

CENSUS PLUS: AN ALTERNATIVE COVERAGE METHODOLOGY

Eric Schindler and Alfredo Navarro¹
Bureau of the Census, Washington, DC 20233

KEY WORDS: Post Enumeration Survey, Synthetic Adjustment

I. INTRODUCTION

The 1990 Post Enumeration Survey (PES) used capture-recapture or Dual System methods to estimate coverage in the 1990 Census. CensusPlus is an alternative coverage measurement method in which, after completion of the normal census operations, a sample of blocks is revisited. This second collection effort applies intensive independent and dependent methods, including matching to the original Census forms, to obtain the best possible count of usual residents in the sample blocks on Census Day. Final estimates are based on the total number of usual residents found in the sample blocks in either the original Census or the reenumeration. Unlike the Dual System estimate where the so-called "fourth cell" estimates usual residents not found in either enumeration, there is no attempt to estimate persons missed in both enumerations. It is therefore very important for CensusPlus to locate all usual residents in the sampled blocks on Census Day. See Wright (1993) for a complete theoretical discussion of CensusPlus.

This paper describes an empirical study in which CensusPlus estimation procedures are applied to the 1990 PES data. Coverage rates and population estimates are calculated and compared for the two estimation methods for the 357 PES poststrata, for groups of poststrata collapsed on age and sex, and for the PES blocks. Estimated CensusPlus standard errors are comparable to those from the Dual System. Since the more stringent CensusPlus collection procedures could not be applied, it is not surprising that CensusPlus population estimates are lower than corresponding Dual System estimates. Even so, CensusPlus, with the 1990 PES collection methods, captures about 70% of the undercount measured by the PES. Alternative counting methods and more intensive CensusPlus collection procedures could reduce this 30% gap. How far the difference can be narrowed will be determined by the 1995 Census Test. The general

framework of the coverage survey methodology being tested in 1995 will permit both CensusPlus and Dual System estimates. The results of this simulation with the 1990 PES data will be useful in developing the revised methodology and the design of the coverage measurement sample. The results of the 1995 test will be used to develop the coverage methodology for 2000.

II. METHODOLOGY FOR 357 POSTSTRATA

The 1990 PES consisted of two samples in 5180 blocks or clusters of several small blocks, selected by a stratified sample based on region, place size, and predominant race. The E-Sample consisted of those persons in the PES blocks actually enumerated in the Census. Whole person imputations were omitted. The P-Sample consisted of all persons found in an independent reenumeration of the same blocks. The "357 PES Design" defined fifty-one poststratum groups by:

- 51 American Indians living on Reservations
- 49-50 Asian and Pacific Islander owners/renters
- 25-28/31-34 Black owners/renters in large cities in the four Census regions
- 29,30,35,36 Black owners/renters in small cities and non-urban areas
- 37-48 Non-Black Hispanics (similar to Blacks)
- 1-24 All other owners/renters in the three place types in the four Census regions.

Each poststratum group was divided into seven poststrata based on age and sex.

For each poststratum j the following were obtained:

- $CEN_{tot,j}$ = the 100% Census Count
- $CEN_{real,j}$ = the count of persons in the Census with enough data for matching; this excludes persons with only imputed data
- $E_{tot,j}$ = the weighted estimate of E-Sample persons
- $E_{correct,j}$ = the weighted estimate of correctly enumerated E-Sample persons
- P_j = the weighted estimate of P-Sample persons
- M_j = the weighted estimate of matched persons

¹Eric Schindler and Alfredo Navarro are mathematical statisticians in the Decennial Statistical Studies Division of the Census Bureau. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

For each of the 357 PES poststrata, the CensusPlus estimate used was approximately equal to the Census count plus persons found in the P-Sample who could not be matched in the E-Sample minus persons in the E-Sample who were erroneously enumerated, adjusted upwards for persons without sufficient data for matching. The CensusPlus estimate and undercount estimate, as suggested by Wright (1993), for the j -th PES poststratum are given by:

$$CPLUS_j = \frac{CEN_{real,j}}{E_{tot,j}} \times (E_{correct,j} + P_j - M_j)$$

where

$(E_{correct,j} + P_j - M_j)$ is an approximation of the number of usual residents identified by the combined E- and P-samples and

$\frac{(E_{correct,j} + P_j - M_j)}{E_{tot,j}}$ is the ratio of these usual residents to E-sample persons. This ratio is used to adjust $CEN_{real,j}$, the number of persons in the Census with adequate data for matching.

and

$$UC_{CPLUS,j} = \frac{CPLUS_j - CEN_{tot,j}}{CPLUS_j}$$

is the CensusPlus undercount rate.

Variance estimates for the $CPLUS_j$ and $UC_{CPLUS,j}$ for the 357 poststrata were calculated using a Jackknife procedure dropping one block at a time.

The PES Dual System estimates were reproduced using:

$$DSE_j = CEN_{real,j} \times \frac{E_{correct,j}}{E_{tot,j}} \times \frac{P_j}{M_j}$$

which is algebraically equivalent to other forms of the Dual System estimate and parallels the formula for CensusPlus.

Dual System undercount rates were estimated by:

$$UC_{DSE,j} = \frac{DSE_j - CEN_{tot,j}}{DSE_j}$$

Variance estimates of the DSE_j and $UC_{DSE,j}$ were also calculated using the Jackknife procedure.

III. RESULTS FOR 357 POSTSTRATA

For both CensusPlus and Dual System estimation the poststratum level standard error estimates often exceed the estimated undercount rates. Only those poststrata with over 1,750,000 persons have undercount rate standard errors consistently under 2.0%. Poststrata with about 1,000,000 persons have undercount rate standard errors up to 4.0%; those with down to 400,000 persons up to 10.0%; and the smallest poststrata have undercount rate standard errors up to 23%.

Table 1 shows the number of poststrata for which the ratio of the standard error estimates fall into certain ranges. All but 43 of the 357 poststratum the error estimates are within 20% of each other. CensusPlus standard errors average slightly less than the Dual System standard errors. As could be expected, the larger discrepancies between the two standard error estimates occur in smaller poststrata. The standard errors for the two estimators for all poststrata with more than 1,500,000 persons are within 10%; with more than 1,000,000 within 20%. The poststrata with CensusPlus standard errors much less than the Dual System estimate standard errors usually have the larger Dual System undercount rates and somewhat smaller CensusPlus undercount rates. These poststrata are usually for minority persons and renters. CensusPlus estimates fewer persons in these groups than Dual System estimates and shows a lower differential undercount for these groups. However, this is largely caused by the inability of CensusPlus, limited by the PES procedures, to account for many persons missed by both enumerations, but estimated by the Dual System procedures. Implementation of appropriate CensusPlus procedures would tend to increase both the population estimate and the standard error estimate closer to those for the Dual System.

TABLE 1: Ratio of SE(C+)/SE(DSE) for 357 Poststrata (average Poststratum Populations in 1000s)

RATIO	PS	Ave CEN	Ave DSE	Ave C+
0.5-0.7	7	296	342	324
0.7-0.8	28	287	310	300
0.8-0.9	61	419	440	434
0.9-1.0	147	705	713	710
1.0-1.1	92	1,034	1,039	1,038
1.1-1.2	14	476	480	483
1.2-1.6	8	403	414	416

For the 357 poststrata, the CensusPlus estimates using PES procedures vary from 4% higher to 11% percent lower than the Dual System estimates. (The poststratum CensusPlus estimate is higher than the Dual System estimate when the estimated number of matches, M_j , is greater than Census count of real persons, $CEN_{real,j}$. This phenomenon is caused partly by sampling variance and partly by the 1990 PES collection and editing procedures.) Table 2 shows the distribution of the percent differences between the CensusPlus estimates and the Dual System estimates.

TABLE 2: Comparison of CensusPlus and Dual System Estimates - 357 Poststrata (average Poststratum populations in 1000s)

Direction	% DIFF	# PS	Ave CEN	Ave DSE	Ave C+
Census+ is > DSE	2%-4.4%	6	202	214	222
	1%-2%	15	321	325	329
	0.5%-1%	22	468	486	489
	0%-0.5%	54	1161	1164	1165
Census+ is < DSE	0%-0.5%	96	997	998	997
	0.5%-1%	47	486	492	489
	1%-1.5%	33	452	473	467
	1.5%-2%	23	545	572	562
	2%-3%	28	350	370	360
	3%-5%	19	357	389	374
	5%-11%	14	194	225	209

These differences sum to the 0.5% difference between the national CensusPlus undercount rate of 1.1% and the PES undercount rate of 1.6%. The estimates for all poststrata with more than 1,600,000 persons are within 0.5%; with more than 1,000,000 persons within 3.0%. The largest differences are found in the smaller poststrata with relatively large undercounts or overcounts. When the CensusPlus estimate is less than the Dual System estimate, the CensusPlus estimate is said to have a residual undercount. Of the 33 poststrata with residual undercounts over 3%, 32 are for non-Whites, 31 are for renters, 19 are for males ages 18-49, 12 are for blacks in large urban areas, and 8 (out of 8 such poststrata) are for black male renters aged 18-49 in large urban areas. The 10 million persons in these poststrata have a 9.8% Dual System undercount rate but only a 5.2% CensusPlus undercount rate, compared to 1.2% and 0.9% for the remaining 240 million persons. These 33 poststrata account for 40% of the total difference between the Dual System and the CensusPlus estimates. Consistent with these large residual undercounts, the standard errors of the CensusPlus undercount rates for these 33 poststrata average only 90% of the Dual System standard errors (versus 97.5% for the other 324 poststrata). The enhanced CensusPlus procedures must be directed at collecting these hardest to collect groups, for which the PES showed an additional residual undercount when compared to Demographic Analysis.

IV. RESULTS FOR 51 POSTSTRATUM GROUPS AND LARGER GROUPS

The 357 poststrata were collapsed by age and sex to form the 51 basic poststratum groups defined by race/Hispanic origin, region, place type, and tenure. Further collapsing to race/Hispanic origin by tenure groups, as shown in Table 3, was also done.

A 357 by 357 variance/covariance matrix was formed with

$$COV(CPLUS_{i,j}) = \frac{n-1}{n} \sum_{m=1}^{510} (CPLUS_{i,m} - CPLUS_i) \times (CPLUS_{j,m} - CPLUS_j)$$

where:

$CPLUS_{i,m}$ is the Jackknife CensusPlus estimate for poststratum i when block m is removed and

n is the number of blocks with persons in both poststratum i and poststratum j

The seven age/sex CensusPlus estimates within each of the 51 poststratum groups were added as follows

$$CPLUS_k = \sum_{j=1}^7 CPLUS_{(k,j)}$$

The variance is given by

$$VAR(CPLUS_k) = \sum_{j=1}^7 VAR(CPLUS_{(k,j)}) + 2 \sum_{j=1}^6 \sum_{l=j+1}^7 COV(CPLUS_{(k,j)}, CPLUS_{(k,l)})$$

The undercount rate is given by

$$UC_{CPLUS,k} = \frac{CPLUS_k - CEN_{tot,k}}{CPLUS_k}$$

$$\text{where } CEN_{tot,k} = \sum_{j=1}^7 CEN_{tot,(k,j)}$$

and its variance is approximated by

$$VAR(UC_{CPLUS,k}) \approx \frac{CEN_{tot,k}^2}{CPLUS_k^4} \times VAR(CPLUS_k)$$

Analogous estimates are made for the Dual System estimates for the 51 poststratum groups and larger combinations.

The results are shown in Table 3. Because the 51 poststratum groups are larger than the original 357 poststrata, the coefficients of variation are somewhat lower. The standard errors of the undercount rates are generally less than 2.0% except for American Indians living on reservations, two of the six Black renter groups, and most Hispanics.

TABLE 3: Standard Errors of Undercount Rates for 51 Poststratum Groups

Groups	#PS	C +ave	C +max	DSEave	DSEmax
Wh/Ot Owner	12	0.68%	1.17%	0.66%	1.19%
Wh/Ot Renter	12	1.74%	4.49%	1.74%	4.10%
API	2	1.73%	1.99%	2.00%	2.49%
AmInd	1	4.11%		4.58%	
Black Owner	6	1.36%	2.00%	1.41%	1.90%
Black Renter	6	2.06%	4.09%	2.41%	5.41%
Hisp Owner	6	2.22%	4.29%	2.18%	4.34%
Hisp Renter	6	3.24%	6.36%	3.15%	5.11%

The standard errors are generally within 20% of each other and within a few percent for those poststratum groups with 7,000,000 or more persons.

Many of the CensusPlus standard error estimates which are lower than the Dual System standard error estimates are for renters. For example, the standard error estimate for all Black renters is 0.74% for CensusPlus compared with 0.83% for Dual System estimation. (See Table 4 below.) As in the case for the individual poststrata, this difference is caused by lower undercount estimates because the CensusPlus estimator with the PES procedures does not account for many of the hardest to count persons missed in both enumerations but covered by the Dual System.

For the PES universe (which excludes over four million persons in group quarters and remote areas of Alaska), the estimated CensusPlus undercount rates for Blacks, Non-Black Hispanics, and Others are 3.06%, 4.44%, and 0.44%, respectively. There was a Black vs. Other undercount differential of 2.62% and a Non-Black Hispanic vs. Other differential of 4.00%. (A differential undercount becomes a differential "add rate" in the context of an adjusted or "one-number" Census.) The Dual System estimation undercount rates are 4.57%, 4.99%, and 0.77%, and the differential undercounts are 3.81% and 4.23%, respectively. Most of the reduction in the estimated differential undercount for Blacks can be attributed to Black renters, and, as discussed above, it is caused by the fact that the CensusPlus estimator with the PES procedures does not account for many of the Black renters, who are missed in both enumerations but counted in the fourth cell of the Dual System estimation. The 16 million Black renters accounted for 40% of the difference between the Dual System and CensusPlus estimates. Because CensusPlus, as implemented for this study, does not estimate that these Black renters exist, it only appears that they are not being missed by the Census. Without the appropriate enhanced CensusPlus procedures, a

smaller CensusPlus coverage adjustment is calculated for these groups, but the estimate is short of the actual population, as estimated by the 1990 PES or Demographic Analysis.

V. RESULTS FOR PES BLOCKS

If the 1990 PES had been used for adjustment purposes, synthetic estimates would have been made at the block level by applying each person's PES poststratum adjustment factor and adding the number of adjusted persons in the block. Controlled rounding and imputation would be used to produce whole persons with complete characteristics. Synthetic CensusPlus and Dual System estimates were made for the 5180 PES blocks. Also, for each block the information required to make a direct block level CensusPlus or Dual System estimate was available. (These data, while adequate for producing estimates of the truth at the poststratum level, are less suitable for estimating the truth at the PES block level, but they are the best we have.) Six undercount rates can be estimated by comparing the synthetic or direct block level Dual System estimate or CensusPlus estimate as the target or actual population with the Census and/or the synthetic estimates as the estimated population.

Table 5 shows the extent of the dispersion of the estimated block undercount rates for the 1101 PES blocks with 100 or more E-sample persons. The number of blocks with estimated overcounts or undercounts exceeding 10% or 20% is shown for each of the six undercount rates.

The first four rows of Table 5 show that the estimates of the Census are much closer to the synthetic estimates than to the direct block level estimates. The worst 1% of the Census counts differ from the synthetic CensusPlus estimates by 8% to 10%, and from the Dual System estimates by 10% to 12%. However, even for these large blocks about 1% of the blocks have Census counts 30% or more higher than the direct CensusPlus or Dual System estimates and another 1% have Census counts 30% or more lower than the direct CensusPlus or Dual System estimates. That is, the Census count is relatively close to the synthetic CensusPlus and Dual System estimates, but can differ substantially from the direct block level CensusPlus or Dual System estimate.

TABLE 5: Overcount and Undercount Estimates for 1101 Blocks with 100 or more E-Sample Persons

BASIS FOR ESTIMATES Estimate Target		Overcounts			Undercounts		
		> 20%	10%-20%	0%-10%	0%-10%	10%-20%	> 20%
CEN	SynC+	0	0	179	922	0	0
CEN	SynDSE	0	0	158	931	12	0
CEN	DirC+	24	28	336	603	92	18
CEN	DirDSE	16	24	304	606	118	33
SynC+	DirC+	28	48	454	433	65	11
SynDSE	DirDSE	21	50	495	495	84	18

Despite the possibly high errors in the direct estimates, the last four rows of Table 5 are more likely to reflect reality for the block than the first two. Synthetic adjustment makes relatively minor adjustments to the estimates of block population in order to achieve good estimates at higher levels of aggregation, centering the skewed distribution from rows 3 and 4 in rows 5 and 6. However, for many PES blocks, the synthetic estimates still miss the directly estimated PES block populations by large margins.

Assuming the direct estimates to be the truth, synthetic CensusPlus adjustment "improves" the estimates for two-thirds of the blocks compared to using the 1990 Census count. Synthetic Dual System adjustment is closer than the unadjusted data to the direct Dual System estimate for 60% of the blocks.

Usually the Census count and the two synthetic estimates are fairly close to each other. Similarly, the two direct estimates. However, the two sets of estimates differ substantially for a large number of blocks. Despite the large errors involved in direct Dual System or CensusPlus estimation at the block level, these results seem to confirm previous results that synthetic adjustment at low levels of aggregation does not produce accurate population estimates.

There are some indications that synthetic estimation may have same problems at the tract level. That is, synthetic adjustment at the tract level would likely produce estimates not very close to those produced by either direct CensusPlus or direct Dual System estimation.

VI. LIMITATIONS

It was necessary to use the PES files to simulate the CensusPlus procedures. More appropriate and intensive CensusPlus procedures might find some persons dropped from the E-Sample for Dual System estimation and, for this study, CensusPlus estimation. Improved procedures would also locate some persons missed by

both the Census and the 1990 PES, but accounted for in the Dual System estimate. The comparison of population or undercount rate estimates are therefore less reliable than we would like.

The PES obtained substantial variance reduction by searching surrounding blocks for matches or evidence of erroneous enumerations. This information was used for these empirical CensusPlus estimates but would not be available under the current CensusPlus design options. Since CensusPlus concentrates on usual residents in the PES block on Census Day, dropping the surrounding block search, combined with the more intensive within PES block procedures, may have a smaller effect on the CensusPlus estimate than it does on the Dual System estimate. It is believed that the results are adequate for the comparison of standard errors.

VII. CONCLUSIONS

Estimates of standard errors from CensusPlus and Dual System estimation are close to one another with no clear advantage for either method. Thus, from the viewpoint of standard errors, CensusPlus can provide a reasonable alternative to the 1990 Post Enumeration Survey. For either approach, the standard errors of the undercount rates exceed 2.0% for many of the 357 poststrata and for some of the 51 minority poststratum groups. Increased sample size would be needed to improve these standard errors for either CensusPlus or Dual System estimation.

As expected, the simulated CensusPlus estimates in this study measure a smaller undercount than the Dual System estimates, especially for the hardest to collect demographic groups. In a one-number Census environment where only coverage adjusted estimates are released, CensusPlus estimation based on the 1990 PES procedures would leave a residual undercount compared to Dual System estimation, especially for minority renters where Dual System estimation already had a residual undercount compared to Demographic Analysis. CensusPlus estimates closer to the Dual System estimates can only be obtained by improved procedures being developed and test in the 1995 Census test.

VIII. REFERENCES

Wolter, Kirk M. and Causey, Beverley D. (1991), "Evaluation of Procedures for Improving Population Estimates for Small Areas," Journal of the American Statistical Association, 86, 278-284.

Wright, Tommy, (1993) Oak Ridge National Laboratory and Bureau of the Census, CensusPlus: A Sampling & Prediction Approach for the 2000 Census of the United States, internal Census Bureau document.

TABLE 4: CensusPlus Estimates and Dual System Estimates (Standard Errors in Parentheses)

Poststratum	1990 Census	Dual System Estimate			CensusPlus Estimate			Residual Undercount
		Estimate	UC	Diff UC	Estimate	UC	Diff UC	
Wh/Ot Owner	135050621	134696664 (308868)	-0.26% (0.23%)	-1.03%	134677092 (311743)	-0.28% (0.23%)	-0.72%	0.02%
Wh/Ot Renter	51727331	53358400 (273889)	3.06% (0.50%)	2.29%	52903672 (263012)	2.22% (0.49%)	1.78%	0.84%
API Owner	4113667	4054974 (60289)	-1.45% (1.51%)	-2.22%	4023932 (57708)	-2.23% (1.47%)	-2.67%	0.78%
API Renter	3113817	3346850 (89654)	6.96% (2.49%)	6.19%	3229549 (66643)	3.59% (1.99%)	3.15%	3.37%
AI on Reserv	373151	425110 (22161)	12.22% (4.58%)	11.45%	410746 (18600)	9.15% (4.11%)	8.71%	3.07%
White/Other (Total)	194378587	195881998 (428516)	0.77% (0.22%)	0.00%	195244992 (421174)	0.44% (0.21%)	0.00%	0.33%
W/O Renter Large Urban NE	6719384	6798119 (94751)	1.16% (1.37%)	0.39%	6763663 (95890)	0.66% (1.41%)	0.22%	0.50%
Black Owner	13420911	13730533 (79025)	2.26% (0.56%)	1.49%	13673604 (77820)	1.85% (0.56%)	1.41%	0.41%
Black Renter	15585108	16664898 (147022)	6.48% (0.83%)	5.71%	16249424 (125850)	4.09% (0.74%)	3.65%	2.39%
Black (Total)	29006019	30395432 (167380)	4.57% (0.53%)	3.80%	29923028 (144046)	3.06% (0.47%)	2.62%	1.51%
Black Renter Large Urban NE	3165968	3455043 (60585)	8.37% (1.61%)	7.60%	3298155 (54038)	4.01% (1.57%)	3.57%	4.36%
Hisp Owner	9411952	9586234 (65855)	1.82% (0.67%)	1.05%	9584441 (67444)	1.80% (0.69%)	1.36%	0.02%
Hisp Renter	11546927	12474296 (155886)	7.43% (1.18%)	7.66%	12348941 (151754)	6.49% (1.15%)	6.05%	0.96%
N/B Hispanic (Total)	20958879	22060530 (180114)	4.99% (0.78%)	4.22%	21933382 (174535)	4.44% (0.76%)	4.00%	0.55%
Hisp Renter Large Urban S	2062504	2274880 (63529)	9.34% (2.53%)	8.57%	2215186 (48271)	6.89% (2.03%)	6.45%	2.45%
US TOTALS	244343485	248337959 (487440)	1.61% (0.19%)	N/A	247101402 (480451)	1.12% (0.19%)	N/A	0.49%

KEY: Wh/Ot = Non-Hispanic White/Other Hisp = Non-Black Hispanic
API = Non-Hisp Asian/Pacific Islander AI = Non-Hisp American Indians living on Reservations
DIFF UC = Undercount Rate - Undercount Rate for all non-Black, non-Hispanics
Residual Undercount = Dual System Undercount - CensusPlus Undercount