rates

	Correct Enumeration Rate	Match Rate	Correction Ratio
Resolved Cases Prior to Estimation	0.9743	0.9406	1.0358
Total	0.9528	0.9159	1.0403

(Source: Mule, 2001)

As expected, resolved people tend to have higher match and correct enumeration rates than people involved in one of the estimation steps. For example, people requiring the targeted extended search tend to be in areas with high rates of geocoding error, which makes them more difficult to match. The correction ratio increases slightly due to the estimation steps, meaning the match rate dropped relatively more than the correct enumeration rate.

What are the estimates of ratio bias in the A.C.E.?

For each post-stratum, the A.C.E. estimate is a non-linear function of estimated totals from the P sample and E sample, and thus is potentially subject to ratio bias. Large ratio biases indicate that collapsing may be required for some post-strata. Table 13 gives the estimated ratio bias for the race/Hispanic origin domains, and the figure in Appendix 2 shows the distribution of relative ratio bias for the 416 post-strata that remained for A.C.E. estimation following collapsing due to sample size or variance issues (Haines, 2000).

Table 13. Ratio Bias by Race/Hispanic Origin Domain

Domain	Ratio Bias
American Indian on Reservations	0.067%
American Indian off Reservations	0.103%
Hispanic	0.019%
Non-Hispanic Black	0.015%
Native Hawaiian or Pacific Islander	0.120%
Non-Hispanic Asian	0.016%
Non-Hispanic White or Other Race	0.011%

These results suggest that ratio bias is negligible in the A.C.E. because the ratio bias estimate for each domain is less than 0.2 percent and most are less than 0.1 percent (Fay, 1999).

What is the regional variation of late census adds and census people with insufficient information?

Some people were added to the census too late to be in the E sample but are included in the coverage correction factors and A.C.E. synthetic estimates. Likewise, census people who are not data defined have insufficient information for matching and are excluded from the E sample but included in the synthetic estimates. The tables in Appendix 3 show the regional distribution of late adds and insufficient information people for the 64 major post-stratum groups.

Person Matching

The person matching phase of the A.C.E. is the process in which the people in interviewed housing units, the P sample, are matched to census person records, the E sample, in the same sample areas. Person matching is divided into before follow-up matching, person follow-up, and after follow-up coding. Before follow-up is a computer and clerical process to determine which person records match, which are duplicates, and which require follow-up. Person follow-up is a field operation to resolve the match or resident status of certain P-sample non-matches and the enumeration status for E-sample non-matches. After follow-up uses the results of the field follow-up to determine the final status of each person. All results presented in this document are after follow-up results because these provide the direct input for A.C.E. estimation. See Childers et al (2001) for detailed before follow-up matching results and person follow-up results.

What are the results of person matching?

Tables 14A and 14B summarize the person matching results for the total population. These results are collapsed over the full set of match codes and exclude non-sampled targeted extended search people. Childers et al (2001) includes more detail on person match codes and results.

Table 14A. Distribution of A.C.E. Person Matching Results for the P Sample (unweighted)

	Total People	Matches	Non-Matches	Unresolved	Removed
Number	653,338	578,695	54,424	7,826	12,393
Rate	100.0%	88.6%	8.3%	1.2%	1.9%

(Source: Childers et al, 2001)

Table 14B. Distribution of A.C.E. Person Matching Results for the E Sample (unweighted)

	Total	Correctly Enumerated	Erroneously Enumerated	Unresolved
Number	704,602	652,390	31,064	21,148
Rate	100 0%	92.6%	4.4%	3.0%

(Source: Childers et al, 2001)

How do person matching results vary among demographic groups?

People who match between the P sample and E sample were usually counted correctly in the census. Except in certain cases, non-matched people in the P sample and E sample are followed up to confirm their match status or enumeration status. P-sample people were missed by the census if they were field-verified as Census Day residents. E-sample people were correctly counted if they were field-verified as residents, otherwise they were erroneously enumerated. The status of some people in the P sample and E sample remains unresolved despite follow-up and requires imputation in A.C.E. missing data processing. Table 15 presents the after follow-up person match results for the total population and major race/Hispanic origin groups. Person matching occurs before A.C.E. missing data processing and post-stratification. This means that race or Hispanic origin could be blank, and also that these groups do not correspond exactly to the A.C.E. race/Hispanic origin domains created for A.C.E. estimation.

Table 15. Person Matching Results for Race/Hispanic origin groups (unweighted)

Race/Hispanic Origin Group	Percent of P-sample Not Matched[1]	Percent of E-sample Erroneous Enumerations[1]
Blank	16.6	6 6
American Indian on Reservations	12.4	3.0
American Indian off Reservations	12.2	5.0
Hispanic	12.4	7.3
Black	13.2	6.4
Native Hawaiian or Pacific Islander	15.1	3.7
Asian	9.6	4.5
White	6.3	3.5
Other Race	11.2	6.8
Multiple Race	9.3	3.5
Total	8.6	4.6

[1]These are percentages of resolved people before A.C.E. missing data processing and post-stratification.
(Source: Childers et al, 2001)

Not surprisingly, these results are correlated with the coverage rates shown earlier. The groups that traditionally have the lowest undercount rates also generally have lower non-match and lower erroneous enumeration rates.

What are the P-sample match results by mover status?

The A.C.E. P-sample consists of nonmovers, outmovers, and those with unresolved mover status. Table 16 decomposes the P-sample match results for the A.C.E. by mover status.

Table 16. Person Matching Results by Mover Status (unweighted)

Mover Status	Percent of P-sample Not Matched[1]
Nonmover	8.0
Outmover	23.6
Unresolved Mover Status	22.6
Total	8.6

[1]These are percentages of resolved people before A.C E. missing data processing and post-stratification.
(Source: Childers et al, 2001)

As expected, the nonmatch rates are significantly higher for outmovers and people with unresolved mover status. Nonmovers have not changed their residence, and thus it is easier to find their corresponding E-sample records and match them. Outmovers and unresolved people comprise only about 5 percent of the P sample, thus the overall match rate is largely determined by the nonmovers (Childers et al, 2001).

What are the types of E-Sample erroneous enumerations?

As in the 1990 PES, many different E-sample person matching outcomes are classified as erroneous enumerations for estimation purposes. Table 17 shows the weighted rates of each type of erroneous enumeration as a percentage of the E sample.

Table 17. Types of Erroneous Enumerations

Erroneous Enumeration	Percent of E Sample[1]
Insufficient Information	2.0%
Duplicate	0.9%
Fictitious	0.3%
Geocoding Error	0.3%
Other Residence	1.1%
Total	4.6%

[1]These are percentages of resolved people before A.C.E. missing data processing and post-stratification.
(Source: Childers et al, 2001)

How consistent are P-sample and E-sample responses?

People match between the P sample and E sample based on identifiers and characteristic data, such as name, race, and sex. Ideally, a matched person will provide the same response on the census form, the source of the E sample, and in the A.C.E. interview, the source of the P sample. This response consistency ensures that the person is placed in the same post-stratum in the P and E samples. However, some data are inconsistent due to imputation or to reporting errors. Table 18 shows the level of consistency between certain variables used for post-stratification: tenure, race/Hispanic origin domain, and age/sex. The other post-stratification variables, such as return rate, are processing or geographic variables that are always consistent for a matched person. Inconsistency has the potential effect of increasing the heterogeneity within a post-stratum, weakening the assumption that people in the same post-stratum have the same probability of correct census enumeration. Heterogeneity increases if the inconsistency occurs between two post-strata with very different coverage properties or if the inconsistency rates are high.

Table 18 includes only people who match in the P sample and E sample because the data for non-matches are not available. Data from 1990 are also not readily available. Matched cases are decomposed by imputation status to shed more light on the source of inconsistency. For imputed cases, inconsistency is usually introduced by the imputation procedure. For non-imputed cases, inconsistency is attributable to factors such as the data collection mode, recall bias, proxy responses, or data capture difficulties. Farber (2001) presents further consistency results.

Table 18. Consistency of Matching P-Sample and E-Sample Post-Stratification Variables

Variable	Matched People	Consistent People	Inconsistent People	
			Number	Percent
Tenure	549,645	523,830	25,815	4.70%
Non-Imputed	520,715	501,186	19,529	3.75%
Imputed	28,930	22,644	6,286	21.73%
Age/Sex	549,645	521,484	28,161	5.12%
Non-Imputed	511,792	497,176	14,616	2.86%
Imputed	37,853	24,308	13,545	35.78%
Race/Origin Domain	549,645	528,371	21,274	3.87%
Non-Imputed	505,389	489,299	16,090	3.18%
Imputed	44,256	39,072	5,184	11.71%

(Source: Farber, 2001)

Overall, the rates of inconsistency are low for the three post-stratification variables. Even though the imputed people have high rates of inconsistency, their small number compared to the number of non-imputed people means that they should have little effect on the A.C.E. estimates. The inconsistency observed in these post-stratification variables might have the effect of lowering the measured undercount for some levels of these variables, but the undercount is probably closer to the true undercount than a post-stratification design with coarser post-stratification variables.

What are the results of person matching quality assurance?

The person match QA procedures checked the change rates for two staff levels: clerks and technicians. Due to their extensive training and experience, the analysts are excluded from this change rate analysis. The change rate is the proportion of person match codes for the given staff level that were changed during review at a higher staff level. Note that not all cases were reviewed by higher levels. The QA checked a certain proportion of cases as a standard, with greater rates where problems were suspected. Table 19 shows the lower and upper limits for the after follow-up change rates for clerks and technicians. These limits are shown because three different models are used to estimate the change rate. The lower limit is the lowest of the three estimates, while the upper limit is the highest. Childers et al (2001) contain more details and results of person match QA.

Table 19. A.C.E. Person Match After Follow-up Change Rates

Staff Level	Change Rate	
	Lower Limit	Upper Limit
Clerk	0.11%	0.95%
Technician	0.13%	0.71%

(Source: Childers et al, 2001)

These change rates indicate low levels of matching error. A number of operational improvements were implemented since the 1990 Census to reduce matching error. At the maximum, the change rate for any level was less than one percent. Because the change rate is an overestimate of matching error, the true matching error was even lower than the estimated change rates.

What are the results of person follow-up quality assurance?

The person follow-up QA plan included a recontact of followed-up people to verify the initial contact by an enumerator. Each enumerator had a certain proportion of households recontacted as a standard, with higher rates of QA where problems were suspected. Table 20 gives the outcomes of the QA recontact. Passed cases had been correctly followed up, failed cases appeared to have been followed up but actually had not, and noninterviewed cases were not

successfully recontacted. Ineligible cases had more than one respondent on the person follow-up form or were selected after the QA cutoff date. Childers et al (2001) contain more details and results of person follow-up QA.

Table 20. Outcome of Cases in Person Follow-up Quality Assurance

Outcome	Randomly Selected		Targeted	
	Number	Percentage of Total QA Cases	Number	Percentage of Total QA Cases
Pass	3,899	43.67%	4,067	45.55%
Fail	40	0.45%	84	0.94%
Noninterview	271	3.04%	212	2.34%
Not Eligible	281	3.15%	75	0 84%
Totals	4,210	50.30%	4,363	49.70%

(Source: Childers et al, 2001)

Accuracy and Coverage Evaluation Missing Data Processing

The A.C.E. missing data process consists of three basic steps. The first is an adjustment of the weights of interviewed housing units to account for whole household noninterviews. Then missing data for individual P-sample person characteristics are imputed. Finally, the missing data process imputes a resolution for unresolved cases, such as unresolved P-sample residents and possible P-sample matches or E-sample correct enumerations. The results are then used in post-stratification and DSE. Cantwell et al (2001) provides further results and information on missing data processing.

How much data were missing in the A.C.E.?

The quality of the census and A.C.E. data is affected by the amount of respondent-provided information: lower amounts of imputed data generally mean higher quality data. The rates of missing or unresolved resident status, match status, and enumeration status are shown in Table 21A, and the rates of missing data in the post-stratification variables are in Table 21B. Note that E-sample people cannot have an unresolved resident or match status and P-sample people cannot have an unresolved enumeration status. See Cantwell et al (2001) for further details on A.C.E. missing data. Similar results from the 1990 PES are also given where available. Note the P sample in the A.C.E. consists of non-mover and out-mover households, while the 1990 PES P sample included non-movers and in-movers. This procedural difference could contribute to differences in unresolved and missing data rates between 1990 and 2000.

Table 21A. Overall A.C.E. and 1990 PES Unresolved Rates (unweighted)

Unresolved Characteristic	2000 A.C.E		1990 PES	
	P Sample	E Sample	P Sample	E Sample
Resident Status	2.3%	Not Applicable	Not Applicable	Not Applicable
Match Status	1.2%	Not Applicable	1.9%	Not Applicable
Enumeration Status	Not Applicable	3.0%	Not Applicable	2.4%

(Source: Cantwell et al, 2001)

Table 21B. Overall A.C.E. and 1990 PES Missing Data Rates (weighted)

Missing Characteristic	2000 A.C.E.		1990 PES	
	P Sample	E Sample	P Sample	E Sample
Race	1.4	3.2	2.5	11.8
Hispanic Origin	2.3	3.4	Not Available	Not Available
Age	2.4	2.9	0.7	2.4
Sex	1.7	0.2	0.5	1.0
Tenure	1.9	3.6	2.3	2.5

(Source: Cantwell et al, 2001)

Few people had unresolved status of any kind in the A.C.E., meaning the imputation methodology should have little effect on the estimates. The A.C.E. estimates are based largely on non-imputed resident status, match status, and enumeration status.

Table 21B shows that the rates of missing data were low in the A.C.E., similar to the 1990 PES. There were few problems gathering complete information about respondents in successful A.C.E. interviews for the P sample or from census forms for the E sample. This suggests that the post-stratification results accurately reflect respondents' true characteristics, and should help to reduce heterogeneity in the post-strata. The imputation method determines the post-strata for only a small number of people.

How much weight had been given to noninterviewed housing units and how was it redistributed to interviewed housing units in the P sample?

The noninterview adjustment is the process in which the weights of noninterviewed P-sample housing units are spread proportionally to successfully interviewed housing units. The weights are spread to housing units that have similar characteristics to minimize bias. Table 22 gives summary statistics of the noninterviewed housing units for the A.C.E. and 1990 PES. For A.C.E., noninterviews can refer either to interview day or Census Day, so each is listed. The 1990 PES estimation methodology did not require the reconstruction of the Census Day

household, thus Census Day noninterview statistics are not applicable for the 1990 PES. Weighted results use the final P-sample weights, including all A.C.E. sampling and missing data steps.

Table 22. Summary Statistics of A.C.E. Noninterviewed Housing Units

		A.C.E.		1990 PES
	Noninterviewed Housing Units	Unweighted	Weighted	Unweighted
Interview Day	Number	3,052	1,196,445	2,246
	Percent of Occupied Housing Units	1.1%	1.2%	1.6%
Census Day	Number	7,794	2,909,466	Not Applicable
	Percent of Occupied Housing Units	3.0%	2.9%	Not Applicable

(Source: Cantwell et al, 2001)

Because the percentage of noninterviewed housing units is small, the noninterview adjustments to interviewed housing units are also small, as explained in Cantwell et al (2001). For Census Day noninterview adjustment, more than half of the interviewed housing units received no additional weight, indicating that those noninterview adjustment cells contained zero noninterviewed housing units. About 80 percent of the Census Day interviews received an additional 5 percent or less of weight, and more than 90 percent received 10 percent or less extra weight to account for noninterviewed housing units. The maximum adjustment was 3, meaning the weight was tripled. However, only 2 out of the more than 250,000 Census Day interviewed housing units received this maximum noninterview adjustment.The interview day noninterview adjustment was overall smaller. About 75 percent of the interview day interviewed housing units received no additional weight. Overall, the effects of the noninterview adjustment were small. The weights of interviewed housing units were changed little by the additional weight they received to account for noninterviewed housing units.

What is the distribution of P-sample resident status before and after resident status imputation?

For some interviewed P-sample housing units, the resident status of a household member remains undetermined after person follow-up, and is imputed in the A.C.E. missing data process. Table 23 shows the weighted distribution of resident status before and after imputation for P-sample people. Resident status imputation was not done in the 1990 PES.

Table 23. Results of A.C.E. Resident Status Imputation (weighted)
(Weights are final estimation weights.)

	Resident	Non-Resident	Unresolved		Resident Rate for Resolved Cases
			Insufficient Info.	Other	
Before Imputation	96.9%	0.9%	1.2%	1.1%	99.1%
After Imputation	98.6%	1.4%	0.0%	0.0%	98.6%

(Source: Cantwell et al, 2001 and A.C.E. missing data output files)

The resident rate drops following imputation because the imputation method imputes lower residence probabilities for unresolved people. Both residents and non-residents play a role in the imputation method, and the zero resident rate of non-residents lowers the imputed resident rates.

What is the distribution of P-sample match status before and after match status imputation?

As with resident status, the match status of a P-sample person may be unresolved despite the attempts of field follow-up and requires imputation. The distribution of weighted P-sample match status before and after imputation is shown in Table 24.

Table 24. Results of A.C.E. Match Status Imputation (weighted)
(Weights are final estimation weights.)

		Match	Non-Match	Unresolved		Match Rate for Resolved Cases
				Insufficient Info.	Other	
A.C.E.	Before Imputation	91.0%	8.0%	1.0%	0.0%	91.9%
	After Imputation	91.8%	8.2%	0.0%	0.0%	91.8%

(Source: Cantwell et al, 2001)

As with resident status, the match rate drops after imputation because of the role played by non-matches and their zero match probability.

What is the distribution of E-sample enumeration status before and after enumeration status imputation?

An E-sample person may have an unresolved enumeration status after person follow-up. The distribution of weighted enumeration status is shown in Table 25 before and after imputation.

Table 25. Results of A.C.E. Enumeration Status Imputation (weighted)

	Correct	Erroneous	Unresolved	CE[1] Rate for Resolved Cases
Before Imputation	93.4%	4.0%	2.6%	95.9%
After Imputation	95.3%	4.7%	0.0%	95.3%

[1] CE means correct enumeration.
(Source: Cantwell et al, 2001)

The correct enumeration rate is also lowered by the role of erroneous enumerations and their zero CE rate in the imputation method.

Targeted Extended Search

For some block clusters, the A.C.E. included targeted extended search (TES), an extension of the search area to one surrounding ring of census blocks for person matching. The goal of TES is to reduce the variance of the estimates by reducing the number of unresolved and non-matched cases caused by geocoding error. The TES operation also reduces the occurrence of P-sample geocoding error. Some of the TES block clusters are targeted with certainty due to a high proportion of weighted or unweighted geocoding error or A.C.E. non-matches identified during the initial housing unit matching phase, or because they were relisted clusters. Other TES clusters are selected at random.

What effect does TES have on the correct enumeration rate and match rate?

Table 26 shows the national P-sample match rate and E-sample correct enumeration rate with TES and what these rates would have been if TES had not been implemented.

Table 26. A.C.E. Rates by TES Status (weighted)

	Correct Enumeration Rate	Match Rate
Rates with TES	95.28%	91.58%
Rates without TES	92.35%	87.74%
Difference	2.93%	3.84%

(Source: Navarro and Olson, 2001)

The TES operation had the desired effect. It successfully raised the match and correct enumeration rates by searching for and often finding matches outside of selected clusters. The increase in these rates also contributes to lower A.C.E. variances.

What about balancing error?

Theoretically, the errors of inclusion and exclusion should balance. An error of inclusion occurs when a person who actually lives outside an A.C.E. selected cluster is wrongly counted in the cluster. An error of exclusion occurs when a person who actually lives inside an A.C.E. selected cluster is wrongly counted outside the cluster. Without TES, these errors balance and the DSE is not biased, but its variance increases due to lower match and correct enumeration rates. The TES operation is designed to reduce the A.C.E. variance without introducing balancing error. To avoid balancing error, TES uses overlapping search areas of one surrounding ring of census blocks.

The greater increase in the match rate, about 3.8 percent, over the correct enumeration rate increase, about 2.9 percent, suggests that some aspect of the A.C.E. is out of balance or that TES is correcting potential P-sample geocoding error. The data do not exist at this time to determine if TES or some other process was the cause of this imbalance. Navarro and Olson (2001) includes more details on balancing error.

Accuracy and Coverage Evaluation Interviewing

The A.C.E. data are collected via computer assisted personal interviewing (CAPI) conducted at a probability sample of housing units. An A.C.E. interview is attempted at all housing units in the sample. While successful interviews with a household member are obtained at most of the housing units, some interviews are conducted with proxy respondents, such as building managers or neighbors, and some interviews are not successful and the housing unit is nonrespondent. The total interviewing workload was 300,913 housing units in the United States.

What is the distribution of interviewing results?

The results of A.C.E. interviewing for interview day and Census Day are listed in the Table 27. The interview day outcome refers to the success of getting an interview during the A.C.E. process with a current resident or with a proxy. The Census Day outcome shows whether the Census Bureau was successful in getting information about the Census Day residents from the interview day resident or proxy. Byrne et al (2001) includes further person interviewing results, and Childers et al (2001) has detailed explanations of the interview outcomes.

Table 27. A.C.E. Interview Results (unweighted)

Interview Outcome	Interview Day		Census Day	
	Housing Units	Percent	Housing Units	Percent
Interviews	264,103	87.8	254,175	84.5
Complete interview with household member	249.854	83.0	233,327	77.5
Complete interview with proxy	12,317	4.1	18,335	6.1
Sufficient partial interview	1,932	0.6	2,513	0.8
Noninterviews	3,052	1.0	7,794	2.6
Field noninterview	373	0.1	2,667	0.9
Insufficient information for all people in household	2,196	0 7	2,418	0.8
No interview day/Census Day residents - converted to noninterview	483	0.2	2,709	0.9
Vacants	29,662	9.9	28,472	9.5
Vacant on interview day/ Census Day[1]	29,662	9.9	23,911	7.9
No Census Day residents - vacant[2]	Not Applicable	Not Applicable	4,561	1.5
Deletes - Not a housing unit on interview day/Census Day	4,096	1.4	10,472	3.5
Total	300,913	100	300,913	100

[1]"Vacant on Census Day" means that the interviewer determined the Census Day vacancy status.
[2] "No Census Day residents" means that the household should have been counted somewhere else on Census Day. These are whole household nonresidents. See Childers and Fenstermaker (2000) for more information.
(Source: Childers et al, 2001)

The overall interview rate is the percentage of successful interviews out of the total number of occupied housing units. Vacant and delete housing units are excluded. Overall, the unweighted interview rates are:

- 98.9 percent for interview day
- 97.0 percent for Census Day

As Childers et al (2001) points out, the 98.9 percent interview day rate exceeded the expected interview rate of 98 percent. The interview day interview rate in the 1990 PES was 98.4 percent unweighted (Childers et al, 2001). The different estimation methodology of the 1990 PES did not require the calculation of a 1990 Census Day interview rate.

How many interviews were proxy interviews?

The A.C.E. interviewing resulted in many complete interviews, but not all of these were with a member of the household. Also, some interviews could only be partially completed or not conducted at all. The distribution of outcomes for the interview workload is given in Table 28.

Table 28. Distribution of A.C.E. Interviewing Outcomes by Respondent Type (unweighted)

	Total Workload	Complete Interview	Partial Interviews	Refusal, No Knwl Resp or Lang. Barrier[1]	Vacant on Interview Day	Nonexistent on Interview Day
Total	300,913	252,562	14,220	373	29,662	4096
Householder	252,598	242,191	10,392	6	8	1
Proxy	48,315	10,371	3,828	367	29,654	4,095

[1]Refusal, No Knowledgeable Respondent, or Language Barrier.
(Source: Byrne et al, 2001)

These results show that the A.C.E. data were collected largely from knowledgeable respondents. Householders or household members took part in more than 80 percent of the interview workload, while even proxy respondents were usually able to provide at least a partial interview or identify a unit as vacant or nonexistent.

How does the distribution of interviews differ by phase of interview?

The A.C.E. interviewing was conducted in two stages: telephone and personal visit. Initial A.C.E. interviewing used the telephone. When a census mailback response from certain types of housing units in the A.C.E. sample areas included a telephone number, a telephone interview was attempted. Telephone interviewing potentially resulted in less recall bias and fewer movers. About 30 percent of the total interviewing workload was completed by telephone. The second stage of interviewing was personal visit interviewing. Table 29 shows the distribution of the interview workload by stage.

Table 29. Distribution of the A.C.E. Sample by Stage of Interview and Mover Status

Stage	Number of Interviews	Percent of Total
Telephone	88,573	29.4%
Personal Visit	212,340	70.6%
Total	300,913	100.0%

(Source: Byrne et al, 2001)

Almost 30 percent of the interviewing workload was completed by telephone, which exceeded expectations. Because these interviews were close to Census Day, the collected data is likely of high quality due to lower recall bias. The telephone stage also reduced the workload for the personal visit stage, which in turn possibly reduced costs and gave interviewers more time to collect quality data from household members or knowledgeable proxy respondents.

How many interviews went to nonresponse conversion?

Nonresponse conversion (NRCO) is the process in which the best interviewers attempt to obtain interviews at housing units where previous efforts have not been successful. Table 30 shows the number of cases that entered NRCO and their outcomes.

Table 30. A.C.E. Nonresponse Conversion Results

	Interview[1]	Refusal	Vacant	Non-Existent	Total NRCO Cases
Number of Cases	8,264	217	1,110	144	9,735
Percent of NRCO Cases	84.9%	2.2%	11.4%	1.5%	100%

[1]Includes complete and partial interviews.
(Source: Byrne et al, 2001)

Many of the NRCO cases had been refusals during the initial attempts of interviewing. Thus the conversion of about 85 percent of these cases to successful interviews represents a large reduction in the overall number of refusals in the A.C.E. interviewing process.

What were the results of interviewing quality assurance?

The interviewing QA involved a reinterview at a sample of housing units to verify that the unit had been interviewed. A random sample of about 5 percent of the housing units comprised one part of the QA, while a targeted sample of about another 5 percent was available as needed to handle potential problems. Table 31 shows the rates and outcomes of the interviewing QA procedures overall and for each phase of interviewing.

Table 31. A.C.E. Person Interviewing Quality Assurance Results

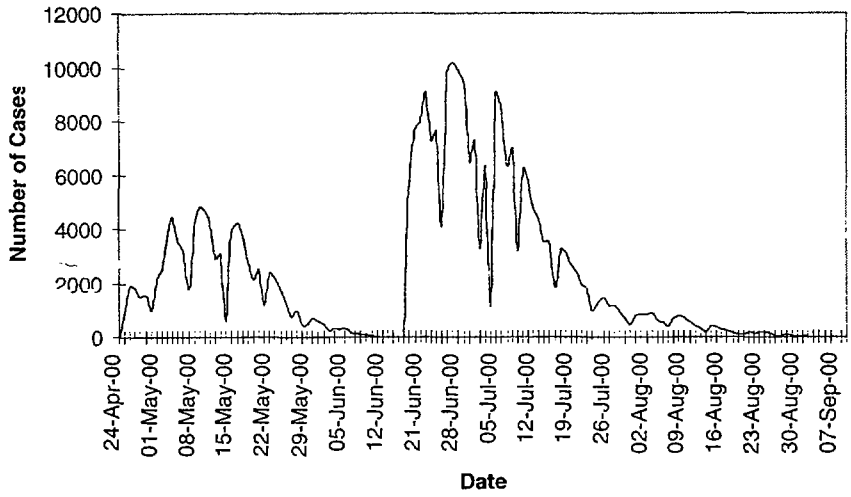
Phase	QA Outcome	Random QA	Targeted QA
Overall	Pass	14,707 (99.8%)	19,951 (99.0%)
	Fail	19 (0.1%)	171 (0 8%)
	Undetermined	14 (0.1%)	33 (0.2%)
	Total	14,740 (100%)	20,155 (100%)
Telephone	Pass	4,398 (100.0%)	4,622 (99.5%)
	Fail	2 (0.0%)	17 (0.4%)
	Undetermined	0 (0.0%)	5 (0.1%)
	Total	4,400 (100%)	4,644 (100%)
Personal Visit	Pass	10,309 (99.7%)	15,329 (98 8%)
	Fail	17 (0.2%)	154 (1.0%)
	Undetermined	14 (0 1%)	28 (0 2%)
	Total	10,340 (100%)	15,511 (100%)

(Source: Byrne et al, 2001)

How much time lapsed between Census Day and the collection of the A.C.E. data?

The quality of interviews can vary depending on the date when the interview was conducted. For example, interviewers were trained not to use proxy respondents during the first three weeks a unit is assigned for interviewing. Thus earlier interviews may be of higher quality because they generally were conducted with a household member. Likewise, interviews conducted further from Census Day may be more prone to recall bias or might involve more movers. Figure 1 shows the completion of the interview workload by date. Table 4.1 in Appendix 4 gives similar information but with a decomposition by interview phase, telephone or personal visit.

Figure 1. Person Interviewing Workload by Date



Although interviewing in some areas was completed as late as September, about 99 percent of the workload was completed by the end of the first week in August, about four months after Census Day.

P-Sample and E-Sample Weights

The housing units and people in the P sample and E sample are weighted to account for people not in the A.C.E. survey. The size, distribution, and computation of the weights potentially affect the variance and bias of the A.C.E. estimates.

How many clusters had their weights trimmed and by how much?

Weight trimming is the procedure to control the influence of outlier clusters by reducing the weight applied to the people in the cluster. Weight trimming may introduce bias into the estimates but it also reduces variance, with the overall goal of reducing total mean squared error. Clusters were identified for weight trimming by computing an estimate of weighted net error, the difference between the weighted omissions and weighted erroneous enumerations. The weight of a cluster on an American Indian reservation was trimmed if its weighted net error exceeded 6,250, while a cluster net error had to exceed 75,000 for the balance of the United States to be trimmed. The list of clusters that required weight trimming is provided below with their weighted net errors before and after trimming, along with the national total weighted net error before and after trimming. The cluster codes in Table 32 have been changed from the actual codes to prevent disclosure.

Table 32. A.C.E. Weighted Net Errors

Cluster Code	Estimated Weighted Net Error Before Trimming	Weighted Net Error After Trimming
1	77,975	75,000
All Clusters	10,700,474	10,703,449

Note that the total weighted net error actually increased after weight trimming. This occurred because of its computation, which is the absolute difference of the total weighted omission and the total weighted erroneous enumerations. Weight trimming reduced both components of the difference in equal proportion but in unequal absolute numbers, which caused the difference to increase.

How much weight variation is there?

The figures in Appendix 5 show the amount of variation in the final person weights for the A.C.E. and the 1990 PES. The A.C.E. has lower weight variation. The maximum weight in the A.C.E. was below 6,000, while the 1990 PES had some people with weights over 20,000. About 99 percent of the A.C.E. P-sample and E-sample weights are lower than 800, compared to about 75 percent of the 1990 PES P-sample and E-sample weights. The A.C.E. also has fewer people with extremely small weights.

References

Baumgardner, S., Moul, D., Pennington, R., Piegari, R., Stackhouse, H., Zajac, K., Alberti, N., Reichert, J., Treat, J., (2001), "Quality of Census 2000 Processes," DSSD Census 2000 Procedures and Operations Memorandum Series B-3*, February 28, 2001.

Bell, W. (2001), "Accuracy and Coverage Evaluation: Correlation Bias," DSSD Census 2000 Procedures and Operations Memorandum Series B-12*, February 28, 2001.

Bureau of the Census (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," Internal Census Bureau Memorandum.

Byrne, R., Imel, L., Ramos, M., and Stallone, P., (2001), "Accuracy and Coverage Evaluation: Person Interviewing Results," DSSD Census 2000 Procedures and Operations Memorandum Series B-6*, February 28, 2001.

Cantwell, P., McGrath, D., Nguyen, N., and Zelenak, M.F., (2001), "Accuracy and Coverage Evaluation: Missing Data Results," DSSD Census 2000 Procedures and Operations Memorandum Series B-7*, February 28, 2001.

Childers, D., Byrne, R., Adams, T., and Feldpausch, R., (2001), "Accuracy and Coverage Evaluation: Person Matching and Follow-up Results," DSSD Census 2000 Procedures and Operations Memorandum Series B-6*, February 28, 2001.

Childers, D., and Fenstermaker, D., "Accuracy and Coverage Evaluation: Overview of the Design," DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-2, January 11, 2000.

Davis, P., "Accuracy and Coverage Evaluation: Dual System Estimation Results," DSSD Census 2000 Procedures and Operations Memorandum Series B-9*, February 28, 2001.

Farber, J., "Accuracy and Coverage Evaluation: Consistency of Post-Stratification Variables," DSSD Census 2000 Procedures and Operations Memorandum Series B-10*, February 28, 2001.

Fay, R., "Accuracy and Coverage Evaluation Survey: Plan and Initial Results for Evaluating Ratio Bias for Dual System Estimates Under Poststratification Alternatives," DSSD Census 2000 Procedures and Operations Memorandum Series Q-8, September 22, 1999.

Griffin, R., and Malec, D., "Accuracy and Coverage Evaluation: Assessment of Synthetic Assumptions," DSSD Census 2000 Procedures and Operations Memorandum Series B-14*, February 28, 2001.

Haines, D., "Accuracy and Coverage Evaluation Survey: Final Post-stratification Plan for Dual System Estimation", DSSD Census 2000 Procedures and Operations Memorandum Series Q-25, April 20, 2000.

Miskura, S., "Results of Reinstatement Rules for the Housing Unit Duplication Operations," Census 2000 Informational Memorandum no. 82, November 21, 2000.

Mule, V., "Accuracy and Coverage Evaluation: Decomposition of Dual System Estimation Components," DSSD Census 2000 Procedures and Operations Memorandum Series B-8*, February 28, 2001.

Navarro, A., and Olson, D., "Accuracy and Coverage Evaluation: Effect of Targeted Extended Search," DSSD Census 2000 Procedures and Operations Memorandum Series B-18*, February 28, 2001.

Robinson, J. G. (2001), "Accuracy and Coverage Evaluation: Demographic Analysis Results," DSSD Census 2000 Procedures and Operations Memorandum Series B-4*.

Thompson, J. (1992), "CAPE Processing Results," Internal Census Bureau Memorandum.

Appendices

Race/Hispanic Origin Domains

The race/Hispanic origin domain assignment generally follows the guidelines listed below, but it is essential to see Haines (2000) for the complete set of rules used to classify people into one of the seven domains. The race/Hispanic origin domain assignment is hierarchical.

Domain 1 (American Indian or Alaska Native on reservations) includes:

- All people on a reservation with American Indian or Alaska Native either as their single race or as one of multiple races, regardless of their Hispanic origin.

Domain 2 (American Indian or Alaska Native off reservations) includes:

- All people in Indian Country[1] but not on a reservation with American Indian or Alaska Native either as their single race or as one of multiple races, regardless of their Hispanic origin.

- All non-Hispanic people not in Indian Country with American Indian or Alaska Native as their single race.

Domain 3 (Hispanic) includes:

- All Hispanic people in Indian Country, excluding those with American Indian or Alaska Native either as their single race or as one of multiple races.

- All Hispanic people not in Indian Country, excluding those who live in the state of Hawaii and have Native Hawaiian or Pacific Islander as a single race or as one of multiple races.

---

[1] Indian Country is land considered (either wholly or partially) on an American Indian reservation/trust land, Tribal Jurisdiction Statistical Area, Tribal Designated Statistical Area, or Alaska Native Village Statistical Area. For Census 2000, Tribal Jurisdiction Statistical Area has been formally renamed as Oklahoma Tribal Statistical Area.

Domain 4 (Non-Hispanic Black) includes:

- All non-Hispanic people with Black as their only race.
- All non-Hispanic people with the race combination of Black and American Indian or Alaska Native who do not live in Indian Country.
- All people with the race combination of Black and another single race group (Native Hawaiian or Pacific Islander, Asian, White, or "Some other race"), excluding those who live in the state of Hawaii and are Native Hawaiian or Pacific Islander in addition to Black.

Domain 5 (Native Hawaiian or Pacific Islander) includes:

- All non-Hispanic people with the single race Native Hawaiian or Pacific Islander.
- All non-Hispanic people with the race combination of Native Hawaiian or Pacific Islander and American Indian or Alaska Native who do not live in Indian Country.
- All non-Hispanic people with the race combination of Native Hawaiian or Pacific Islander and Asian.
- All people living in the state of Hawaii with Native Hawaiian or Pacific Islander race, regardless of their Hispanic origin and whether they identify with a single race or multiple races.

Domain 6 (Non-Hispanic Asian) includes:

- All non-Hispanic people with Asian as their single race.
- All people with the race combination of Asian and American Indian or Alaska Native who do not live in Indian Country.

Domain 7 (Non-Hispanic White or "Some other race") includes:

- All non-Hispanic people self-identifying as either White or "Some other race" as their single race, or self-identifying as both White and "Some other race."
- All non-Hispanic people with the race combination of American Indian or Alaska Native and White or "Some other race" who do not live in Indian Country.
- All non-Hispanic people with the race combinations of Asian and White or "Some other race."
- All non-Hispanic people with the race combination of Native Hawaiian or Pacific Islander and White or "Some other race," excluding those who live in the state of Hawaii.
- All non-Hispanic people with three or more races who live in Indian Country, excluding those with American Indian or Alaska Native as one of the races.
- All non-Hispanic people with three or more races and who do not live in Indian Country, excluding those who live in Hawaii and have Native Hawaiian or Pacific Islander as one of the races.

Figure 2.1. Distribution of A.C.E. Estimated Ratio Bias for the 416 Collapsed Post-Strata

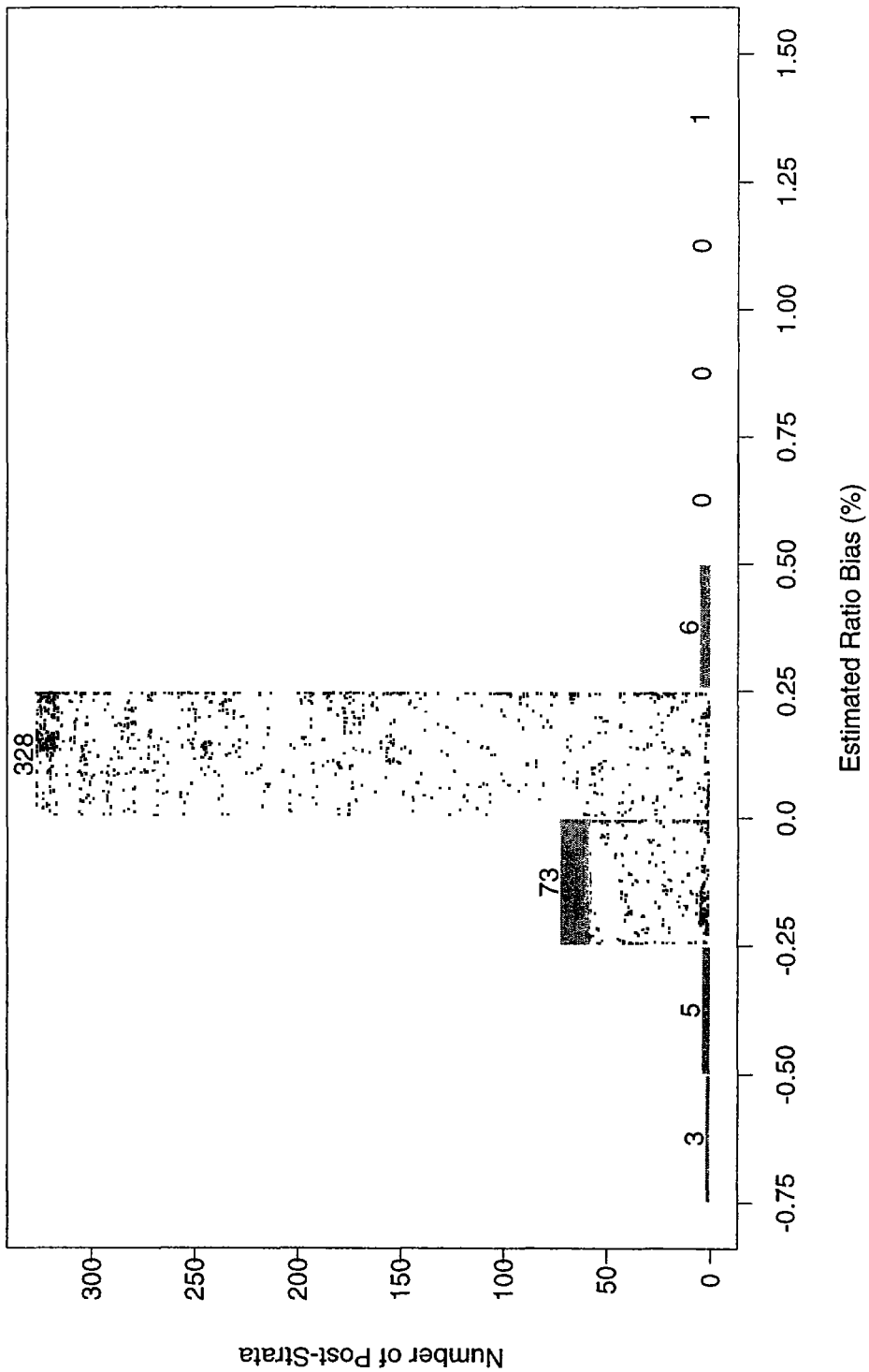


Table 3.1: Census 2000 A.C.E. 64 Post-Stratum Groups by region - Percent Late Adds

Race/Hispanic Origin Domain Number*	Tenure	MSA/TEA	High Return Rate				Low Return Rate			
			NE	MW	S	W	NE	MW	S	W
Domain 7 (Non-Hispanic White or "Some other race")	Owner	Large MSA MO/MB	05	0.2	02	03	21	14	03	05
		Medium MSA MO/MB	04	02	0.3	03	10	06	04	04
		Small MSA & Non-MSA MO/MB	04	03	03	03	08	04	04	0.5
		All Other TEAs	13	13	16	14	19	19	23	22
	Non-Owner	Large MSA MO/MB	12	08	05	0.5	17	16	07	06
		Medium MSA MO/MB	13	07	06	06	1.8	09	06	06
		Small MSA & Non-MSA MO/MB	13	07	07	0.7	15	07	07	07
		All Other TEAs	32	2.2	24	27	38	25	28	35
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB	08	06	03	03	32	17	04	0.3
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	10	04	13	07	0.8	07	23	14
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	09	1.0	06	05	16	17	08	06
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	15	0.6	12	0.7	19	06	17	1.2
		All Other TEAs								
Domain 3 (Hispanic)	Owner	Large MSA MO/MB	1.2	05	0.4	04	3.3	27	06	05
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	11	0.6	12	09	12	1.1	22	24
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	12	10	06	05	19	22	05	06
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	2.1	10	1.4	1.4	17	09	21	3.3
		All Other TEAs								

		NE	MW	S	W
Domain 5 (Native Hawaiian or Pacific Islander)	Owner	16	05	04	09
	Non-Owner	12	06	06	10
Domain 6 (Non-Hispanic Asian)	Owner	18	05	04	05
	Non-Owner	13	08	05	06
American Indian or Alaska Native – **Domain 1** (On Reservation)	Owner	16	08	15	09
	Non-Owner	08	0.7	14	11
American Indian or Alaska Native – **Domain 2** (Off Reservation)	Owner	14	07	15	09
	Non-Owner	17	10	13	10

- For Census 2000, persons can self-identify with more than one race group For post-stratification purposes, persons are included in a single Race/Hispanic Origin Domain This classification does not change a person's actual response Further, all official tabulations are based on actual responses to the census

Table 3.2: Census 2000 A.C.E. 64 Post-Stratum Groups by Region - Percent iis

Race/Hispanic Origin Domain Number*	Tenure	MSA/TEA	High Return Rate				Low Return Rate			
			NE	MW	S	W	NE	MW	S	W
Domain 7 (Non-Hispanic White or "Some other race")	Owner	Large MSA MO/MB	09	09	10	11	26	26	18	18
		Medium MSA MO/MB	08	0.8	11	12	18	16	16	19
		Small MSA & Non-MSA MO/MB	08	09	09	11	12	18	1.5	19
		All Other TEAs	16	11	11	28	16	15	15	25
	Non-Owner	Large MSA MO/MB	21	2.0	27	19	44	54	39	31
		Medium MSA MO/MB	16	16	24	23	35	30	34	32
		Small MSA & Non-MSA MO/MB	14	16	2.0	1.9	2.2	33	28	28
		All Other TEAs	21	16	18	37	3.0	21	25	39
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB	27	24	24	25	53	49	33	3.4
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	26	23	22	29	28	34	33	52
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	44	33	35	31	66	60	4.4	45
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	25	30	29	32	37	59	39	56
		All Other TEAs								
Domain 3 (Hispanic)	Owner	Large MSA MO/MB	25	25	26	40	5.9	53	42	50
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	24	23	40	39	3.8	3.5	4.1	45
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	36	32	32	38	54	5.7	40	4.3
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	25	2.8	37	41	39	4.0	49	54
		All Other TEAs								

			N	MW	S	W
Domain 5 (Native Hawaiian or Pacific Islander)		Owner	33	37	28	37
		Non-Owner	51	43	4.3	36
Domain 6 (Non-Hispanic Asian)		Owner	28	2.3	22	24
		Non-Owner	38	33	3.3	31
American Indian or Alaska Native	**Domain 1** (On Reservation)	Owner	88	49	42	50
		Non-Owner	67	47	37	46
	Domain 2 (Off Reservation)	Owner	27	21	19	30
		Non-Owner	36	27	2.7	32

•For Census 2000, persons can self-identify with more than one race group For post-stratification purposes, persons are included in a single Race/Hispanic Origin Domain This classification does not change a person's actual response Further, all official tabulations are based on actual responses to the census

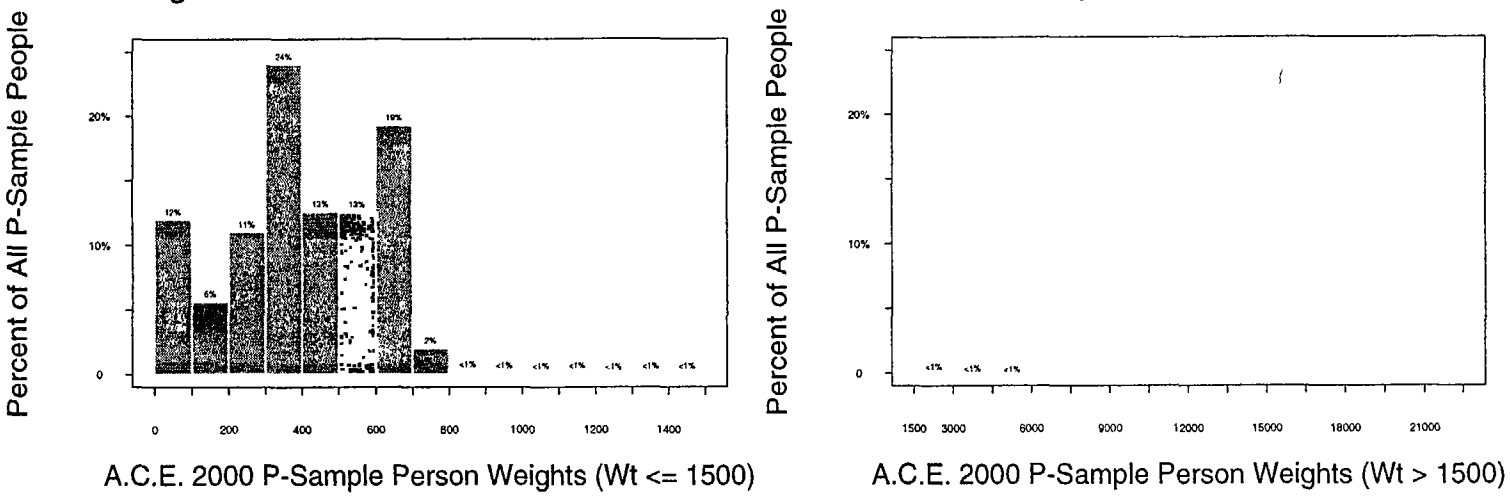
Timing of the A.C.E. Interviewing

Table 4.1 gives the distribution of the A.C.E. interview workload by date of completion and by phase. Because the goal of the A.C.E. interview is to reconstruct the Census Day household, the timing of the interview is important to reduce recall bias and to minimize the number of people who have moved since Census Day, April 1, 2000. About 90,000 interviews, or 30 percent of the total workload, were conducted in the telephone phase, which finished the week of June 11, 2000. By the week of July 16, over 90 percent of all interview were done. Byrne et al (2001) has more details, including the timing for each A.C.E. regional office.

Table 4.1. Distribution of Personal Interviewing Workload by Interview Week- Unweighted

Phase	Week Starting On	Number	Cumulative Percent
All	Total	300,913	100
Telephone	Total Telephone	88,573	29.4
	April 23, 2000	7,699	2.6
	April 30, 2000	20,590	9.4
	May 7, 2000	25,638	17.9
	May 14, 2000	19,728	24.5
	May 21, 2000	10,497	28.0
	May 28, 2000	3,232	29.0
	June 4, 2000	1,154	29.4
	June 11, 2000	35	29.4
Personal Visit	Total Personal Visit	212,340	70.6
	June 18, 2000	45,204	44.5
	June 25, 2000	57,241	63.5
	July 2, 2000	41,642	77.3
	July 9, 2000	31,344	87.7
	July 16, 2000	17,038	93.4
	July 23, 2000	7,764	96.0
	July 30, 2000	5,057	97.7
	Aug 6, 2000	3,982	99.0
	Aug 13, 2000	1,756	99.6
	Aug 20, 2000	939	99.9
	Aug 27, 2000	336	100.0
	Sept 3, 2000	36	100.0
	Sept 10, 2000	1	100.0

Figure 5.1. Distribution of A.C.E. 2000 and 1990 PES P-Sample Person Weights



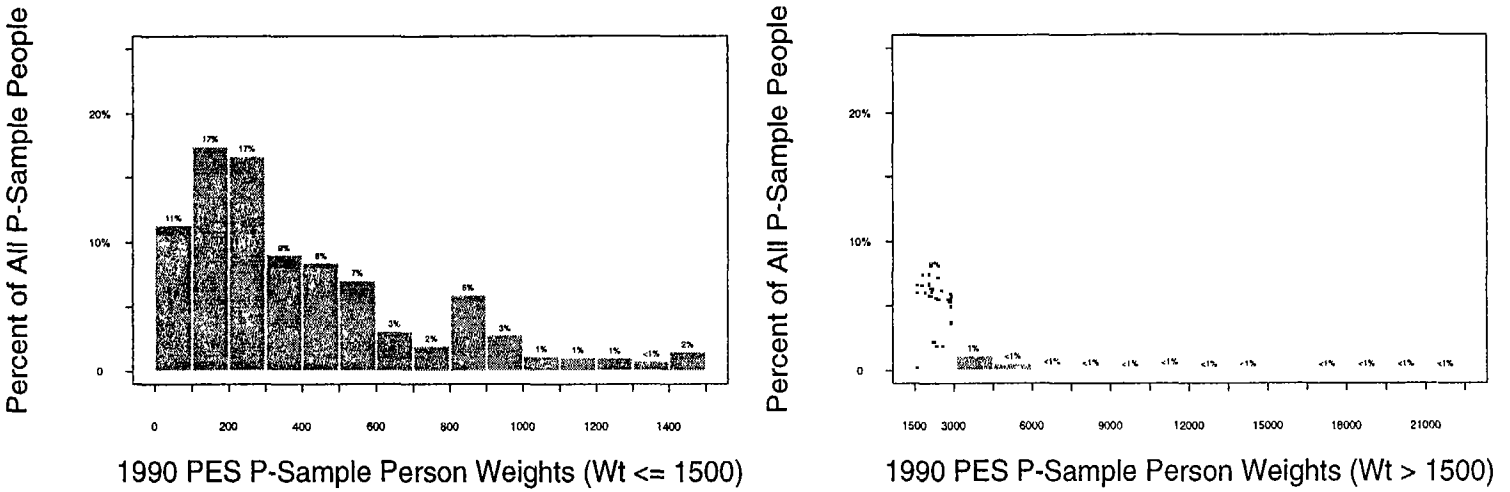


Figure 5.2. Distribution of A.C.E. 2000 and 1990 PES E-Sample Person Weights