VARIANCE ESTIMATION FOR THE 1995 CENSUS TEST: METHODOLOGY AND FINDINGS

Robert E. Fay and Machell Kindred Town
U.S. Department of Commerce
Bureau of the Census

Prepared for Presentation at the 1996 Annual Research Conference
Bureau of the Census
Arlington, VA, March 17-21, 1996

VARIANCE ESTIMATION FOR THE 1995 CENSUS TEST: METHODOLOGY AND FINDINGS

Robert E. Fay and Machell Kindred Town
Bureau of the Census

ABSTRACT

The 1995 Census Test evaluated two fundamental changes in decennial census design: sampling for Nonresponse Followup (NRFU) and integrated coverage measurement (ICM). Sampling of nonresponding households after the mail phase of the census will reduce the workload and cost of the census. Housing units not returning their census forms by mail are classified by the Postal Service as occupied or vacant; a sample of occupied nonresponding households is then selected and used to estimate for the universe of nonresponding households. ICM attempts to mitigate differential undercoverage experienced in past censuses by incorporating estimates of omitted persons directly into the final census products. A subsample will be drawn to estimate the residual undercoverage of the census, and estimates of the undercoverage will be integrated into the final count.

Although the use of sampling for characteristics has been a feature of modern censuses, the 2000 Census will be the first to use sampling to determine the number of persons. Consequently, estimation of variance for census data products has become a matter of increased importance. The paper will summarize the methodology implemented for the 1995 Census Test, aspects of which were previously presented (Town and Fay 1995). The paper elaborates features of the application, such as adaptations to the analysis of experimental panels assessing the effect of block vs. unit sampling in NRFU estimation. The paper then reports the findings from the production variance estimation for the 1995 Census Test. These findings will include comparison of estimated variances with the target variances assumed in the initial design. The test will provide important evidence on the relative levels of variance for a "Census Plus" strategy in the ICM vs. the dual-system approach incorporated in a number of previous attempts to measure census undercount. Finally, the analysis identifies features of sample design likely to affect the quality of variance estimation in 2000.

KEYWORDS

Census adjustment, imputation, jackknife, census undercount.

1. INTRODUCTION

The Census Bureau recently released the proposed design for the 2000 census (U.S. Bureau of the Census 1996a, 1996b). The 2000 census, like censuses for the past several decades, will contact respondents initially through the mail. An innovation for 2000, however, will be additional mail contact with reminders and a replacement questionnaire for those not responding initially, in order to increase mail response. Following practice in previous censuses, there will be a subsequent attempt to contact most nonresponding households in person. After areas reach a proposed level of 90 percent response, the remaining households in 2000 will be revisited on a sample basis to complete the nonresponse follow up (NRFU) phase of the census. The term "truncated census" has been applied to the four-step process of 1) mailing census questionnaires; 2) attempting to follow up most or all nonresponding households; 3) sampling the remaining nonrespondents for intense completion efforts; and 4) measuring and correcting census coverage errors through Integrated Coverage Measurement (ICM).

A sample of blocks will be selected for ICM. Both mail and nonmail households in the ICM sample will be eligible for a detailed interview to determine omissions and erroneous inclusions of persons from the census. Within sampled blocks, ICM interviewers will also check for omissions of housing units from the census and any missed persons living in them. The results from the ICM are to be integrated into the census results, so that all statutory obligations for the 2000 census will be met with census estimates corrected for estimated omissions, providing a "one-number census."

The 1995 Census Test was designed to test major features of this new approach to taking the census. The test was conducted in Oakland, California, Paterson, New Jersey, and six parishes (county equivalents) in Northwest Louisiana. After mail response, samples of nonrespondents were selected without the intermediate step of attempting to contact all nonresponding households as in the truncated census. (This approach has gained the designation "direct sampling" at the Census Bureau, referring to 1) mailing census questionnaires; 2) sampling the remaining nonrespondents for intense completion efforts; and 3) measuring and correcting census coverage errors through ICM.) U.S. Bureau of the Census (1996c) provides an overview of census operations in the test. ICM operations and adjustments were employed in all three sites.

Previously (Town and Fay 1995), we described general features of the sample design for the 1995 Census Test. We examined alternative approaches to estimating the variance of site-level estimates and reported results from a Monte Carlo study using data from the 1990 census for Paterson and the 1990 Post-Enumeration Survey (PES). The results of the study favored a jackknife estimation approach, modified to accommodate the design. Data from the 1995 test subsequently have become available. Section 2 reviews this issue, summarizing the sample design and noting additional features as necessary to describe the methodology for the variance calculation for the site-level estimates. Section 3 reports the findings from the test and further analyzes the sources of variance in the site-level variance.

In addition to the site-level estimates, this paper begins to address variance estimation at lower levels, such as blocks, tracts, and the other units of census aggregation. Here, the estimation issues are complex. In this paper, we will focus on only one such issue. One of two experimental panels in the Oakland site employed unit sampling for NRFU. Specifically, in the 1995 test, each block with any nonresponding units was allocated at least one sample hit. (We will use the term unit sample to refer to any nonclustered sample of the nonresponding addresses.) The other panel in Oakland and the Paterson and Northwest Louisiana sites were sampled through block sampling. We focus our attention on variance estimation for the unit sample in Oakland, because it now appears a more likely precursor for NRFU sampling and estimation in the 2000 census than the block samples implemented in the majority of the test site.

The concluding section discusses implications and limitations of these findings, and directions for further research.

This Proceedings version largely represents the conference version, but the text will note instances where new results have been included. We have attempted not to expand the scope of the paper but only to supplement the evidence available at the time of presentation.

2. VARIANCE METHODOLOGY FOR SITE-LEVEL ESTIMATES

2.1 Sample Design and Estimation

Our previous paper (Town and Fay 1995) describes general features of the sample design for the tests, citing internal documentation for additional information. We will first furnish here a brief summary of important aspects of the design.

In all three sites two samples were drawn:

1) An ICM sample of block clusters was selected. In each sampled block cluster, all mail nonrespondents, except for identified vacants, were followed up through standard NRFU operations. Subsequently, both mail and nonmail households became eligible for ICM interview. In Louisiana, only the "Census Plus" (C+) strategy was implemented, and adjustment factors were built into the estimates from the test census. In Oakland and Paterson, data were collected from ICM sample households enabling the calculation of both "Census Plus" and dual-system estimates (DSE). The latter methodology had been employed in a number of previous survey evaluations of census undercoverage, including the 1990 Post Enumeration Survey (PES). In Oakland and Paterson, adjustment factors based on DSE provided the final site-level estimates from the test.

2) From the remaining universe, a NRFU sample of nonresponding addresses was drawn. In Paterson and Louisiana, a block sample was drawn after stratifying the non-ICM blocks by characteristics including the level of nonresponse. Oakland was divided into two panels: one receiving the block sample design used in Paterson and Louisiana; the other, a unit sample drawing a systematic sample of housing units within blocks. Except for the manner of sample selection, the intention was to administer the NRFU operations in the same manner as in ICM sample block clusters, but without subsequent ICM operations.

The ICM design was consequently quite clustered, since the target size of a block cluster was 30 housing units. The number of sampled block clusters was approximately 150 in Oakland, 100 in Paterson, and 75 in Northwest Louisiana. The earlier paper (Town and Fay 1995) includes details of the stratification and sample selection. Figure 1 attempts to clarify the relationship between the ICM and NRFU samples.

The design of the test reflected an effort to obtain critically important evidence on the effect of major alternatives for 2000. A primary objective was to test C+ as an estimation methodology. This approach attempts to measure persons omitted from the census through a second interview. One advantage of C+ is its reliance on standard survey estimation.

Although it was anticipated that C+ estimates might fall slightly below DSE, differences between the DSE and C+ outcomes for the 1995 test were larger than expected. In short, results from DSE were within the realm of past experience, providing an indication of undercounts of minority populations. Repeating past findings, DSE detected only modest difference in the coverage of adult men and women, whereas demographic analysis at the national level has generally been held to provide strong evidence of such a differential (Robinson *et. al.* 1993) in past decennial censuses. The results of C+, however, were so disappointing as to suggest initial failure. Possible improvements in the methodology for C+ will be tested in 1996.

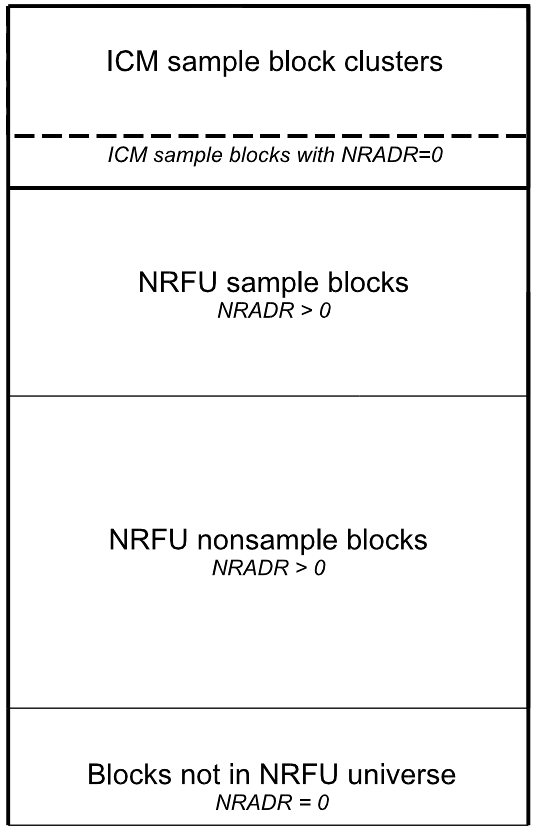

Fig. 1. Relationship between ICM and NRFU sampling. ICM block clusters were first selected, prior to the census mailout. The sample ICM clusters consequently included some blocks with complete mail response, for which the number of nonresponding addresses, NRADR, was 0. After mail response, blocks with NRADR > 0 constituted the universe from which the NRFU sample was selected, using NRADR as one of the variables affecting stratification.

The Oakland site provided a comparison of block vs. unit sampling for nonresponse. There is a technical consensus with the Census Bureau that current ICM methodology virtually requires that the ICM sample blocks be completely interviewed for NRFU in order to permit later matching of ICM respondents. Otherwise, ICM respondents not matching to the census could not be easily distinguished from the generally larger number of mail nonrespondents.

With complete nonresponse followup required for ICM sample blocks, the dilemma is then to choose between:

a) block sampling for the rest of the NRFU sample, in order to insure consistency between the NRFU estimates and the adjustment factors derived from the ICM sample; vs.,

b) unit sampling of nonresponding housing units, which previous studies have shown has a considerably lower variance.

The findings from the 1995 test indicated no consistent difference in the level of the estimates between the two panels (Treat 1996). Indeed, the overall estimates of population total were extremely close between the two panels. Consequently, a technical recommendation to employ unit sampling for nonresponse in 2000 is now under serious review, although the eventual sample design may vary considerably from the version implemented in Oakland.

Estimation for the 1995 test might be described as a "top-down" approach. Traditional survey estimators, combining NRFU results with ICM findings, were used at the site level. To produce detailed data at the block, block group and other such levels, imputation was employed. For the unit sample in Oakland, the imputation for followup nonresponse amounted to a relatively simple duplication of sample households. Since the sampling interval was approximately 1 in 3.5, each sample household was duplicated 2 or 3 times. A more complex procedure was required for block sampling. To resolve differences between the site-level estimates and estimates formed by adding up the observed and imputed data, the imputation was followed by procedures that either augmented or reduced the imputations to meet the overall targets defined by the site-level post-NRFU estimates.

After imputation and imposition of site-level constraints, then, the final post-NRFU estimates were obtained by adding up, without weights, the number of persons on the imputed file. ICM adjustment factors were then applied at the block level and the resulting fractional people distributed according to a controlled rounding algorithm within each block. Because the imputations were constrained to sum to the survey estimates at the site level, however, the discussion of variance estimation for the site-level estimates may ignore the details of the imputation methodology.

Site-level estimation was performed in two steps. The first employed sample data from NRFU, including the data from nonresponse followup from ICM sample blocks, to produce a post-NRFU estimate of the population. The second adjusted the post-NRFU estimate by the ICM findings, producing DSE estimates in the two urban sites and C+ estimates in all three.

The post-NRFU site-level estimate gave unit weight to all counts from self-response (mail returns, Be Counted forms, Reverse Computer-Assisted Telephone Interviews) and a survey weight to estimates from the NRFU sample, including followup cases in ICM blocks, based on the unconditional probabilities of selection into the NRFU and ICM samples. In each site, poststratification and ratio estimation were employed to reflect the total number of nonresponding addresses. In Oakland, ratio estimation proceeded separately for the two experimental panels, stratifying the unit-sampling panel by number of nonresponding addresses in the block. In Louisiana, the ratio estimation was elaborated into a two-way rake (iterative proportional adjustment) in order to control to site-level totals of nonresponding addresses by poststrata and to broad groupings of poststrata by parish, for purposes of parish-level estimation.

Following the general design of previous censuses, DSE employed results from two overlapping samples of households within ICM sample blocks: a P-sample to estimate omitted persons and an E-sample to estimate erroneous enumerations. C+ produced a resolved roster (R-sample) for each household, which was created after reconciling the initial P-sample roster with the respondent to resolve differences from the census enumerations.

2.2 Reflecting the Relationship Between ICM and NRFU Sampling in Variance Estimation

The methodology for site-level variance involved two relatively novel methodological aspects, which were previously discussed in Town and Fay (1995) and Fay and Train (1995). We summarize these two aspects in this section and in section 2.3.

To review, the ICM sample, s_1, with sampling fraction f_1, was selected first and received both the NRFU and ICM treatments. The remaining blocks were restratified, using information unavailable when the ICM sample was selected, and sampled into s_2, with sampling fraction f_2, for additional NRFU treatment. Post-NRFU estimates, X, used both samples with weights based on unconditional probabilities of selection. The problem is similar but not identical to double sampling. Variances for post-NRFU estimates, X, ICM estimates, Y, and any covariances between them are required. Following the development in Town and Fay (1995), we first discuss the situation of stratified simple random sampling.

Variances for estimates based only on the ICM sample, s_1, are straightforward. We employed a stratified jackknife incorporating the finite population corrections for variances for estimates, Y, based on s_1. Empirical evidence presented in Town and Fay (1995) favored this choice over half-sample replication and another replication method to be described later in this section.

We summarize the derivation in Town and Fay (1995) of a variance estimator for characteristics based on both s_1 and s_2. Let $U - s_1$ denote the non-ICM blocks, which is the universe from which the stratified NRFU sample, s_2, is selected. Under standard assumptions, two unbiased estimates of a finite population total are possible. One uses the data from s_1 only,

$$\hat{X}_1 = \sum_{i \in s_1} f_1^{-1} x_i$$

while the second conditions on s_1,

$$\hat{X}_2 = \sum_{i \in s_1} x_i + \sum_{i \in s_2} f_2^{-1} x_i$$

The data from the two samples are combined so that each sample x_i is given weight f^{-1}, where

$$f = f_1 + (1 - f_1) f_2$$

that is,

$$\hat{X} = (1 - f_2) \left(\frac{f_1}{f} \right) \hat{X}_1 + \left(\frac{f_2}{f} \right) \hat{X}_2$$

Figure 2 illustrates this decomposition.

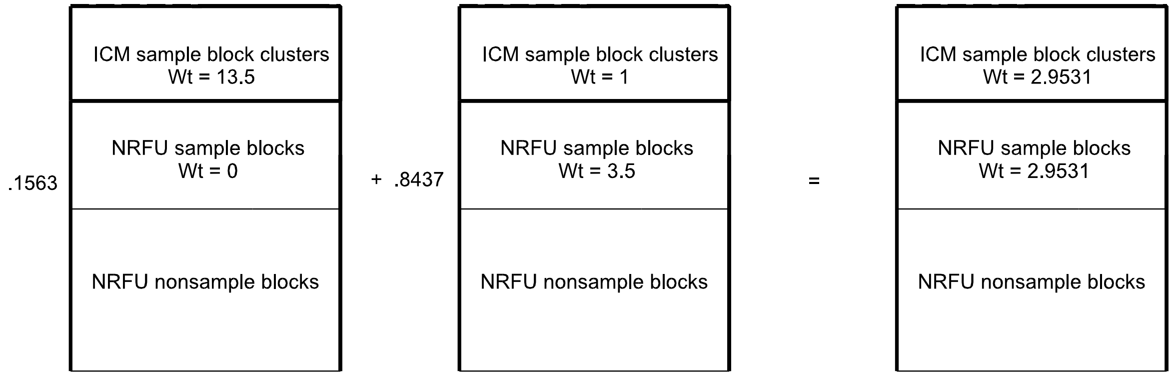

Fig. 2. Illustration of the decomposition of the NRFU estimator into $\hat{X}_1$ based on the ICM only and $\hat{X}_2$ using the ICM sample as self-representing and employing the NRFU sample to estimate the balance of the universe. The weights are illustrative but were among the combinations appearing in Oakland.

We have

$$\begin{aligned} Cov(\hat{X}_1, \hat{X}_2) &= E(Cov(\hat{X}_1, \hat{X}_2 \mid s_1)) \\ &\quad + Cov(E(\hat{X}_1 \mid s_1), E(\hat{X}_2 \mid s_1)) \\ &= 0 \end{aligned}$$

since

$$\begin{aligned} Var(\hat{X}_1 \mid s_1) &= 0 \\ Var(E(\hat{X}_2 \mid s_1)) &= 0 \end{aligned}$$

Thus, a jackknife estimate of variance may be formed from

$$v_J(\hat{X}) = (1 - f_1)(1 - f_2)^2 v_J(X_1^*) + (1 - f_2) v_J(X_2^*) \qquad (1)$$

where $v_J(X_1^*)$ represents the usual jackknife variance estimate for X_1^*, the portion of the estimated total $\hat{X}$ based on s_1, and where, analogously, $v_J(X_2^*)$ represents the usual jackknife variance estimate for X_2^*, the portion of the estimated total $\hat{X}$ based on s_2. In other words, both X_1^* and X_2^* are based on applying the unconditional weight, f^{-1}, to the sample data from s_1 and s_2, respectively.

Eq. (1) was implemented through replication, using the program VPLX (Fay 1995a). The variance of each characteristic, Y, was estimated through

$$V(Y) = \sum_{r} b_r (Y_r - Y)^2 \tag{2}$$

where b_r are coefficients determined by the sample design and Y_r are estimates for replicate samples r. In turn, the replicate estimates were defined by replicate weights identifying how each observation should contribute to the replicate estimate

$$Y_r = \sum_{i} w_{ir1} \, y_i$$

One set of replicate weights { w_{ir1} } is defined for post-NRFU estimation along the lines of *eq.* (1). Mail response cases and other observations receiving unit weight in the census post-NRFU received replicate weights of 1. A second set of replicate weights { w_{ir2} } for each observation in s_1 provided the basis for estimates of variance for ICM characteristics, Y, for an appropriate choice of b_r.

The replicates were set up to maintain the covariance between the two samples in the following manner:

1) Replicate weights corresponding to the stratified jackknife were set up first for s_1.
2) Replicates were then set up for s_2, to follow the replicates for s_1. Cases in s_2 received their full sample weight over the range of replicates set up for s_1. Similarly, cases in s_1 were given their full sample weight over the range of replicates set up for s_2.

Over the range of replicates corresponding to s_1, { w_{ir1} } and { w_{ir2} } were constructed to give the same values of b_r. Thus, *eq*. (2) could be used for all site-level characteristics, including those, such as the DSE, combining estimates from the two samples.

2.3 Variance Estimation for Systematic Sampling

Although the motivation for the variance estimator is in terms of stratified sampling without replacement, both s_1 and s_2 were drawn through systematic sampling with equal probabilities of selection within strata. In Town and Fay (1995) samples drawn from census results in Paterson according to the 1995 sample design indicated no clear pattern of bias in the variance estimator based on the jackknife, and some advantage in performance over another variance estimator designed specifically for systematic sampling (Fay and Train 1995). We next describe this variance estimator, which we used for the unit sample panel in Oakland.

Among the estimators that Wolter (1985, ch. 7) studied, two estimators, both based on squared differences between neighboring sample cases, did relatively well as general solutions. Expressed as estimators of the variance of the estimated population total

$$\hat{Y}_0 = \sum_{i=1}^{n} w_i \, y_i$$

they were

$$v_2 = (1-f)\frac{n}{2(n-1)}\sum_{i=2}^{n}(w_i y_i - w_{i-1}y_{i-1})^2$$

and

$$v_3 = (1-f)\sum_{i=1}^{n/2}(w_{2i}y_{2i} - w_{2i-1}y_{2i-1})^2$$

where y_i, i = 1,..., n, represents a systematic sample from an ordered population, and f denotes the sampling fraction n/N. Estimator v_3 assumes that n is an even number. Both estimators employ squared differences of neighboring observations to estimate variation; v_3 by comparing n/2 distinct pairs, while v_2 compares each sample observation, except for j = 1 and j = n, to two others.

Following the development in Fay and Train (1995), consider instead the following modification to v_2:

$$v_{2m} = 1/2\,(1-f)\left[(w_n y_n - w_1 y_1)^2 + \sum_{i=2}^{n}(w_i y_i - w_{i-1}y_{i-1})^2\right]$$

This estimator adds a comparison of the first and last sample case. In applications where the order of the sort is highly informative and $w_1 y_1$ and $w_n y_n$ are likely to be highly dissimilar, this step cannot be taken lightly. It is used here, however, to establish a link between v_2 and the successive difference replication method.

Plackett and Burman (1946) provided a method of constructing orthogonal matrices, $A = \{a_{ij}\}$ of order 4k such that $AA' = 4kI$, with each a_{ij} = 1 or -1, for most values of k up to 100 or more. These methods are implemented in VPLX. In turn, most of the matrices constructed in this manner have a first row consisting entirely of 1's. Let 4k, at least n+2, be the order of such a matrix, A. Then, for each replicate r = 1,...4k, assign to each observation y_i the replicate factor

$$f_{ir} = 1 + (2)^{-3/2}a_{i+1,r} - (2)^{-3/2}a_{i+2,r}$$

for $i < n$, and

$$f_{nr} = 1 + (2)^{-3/2}a_{n+1,r} - (2)^{-3/2}a_{2,r}$$

The replicate factors are related to the replicate weights through $w_{ir} = f_{ir}\,w_i$ where w_i represents the original weight. In turn these replicate factors define a set of replicate totals

$$\hat{Y}_r = \sum_{i=1}^{n} f_{i,r}\,w_i\,y_i$$

Fay and Train (1995) show that the resulting replicate variance estimate,

$$v_{r2m} = 4\,(4k)^{-1}\,(1-f) \sum_{r=1}^{4k} (\hat{Y}_j - \hat{Y}_0)^2$$

is identical to v_{2m}. This is an interim result, however, since the purpose is to motivate the *successive difference replication* variance estimator, v_{r2}, obtained by defining

$$f_{ir} = 1 + (2)^{-3/2} a_{i+1,r} - (2)^{-3/2} a_{i+2,r}$$

for all *i*. Consequently, y_1 and y_n no longer share a row of the Hadamard matrix, and each end point is compared directly to only one other observation rather than two.

Strictly speaking, the variance estimator is not unbiased for simple random samples, but Monte Carlo studies suggest a small bias, less than 1 percent, even for relatively small *n*, when estimating variances for ratio estimates. (For simple unbiased estimates, the variance estimator may have an upward bias resulting from the treatment of the end points.) Applied to the highly clustered Paterson data, Town and Fay (1995) found this estimator did not preform as well as the jackknife. Consequently, this estimator was used only for NRFU estimation from the unit sample in Oakland. Here, we reasoned that the application was closer to other empirical tests that have favored v_{r2}, such as the application to the Current Population Survey.

2.4 Limitations of the Variance Estimates

The variance strategy that we have described closely respects the sample design. We noted in passing, however, that we selected a jackknife procedure based on stratified simple random sampling without replacement, based on favorable empirical performance reported in Town and Fay (1995). Thus, the variance estimates do not presume any advantage to the particular form of systematic sampling employed to draw the samples for the ICM and NRFU block samples.

We did not attempt in this research to reflect additional variance due to missing data for the population count. In particular, the ICM estimates included imputations for missing data. One source was particularly notable: a followup interview for specific classes of ICM cases, such as whole households not matching the census, was attempted only in one of the two panels of the design and the followup results imputed to the other panel; these imputations were employed in the DSE estimates. Thus, the variance results we report for the 1995 test are likely to be underestimates.

3. VARIANCE RESULTS FOR THE 1995 TEST

Table 1 presents the primary site-level variance results.

Table 1. Site-Level Preliminary Estimates and Estimated Standard Errors, 1995 Test
(Excluding the Group Quarters Population)

	Post-NRFU	C+	DSE
Oakland	332,734 (1,264)	334,482 (5,681)	361,538 (6,738)
Paterson	127,950 (1,477)	132,337 (2,164)	145,504 (2,746)
Northwest LA	114,163 (808)	116,156 (1,695)	

Note: The population estimates shown are not official. They omit the group quarters population, which was excluded from NRFU sampling and ICM estimation. In addition, they may disagree by a few persons with final census results for the sites, excluding group quarters, because of the effect of rounding in the computations. No DSE estimates are available for the LA site.

At the site level, the standard errors for either the C+ or DSE estimates are considerably larger for the post-NRFU estimates, indicating that the ICM component of the estimation is the dominant contributor to variance at the site level.

The standard errors for the DSE are somewhat larger than those for C+. Since the DSE estimates are so much larger, however, the variance advantage of C+ is, at this point, inconsequential. Variance comparisons between the two ICM methodologies will only be of interest when there is much less of a difference in the level of the estimates. It is unclear whether C+ would continue to show a variance advantage over DSE if it produced comparable population estimates.

Table 2. Site and Parish Level Estimates for Northwest Louisiana

	Post-NRFU	C+
Northwest LA site	114,163 (808)	116,156 (1,695)
Bienville Parish	15,078 (369)	15,380 (424)
Desoto Parish	24,102 (345)	24,546 (516)
Jackson Parish	15,004 (358)	15,218 (398)
Natchitoches Parish	34,788 (464)	35,417 (641)
Red River Parish	9,259 (195)	9,424 (244)
Winn Parish	15,932 (326)	16,171 (385)

Note: The population estimates shown are not official. They omit the group quarters population, which was excluded from NRFU sampling and ICM estimation. In addition, they may disagree by a few persons with final census results for the sites, excluding group quarters, because of the effect of rounding in the computations.

Table 2 compares the site and parish level results for Louisiana. Because the ICM adjustments were computed at the site level and then applied to each parish separately, the relative contributions of NRFU and ICM estimation shift at the parish level, where now the NRFU contribution is a significant or primary source of the estimated total variance. It should be noted, however, that the estimated variances for the ICM adjustments do not take into account the effect of differences between the site level rates of undercoverage, which formed the basis for the ICM adjustments, and the underlying real rates at the parish level. The assumption that adjustments computed at higher geographic levels may be applied to lower ones has been termed the *synthetic assumption* in the literature on the undercount. Given that assumption, the variance estimates show that the NRFU sampling contributes half or more of the overall variance at the parish level.

As previously noted, Treat (1996) reported insignificant differences between the estimates from the block and unit panels in Oakland. Table 3 shows separate estimates by panel and estimated standard errors. As described in Section 2.2, the production estimate employed both the ICM and NRFU sampling results. The alternative estimator, $\hat{X}_2$, treats the ICM sample as self-representing and weights only the NRFU sample. For this second approach, the ratio of estimated standard errors is about 2.6, which represents a ratio of variances of about 6.9.

Table 3. Estimated Persons and Standard Errors, for Oakland by Panel

	Panel 1: Block	Panel 2: Unit
Production estimates	160,894 (1,154)	160,878 (644)
NRFU sample only	160,728 (1,246)	161,388 (476)

Note: The population estimates shown are not official. They omit the group quarters population, which was excluded from NRFU sampling and ICM estimation. In addition, they may disagree by a few persons with final census results for the sites, excluding group quarters, because of the effect of rounding in the computations. Some blocks were assigned to another study and excluded from the analysis of the panels.

Subsequent to the conference presentation, we pursued some additional research to verify these findings and to account for the large design effect of block sampling in Oakland. The findings are included in the remainder of this section.

Although the Postal Service was seen as the primary identifier of vacant housing units, the majority of vacants and deleted units were identified during NRFU. Consequently, the estimation task of NRFU may be described as estimating: 1) the number of occupied units among the nonresponding addresses, and 2) the average number of persons per occupied unit. In other words:

$$\text{\# NRFU persons} = (\text{\# occ hu}) \times (\text{\# persons})/(\text{occ hu})$$

Table 4 shows that this decomposition helps to clarify the sources of variance in the estimate of total persons. The coefficient of variation of each component is shown separately. For block sampling, estimating the number of occupied units is almost as large a challenge as estimating the average number of persons, whereas this is less a concern for unit sampling. In other words, the design effect for estimating the occupancy rate is particularly high for block sampling, with a ratio of standard errors of about 3.4, compared to a ratio of about 2.4 for estimating the average number of persons.

The table next shows the resulting coefficient of variation if the two components were statistically independent, followed by the actual estimated coefficients of variation for estimated persons. In both cases, the actual is barely greater than the result under independence, showing that the total variance can be usefully decomposed into the problems of estimating occupied units and the average number of persons per occupied unit.

Table 4. Estimated Percent Coefficients of Variation for the Occupancy Rate and Average Number of Persons per Occupied Unit for the Oakland NRFU Universe

	Panel 1: Block	Panel 2: Unit
# occ hu	0.51	0.15
# persons/occ hu	0.57	0.24
Under independence	0.77	0.28
Observed for total persons	0.78	0.29

Note: The table reports the c.v.'s of the separate components, the resulting c.v.'s under independence, and the estimated c.v.'s for total persons.

Figure 3 employs this approximate decompostion to study the effect of cluster size.

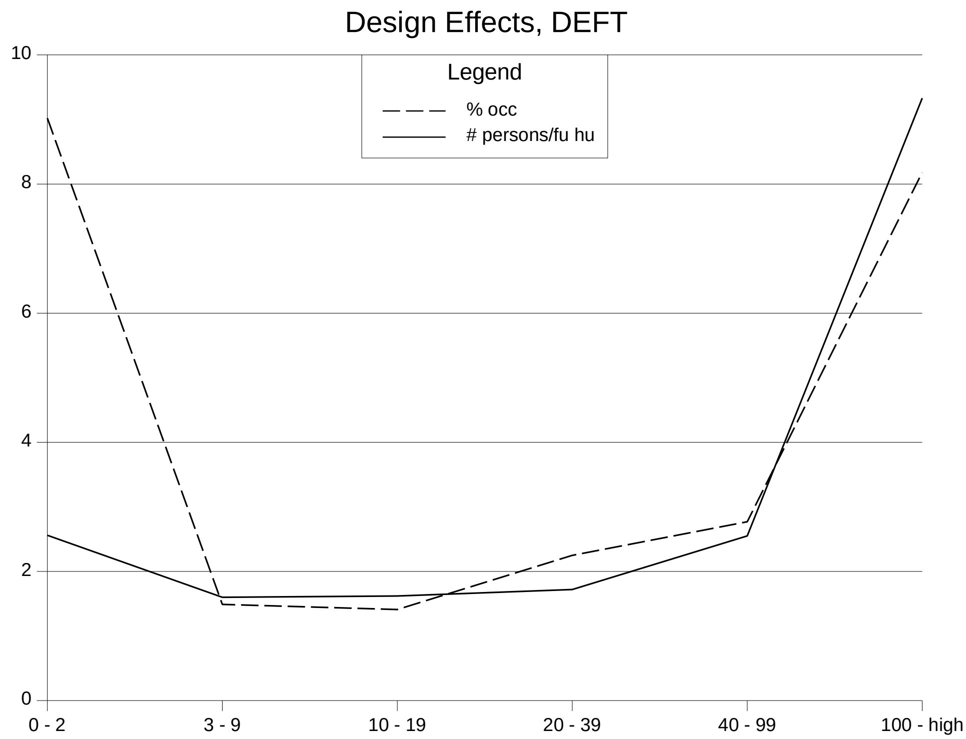

Fig. 3. Ratios of estimated standard errors (deft) for block vs. unit sampling in Oakland, using the NRFU sample only,for blocks grouped by NRADR, the number nonresponding addresses. Because blocks with one nonresponding address were always selected in the unit sample, the results for the 0-2 interval are not meaningful. Otherwise, the comparison shows slowly rising deft in the range of 3-99 nonresponding addresses, but a sharp increase at the upper end for NRADR > 100. The results for the upper end could be sensitive to the outcome in only a few blocks.

The number of nonresponding addresses varied widely by block in Oakland. Although the production estimation did not formally poststratify by the number of nonresponding addresses, a calculation of this sort helps to illustrate the variation in the design effects by block. As noted in the note to the figure, results for 0-2 should be discounted. The results suggest that the Oakland findings may be particularly affected by high design effects estimated at the upper end. Even without the extreme findings at the upper end, however, deft's approaching or above 2 for blocks in the 20-39 and 40-99 ranges indicate that the variance advantage of unit sampling is considerable and persistent.

4. VARIANCE ESTIMATION FOR UNIT SAMPLING

Fay (1995b) described a variance estimator for nearest neighbor imputation appropriate for a simple random sample of size n from a finite population of size N without replacement. We first describe this general form for estimating the variance of the sample mean and then discuss its application to imputation for unit sampling.

For each nonrespondent k, let $nn1(k)$ be the closest responding observation. Closeness may be based on observed characteristics but not the value of the unobserved y. For example, we may use the sort order of census id's within blocks and address range areas (ARA's). We then impute the missing value from the nearest neighbor. In order to obtain a variance estimate, let *nn2(k)* be the second nearest neighbor; in other words, the observation that would have been used for imputation if *nn1(k)* were deleted from the sample. For each respondent k, let $nn1^{-1}(k)$ be the set of nonrespondents for which k is the closest responding observation. Similarly, define $nn2^{-1}(k)$ to be the set for which k is the second nearest neighbor. Finally, for each respondent pair $k, k' \in r$, let $nnp^{-1}(k, k')$ be the set of nonrespondents for which k and k' are the 1st and 2nd nearest neighbors, respectively. Like the estimator proposed by Rao and Shao (1992), the variance estimator begins by modifying the usual jackknife replicates. Define adjusted replicate values

$$\begin{aligned} \bar{y}^{b}(-k) &= \left(\frac{1}{n-1}\right)\left[n\bar{y}_s - y_k + \sum_{j \in nn1^{-1}(k)} (y_{nn2(j)} - y_k)\right] \quad \text{if } k \in A_r \\ &= \left(\frac{1}{n-1}\right)\left[n\bar{y}_s - y_{.k}\right] \quad \text{if } k \in A_{nr} \end{aligned} \tag{3}$$

where A_r and A_{nr} are the sets of respondents and nonrespondents, respectively. Define a second set of adjusted replicate values

$$\begin{aligned} \bar{y}^{c}(-k) &= \bar{y}_s + \left(\frac{1}{n-1}\right) \sum_{j \in nn1^{-1}(k)} (y_{nn2(j)} - y_k) \quad \text{if } k \in A_r \\ &= \bar{y}_s \quad \text{if } k \in A_{nr} \end{aligned} \tag{4}$$

The proposed variance estimator is:

$$\begin{aligned} v_6 = {} & \frac{n-1}{n}\,\frac{N-n}{N} \sum_{k \in s} \left[\bar{y}^{b}(-k) - \bar{y}\right]^2 + \frac{n-1}{n}\,\frac{n}{N} \sum_{k \in s} \left[\bar{y}^{c}(-k) - \bar{y}\right]^2 \\ & - \frac{1}{2n^2} \sum_{k, k' \in r} \left[\sum_{j \in nnp^{-1}(k,k')} \left(y_{k'} - y_k\right)\right]^2 - \frac{1}{2n^2}\,\frac{N-2n}{N} \sum_{j \in A_{nr}} \left[y_{nn2(j)} - y_{nn1}\right. \end{aligned} \tag{5}$$

The first of the four major terms on the right-hand side of (5) approximates the effect of each $k \in A_r$ by substituting a second nearest neighbor in its place for each case for which k provided an imputation. The second major term compensates for the effect of the finite population correction in the first term, to capture the full effect of each $k \in A_r$ on the imputed values. The sum of the first and second terms overestimates some variances and covariances; the other two terms correct these effects. The third term corrects estimated covariances between cases with missing data sharing the same 1st and 2nd nearest neighbors. An additional correction to the estimated variance for each imputed case is incorporated into the last term, although the term also estimates the variance in predicting missing y's in the finite population.

Estimator (5) relies on model assumptions:

$$\begin{aligned} E_\xi(y_k) &= E_\xi(y_{nn1(k)}) \\ V_\xi(y_k \mid x_k) &= 1/2\; E_\xi(y_{nn1(k)} - y_{nn2(k)})^2 \\ Cov_\xi(y_k, y_{k'} \mid x_k, x_{k'}) &= 0\,,\; k \neq k' \end{aligned} \tag{6}$$

Note that (6) does not assert a specific functional relationship between the *x*'s and the *y*'s or their variance. Because matching is frequently imperfect, model (6) must be regarded as an idealization.

In application to census imputation for NRFU, the first term of (5) drops out, and the role of *n* becomes indeterminate. Instead, we re-express (4) and (5) in terms of estimates of total

$$\begin{aligned} Y^{c}(-k) &= Y + \sum_{j \in nn1^{-1}(k)} (y_{nn2(j)} - y_k) \quad \text{if } k \in A_r \\ &= Y \qquad\qquad\qquad\qquad\qquad\quad\ \text{if } k \in A_{nr} \end{aligned} \tag{7}$$

In other words, for each donor, *eq.* (7) creates a replicate estimate substituting second nearest neighbors.

The adapted form of (5) is:

$$\begin{aligned} V(Y) = \sum_{k \in A_r} \left[Y^{c}(-k) - Y \right]^2 - \frac{1}{2} \sum_{k,k' \in A_r} \left[\sum_{j \in nnp^{-1}(k,k')} \left(y_{k'} - y_k \right) \right]^2 \\ + \frac{1}{2} \sum_{j \in A_{nr}} \left[y_{nn2(j)} - y_{nn1(j)} \right]^2 \end{aligned} \tag{8}$$

In spite of its complex appearance, *eq.* (8) describes an estimator easily computed through replicate weights. In application to the census, the replicate weights would take the simple form of 1's and 0's.

Inspection of (5) or (8) reveals that these variance estimators derive some inspiration from the estimator of Rao and Shao (1992) for the hot deck, but there are important differences. The Rao and Shao estimator was based on hot deck cells with appreciable numbers of donor cases, but no such restriction is place here. Rao and Shao also required that imputations be selected from the hot deck with replacement, and this requirement added considerable robustness against failure of model assumptions. The assumptions of *eq.* (6) can be regarded as a nonparametric model of closeness. It is possible to create populations for which the 1st nearest neighbor is much closer to the true value than the 2nd, and for these populations the variance estimator tends to have a significant upward bias. On the other hand, the matching rules do not have to reflect an additional random selection from a larger collection of donors, unlike the Rao-Shao estimator.

We constructed an empirical test of the variance estimator based on the same census population for Paterson as in Town and Fay (1995). We extracted mail nonresponse addresses from the 1990 census, including delete cases, sorted by ARA (similar to census tracts), block, and census id. There were 33 ARA's in this population, with 1033 blocks containing one or more nonresponding addresses. We did not fully replicate the details of the Oakland sampling; instead we considered systematic sampling in each ARA, with independent starts in each. We used intervals of 3, 5, 7, 9, and 11, defining nearest neighbor simply by position in the sort order. Imputation was constrained to remain within ARA's, but not blocks. Two forms of Monte Carlo were performed. In the first, each of the possible systematic samples in each ARA was evaluated separately (for example, 3 possible samples for a sampling interval of 3, *etc.*) and summed to a site-level result. The first allowed a rapid evaluation of the bias. The second actually drew full samples in order to measure the coefficient of variation of the site-level variance estimate, and was based on a larger sample.

Table 5. Results of a Small Monte Carlo Study to Evaluate the Performance of the Nearest Neighbor Imputation Variance Estimate for Systematic Sampling and Imputation

Sampling interval	Bias	CV
3	12.5	22
5	-2.4	
7	18.6	
9	1.2	
11	17.5	31

The results in this table, although not perfect, show this approach to be a strong contender. Its advantages are the simplicity and flexibility of estimation that the nearest neighbor imputation allows, the production of a file (before ICM results are considered) that looks like a complete count file, and the ability to compute adequate measures of precision.

Fig. 4-6 provide summaries for the 33 Address Range Areas (ARA's). ARA's in Paterson essentially correspond to census tracts, representing groupings of blocks. Fig. 4 shows that the estimated variance increases quite predictably with the size of the estimate in the ARA. Fig. 5 shows that the coefficients of variation of the estimates are relatively acceptable at the ARA level, suggesting the possibility of possibly publishing direct variance estimates at this level, probably in conjunction with variance generalization. Fig. 6 shows that the actual variance, based on only 3 possible samples, cannot be predicted with precision by this approach.

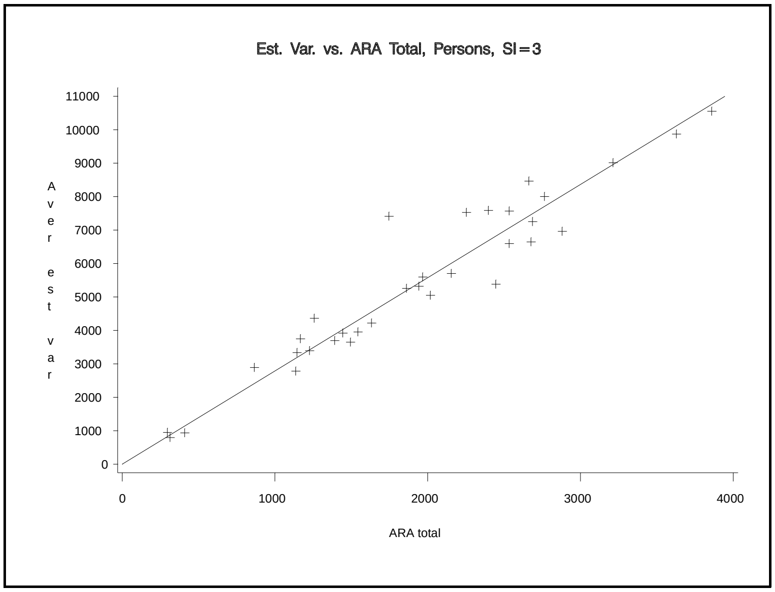

Fig. 4. Average variance estimate for each ARA compared to the total number of persons in the followup universe, when SI=3. A least squares line is shown. The average variance estimate exhibits quite stable behavior as a function of the followup population.

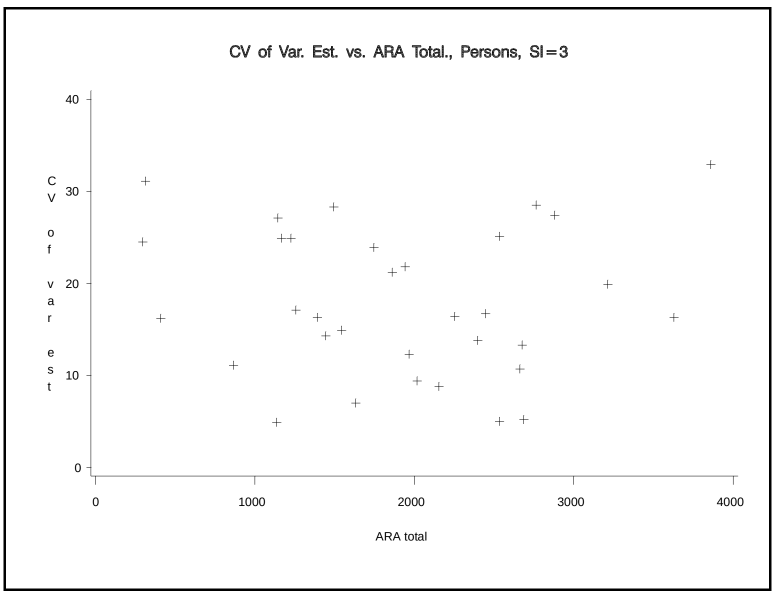

Fig. 5. CV of the variance estimator over sampling for each ARA, compared to the total number of persons in the followup universe. Because there are only 3 possible samples in each ARA when SI=3, the distribution of CV's is scattered, even for ARA's with large followup populations.

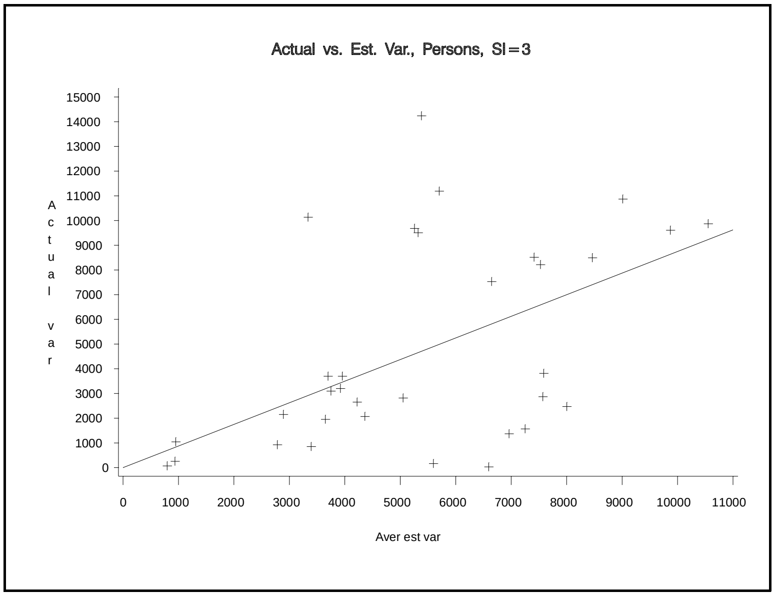

Fig. 6. Actual variance by ARA compared to the average variance estimate, when SI=3. A least squares line, with slope .875, is shown. Because the actual variances are based on only 2 degrees of freedom in each ARA, there is wide variation about the regression line.

Subsequent to the conference, additional computations were performed to suplement these findings. Additional populations were defined by randomly reducing the universe by 10 percent, roughly equivalent to the effect of a modest increase in response.

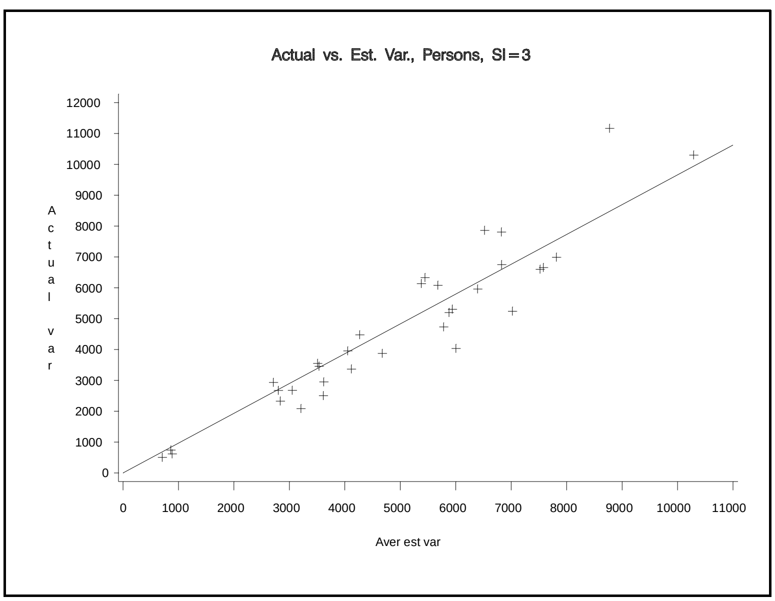

Fig. 7. Actual variance by ARA compared to the average variance estimate, when SI=3, averaged over a series of populations based on randomly deleting 10%. A least squares line, with slope .966, is shown.

In general, this approach introduces many more possible samples than the few provided with systematic sampling. In other words, an interpretation is that the actual variance of systematic sampling, based on few possible samples, is intrinsically difficult to predict, but the variance estimator does successfully indicate the average variance for a set of similar populations formed by random deletion.

5. CONCLUSIONS

Variance measures the consequences of survey design and estimation. Thus, it is appropriate for a paper on variance estimation to offer a number of conclusions addressing these two related issues, as well as variance estimation *per se*. In most cases, further work is warranted to confirm or dispute what we report here.

We believe that the effect of sampling variance will be an important factor affecting public acceptance of the 2000 census. If the sampling variances are too high, then comparison of census estimates with other sources of information, including the Census Bureau's postcensal estimates, will undermine the acceptance of the 2000 results. At higher levels of geography, we expect the effect of ICM estimation will have the larger effect on the variance, whereas NRFU sampling and estimation will be important at lower levels of geography.

One of the important findings from the census test is empirical support for combining unit sampling for NRFU with block sampling for ICM. This conclusion, combined with considerable evidence that unit sampling will offer substantial variance advantages over block sampling, strongly favors the future use of unit sampling.

Several previous methodological studies have been based on block sampling. These were both important and useful efforts. Unit sampling, however, fundamentally changes everything. Unit sampling does not exclude these methodologies from further consideration, but it does require that evaluation of the properties of each estimator previously based on block sampling be revisited.

Sections 2 and 3 documented the methodology we employed for site-level estimates from the 1995 test. We do not see these methods as a blueprint for 2000. In particular, part of the complexity of the methodology arose from representing the joint use of ICM and NRFU samples for NRFU estimation. In the future, we agree with Zaslavsky's suggestion (1995) that we look for less complex methods for 2000. In fact, given the much larger size of the NRFU sample compared to the ICM sample, we recommend that all followup interviews in ICM blocks be self-representing. This move would simplify the variance estimation enormously. With the 2000 sample sizes, no other use of the ICM data for NRFU estimation could possibly improve on this simple approach by more than a trivial amount. As suggested by the variance calculation in Oakland, weighting by unconditional probabilities of selection may prove worse than making ICM blocks self-representing for NRFU estimation, when unit sampling is used for NRFU.

The scope of variance estimates produced so far for the 1995 test falls short of requirements for 2000. Indeed, we have not yet addressed the problem of variance estimates for lower levels of geography for the block sample. As much as it would be desirable to do so, we believe that it is important to focus now on the probable design for 2000.

The results reported in Section 4 for variance estimation with nearest neighbor imputation are quite promising, even though fine tuning and further assessment are required. Potentially, this approach opens up a wide variety of imputation strategies meriting further assessment as census estimators. For example, the study reported in Section 4 used only proximity in the file to other nonresponding addresses, but other strategies could be evaluated, such as considering characteristics of neighboring responding units.

This experience suggests the possibility of employing some form of unit imputation for NRFU and estimating entirely "bottom-up" for NRFU. Much further work is required to evaluate this strategy in comparison to alternatives but, if this strategy is selected, we observe that variance estimation will become far simpler than under estimation strategies that first estimate NRFU aggregates and then constrain an imputation procedure to these estimated aggregates.

A goal for 2000 would be to base variance estimation entirely on assigning each case replicate weights. Imputations made from NRFU would carry replicate weights based on refining the methods described in Section 4. Cases included in the file on the basis of ICM findings would have replicate weights reflecting the effects of the ICM design and imputation. Once the replicate weights have been created, mass calculation of variances for census characteristics will become relatively easy.

Further work remains on how information on variance should be presented to the public. For example, variance generalization will almost surely have some role, certainly at the block level, but research should examine the quality of the generalized variances and their interpretation. For example, empirical work is required in order to assess the performance of confidence intervals based on generalized variances.

[1]This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributed to the authors and do not necessarily reflect those of the Census Bureau. We thank George Train for computational contributions and Mary Ann Cochran for editorial assistance.

REFERENCES

FAY, R.E. (1995a), "VPLX: Variance Estimation for Complex Samples," unpublished program documentation, available at http://www.census.gov.

______ (1995b), "Replication-Based Variance Estimators for Imputed Survey Data from Finite Populations," unpublished manuscript.

FAY, R. E. and TRAIN, G. (1995), "Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties," *Proceedings of the Government Statistics Section,* Alexandria, VA: American Statistical Association, pp. 154-159.

PLACKETT, R.L. and BURMAN, J.P. (1946), "The Design of Optimal Multifactorial Experiments," Biometrika, **33**, 305-325.

RAO, J.N.K. and SHAO, J. (1992), "Jackknife Variance Estimation with Survey Data Under Hot Deck Imputation," *Biometrika,* **79**, 811-822.

ROBINSON, J. G., AHMED, B., DAS GUPTA, P. and WOODROW, K. (1993), "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis," *Journal of the American Statistical Association*, **88**, 1061-1071.

TOWN, M. K., and FAY, R.E. (1995), "Properties of Variance Estimators for the 1995 Census Test," *Proceedings of the Survey Research Methods Section*, Alexandria, VA: American Statistical Association, pp. 724-729.

TREAT, J. (1996), "Analysis Comparing the NRFU Block Sample and the NRFU Unit Sample Nonresponse Followup Evaluation," 1995 Census Test Results Memorandum No. 31, U.S. Bureau of the Census.

U.S. BUREAU OF THE CENSUS (1996a), The Plan for Census 2000, report issued February 28, 1996.

______ (1996b), The Plan for Census 2000, revised report dated April 5, 1996.

______ (1996c), "Documentation of the Design of the 1995 Integrated Coverage Measurement," 1995 Census Test Memorandum Series IP-MD-40, from David C. Whitford to Ruth Ann Killion.

WOLTER, K. (1985), *Introduction to Variance Estimation*, New York: Springer-Verlag.

ZASLAVSKY, A. M. (1995), "Discussion: Sampling from the 1995 Census Test Buffet," *Proceedings of the Survey Research Methods Section*, Alexandria, VA: American Statistical Association, pp. 748-750.