Executive Steering Committee
For A.C.E. Policy II
(ESCAPII)
Report 14
September 19, 2001

Conditioning of Census 2000 Data Collected in Accuracy and Coverage Evaluation Block Clusters

Katie Bench
Planning, Research,
and Evaluation Division

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

CONDITIONING OF CENSUS 2000 DATA COLLECTED IN ACCURACY AND COVERAGE EVALUATION BLOCK CLUSTERS

EXECUTIVE SUMMARY

Did the Accuracy and Coverage Evaluation (A.C.E.) contaminate Census 2000 data collected in A.C.E. blocks?

The evidence suggests that contamination bias is not a problem. Globally, we did not find evidence of contamination bias for high-level proportions and averages. We computed a t-statistic to see if the ratio $N_{c,ace}/N_c$ was significantly different from one for the nation and the 36 evaluation poststrata. None of these t-tests were significant. In addition, the t-tests used to detect significant differences between A.C.E. and non-A.C.E. proportions for the second and third fundamental indicators yielded little to no evidence of contamination.

The study also broke the data down to very detailed cells. These cells were demographic, geographic, and response related indicators of contamination broken down by the 36 evaluation poststrata, and region and Type of Enumeration Area (TEA). No systematic error was detected in these cells, although the number of significant results were somewhat above chance levels. Many of them were not considered as significant when we drew conclusions. This happened under two circumstances. First, there were several proportions that were close to 0 or 1. We regarded t-tests that used these small or large proportions to be unreliable because design based estimation procedures underestimate the variances for small proportions. Second, some of the differences between A.C.E. and non-A.C.E. proportions were extremely small. So, while these difference were mathematically significant they were not practically significant.

Based on these findings the total error model will not adjust for contamination bias, and concerns about contamination should not affect the adjustment decision. This finding is consistent with the earlier assumption that contamination bias would not occur during Census 2000. The discussion below provides more understanding of the study.

Background of Study

The A.C.E. uses the dual-system estimation method. The dual system estimation method assumes there are two independent lists of the population. The first list is the original census enumerations, and the second is a list of those covered by the sampling frame for the A.C.E. sample.

The independence assumption can fail due to causal dependence, or conditioning of Census 2000 data collected in A.C.E. block clusters. This can also be referred to as contamination. Contamination occurs when the event of an individual's inclusion or exclusion from one list affects the probability of their inclusion in the other list. Research undertaken in the 1990 census and on test censuses leading up to Census 2000 mostly show that we have not experienced

contamination in the past between the census and the coverage measurement survey. One paper found some possible evidence of contamination, and another found the update/leave Type of Enumeration Areas to be a weak area of concern for contamination.

Methods of Study

The Whole Group Analysis (WGA) tries to determine if the A.C.E. contaminated Census data collected in A.C.E. Blocks. To determine the potential existence of contamination, the WGA aggregates census data in A.C.E. blocks to the national, evaluation poststrata, or regional and TEA levels. Then the WGA compares census data in A.C.E. blocks to census data in non-A.C.E. blocks to see if significant differences exist. We examined the data to see if there was any firm evidence of contamination for:

- Three fundamental indicators.
- Demographic, geographic, and response related indicators.

The first fundamental indicator of contamination is the ratio $N_{c,ace}/N_{c.}$ To detect contamination we performed a t-test to see if the ratio differed significantly from one. We calculated this ratio at the national and evaluation poststrata levels. $N_{c,ace}$ is the sample-weighted number of census enumerations in the A.C.E., and N_c is the census count from all clusters.

The second fundamental indicator of contamination is the average number of persons per block. This average helps determine if the A.C.E. affected the census count in A.C.E. blocks. To detect contamination we calculated this average for A.C.E. and non-A.C.E. blocks at the 36 evaluation poststrata level, but not at the national level, and performed t-tests.

The third fundamental indicator of contamination is the average number of housing units per block. This average helps determine if the A.C.E. affected the census housing unit count in A.C.E. blocks. To detect contamination we calculated this average for A.C.E. and non-A.C.E. blocks at the 36 evaluation poststrata level, but not at the national level, and conducted t-tests.

We further investigated the presence of contamination with demographic, geographic, and response related indicators at the evaluation poststrata level and region and TEA levels. We compared census data in A.C.E. block clusters to census data in non-A.C.E. block clusters by performing t-tests for the difference between various proportions in A.C.E. block clusters and proportions in non-A.C.E. block clusters. For example, we calculated the proportion of census people in A.C.E. blocks who are of the Black or African American race, and the proportion of census people in non-A.C.E. blocks who are of the Black or African American race. We subtracted the proportion for non-A.C.E. blocks from the proportion for A.C.E. blocks and did a t-test on this difference to see if it was significantly different from zero. One set of t-statistics were calculated by the 36 evaluation poststrata, and the other set of t-statistics were calculated by region and TEA. We also calculated the t-statistics by the 16 preliminary evaluation poststrata.

1. BACKGROUND

This report provides information to help determine if the Accuracy and Coverage Evaluation (A.C.E.) contaminated census data collected in A.C.E. blocks.

The A.C.E. uses the dual-system estimation method. The dual system estimation method assumes there are two independent lists of the population. The first list is the original census enumerations, and the second is a list of those covered by the sampling frame for the A.C.E. sample (Hogan, 2000).

The independence assumption can fail due to causal dependence, or contamination, between the two lists. Contamination occurs when the event of an individual's inclusion or exclusion from one list affects the probability of their inclusion in the other list (Mulry and Spencer, 1991). Research undertaken in the 1990 census and on test censuses leading up to Census 2000 mostly show that we have not experienced contamination in the past between the census and the coverage measurement survey (Davis, 1990; Hawala, 1999). One paper found some possible evidence of contamination, and another found the update/leave Type of Enumeration Areas to be a weak area of concern for contamination (Griffiths, 1996; Bench, Kearney, and Petroni, 2000). Brief descriptions of these studies follow:

- The 1990 Post Enumeration Survey (PES) Evaluation Project, P14, Part I: *Independence of the Census and the P-sample: Comparison of Blocks* by Mary Davis conducted a paired block analysis that looked at the difference between PES blocks and non-PES blocks. This analysis first drew a sample of PES blocks paired with comparable non-PES blocks; second, for each block, aggregated census data from person/housing unit level records to block level records; third, tested preliminary variables for relevance, completeness, and redundancy; and finally, tested the resulting data for the difference between PES and non-PES blocks with paired t-test comparisons and Wilcoxon signed rank tests. After testing a variety of census data, Davis concluded that no differences attributable to the PES were found between surveyed and nonsurveyed blocks.

- The Census 2000 Dress Rehearsal Evaluation Memorandum C-2: *Contamination of Initial Phase Data Collected in ICM Block Clusters* by Sam Hawala conducted a paired block analysis that compared Integrated Coverage Measurement (ICM/PES) blocks with non-ICM/PES blocks. The study was based on the approach used in the 1990 PES Evaluation Project P14, except that Hawala paired block clusters that were in ICM/PES with up to four block clusters that were not in the ICM/PES. Hawala found very few significant differences in population coverage and no significant differences in housing unit status and respondent reaction indicators. He concluded that there was no evidence of contamination between the two operations.

- The 1995 Census Tests ICM Evaluation Project 11: *The Contamination Study* by Richard Griffiths conducted a paired block analysis similar to the 1990 and Dress Rehearsal studies that compared ICM and non-ICM blocks for the four ICM sampling strata. He

found differences in mail response rates between ICM and non-ICM for blocks with a high concentration of Asian-Pacific Islanders. This indicated possible contamination in only one of the four strata. He concluded the evidence of contamination was somewhat weak, and served better as a warning than a call to arms.

- The 2000 American Statistical Association paper by Katie M. Bench, Anne T. Kearney, and Rita J. Petroni titled, *An Investigation into the Independence Between the Census 2000 Dress Rehearsal and the Integrated Coverage Measurement/Post Enumeration Survey*, compared aggregated data from A.C.E. blocks to non-A.C.E. blocks. They found that Update/Leave TEAs were a weak area of concern for contamination. There were few significant results, and most of these were found in Update/Leave TEAs.

2. METHODS

To determine the potential existence of contamination in Census 2000, we performed the Whole Group Analysis (WGA) (Bench, 2001; Kearney, 2001; Bench, Kearney, and Petroni, 2000). The WGA aggregates census data in A.C.E. blocks to the national, evaluation poststrata, or the region and Type of Enumeration Area (TEA) level, and census data in non-A.C.E. blocks to the national, evaluation poststrata, or the region and TEA level. We then compare census data from A.C.E. blocks to census data from non-A.C.E. blocks to see if significant differences exist. The WGA approaches detecting contamination bias from a global hypothesis. We first examined the three fundamental indicators of contamination, and then looked at demographic, geographic, and response related indicators.

If we detect contamination, we will compute contamination bias and incorporate it in the total error model. Since contamination bias is defined as $DSE*(1-N_{c,ace}/N_c)$, we viewed the ratio $N_{c,ace}/N_c$ as the first fundamental indicator of contamination. T-tests were performed to see if this ratio differed significantly from one at the national and 36 evaluation poststrata levels (defined in Appendix A). $N_{c,ace}$ is the sample-weighted number of census enumerations in the A.C.E., and N_c is the census count from all clusters.

The second fundamental indicator of contamination is the average number of persons per block. This average helps determine if the A.C.E. affected the census count in A.C.E. blocks. To detect contamination we calculated this average for A.C.E. and non-A.C.E. blocks at the 36 evaluation poststrata level, but not at the national level, and performed t-tests.

The third fundamental indicator of contamination is the average number of housing units per block. This average helps determine if the A.C.E. affected the census housing unit count in A.C.E. blocks. To detect contamination we calculated this average for A.C.E. and non-A.C.E. blocks at the 36 evaluation poststrata level, but not at the national level, and computed t-tests.

We further investigated the presence of contamination with demographic, geographic, and response related indicators at the evaluation poststrata level and region and TEA levels. We

compared census data in A.C.E. block clusters to census data in non-A.C.E. block clusters by performing t-tests for the difference between the proportions in A.C.E. block clusters and the proportions in non-A.C.E. block clusters. For example, we calculated the proportion of census people in A.C.E. blocks who are of the Black or African American race, and the proportion of census people in non-A.C.E. blocks who are of the Black or African American race. We subtracted the proportion for non-A.C.E. blocks from the proportion for A.C.E. blocks and did a t-test on this difference to see if it was significantly different from zero. One set of t-statistics were calculated by the 36 evaluation poststrata, and the other set of t-statistics were calculated by region and TEA. We are also calculating the t-statistics by the 16 preliminary evaluation poststrata (defined in Appendix B). These results are located in Appendix C.

For this analysis A.C.E. and non-A.C.E. block clusters were defined as follows:

- A.C.E. block clusters - blocks that remained in the A.C.E. sample after A.C.E. sample reduction. However, blocks subsampled out during small block subsampling, and housing units subsampled out after E-sample subsampling received a weight of zero.

- Non-A.C.E. block clusters - blocks not sampled for the initial A.C.E. sample.

- Note, blocks that were sampled for the initial A.C.E. sample, but did not remain in the A.C.E. sample after A.C.E. sample reduction were not included in this analysis. These blocks were not part of the A.C.E. block clusters or the non-A.C.E. block clusters.

We used VPLX and a stratified jackknife estimation to calculate the t-statistics. However, we did not calculate standard errors for the proportions from non-A.C.E. blocks. Since the non-A.C.E. blocks are close to the whole population, the standard errors for these proportions would have been very close to zero. So, we treated these proportions as constants. In addition, design based estimation procedures underestimate the variances for small proportions. Therefore, we may find more significant differences than we otherwise would.

3. RESULTS

A small portion of the t-tests results obtained through the Whole Group Analysis are given in this section. The first sub-section presents the significant results of tests calculated at the national level, by the 36 evaluation poststrata, and by state. The second sub-section presents the significant results of tests calculated by region. The third sub-section presents the significant results of tests calculated by region and TEA.

3.1 Significant Results for T–tests Calculated for the Nation, 36 Evaluation Poststrata, and States

Sections 3.1.1, 3.1.2, and 3.1.3 contain the significant results of t-tests used to detect contamination at the national level, evaluation poststrata level, and state level. For the t-tests calculated by the 36 evaluation poststrata and state, we performed t-tests for 35 variables (Appendix D; Bench, 2001). We calculated a separate t-statistic for each of the 36 evaluation poststrata for 32 of these variables. The remaining three variables had a t-statistic calculated for each state. We used the False Discovery Rate multiple comparison procedure (FDR) (Bench, Kearney, Petroni, 2000; Benjamini and Hochberg, 1995) to test each of these 35 variables separately for significant results.

3.1.1 T-test to see if $N_{c,ace}/N_c$ differs from one

We computed the ratio of $N_{c,ace}/N_c$ fundamental indicator at the national level, and for each of the 36 evaluation poststrata. The t-test for the national level ratio was not significant, and neither were the t-tests for the 36 evaluation poststrata ratios. Since no t-statistics were significant, I did not include these numbers in this memorandum.

3.1.2 T-test for the Average Number of Persons per Block

We computed the average number of persons per block to help determine if the A.C.E. affected the census count in A.C.E. blocks. We calculated this average for A.C.E. and non-A.C.E. blocks at the 36 evaluation poststrata level, but not at the national level. None of the t-tests for 36 evaluation poststrata were significant. Since none of the differences between the average for A.C.E. blocks and the average for non-A.C.E. blocks were significant, I did not include these numbers.

3.1.3 T-tests for the difference between A.C.E. proportions and non-A.C.E. proportions

Tables 1 through 3 below display the results of the t-tests calculated for the average number of housing units per block, and demographic, geographic, and response related indicators by the 36 evaluation poststrata and state. Table 1 shows the number of variables tested and how many comparisons were made for each variable. Table 2 shows, for each variable tested by the 36 evaluation poststrata or state, the A.C.E. and non-A.C.E. proportions that are significantly different. Table 3 contains the significant differences for each variable in Table 2 whose proportions were not close to zero or one.

When the proportions we tested were close to the end points, that is zero or one, we suspect that their standard errors and hence the associated t-statistics may be unreliable. T-statistics based on such proportions are not considered when we draw conclusions. Tables E-1 and E-2 in appendix E show, for each of the variables tested by the 36 evaluation poststrata, which of the 35 A.C.E. and non-A.C.E. proportions are close to zero or one. They also show the A.C.E. and non-A.C.E. proportions that are significantly different.

Table 1. Number of Comparisons

Number of Variables	Number of Comparisons
35	36
3	51

Table 2. Variables with Significantly Different Proportions or Averages for Specific Evaluation Poststrata or State

Variable	Evaluation Poststrata (if applicable)	State (if applicable)	A.C.E. Proportion/ Average***	Non-A.C.E. Proportion/ Average***
Age group 4 (50+)	28		0.16562	0.18372
Black	3		**0.00019	**0.00074
	25		0.98240	0.97830
Pacific Islander	2		**0.00023	**0.00059
	14		**0.00002	**0.00010
	23		**0.00005	**0.00050
	34		0.02839	0.03687
Asian	5		**0.00057	**0.00120
	6		**0.00087	**0.00185
	34		0.96646	0.95747
Be Counted	7		**0.00016	**0.00087
	17		**0.00053	**0.00107
1 UBSA	9		0.96350	0.94473
	11		**0.98710	**0.98073
	32		0.11826	0.16743
2 UBSA	28		0.02459	0.03619
3 to 10 UBSA	11		**0.00152	**0.00381
	30		**0.00544	**0.01009
11+ UBSA	9		0.01628	0.03194
	11		**0.00022	**0.00115
Proportion People in an Evaluation	25		**0.02978	**0.03367
Poststratum	36		**0.00686	**0.00771
Update Enumerate	24		**0.00032	**0.00341
	35		**0.00009	**0.00039
Avg. Housing Units per Block		Florida	24.39683	30.00550
		West Virginia	57.05223	17.21393

** We are not considering these proportions as significantly different because they are close to 0 or 1. Under these circumstances, we regard the t-statistics as unreliable.
*** Base totals for these proportions/Averages are in Table F-1 of Appendix F.

Table 2 shows that the FDR procedure yielded ten significant results. There are more than ten proportions/averages with significant differences listed in the last two columns of the table, but we are not considering t-statistics for the proportions marked with ** as significantly different.

These proportions are either very close to zero or very close to one, so we are suspicious of the t-statistics.

Table 3. Variables with Significant Differences for Specific Evaluation Poststrata or State

Variable	Evaluation Poststrata (if applicable)	State (if applicable)	Significant Difference	Standard Error	P-value	Critical P-value
Age Group 4 (50+)	28		-0.01810	0.00439	0.00004	0.00278
Black	25		0.00410	0.00146	0.00516	0.00556
Pacific Islander	34		-0.00848	0.00275	0.00208	0.01111
Asian	34		0.00899	0.00317	0.00459	0.00833
1 UBSA	9		0.01877	0.00633	0.00302	0.00556
11+ UBSA	9		-0.01567	0.00486	0.00127	0.00556
2 UBSA	28		-0.01160	0.00283	0.00004	0.00278
1 UBSA	32		-0.04918	0.01673	0.00329	0.00833
Avg. Housing Units per Block		Florida	-5.60867	1.74212	0.00128	0.00392
		West Virginia	39.83830	1.71669	0.00067	0.00196

Table 3 shows the same results as Table 2, but lists the significant difference and critical p-values instead of the proportions/averages. This table only lists significant differences for variables that had A.C.E. and non-A.C.E. proportions bigger than 1.5 percent or smaller than 98.5 percent.

In conclusion, table 3 shows ten significant differences, but no systematic error was detected, although the number of significant results were somewhat above chance levels. Three of the ten significant differences are less than one percent, and on their A.C.E. and non-A.C.E. proportions are on the border of being considered too close to zero or one. An additional two variables also have A.C.E. and non-A.C.E. proportions on the border of being considered too close to zero or one. The variables Black, Pacific Islander, and Asian all have the significant differences less than one percent, and proportions on the border of being considered too close to zero or one. The variables 2 UBSA and 11+ UBSA both have proportions on the border of being considered too close to zero or one. In addition, the variables 1 UBSA and 11+ UBSA are from the same distribution, and both significant for evaluation poststrata 9. They are measuring different aspects of the same thing, and could probably be considered as one significant result. This means that we are only totally confident in five of the ten significant differences (Age group 4, UBSA1 for two evaluation poststrata, Avg. Housing Units per Block for Florida and West Virginia).

3.2 Significant Results for t-tests Calculated by Region but Collapsed Across TEA

This section contains the significant results for t-tests calculated for demographic, geographic, and response related indicators by region, but collapsed across TEA. These t-tests, looked for significant differences in A.C.E. and non-A.C.E. blocks for 29 variables. We divided these 29 variables into 9 variable groups, and applied the FDR procedure separately within each variable group to determine significant results. The nine variable groups are as follows:

1. Average number of persons per occupied housing unit (NP)
2. Proportion of housing units in Nonresponse Followup (NRU)
3. Proportion of housing units in Coverage Improvement Followup (CIU)
4. Proportion of housing units in Coverage Edit Followup (CEU)
5. Proportion of the following items edited or imputed: Hispanic origin (Hispanic origin Edited), sex (Sex Edited), race (Race Edited), tenure (Tenure Edited), and relationship (Relationship Edited)
6. Proportion non-relative (Non-relative), proportion other relative (Other Relative), proportion male (Male), proportion renter (Renter), proportion Hispanic (Hispanic), proportion African American or Black (Black), proportion Asian (Asian), proportion Native American (Native American), proportion Native Hawaiian or other Pacific Islander (Pacific Islander)
7. Average number of data defined persons per occupied housing unit (DDP), proportion long form (Long Form), proportion be counted form (Be Counted Form)
8. Proportion of the following number of units at the basic street address: 1 (1 UBSA), 2 (2 UBSA), 3 to 10 (3 to 10 UBSA), and 11 or more (11+ UBSA)
9. Proportion of people in the following age groups: 0 - 17 years of age (Age group 1), 18 to 29 years of age (Age group 2), 30 to 49 years of age (Age group 3), 50 plus years of age (Age group 4)

We did not compute t-statistics by region collapsed across TEA for seven variables. The seven excluded variables are:

- Housing units per block. T-tests are only calculated by state for this variable.
- Persons per block. T-tests are only calculated by the 36 evaluation poststrata for this variable.
- Proportion of people in each evaluation poststrata. T-tests are calculated by TEA and collapsed across TEA for this variable. The results collapsed across TEA were in section 3.1.2.
- Proportion of people in Mailout Mailback TEA, Update Leave TEA, List Enumerate TEA, and Update Enumerate TEA. T-tests for these four variables are included in section 3.3.

Tables 4 and 5 display the results of the t-tests performed at the national and regional level collapsed across TEA. Table 4 shows which variables have significantly different A.C.E. and non-A.C.E. proportions. The proportions close to zero or one are indicated with **. Table 5 contains the significant differences for each variable in Table 4 whose proportions are not marked with **.

Table 4. Variables with Significantly Different Proportions Collapsed Across TEA for a Specific Region

Region	Variable	A.C.E. Proportions***	Non-A.C.E. Proportions***
National	Relationship Edited	0.03637	0.03788
	Pacific Islander	**0.00187	**0.00217
	Other Relative	0.05552	0.05731
	Native American	**0.01030	**0.01143
	Age group 1 (0-17)	0.25990	0.26299
South	Black	0.17676	0.18824
	Other Relative	0.05705	0.05990
	Hispanic	0.10284	0.11660
	Pacific Islander	**0.00050	**0.00079
West	Native American	0.02014	0.02247

** We are not considering these proportions as significantly different because they are close to 0 or 1. Under these circumstances, we regard the t-statistics as unreliable.
*** Base totals for these proportions are in Table F-2 of Appendix F.

Table 5. Variables with Significant Differences Collapsed Across TEA for a Specific Region

Region	Variable	Significant Difference	Standard Error	P-value	Critical P-value
National	Relationship Edited	-0.00151	0.00058	0.00941	0.02000
	Other Relative	-0.00180	0.00069	0.00914	0.02222
	Age group 1 (0-17)	-0.00309	0.00136	0.02298	0.02500
South	Black	-0.01148	0.00568	0.04318	0.04444
	Other Relative	-0.00285	0.00112	0.01120	0.03333
	Hispanic	-0.01375	0.00479	0.00405	0.02222
West	Native American	-0.00233	0.00084	0.00538	0.01111

The results in tables 4 and 5 do not show much evidence of contamination. Only two out of the four regions have any significant results. The West has only one significant difference and it is less than 0.01. The South has three significant differences which are all from the same variable group. Since we used the FDR procedure to test separately for significance in each variable group, three out of nine variables with significant results is a concern. However, the difference

for Other Relative is less than 0.01, and the difference for Black is barely significant. This p-value might have been slightly bigger and hence not significant if we had calculated the small standard errors for non-A.C.E. estimates. At the national level Other Relative, Relationship Edited, and Age group 1 have significant results which are all less than 0.01. Furthermore, out of these three variables only Other Relative has a significant result in one of the four regions. That region is the South. When we consider the significant differences that are small and barely significant, the significant difference for Hispanic in the South is the only significant result that we are concerned with, so there does not seem to be much evidence of contamination for the Nation and regions collapsed across TEA.

3.3 Significant Results for T-tests Calculated by Region and TEA

Sections 3.3.1 through 3.3.5 contain the significant results of t-tests calculated for demographic, geographic, and response related indicators by region and TEA. For the t-tests calculated for the nation by TEA, we performed t-tests for 69 variables. These variables include:

- The 29 tested in section 3.2
- Proportion of people in the Mailout Mailback TEAs
- Proportion of people in the Update Leave TEAs
- Proportion of people in the List Enumerate TEAs
- Proportion of people in the Update Enumerate TEAs
- The Proportion of people in each of the 36 evaluation poststrata.

For the t-tests calculated by the four regions and TEA, we performed t-tests for 33 variables. These 33 variables are the same tested at the national and TEA level, except for the proportion of people in the 36 evaluation poststrata. We calculated a separate t-statistic for each of the four TEAs for each variable tested at the national and region level, then used the FDR procedure to test each of these variables separately for significance.

3.3.1 T-test for the difference between two proportions at the national and TEA level

Tables 6 and 7 display the results of the t-tests performed at the national and TEA level. Table 6 shows which variables have significantly different A.C.E. and non-A.C.E. proportions. The proportions close to zero or one are indicated with **. Table 7 contains the significant differences for each variable in Table 6 whose proportions are not marked with **.

Table 6. National - Variables with Significantly Different Proportions for a Specific TEA

Variable	TEA	A.C.E. Proportion***	Non-A.C.E. Proportion***
Relationship Edited	Mailout Mailback	0.03657	0.03826
Other Relative	Mailout Mailback	0.05840	0.06008
	List Enumerate	0.02270	0.03207
Black	Update Enumerate	**0.00633	**0.01640
Native American	Mailout Mailback	**0.00778	**0.00858
	List Enumerate	0.01656	0.02822
Pacific Islander	Mailout Mailback	**0.00208	**0.00234
	Update Leave	**0.00097	**0.00145
	Update Enumerate	**0.00032	**0.00074
1 UBSA	Update Enumerate	0.94053	0.89200
3 to 10 UBSA	Update Enumerate	0.02126	0.03770
11+ UBSA	Update Enumerate	0.01519	0.04547
People in evaluation poststratum 4	Mailout Mailback	0.03247	0.03703
People in evaluation poststratum 11	Update Leave	0.19183	0.21736
People in evaluation poststratum 24	Update Enumerate	**0.00119	**0.01026
People in evaluation poststratum 25	Mailout Mailback	0.03678	0.04132
People in evaluation poststratum 31	Mailout Mailback	0.04378	0.04815
People in evaluation poststratum 35	Update Enumerate	**0.00027	**0.00096
People in evaluation poststratum 36	List Enumerate	0.01280	0.02360
	Mailout Mailback	**0.00408	**0.00457

** We are not considering these proportions as significantly different because they are close to 0 or 1. Under these circumstances, we regard the t-statistics as unreliable.
*** Base totals for these proportions are in Table F-3 of Appendix F.

Table 7. National - Variables with Significant Differences for a Specific TEA

Variable	TEA	Significant Difference	Standard Error	P-value	Critical P-value
Relationship Edited	Mailout Mailback	-0.00169	0.00063	0.00723	0.02500
Other Relative	Mailout Mailback	-0.00168	0.00080	0.03531	0.05000
	List Enumerate	-0.00937	0.00390	0.01632	0.02500
Native American	List Enumerate	-0.01166	0.00516	0.02382	0.05000
1 UBSA	Update Enumerate	0.04854	0.01488	0.00111	0.02500
3 to 10 UBSA	Update Enumerate	-0.01644	0.00591	0.00539	0.02500
11+ UBSA	Update Enumerate	-0.03028	0.01013	0.00279	0.02500
People in evaluation poststratum 4	Mailout Mailback	-0.00456	0.00199	0.02202	0.02500
People in evaluation poststratum 11	Update Leave	-0.02554	0.01051	0.01515	0.02500
People in evaluation poststratum 25	Mailout Mailback	-0.00454	0.00164	0.00562	0.02500
People in evaluation poststratum 31	Mailout Mailback	-0.00436	0.00189	0.02094	0.02500
People in evaluation poststratum 36	List Enumerate	-0.01080	0.00476	0.02329	0.05000

There are twelve significant differences at the National and TEA levels. However, six of these are less than 0.01, and an additional three are from the same distribution. 1 Unit at Basic Street Address (UBSA), 3 to 10 UBSA, and 11+ UBSA in the Update Enumerate TEA are all significant. These come from the same distribution of UBSA. They are measuring different aspects of the same thing, and could probably be considered one significant result. When we consider the six small significant differences and the UBSA variables as a group, we have four significant results that concern us. However, no systematic error was detected, although the number of significant results were somewhat above chance levels.

3.3.2 T-test for the difference between two proportions for the Northeast by TEA

Tables 8 and 9 display the results of the t-tests performed for the Northeast at the TEA level. Table 8 shows which variables have significantly different A.C.E. and non-A.C.E. proportions. The proportions close to zero or one are indicated with **. Table 9 contains the significant differences for each variable in Table 8 whose proportions are not marked with **.

Table 8. Northeast - Variables with Significantly Different Proportions for a Specific TEA

Variable	TEA	A.C.E. Proportion***	Non-A.C.E. Proportion***
HU in NRU	Update Leave	0.32308	0.35039
Hispanic	Update Leave	0.01348	0.02001
Black	Update Leave	**0.00793	**0.01291
	Update Enumerate	0.00404	0.03020
Native American	Mailout Mailback	**0.00395	**0.00471
	List Enumerate	**0.00606	**0.01211
Be Counted	Update Enumerate	**0.00058	**0.00280
1 UBSA	Update Enumerate	0.98894	0.92972
2 UBSA	Update Enumerate	0.00545	0.02701
3 to 10 UBSA	Update Enumerate	0.00561	0.02899

** We are not considering these proportions as significantly different because they are close to 0 or 1. Under these circumstances, we regard the t-statistics as unreliable.
*** Base totals for these proportions are in Table F-4 of Appendix F.

Table 9. Northeast - Variables with Significant Differences for a Specific TEA

Variable	TEA	Significant Difference	Standard Error	P-value	Critical P-value
HU in NRU	Update Leave	-0.02731	0.01111	0.01397	0.02500
Hispanic	Update Leave	-0.00653	0.00255	0.01039	0.02500
Black	Update Enumerate	-0.02615	0.00319	0.00000	0.02500
1 UBSA	Update Enumerate	0.05922	0.00580	0.00000	0.02500
2 UBSA	Update Enumerate	-0.02156	0.00404	0.00000	0.02500
3 to 10 UBSA	Update Enumerate	-0.02338	0.00374	0.00000	0.02500

There are six significant differences at the Northeast and TEA levels. However, the difference for Hispanic is less than 0.01, and 1 UBSA, 2 UBSA, and 3 to 10 UBSA are related. These three variables are from the same distribution, and measure different aspects of the same thing. They could easily be grouped and considered as one significant result. In addition, the A.C.E. proportions for Black, 2 UBSA, and 3 to 10 UBSA are all less than 0.01, but they are included in Table 9 because their non-A.C.E. proportions are all around 0.03, and the differences between the A.C.E. and non-A.C.E. proportions are all between 0.02615 and 0.02156. However, since we assumed standard errors were zero for the non-A.C.E. proportions, and the A.C.E. proportions are all small, the t-statistics may still be unreliable. Considering these facts, there are only two or three significant results that we are concerned with, and no systematic error was detected, although the number of significant results were somewhat above chance levels in the

Northeast.

3.3.3 T-test for the difference between two proportions for the Midwest by TEA

Tables 10 and 11 display the results of the t-tests performed for the Midwest at the TEA level. Table 10 shows which variables have significantly different A.C.E. and non-A.C.E. proportions. The proportions close to zero or one are indicated with **. Table 11 contains the significant differences for each variable in Table 10 whose proportions are not marked with **.

Table 10. Midwest - Variables with Significantly Different Proportions for a Specific TEA

Variable	TEA	A.C.E. Proportion***	Non-A.C.E. Proportion***
Hispanic Origin Edited	Update Enumerate	0.04624	0.07053
Other Relative	List Enumerate	0.00378	0.01918
Native American	List Enumerate	0.00378	0.02874
Pacific Islander	Update Leave	**0.00010	**0.00035
Be Counted Form	Update Leave	**0.00067	**0.00137
Long Form	List Enumerate	0.52841	0.41413

** We are not considering these proportions as significantly different because they are close to 0 or 1. Under these circumstances, we regard the t-statistics as unreliable.
*** Base totals for these proportions are in Table F-5 of Appendix F.

Table 11. Midwest - Variables with Significant Differences for a Specific TEA

Variable	TEA	Significant Difference	Standard Error	P-value	Critical P-value
Hispanic Origin Edited	Update Enumerate	-0.02429	0.01029	0.01829	0.02500
Other Relative	List Enumerate	-0.01540	0.00416	0.00021	0.02500
Native American	List Enumerate	-0.02495	0.00415	0.00000	0.02500
Long Form	List Enumerate	0.11428	0.03932	0.00366	0.02500

The Midwest has only four significant differences. The A.C.E. proportions for Other Relative and Native American are less than 0.01, but they are included in Table 11 because the non-A.C.E. proportions for these variables are 0.01918 and 0.02874, and their differences are -0.01540 and -0.02495. However, since we assumed standard errors were zero for the non-A.C.E. proportions, and the A.C.E. proportions are all small, the t-statistics may be unreliable. So, there seems to be no evidence of contamination in the Midwest.

3.3.4 T-test for the difference between two proportions for the South by TEA

Tables 12 and 13 display the results of the t-tests performed for the South at the TEA level. Table 12 shows which variables have significantly different A.C.E. and non-A.C.E. proportions. The proportions close to zero or one are indicated with **. Table 13 contains the significant differences for each variable in Table 12 whose proportions are not marked with **.

Table 12. South - Variables with Significantly Different Proportions for a Specific TEA

Variable	TEA	A.C.E. Proportion***	Non-A.C.E. Proportion***
Renter	List Enumerate	0.15362	0.25765
Non Relative	Update Leave	0.03140	0.03446
Asian	Update Enumerate	**0.00053	**0.00340
Pacific Islander	Mailout Mailback	**0.00060	**0.00091
	Update Leave	**0.00026	**0.00048
1 UBSA	Update Enumerate	0.95942	0.83145
3 to 10 UBSA	Update Enumerate	0.02027	0.05473

** We are not considering these proportions as significantly different because they are close to 0 or 1. Under these circumstances, we regard the t-statistics as unreliable.
***Base totals for these proportions are in Table F-6 of Appendix F.

Table 13. South - Variables with Significant Differences for a Specific TEA

Variable	TEA	Significant Difference	Standard Error	P-value	Critical P-value
Renter	List Enumerate	-0.10403	0.01325	0.00000	0.02500
Non-relative	Update Leave	-0.00306	0.00116	0.00852	0.02500
1 UBSA	Update Enumerate	0.12797	0.01682	0.00000	0.02500
3 to 10 UBSA	Update Enumerate	-0.03446	0.01470	0.01908	0.02500

The South has only four significant differences. The difference for Non-relative is less than 0.01. In addition, 1 UBSA and 3 to 10 UBSA are from the same distribution. They are measuring different aspects of the same thing, and could probably be grouped and considered as one significant result. Based on these significant results, there seems to be no evidence of contamination in the South.

3.3.5 T-test for the difference between two proportions for the West by TEA

Tables 14 and 15 display the results of the t-tests performed for the West at the TEA level. Table 14 shows which variables have significantly different A.C.E. and non-A.C.E. proportions. The proportions close to zero or one are indicated with **. Table 15 contains the significant

differences for each variable in Table 14 whose proportions are not marked with **.

Table 14. West - Variables with Significantly Different Proportions for a Specific TEA

Variable	TEA	A.C.E. Proportion***	Non-A.C.E. Proportion***
Other Relative	List Enumerate	0.02005	0.03679
Hispanic	Update Enumerate	0.07672	0.11151
Native American	Mailout Mailback	**0.01394	**0.01525
	List Enumerate	0.02751	0.04845
Be Counted Form	Update Enumerate	**0.00146	**0.00348

** We are not considering these proportions as significantly different because they are close to 0 or 1. Under these circumstances, we regard the t-statistics as unreliable.
*** Base totals for these proportions are in Table F-7 of Appendix F.

Table 15. West - Variables with Significant Differences for a Specific TEA

Variable	TEA	Significant Difference	Standard Error	P-value	Critical P-value
Other Relative	List Enumerate	-0.01674	0.00629	0.00774	0.02500
Hispanic	Update Enumerate	-0.03479	0.01362	0.01062	0.02500
Native American	List Enumerate	-0.02095	0.00671	0.00181	0.02500

The West has only three significant differences, which yields no evidence of contamination in the West.

4. CONCLUSIONS

The evidence suggests that contamination bias is not a problem. Globally, we did not find evidence of contamination bias for high-level proportions and averages. We computed a t-statistic to see if the ratio $N_{c,ace}/N_c$ was significantly different from one for the nation and the 36 evaluation poststrata. None of these t-tests were significant. In addition, the t-tests used to detect significant differences between A.C.E. and non-A.C.E. proportions for the second and third fundamental indicators yielded little to no evidence of contamination.

The study also broke the data down to very detailed cells. These cells were demographic, geographic, and response related indicators of contamination broken down by the 36 evaluation poststrata, and region and Type of Enumeration Area (TEA). No systematic error was detected in these cells, although the number of significant results were somewhat above chance levels. Many of them were not considered as significant when we drew conclusions. This happened under two circumstances. First, there were several proportions that were close to 0 or 1. We regarded t-tests that used these small or large proportions to be unreliable because design based estimation procedures underestimate the variances for small proportions. Second, some of the

differences between A.C.E. and non-A.C.E. proportions were extremely small. So, while these difference were mathematically significant they were not practically significant.

Based on these findings the total error model will not adjust for contamination bias, and concerns about contamination should not affect the adjustment decision. This finding is consistent with the earlier assumption that contamination bias would not occur during Census 2000.

5. REFERENCES

Bench, Katie, *A.C.E. Evaluation N1: Contamination of Census Data Collected in A.C.E. Clusters - Specifications for Calculation of Estimates for the Whole Group Analysis,* Planning, Research, and Evaluation Division TXE/2010 Memorandum Series: CM-CON-S-01, Bureau of the Census, February 1, 2001.

Bench, Katie. M., Kearney, Anne. T., and Petroni, Rita. J., *An Investigation into the Independence Between the Census 2000 Dress Rehearsal and the Integrated Coverage Measurement/Post Enumeration Survey,* American Statistical Association 2000 Proceedings of the Section on Survey Research Methods, 2000, p. 453-458.

Benjamini, Y. and Hochberg, Y, "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing," *J.R. Statist. Soc. B*, 57, No. 1, 1995, pp. 289-300.

Davis, M., *Final Report for Post Enumeration Survey (PES) Evaluation Project P14, Part I: Independence of the Census and P Sample: Comparison of Blocks*, Coverage Studies and Evaluation Memorandum Series #P-4, Bureau of the Census, 1990.

Griffiths, R., *Results from the 1995 Census Test: The Contamination Study - Project P11*, DSSD 1995 Census Test Memorandum Series #U-7, Bureau of the Census, 1996.

Hawala, S., *Contamination of Initial Phase Data Collected in Integrated Coverage Measurement Block Clusters,* Census 2000 Dress Rehearsal Evaluation Results Memo. Series # C-2, Bureau of the Census, 1999.

Hogan, H., *Accuracy and Coverage Evaluation: Theory and Application*. Report prepared for the February 2-3, 2000 Dual System Estimation Workshop of the National Academy of Science panel to review the 2000 Census, 2000.

Kearney, Anne, *Study Plan for N1: Contamination of Census 2000 Data Collected in Accuracy and Coverage Evaluation Block Clusters,* Bureau of the Census, February 15, 2001.

Mulry, M. H. and Spencer B. D., *Total Error in PES Estimates of Population*, Journal of the American Statistical Association, 86, 1991, p. 839-863.

Appendix A

Census 2000 A.C.E. - 36 Evaluation Post-Stratum Groups

Race/Hispanic Origin Domain Number*		Tenure	MSA/TEA	High Return Rate				Low Return Rate			
				N	M	S	W	N	M	S	W
Domain 7 (Non-Hispanic White or "Some other race")		Owner	Large MSA MO/MB	1		2		3		4	
			Medium MSA MO/MB	5		6	7				
			Small MSA & Non-MSA MO/MB	8		9		10			
			All Other TEAs	11		12		13 part		14	13
		Non-owner	Large MSA MO/MB	15				16			
			Medium MSA MO/MB	17				18			
			Small MSA & Non-MSA MO/MB	19				20			
			All Other TEAs	21							
Domain 4 (Non-Hispanic Black)		Owner	Large MSA MO/MB	22				23			
			Medium MSA MO/MB								
			Small MSA & Non-MSA MO/MB	24							
			All Other TEAs								
		Non-owner	Large MSA MO/MB	25				26			
			Medium MSA MO/MB								
			Small MSA & Non-MSA MO/MB	27							
			All Other TEAs								
Domain 3 (Hispanic)		Owner	Large MSA MO/MB	28				29			
			Medium MSA MO/MB								
			Small MSA & Non-MSA MO/MB	30							
			All Other TEAs								
		Non-owner	Large MSA MO/MB	31				32			
			Medium MSA MO/MB								
			Small MSA & Non-MSA MO/MB	33							
			All Other TEAs								
Domain 5 (Native Hawaiian or Pacific Islander)		Owner		34 part							
		Non-owner		35 part							
Domain 6 (Non-Hispanic Asian)		Owner		34 part							
		Non-owner		35 part							
American Indian or Alaska Native	**Domain 1** (On Reservation)	Owner		36							
		Non-owner									
	Domain 2 (Off Reservation)	Owner									
		Non-owner									

Appendix B

Census 2000 A.C.E. - 16 Preliminary Evaluation Post-Stratum Groups

<table>
<tr><th colspan="2" rowspan="2">Race/Hispanic Origin
Domain Number*</th><th rowspan="2">Tenure</th><th rowspan="2">MSA/TEA</th><th colspan="4">High Return Rate</th><th colspan="4">Low Return Rate</th></tr>
<tr><th>N</th><th>M</th><th>S</th><th>W</th><th>N</th><th>M</th><th>S</th><th>W</th></tr>
<tr><td colspan="2" rowspan="5">Domain 7
(Non-Hispanic White or
“Some other race”)</td><td rowspan="3">Owner</td><td>Large MSA MO/MB</td><td colspan="2" rowspan="2">1</td><td colspan="2" rowspan="2">2</td><td colspan="2" rowspan="2">3</td><td colspan="2" rowspan="2">4</td></tr>
<tr><td>Medium MSA MO/MB</td></tr>
<tr><td>Small MSA & Non-MSA MO/MB</td><td colspan="4">5</td><td colspan="4">6</td></tr>
<tr><td>All Other TEAs</td><td colspan="8">7</td></tr>
<tr><td rowspan="4">Non-owner</td><td>Large MSA MO/MB</td><td colspan="4" rowspan="2">8</td><td colspan="4" rowspan="2">9</td></tr>
<tr><td>Medium MSA MO/MB</td></tr>
<tr><td>Small MSA & Non-MSA MO/MB</td><td colspan="8" rowspan="2">10</td></tr>
<tr><td>All Other TEAs</td></tr>
<tr><td colspan="2" rowspan="5">Domain 4
(Non-Hispanic Black)</td><td rowspan="3">Owner</td><td>Large MSA MO/MB</td><td colspan="4" rowspan="2">11 part</td><td colspan="4" rowspan="2">12 part</td></tr>
<tr><td>Medium MSA MO/MB</td></tr>
<tr><td>Small MSA & Non-MSA MO/MB</td><td colspan="8" rowspan="2">13 part</td></tr>
<tr><td>All Other TEAs</td></tr>
<tr><td rowspan="4">Non-owner</td><td>Large MSA MO/MB</td><td colspan="4" rowspan="2">14 part</td><td colspan="4" rowspan="2">15 part</td></tr>
<tr><td>Medium MSA MO/MB</td></tr>
<tr><td>Small MSA & Non-MSA MO/MB</td><td colspan="8" rowspan="2">16 part</td></tr>
<tr><td>All Other TEAs</td></tr>
<tr><td colspan="2" rowspan="5">Domain 3
(Hispanic)</td><td rowspan="3">Owner</td><td>Large MSA MO/MB</td><td colspan="4" rowspan="2">11 part</td><td colspan="4" rowspan="2">12 part</td></tr>
<tr><td>Medium MSA MO/MB</td></tr>
<tr><td>Small MSA & Non-MSA MO/MB</td><td colspan="8" rowspan="2">13 part</td></tr>
<tr><td>All Other TEAs</td></tr>
<tr><td rowspan="4">Non-owner</td><td>Large MSA MO/MB</td><td colspan="4" rowspan="2">14 part</td><td colspan="4" rowspan="2">15 part</td></tr>
<tr><td>Medium MSA MO/MB</td></tr>
<tr><td>Small MSA & Non-MSA MO/MB</td><td colspan="8" rowspan="2">16 part</td></tr>
<tr><td>All Other TEAs</td></tr>
<tr><td colspan="2" rowspan="2">Domain 5
(Native Hawaiian or Pacific Islander)</td><td colspan="2">Owner</td><td colspan="8">11 part</td></tr>
<tr><td colspan="2">Non-owner</td><td colspan="8">14 part</td></tr>
<tr><td colspan="2" rowspan="2">Domain 6
(Non-Hispanic Asian)</td><td colspan="2">Owner</td><td colspan="8">11 part</td></tr>
<tr><td colspan="2">Non-owner</td><td colspan="8">14 part</td></tr>
<tr><td rowspan="4">American Indian
or
Alaska Native</td><td rowspan="2">Domain 1
(On Reservation)</td><td colspan="2">Owner</td><td colspan="8" rowspan="4">16 part</td></tr>
<tr><td colspan="2">Non-owner</td></tr>
<tr><td rowspan="2">Domain 2
(Off Reservation)</td><td colspan="2">Owner</td></tr>
<tr><td colspan="2">Non-owner</td></tr>
</table>

Appendix C

Table C-1. Proportions for Variables with Significantly Different Proportions for a Specific Preliminary Evaluation Poststrata

Variable	DSSD Evaluation Poststrata	TEA (if applicable)	A.C.E. Proportion	Non-A.C.E. Proportion
Male	8		0.48988	0.48173
Black	3		**0.00019	**0.00074
2 UBSA	01		0.01864	0.02278
	11		0.02327	0.03076
11+ UBSA	7		**0.00103	**0.00235
Update Enumerate	14		**0.00002	**0.00007
Avg Persons per Block	4		8498.68687	10052.51831
	14		33672.29823	37219.16275
Proportion People in an Evaluation	14		**0.00356	**0.00428
Poststratum	14	Mailout Mailback	0.09850	0.10786
	14	Update Enumerate	**0.00027	**0.00096

** We are not considering these proportions as significantly different because they are close to 0 or 1. Under these circumstances, we regard the t-statistics as unreliable.

Table C-2. Significant Differences for a Specific Preliminary Evaluation Poststrata

Variable	DSSD Evaluation Poststrata	TEA (if applicable)	Significant Difference	Standard Error	P-value	Critical P-value
Male	8		0.00815	0.00291	0.00512	0.00625
2 UBSA	01		-0.00414	0.00162	0.01080	0.01250
	11		-0.00749	0.00187	0.00006	0.00625
Avg Persons per Block	04		-1553.83143	577.73167	0.00716	0.01250
	14		-3546.86452	1117.69715	0.00151	0.00625
People in Evaluation Poststratum	14	Mailout Mailback	-0.00936	0.00297	0.00161	0.00625

Table C-3. Base Totals for Variables with Significantly Different Proportions for a Specific Preliminary Evaluation Poststrata

Variable	DSSD Evaluation Poststrata	TEA (if applicable)	A.C.E. Proportion Base	Non-A.C.E. Proportions Base
Male Base	8		19166412.28868	19995449.14934
Race Base	3		4912583.38567	5372788.30050
UBSA Base	1		36342320.62919	35533948.37304
	7		35500274.36540	34636048.09207
	11		23931922.52461	24587933.89716
TEA Base	14		21877859.34924	24168033.07804
Avg. Persons per Block Base	4		843.21068	821.48230
	14		649.72872	649.34381
People in Evaluation	1-16		272555864.92656	273637212.30991
Poststrata Base		Mailout Mailback	220732111.42395	223017069.66353
		Update Enumerate	1412361.25864	1731146.49891

Appendix D

Variables Tested and at What Levels

Variables Tested	Tested at What Levels
Average number of persons per occupied housing unit (NP)	state, region and TEA
Average number of data defined persons per occupied housing unit (DDP)	state, region and TEA
Average number of housing units per block where there is at least one housing unit in the block (Avg. Housing Units per Block)	state
Average number of persons per block (Avg. Persons per Block)	evaluation poststrata
Proportion of housing units in Nonresponse Followup (NRU)	evaluation poststrata, region and TEA
Proportion of housing units in Coverage Edit Followup (CEU)	evaluation poststrata, region and TEA
Proportion of housing units in Coverage Improvement Followup (CIU)	evaluation poststrata, region and TEA
Proportion Renters (Renter)	evaluation poststrata, region and TEA
Proportion of data defined persons on a Be Counted Form (Be Counted Form)	evaluation poststrata, region and TEA
Proportion of data defined persons on a Long Form (Long Form)	evaluation poststrata, region and TEA
Proportion other relative including brother/sister and mother/father (Other Relative)	evaluation poststrata, region and TEA
Proportion nonrelative (Nonrelative)	evaluation poststrata, region and TEA
Proportion male (Male)	evaluation poststrata, region and TEA
Proportion Hispanic (Hispanic)	evaluation poststrata, region and TEA
Proportion Black or African American (Black)	evaluation poststrata, region and TEA
Proportion American Indian/Alaska Native (Native American)	evaluation poststrata, region and TEA
Proportion Asian (Asian)	evaluation poststrata, region and TEA
Proportion Native Hawaiian or Other Pacific Islander (Pacific Islander)	evaluation poststrata, region and TEA
Proportion Tenure Edited or Imputed (Tenure Edited)	evaluation poststrata, region and TEA
Proportion Relationship Edited or Imputed (Relationship Edited)	evaluation poststrata, region and TEA
Proportion Sex Edited or Imputed (Sex Edited)	evaluation poststrata, region and TEA
Proportion Hispanic Origin Edited or Imputed (Hispanic Origin Edited)	evaluation poststrata, region and TEA
Proportion Race Edited or Imputed (Race Edited)	evaluation poststrata, region and TEA
Proportion of 1 unit at basic street address (1UBSA)	evaluation poststrata, region and TEA
Proportion of 2 units at basic street address (2 UBSA)	evaluation poststrata, region and TEA

Variables Tested and at What Levels (continued)

Variables Tested	Tested at What Levels
Proportion of 3 to 10 units at basic street address (3 to 10 UBSA)	evaluation poststrata, region and TEA
Proportion of 11 or more units at basic street address (11+ UBSA)	evaluation poststrata, region and TEA
Proportion of people 0-17 years of age (Age group 1)	evaluation poststrata, region and TEA
Proportion of people 18-29 years of age (Age group 2)	evaluation poststrata, region and TEA
Proportion of people 30-49 years of age (Age group 3)	evaluation poststrata, region and TEA
Proportion of people 50 or more years of age (Age group 4)	evaluation poststrata, region and TEA
Proportion of people in TEA 1 and 6 (Mailout Mailback)	evaluation poststrata, region and TEA
Proportion of people in TEAs 2, 7, and 9 (Update Leave)	evaluation poststrata, region and TEA
Proportion of people in TEAs 3 and 4 (List Enumerate)	evaluation poststrata, region and TEA
Proportion of people in TEAs 5 and 8 (Update Enumerate)	evaluation poststrata, region and TEA
Proportion of people in each evaluation poststratum	collapsed over TEA and TEA

Appendix E

Table E-1. 36 Evaluation Poststrata with Extreme Proportions and Significantly Different Proportions

	Number of Evaluation Poststrata with proportions close to 0 or 1.			Significantly Different A.C.E. and Non-A.C.E. Proportions	
Variable	**A.C.E.**	**Non-A.C.E.**	**Evaluation Poststrata**	**A.C.E.**	**Non-A.C.E.**
Age group 4 (50+)			28	0.16562	0.18372
*Black	26 w/in 0.015***	28 w/in 0.015	3	**0.00019	**0.00074
	9 w/in 0.028	8 w/in 0.034	25	0.98240	0.97830
*Pacific Islander	34 w/in 0.015	35 w/in 0.015	2	**0.00023	**0.00059
	1 w/in 0.028	1 w/in 0.037	14	**0.00002	**0.00010
			23	**0.00005	**0.00050
			34	0.02839	0.03687
*Asian	34 w/in 0.015	34 w/in 0.015	5	**0.00057	**0.00120
	1 w/in 0.034	1 w/in 0.043	6	**0.00087	**0.00185
			34	0.96646	0.95747
*Be Counted	36 w/in 0.015	36 w/in 0.015	7	**0.00016	**0.00087
			17	**0.00053	**0.00107
UBSA = 1	3 w/in 0.015	0 w/in 0.015	9	0.96350	0.94473
	4 w/in 0.050	7 w/in 0.050	11	**0.98710	**0.98073
			32	0.11826	0.16743
UBSA = 2	11 w/in 0.015	7 w/in 0.015	28	0.02459	0.03619
	8 w/in 0.050	12 w/in 0.050			
UBSA 3 to 10	11 w/in 0.015	14 w/in 0.015	11	**0.00152	**0.00381
	7 w/in 0.050	4 w/in 0.050	30	**0.00544	**0.01009
UBSA 11+	6 w/in 0.015	6 w/in 0.015	9	0.01628	0.03194
	10 w/in 0.050	10 w/in 0.050	11	**0.00022	**0.00115
Proportion People in an	8 w/in 0.015	9 w/in 0.015	25	**0.02978	**0.03367
Evaluation Poststratum	25 w/in 0.050	24 w/in 0.050	36	**0.00686	**0.00771
*Update Enumerate	30 w/in 0.015	30 w/in 0.015	24	**0.00032	**0.00341
	4 w/in 0.050	4 w/in 0.050	35	**0.00009	**0.00039

* Over half of the 36 evaluation poststrata have proportions between 0 and 0.015 or 1 and 0.985 for this variable. There are 11 such variables.

** We are disregarding these proportions because they are close to 0 or 1. Under these circumstances, we regard the t-statistics as unreliable.

***For example, for 26 out of the 36 evaluation poststrata, the proportion of people of the Black race in A.C.E. blocks are between 0.0 and 0.015 or 1.0 and 0.985

Table E-2. Variables with Extreme Proportions and No Significant Differences for the 36 Evaluation Poststrata

	Number of Evaluation Poststrata with proportions close to 0 or 1.	
Variable	**A.C.E.**	**Non-A.C.E.**
CEU	***23 w/in 0.050	22 w/in 0.050
CIU	10 w/in 0.015	10 w/in 0.015
	25 w/in 0.050	26 w/in 0.050
*Renter	35 w/in 0.015	35 w/in 0.015
*Hispanic	26 w/in 0.015	26 w/in 0.015
*Indian	33 w/in 0.015	33 w/in 0.015
	3 w/in 0.027	3 w/in 0.031
Non-relative	3 w/in 0.039	3 w/in 0.040
	8 w/in 0.050	7 w/in 0.050
Other Relative	19 w/in 0.050	20 w/in 0.050
*Mailout Mailback	29 w/in 0.015	29 w/in 0.015
	2 w/in 0.050	2 w/in 0.050
*Update Leave	23 w/in 0.015	23 w/in 0.015
	2 w/in 0.050	2 w/in 0.050
*List Enumerate	33 w/in 0.015	33 w/in 0.015
	3 w/in 0.050	3 w/in 0.050
Tenure Edited	13 w/in 0.050	13 w/in 0.050
Relationship Edited	21 w/in 0.050	21 w/in 0.050
Sex Edited	1 w/in 0.015	1 w/in 0.015
	27 w/in 0.050	26 w/in 0.050
Hispanic Origin Edited	14 w/in 0.050	13 w/in 0.050
Race Edited	22 w/in 0.050	24 w/in 0.050

* Over half of the 36 evaluation poststrata have proportions between 0 and 0.015 or 1 and 0.985 for this variable. There are 11 such variables.

***For example, for 23 out of the 36 evaluation poststrata, the proportion of people in housing units that went to CEU in A.C.E. blocks are between 0.0 and 0.05 or 1.0 and 0.95

Appendix F

Table F-1. Base Totals for Variables with Significantly Different Proportions for Specific Evaluation Poststrata

Variable	Evaluation Poststrata	A.C.E. Proportion Base	Non-A.C.E. Proportions Base
Age Group Base	28	9004484.71439	9752661.51154
Race Base	2	11518446.14314	11149819.77218
	3	4912583.38567	5372788.30050
	5	17572707.09374	17296840.59345
	6	12543116.91570	11874843.97713
	14	13099546.99008	11926154.54644
	23	2587192.95233	2853323.49296
	25	8117528.83663	9214643.17332
	34	6469379.16767	6344773.78435
Be Counted Form Base	7	8842843.89477	8149192.19926
	17	10465113.92191	11246189.05411
UBSA Base	9	11880961.87283	11447308.03905
	11	10038009.77366	11036582.79296
	28	9004484.71439	9752661.51104
	30	4291247.19316	4569109.65102
	32	4170648.25894	3777208.77541
TEA Base	24	5190234.77192	5206738.48877
	35	4095773.89489	4215665.90026
People in Evaluation Poststrata Base	1 - 36	272555864.92656	273637212.30991

Table F-2. Base Totals for Variables with Significantly Different Proportions Collapsed Across Tea for a Specific Region

Region	Variable	A.C.E. Proportion Base	Non-A.C.E. Proportions Base
National	Relationship Edited Base	272555864.92656	273637212.30991
	Race Base	272555864.92656	273637212.30991
	Other Relative Base	272555864.92656	273637212.30991
	Age Group Base	272555864.92656	273637212.29585
South	Race Base	98607503.55271	97401581.33954
	Other Relative Base	98607503.55271	97401581.33954
	Hispanic Origin Base	98607503.55271	97401581.33954
West	Race Base	61269971.35808	61715054.16786

Table F-3. National - Base Totals for Variables with Significantly Different Proportions for a Specific TEA

Variable	TEA	A.C.E. Proportion Base	Non-A.C.E. Proportions Base
Relationship Edited Base	Mailout Mailback	220732111.42395	223017069.66353
Other Relative Base	Mailout Mailback	220732111.42395	223017069.66353
	List Enumerate	704392.20130	605129.43202
Race Base	Mailout Mailback	220732111.42395	223017069.66353
	Update Leave	49707000.04266	48283866.71544
	List Enumerate	704392.20130	605129.43202
	Update Enumerate	1412361.25864	1731146.49891
UBSA Base	Update Enumerate	759541.44617	959154.81921
People in Evaluation Poststrata Base	Mailout Mailback	220732111.42395	223017069.66353
	Update Leave	49707000.04266	48283866.71544
	List Enumerate	704392.20130	605129.43202
	Update Enumerate	1412361.25864	1731146.49891

Table F-4. North East - Base Totals for Variables with Significantly Different Proportions for a Specific TEA

Variable	TEA	A.C.E. Proportion Base	Non-A.C.E. Proportions Base
HU in NRU Base	Update Leave	2774173.40784	2992882.64938
Hispanic Origin Base	Update Leave	6164890.75322	6477690.16893
Race Base	Mailout Mailback	44851238.06715	44916394.91678
	Update Leave	6164890.75322	6477690.16893
	List Enumerate	308299.81682	292070.51233
	Update Enumerate	180574.79058	233411.17986
Be Counted Form Base	Update Enumerate	173825.47132	222446.50262
UBSA Base	Update Enumerate	194485.76257	215693.27965

Table F-5. Mid West - Base Totals for Variables with Significantly Different Proportions for a Specific TEA

Variable	TEA	A.C.E. Proportion Base	Non-A.C.E. Proportions Base
Hispanic Origin Edited Base	Update Enumerate	195844.39193	249187.03038
Other Relative Base	List Enumerate	17470.73152	21667.17003
Race Base	Update Leave	8830831.80395	9775018.61438
	List Enumerate	17470.73152	21667.17003
Be Counted & Long Form Base	Update Leave	8727928.08672	9650774.74239
	List Enumerate	16716.94559	20163.55033

Table F-6. South - Base Totals for Variables with Significantly Different Proportions for a Specific TEA

Variable	TEA	A.C.E. Proportion Base	Non-A.C.E. Proportions Base
Renter Base	List Enumerate	9237.99495	10878.35433
Non Relative Base	Update Leave	28249704.95402	25698193.23829
Race Base	Mailout Mailback	69928277.83175	71186916.00828
	Update Leave	28249704.95402	25698193.23829
	Update Enumerate	404808.50130	486437.29224
UBSA Base	Update Enumerate	163501.77797	227642.53137

Table F-7. West - Base Totals for Variables with Significantly Different Proportions for a Specific TEA

Variable	TEA	A.C.E. Proportion Base	Non-A.C.E. Proportions Base
Other Relative Base	List Enumerate	353909.38732	261356.94894
Hispanic Origin Base	Update Enumerate	631133.57484	762110.99643
Race Base	Mailout Mailback	53823355.86444	54358621.52864
	List Enumerate	353909.38732	261356.94894
Be Counted Form Base	Update Enumerate	599365.82267	715969.51134