CONDITIONING OF CENSUS 2000 DATA COLLECTED IN ACCURACY AND COVERAGE EVALUATION BLOCK CLUSTERS[1]

Katie Bench, Anne T. Kearney, U.S. Census Bureau
Katie Bench, 115 Mount Vernon Avenue, Alexandria, VA 22301

Key Words: Causal Dependence, Census, Accuracy and Coverage Evaluation, Multiple Comparisons

1. BACKGROUND

To measure the overall and differential coverage of the U.S. population in Census 2000, the U.S. Census Bureau conducted the Accuracy and Coverage Evaluation (A.C.E.). The A.C.E. included various stages of sampling. An initial A.C.E. sample of block clusters was drawn, and housing units within the sampled block clusters were listed. Then, the A.C.E. sample was reduced through sub-sampling operations (Childers and Fenstermaker, 2000).

The A.C.E. uses dual-system estimation to measure coverage error. The dual system estimation method assumes there are two independent lists of the population. The first list is the original census enumerations in the A.C.E. clusters, 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 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 (Mulry and Spencer, 1991). Research undertaken on 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 an area of weak concern for contamination (Griffiths, 1996; Bench, Kearney, and Petroni, 2000).

This paper provides information to help determine if the Accuracy and Coverage Evaluation contaminated Census 2000 data collected in A.C.E. blocks. It is taken from a more extensive report (see Bench, Report 14, 2001). The A.C.E. could have contaminated the Census during A.C.E. Independent Listing which occurred before Census Day, or during a small fraction of Census follow-up contacts made after the beginning of the A.C.E. survey interviewing. This paper does not provide information to help determine if Census 2000 contaminated A.C.E. data. We did not quantitatively analyze the effect of Census 2000 on the A.C.E. because there was no control group since the Census was conducted in all A.C.E. block clusters. However, in another report, a qualitative analysis is included that provides some information on the effect of Census 2000 on the A.C.E.

2. METHODS

To determine the potential existence of contamination in Census 2000, we performed the Whole Group Analysis (WGA) (Bench, Planning, Research, and Evaluation Division TXE/2010 Memorandum Series: CM-CON-S-01, 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 same levels. 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 by first examining three fundamental indicators of contamination, and then demographic, geographic, and response related indicators.
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. $N_{c,ace}$ is the sample-weighted number of census enumerations in the A.C.E. block clusters, N_c is the census count from all clusters, and DSE is the dual-system estimate (Spencer, 2002).

The second and third fundamental indicators of contamination are the average number of persons per block, and the average number of housing units per block. These averages help determine if the A.C.E. affected the census person and housing unit counts in A.C.E. blocks.
We further investigated the presence of contamination with demographic, geographic, and response related indicators 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 between 18 and 29 years old, and the proportion of census

[1]This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

people in non-A.C.E. blocks who are between 18 and 29 years old.

2.1 Definitions of A.C.E. and non-A.C.E. block clusters

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

1. A.C.E. block clusters - blocks that remained in the A.C.E. sample after all A.C.E. sub-sampling.
2. Non-A.C.E. block clusters - blocks not sampled for the initial A.C.E. sample.

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.

2.2 Estimation of variances

We used VPLX and a stratified jackknife estimator to calculate the t-statistics. However, we did not calculate standard errors for the proportions from non-A.C.E. block clusters. Since the number of non-A.C.E. block clusters is 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.

Note that design based estimation procedures, such as those used in VPLX, underestimate the variances for small and large proportions. Therefore, we may find more significant differences than we otherwise would. As an illustration, 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 are unreliable. T-statistics based on such proportions are not considered when we draw conclusions. We considered proportions within 0.015 of 0 or 1 to be too small or too big. That is, proportions smaller than 0.015 or bigger than 0.985. We believe these cutoffs are conservative; we possibly could have used cutoffs 2 to 3 times as large.

2.3 Multiple comparison procedure

We used the False Discovery Rate (FDR) multiple comparison procedure. The FDR procedure controls for the proportion of errors committed by falsely rejecting the null hypothesis. The FDR has some advantages for our study over other procedures such as the familywise error rate and Bonferroni procedures. For instance, when more of the hypotheses are not true, the potential for increase in power is larger for the FDR procedure, and the power of the FDR procedure is uniformly larger than that of the other methods (Benjamini and Hochberg, 1995).

This paper has two levels of multiple comparisons to be concerned about. The first level is the number of variables we calculated proportions for. The second level is the number of proportions we calculated for each variable. For each variable, we calculated proportions for the 36 evaluation poststrata, each state, or 4 TEA levels.

We took the second level of multiple comparisons into account by using the FDR procedure. We tested each variable separately for significance. For example, we calculated a t-statistic for four TEAs to test the difference between the proportion of people between 18 and 29 years old in A.C.E. blocks, and the proportion of people between 18 and 29 years old in non-A.C.E. blocks for significance. The TEAs are mailout mailback, update leave, update enumerate, and list enumerate. To test the t-statistics for the four TEAs for significance, we applied the FDR procedure to these four t-statistics.

We took into account the first level of multiple comparisons by expecting significant results to number 10 percent of the number of variables tested.

For the three fundamental indicators, we calculated t-tests at the following levels:

- For the first and second fundamental indicators of contamination, we calculated a separate t-statistic for each of the 36 evaluation poststrata (see Bench, Report 14, 2001 for a description of the evaluation poststrata).
- For the third fundamental indicator of contamination, we calculated a separate t-statistic for each state.

For the demographic, geographic, and response related indicators, we calculated t-tests by region and TEA. For the nation by TEA, we calculated separate t-statistics for each of the four TEAs for 69 variables. That is, the 36 poststrata plus the 33 variables listed in the appendix. For the four regions and TEAs, we calculated separate t-statistics for each of the four TEAs for the 33 variables listed in the appendix. We also calculated a set of t-statistics by the 36 evaluation poststrata, and for each region. These results did not change our conclusions regarding contamination and are not included in this paper.

3. RESULTS

Only the significant t-test results obtained through the Whole Group Analysis are given in this section. Section 3.1 presents the significant results of tests for the three fundamental indicators, and section 3.2 presents the significant results of tests for the demographic, geographic, and response related indicators.

3.1 Significant results for the three fundamental indicators

The average number of housing units per block was the only fundamental indicator to have significant results. None of the t-tests for the ratio of $N_{c,ace}/N_c$, or the average number of persons per block were significant. Table 1 shows that two states have significantly different A.C.E. and non-A.C.E. housing units per block averages. These two significant differences indicate possible contamination.

3.2 Significant results for demographic, geographic, and response related indicators

Tables 2 - 6 show significant differences for demographic, geographic, and response related indicator t-tests. Significant differences are shown only for each variable whose proportions were not within 0.015 of 0 or 1.

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

Table 2 shows the twelve significant differences at the national and TEA levels. However, six of these differences are less than 0.01. An additional three are from the same distribution. One Unit at Basic Street Address (UBSA), 3 to 10 UBSA, and 11+ UBSA in the Update Enumerate TEA are all significant. 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 was somewhat above chance levels.

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

Table 3 shows the 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 3 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 0.02156 and 0.02615. 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. 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 was somewhat above chance levels.

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

Table 4 shows that 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, respectively. 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.

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

Table 5 shows that 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, 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.

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

Table 6 shows that the West has only three significant differences, which yield 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 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 TEA. No systematic error was detected in these cells, although the number of significant results were somewhat above chance levels.

But 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 statistically significant they were not practically significant. It should be noted that we only presented the results by region and TEA in this paper, but our conclusions reflect results broken out by the 36 evaluation poststrata as well.

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, *Conditioning of Census 2000 Data Collected in Accuracy and Coverage Evaluation Block Clusters*, Executive Steering Committee for A.C.E. Policy II Report 14, Bureau of the Census, September 19, 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.

Childers, Danny R. and Debbie Fenstermaker, *Accuracy and Coverage Evaluation: Overview of the Design*, DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-02, Bureau of the Census, January 11, 2000.

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.

Spencer, Bruce D., *Total Error Model for Census 2000: How Components of Error Can Be Estimated from the Bureau's Planned Evaluation Studies*, report prepared under contract number 50-YABC-7-66020 for the Bureau of the Census, Abt Associates Inc. and Spencer Statistics, Inc., March 20, 2002.

Table 1. Significant differences for average number of housing units per block for a specific state

Variable	State	Significant Difference	Standard Error	P-value	Critical P-value
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 2. 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[2]	Mailout Mailback	-0.00456	0.00199	0.02202	0.02500
People in evaluation poststratum 11[3]	Update Leave	-0.02554	0.01051	0.01515	0.02500
People in evaluation poststratum 25[4]	Mailout Mailback	-0.00454	0.00164	0.00562	0.02500
People in evaluation poststratum 31[5]	Mailout Mailback	-0.00436	0.00189	0.02094	0.02500
People in evaluation poststratum 36[6]	List Enumerate	-0.01080	0.00476	0.02329	0.05000

Table 3. 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

Table 4. 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
People Enumerated on a Long Form	List Enumerate	0.11428	0.03932	0.00366	0.02500

---

[2] People in evaluation poststratum 4 are non-Hispanic white or "some other race" owners in large and medium metropolitan statistical areas (MSA) and mailout mailback (MO/MB) TEAs with low return rate in the South and West regions.

[3] People in evaluation poststratum 11 are non-Hispanic white or "some other race" owners in large and medium MSA and non-MO/MB TEAs with high return rate in the Northeast and Midwest regions.

[4] People in evaluation poststratum 25 are non-Hispanic black non-owners in large and medium MSA and MO/MB TEAs with high return rate.

[5] People in evaluation poststratum 31 are Hispanic non-owners in large and medium MSA and MO/MB TEAs with high return rate.

[6] People in evaluation poststratum 36 are American Indian or Alaska natives.

Table 5. 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

Table 6. 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

Appendix

Variables Tested	Calculated by
$N_{c,ace}/N_c$	36 evaluation poststrata
Average number of persons per block (Avg. Persons per Block)	36 evaluation poststrata
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 occupied housing unit (NP)	region and TEA
Average number of data defined persons per occupied housing unit (DDP)	region and TEA
Proportion of housing units in Nonresponse Followup (NRU)	region and TEA
Proportion of housing units in Coverage Edit Followup (CEU)	region and TEA
Proportion of housing units in Coverage Improvement Followup (CIU)	region and TEA
Proportion Renters (Renter)	region and TEA
Proportion of data defined persons on a Be Counted Form (Be Counted Form)	region and TEA
Proportion of data defined persons on a Long Form (Long Form)	region and TEA
Proportion other relative including brother/sister and mother/father (Other Relative)	region and TEA
Proportion nonrelative (Nonrelative)	region and TEA
Proportion male (Male)	region and TEA
Proportion Hispanic (Hispanic)	region and TEA
Proportion Black or African American (Black)	region and TEA
Proportion American Indian/Alaska Native (Native American)	region and TEA
Proportion Asian (Asian)	region and TEA
Proportion Native Hawaiian or Other Pacific Islander (Pacific Islander)	region and TEA
Proportion Tenure Edited or Imputed (Tenure Edited)	region and TEA
Proportion Relationship Edited or Imputed (Relationship Edited)	region and TEA
Proportion Sex Edited or Imputed (Sex Edited)	region and TEA
Proportion Hispanic Origin Edited or Imputed (Hispanic Origin Edited)	region and TEA
Proportion Race Edited or Imputed (Race Edited)	region and TEA
Proportion of 1 unit at basic street address (1UBSA)	region and TEA
Proportion of 2 units at basic street address (2 UBSA)	region and TEA
Proportion of 3 to 10 units at basic street address (3 to 10 UBSA)	region and TEA
Proportion of 11or more units at basic street address (11+ UBSA)	region and TEA
Proportion of people 0-17 years of age (Age group 1)	region and TEA
Proportion of people 18-29 years of age (Age group 2)	region and TEA
Proportion of people 30-49 years of age (Age group 3)	region and TEA
Proportion of people 50 or more years of age (Age group 4)	region and TEA
Proportion of people in TEA 1 and 6 (Mailout Mailback)	region and TEA
Proportion of people in TEAs 2, 7, and 9 (Update Leave)	region and TEA
Proportion of people in TEAs 3 and 4 (List Enumerate)	region and TEA
Proportion of people in TEAs 5 and 8 (Update Enumerate)	region and TEA
Proportion of people in each of the 36 evaluation poststrata	TEA (nation wide only)