BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-87/02

REPORT ON STATISTICAL SYNTHETIC ESTIMATION FOR SMALL AREAS

by

Cary T. Isaki, Gregg J. Diffendal, Linda K. Schultz
Statistical Research Division
Bureau of the Census
Room 3134, F.O.B. #4
Washington, D.C. 20233 U.S.A.

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Recommended: Kirk M. Wolter

Report completed: January 28, 1987

Report issued: January 29, 1987

TABLE OF CONTENTS

Report on Statistical Synthetic Estimation for Small Areas

by

C. Isaki, G. Diffendal, L. Schultz

I. Introduction and Executive Summary

A. Introduction

This report summarizes the work of the Census Undercount Adjustment for Small Area group as it pertains to Statistical Synthetic estimation. We report on the results to date. What differentiates the statistical synthetic estimation method of total population of small areas from other synthetic methods is the manner in which the cells or domains for adjustment are constructed. Unlike the demographic analysis synthetic estimator described in Isaki, et.al. (1985) where the cells are defined by age-race-sex, or, synthetic estimators whose domains are defined along administrative/jurisdictional lines, e.g., states/cities, the statistical synthetic estimator uses domains that cross jurisdictional lines but retains characteristics associated with the undercount variable. The basic idea behind the statistical synthetic estimator is to place the requirement of adjusted counts for jurisdictions at a lower level of importance than the requirement for accurate estimation of adjustment factors. By this we mean that although we will need to produce adjusted counts for state A, our adjustment cells need not be defined entirely within state A. Rather, if groups of persons in states A, B and C are felt to be similar in terms of adjustment factors, then the adjustment domain should consist of all such persons in the three states. For example, suppose young Black males in Philadelphia, Baltimore and Washington, D.C. are expected to possess similar adjustment factors. Rather than estimate each factor separately by city, it would be more efficient to estimate a combined factor and use it in synthetic estimation. Further discussion on the motivation of statistical synthetic estimation can be found in Tukey (1981), Tukey (1984), National Research Council (1985), Isaki, et.al. (1986), Tukey (1986) and Schultz, et.al. (1986). The Tukey references mention blocking of the U.S. population according to similarity of undercount rates with the possibility of smoothing the rates prior to application on the census counts. The term block as used here is not the Census Bureau's geographic designation but is terminology gleaned from experimental design. In Isaki, et.al. (1986) several blockings of the U.S. were attempted and the resulting statistical synthetic estimators applied to each of several artificial populations. The Schultz, et.al. (1986) paper extends the Isaki, et.al. (1986) results by considering the effect of sampling error on the statistical synthetic estimator (SSE). The discussion that follows is based on the results of the latter two papers.

B. Executive Summary

The investigation of the statistical synthetic estimation method assumed that unbiased estimates of the population for the constructed adjustment factor domains are available. Given that is true, three SSEs were compared using three artificial populations as standards. The three artificial populations use functions of the variable "substitutions in the census" as a proxy for undercount. The results of the study pertain to the three SSEs and the three artificial populations. When the effects of sampling are considered, the results also pertain to the particular sample design used.

In our investigations, in the absence of sampling error, the SSE termed syn 2 was found to be superior for estimating state total population and for most "race" categories as well. Turning to counties, another SSE, syn DA, was found to be better overall with syn 2 performing better for the counties with the large population. We also considered adjustment of census counts of enumeration districts (EDs) in two states with rather small percent of population in metropolitan areas (and hence not representative of the universe of EDs). The results indicated that the census was superior to the adjusted figures. The adjustment factors, in practice, require estimation via sampling. To investigate the effects of sampling, a simple sample design was constructed and a sample replicate was selected. The sample was used to construct syn 2 and syn DA and the same performance measures used previously in the absence of sampling error were again applied. The measures with sampling error included revealed that syn DA was at least as good as syn 2 and often times did better. Sampling variability of syn 2 needs reduction. As the sample design used was constructed to be equivalent to a 1.1 million person re-enumeration, increasing the sample size appears unlikely. Hence, either gains in sample design efficiency or estimation are likely candidates for improving syn 2. If the variability of syn 2 cannot be reduced, syn DA would be the recommended SSE. Syn DA remains superior to the census for almost all performance measures considered, with or without sampling error.

II. Statistical Synthetic Estimation - No Sampling

A. Background

The motivation for investigating the statistical synthetic estimation method as described below can be found in Tukey (1981) who suggested the blocking of the country so as to form groups of persons with undercount rates as different as possible between blocks but as homogeneous as possible within blocks. To use his example, New York City could be blocked by assigning all Black persons into one block, all Hispanic persons into another block, etc. Persons in rural areas could be placed in a block while persons in small towns placed in another block with such blocks potentially crossing state boundaries. The number of blocks constructed would be limited by the survey design charged to estimate the undercount in each block. A hundred blocks was suggested. These blocks define the categories for which direct estimates of the population are required. They are not sampling strata.

A National Research Council (1985) report on Decennial Census Methodology also recommended research on the topic of blocking. In the report, the panel recommended that several blockings of the U.S. be conducted and the results compared. The idea behind such blocking is to decrease the variance and bias of the usual synthetic estimators (that are constrained by jurisdictional considerations) by grouping persons with similar undercount rates. If this can be accomplished, variance will be reduced because sample allocations can be optimized in an efficient manner and, more importantly, bias can be reduced because the adjustment factors can be formed with fewer constraints.

Once the blocks are formed, statistical synthetic estimation proceeds in the usual manner. Each person in the block counted in the census is adjusted by multiplication by the adjustment factor for the block and the product is summed to the tabulation level of interest. For example, if the level of interest is a county in the state of Montana, and assuming that the state's population is blocked into three blocks (cities, towns and rural blocks), the Montana county's census counts in each of the three blocks are multiplied by the block adjustment factors and summed over the three blocks. The sum represents the statistical synthetic estimate of total population for the county in Montana.

Let F_i denote the true number of persons in the i-th type of block (cities, twins, rural) divided by the corresponding census counts. Let C_{ci} denote the county's census count in the i-th type of block. Then, the synthetic estimator of total population for the county is

$$\hat{Y}_c = \sum_{i=1}^{3} F_i \, C_{ci},$$

B. <u>No Sampling Error</u>

The analysis of the statistical synthetic estimator that follows is completely empirical. First, three blockings were constructed and three artificial populations were created at the enumeration district (ED) level. Then, each statistical synthetic estimator (SSE) was applied toward adjusting the census counts. This was done to provide adjusted census counts at the state, county and ED level.

In this section, it is assumed that the adjustment factors in the model displayed in (1) are measured without error. In section III, an error component due to sampling is considered. We assume that the estimator of the adjustment factor is design unbiased for β. Ideally, the evaluation of the SSE's should use the Post Enumeration Program (PEP) undercount estimates. This was not done primarily due to lack of resources in re-estimating adjustment factors. Another consideration was the lack of a standard with which to compare the resulting small area estimates. We addressed this problem in what follows by creating three artificial populations and used them as standards for comparing the performances of the SSE's and the census. Our results, however, remain specific to the three artificial populations used as a standard.

B.1 Artificial Populations. We briefly describe the construction of the three artificial populations denoted AP1, AP2 and AP3. A more detailed description can be found in Isaki, et.al. (1986). The three artificial populations constructed by age-race-sex at the enumeration district (ED) level are:

i) AP1 = (census - substitutions) + substitutions

ii) AP2 = census + F_{DA1} x substitutions

iii) AP3 = census + F_{DA2} x substitutions

where F_{DA1} and F_{DA2} are defined below.

In all three artificial populations, a function of census substitutions is used as representing the undercount. Census substitutions are the result of imputing people into housing units. For example, people were substituted into the census 1) when no form was completed but people may have lived in the housing unit, 2) when we know only the number of people living in the unit, 3) for machine failure or 4) when field counts for an area (ED or block) were larger than the processed counts. Preliminary analysis using 1980 PEP state data indicated that the census substitution rate was the most important explanatory variable of several types of nonmatch rates in the PEP. The nonmatch rate in the PEP refers to the ratio of estimated total number of persons in the PEP not matched to the census to the PEP estimated total number of persons. Since the nonmatch rate estimates the miss rate of the census (under ideal conditions) and census substitutions were available by age-race-sex at the ED level we focused on census substitutions as a proxy for undercount.

AP1 uses census minus substitutions as the census count and substitutions as the undercount. AP2 and AP3 were formed so that their population counts by age-race-sex at the U.S. level equaled the comparable demographic analysis figure including an assumed 3.5 million illegal aliens (the demographic analysis data were provided by the Census Bureau's Population Division). In both cases the substitution counts are adjusted by factors F_{DA}, the ratio of the differences between the demographic analysis estimate, N_{DA}, and the census to the total of substitutions (F_{DA} = (N_{DA} - census)/substitution). Thirty factors, F_{DA}, are required - five age categories, two sex and three "race" categories; Black, Nonblack Hispanic and Rest are used. Since demographic analysis does not provide an Hispanic category (it provides for Blacks/Nonblacks), AP2 and AP3 differ in how the Hispanic artificial population data are derived. For AP2, we assume that the Hispanics are like the Nonblack population and we used the Nonblack F_{DA} for Hispanics. For AP3, we assume that the Hispanics are like the Black population and used the Black F_{DA} for Hispanics. This latter assumption results in larger undercounts for Hispanics under AP3 than for AP2. Use of the F_{DA} adjustments provide a more pronounced differential undercount, for example, between Black male and female in AP2 than in AP1.

B.2 Estimation. Three statistical synthetic estimators denoted syn DA, syn 1 and syn 2 were constructed and applied to each artificial population. We proceed to describe the construction of adjustment strata (referred to as blocks previously) that identifies the SSE's. The strata for syn DA are defined strictly by 30 age-race-sex categories at the U.S. level. No sub U.S. geographic strata are used. The estimator is related to a demographic analysis synthetic estimator except that the source for direct estimates of the population in each category is obtained via a sample survey rather than via demographic analysis methods. A general form of the statistical synthetic estimator is provided in the Appendix.

The rationale for syn 1 was to separate population groups by age-race-sex and geography. In syn 1, the same "race" and sex categories were used as in syn DA but three rather than five age categories were used. Five geographic groupings were used. The geographic groupings were constructed by assigning each ED in the U.S. an urban (we arbitrarily defined an ED as urban if at least seventy percent of its census population was urban) or non-urban designation. The five area groups were defined on the basis of EDs in district offices (DOs). The DOs are census administrative offices and they were classified as centralized (essentially covering cities), decentralized (remaining mail collection areas) and conventional (personal enumeration). The first group consisted of all urban EDs in DOs with percent Hispanic or Black exceeding 25 percent associated with the 35 of 49 largest Standard Metropolitan Statistical Areas. The DOs in the remaining 14 SMSAs were used to form another group. The second group consisted of urban EDs in DOs surrounding the DOs in group 1 and urban EDs of centralized DOs not assigned to group 1 or the 14 SMSAs previously mentioned. This group was meant to consist of the suburban population. The third group consisted of non-urban EDs in DOs in group 1, non-urban EDs in the 14 SMSAs previously mentioned and non-urban EDs of DOs in group 2. The fourth group consisted of urban EDs of DOs in the 14 SMSAs and the urban EDs in all remaining decentralized DOs. This group consists of areas of mail coverage but not associated with large metropolitan areas of high percent minority. The fifth group consisted of non-urban EDs of the remaining decentralized DOs and all EDs in conventional DOs. In syn 1 there are 90 adjustment strata.

The adjustment strata for syn 2 were constructed by census division (9 of them) and within each division by size of place and race. Five size of place categories and three "race" categories were used. The size of place categories were 1) central cities in SMSAs with population 250,000+ 2) central cities in SMSAs with population less than 250,000 3) population in a SMSA but not in a central city 4) population in cities 10,000 to 50,000 but not in a SMSA and 5) rural areas with population less than 10,000. Because of size and distributional variation, "race" groups were sometimes combined and this resulted in a total of 96 adjustment strata for syn 2. For example, in the New England division, Hispanics and Blacks were combined. In summary, syn 2 emphasized divisional differences by race and size of place. Syn 1 emphasized urban/rural differences by age-race-sex and syn DA used more detailed age categories in an age-race-sex stratification. A detailed description of strata definitions for both syn 1 and syn 2 is presented in the Appendix. Table 1 below summarizes the adjustment strata definitions.

Table 1. Brief Description of Adjustment Strata for Syn 1, Syn 2 and Syn DA

Synthetic Est.	Geographic Level Detail	Person Detail
Syn 1	Group 1. Urban EDs in centralized DOs with population Hispanic or Black exceeding 25% in 35 of 49 largest SMSAs. Group 2. Urban EDs in DOs surrounding DOs in group 1, urban EDs in centralized DOs not assigned to group 1 or the remaining 14 SMSAs. Group 3. Non-urban EDs in DOs in group 1, non-urban EDs in the 14 SMSAs and non-urban EDs of DOs in group 2. Group 4. Urban EDs in the 14 SMSAs and urban EDs in all remaining decentralized DOs. Group 5. Non-urban EDs of remaining decentralized DOs and all EDs in conventional DOs.	Within each of the five groups an adjustment factor is computed for 2 sex by 3 race (Black, Hispanic, Rest) by 3 age groups (0-14, 15-44, 45+) (90 factors)
Syn 2	Within each of the nine census geographic divisions size of place categories (central cities 50,000-250,000/250,000+ in SMSAs, places in SMSA not in central city, cities 10,000-50,000, places 0-10,000) are used. Some place categories are sometimes combined. Occassionally, large cities are treated individually.	Three race (Black, Hispanic, Rest) are used within place groupings. Depending on the place groupings race categories are sometimes combined (96 factors).
Syn DA	U.S. level	2 sex by 3 race (Black, Hispanic, Rest) by 5 age groups (0-14, 15-29, 30-44, 45-64, 65+) (30 factors)

B.3. Description of State Results. We applied each of the three SSEs toward estimating total population and population by "race" for states and counties for each artificial population. The census counts by ED were adjusted and tabulated to the geographic level of interest (state, county or ED). Several summary measures were used in comparing the performances of the SSEs with that of the census. In defining the measures, c represents the census count, e represents the SSE and s represents the artificial population count used as the standard or truth.

The summary measures can be loosely categorized into three types. The first type involves counts of small areas possessing a certain characteristic. For example, the number of adjusted state estimates that are closer to the standard than the census state figures. The second type of measure involves error assessment of the absolute level of the adjustment estimates. Such measures are typified by the mean absolute relative error and the weighted squared relative error. The third type of measure involves error assessment of the proportionate shares derived from the adjustment estimates. Such measures are useful in assessing how well adjustment and the census perform in apportioning shares on the basis of population. The above classification of measures is not mutually exclusive but serves as a rough reminder of the different types of measures.

Summary Measures

1. Number of areas where $ARE(c)_i < ARE(e)_i$

where

$ARE(c)_i = |(c-s)/s|$

c = census count for the area

s = standard count for the area

2. Number of areas where $ADP(c)_i < ADP(e)_i$

where

$$P_i^c = c_i/\sum^N c_i,\ P_i^s = s_i/\sum^N s_i,\ P_i^e = e_i/\sum^N e_i$$

$$ADP(c)_i = |P_i^c - P_i^s|,\ \text{etc.}$$

3. Apportionment

The number of house seats erroneously apportioned on the basis of state estimates of total population using adjustment method e using the artificial population state figure as the truth.

4. $\text{MARE} = \frac{1}{N} \sum_i^N \left| \frac{e_i - s_i}{s_i} \right|$

5. Maximum ARE(e)

6. Median ARE(e)

7. Weighted squared relative error

$$\alpha = \sum_i^N s_i\ [(e_i - s_i) / s_i]^2$$

8. $\text{SADP} = \sum_i^N |P_i^c - P_i^s|$

9. $\text{PI} = \sum_i^N \text{IMPV}_i/M$

$$M = \sum_i^N s_i \qquad \text{IMPV}_i = \begin{cases} s_i & \text{if } |P_i^e - P_i^s| < |P_i^c - P_i^s| \\ 0 & \text{otherwise} \end{cases}$$

10. Weighted squared relative error differences

$$\phi = \sum_{i}^{N} s_i \left[\{(e_i - s_i)/s_i\} - \{(\Sigma e_i - \sum_{i}^{N} s_i)/\sum_{i}^{N} s_i\} \right]^2$$

$$= ([\sum_{i}^{N} e_i]^2/\sum_{i}^{N} s_i) \sum_{i}^{N} (P_i^{c} - P_i^{s})^2/P_i^{s}$$

In the above listing of measures, the first three are of type 1, the next 4 are of type 2 and the last 3 are of type 3. In addition to these measures a set of four criteria of accuracy mentioned in the National Research Council's monograph "Estimating Population and Income of Small Areas" are A) low absolute average error B) low average absolute relative error C) few extreme relative errors and D) absence of bias for subgroups. As criterion A and B are somewhat in contrast (large population areas tend to have errors that dominate A whereas in B the size effect is somewhat muted), the Bureau's primary concern is with criteria B, C and D. The 10 measures of goodness listed above include criterion B and in some respects criterion C. Criterion D, bias, is interpreted as not experiencing an excess of errors of one sign.

We first present the adjustment results for syn 1, syn 2, syn DA and the census for states. Table 2 presents the results, using AP1 as a standard, for total population, Black, Hispanic and the remaining category termed Rest. Succeeding tables provide the results using AP2 and AP3.

Table 2. Measures of Performance of Statistical Synthetic Estimators Compared to the Census at the State Level Using Artificial Population 1 by Total Population and Each of Three Race Groups

A. Total Population

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of states where ARE(c_i) < ARE(e_i)	9	4	7	-
2.	2 - No. of states where ADP(c_i) < ADP(e_i)	14	12	13	-
3.	3 - Apportionment	2	2	2	2
4.	4 - MARE	.0054	.0042	.0052	.0134
5.	5 - Max ARE	.0169	.0147	.0190	.0398
6.	6 - Median ARE	.0050	.0028	.0048	.0121
7.	7 - α	8336	4504	8533	55221
8.	8 - SADP	.0048	.0031	.0048	.0052
9.	9 - PI	.622	.830	.654	-
10.	10 - ϕ	8332	4501	8211	9735

B. Black Population

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of states where ARE(c_i) < ARE(e_i)	9	6	5	-
2.	2 - No. of states where ADP(c_i) < ADP(e_i)	16	17	20	-
3.	4 - MARE	.0084	.0073	.0083	.0208
4.	5 - Max ARE	.0265	.0216	.0267	.0501
5.	6 - Median ARE	.0074	.0068	.0078	.0197
6.	7 - α	2374	1978	2686	20506
7.	8 - SADP	.0077	.0067	.0079	.0079
8.	9 - PI	.501	.566	.362	-
9.	10 - ϕ	2370	1972	2494	2470

C. Hispanic Population

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of states where ARE(c_i) < ARE(e_i)	11	10	2	-
2.	2 - No. of states where ADP(c_i) < ADP(e_i)	15	15	16	-
3.	4 - MARE	.0082	.0071	.0098	.0158
4.	5 - Max ARE	.0429	.0371	.0628	.0668
5.	6 - Median ARE	.0062	.0059	.0072	.0125
6.	7 - α	1234	447	1722	8217
7.	8 - SADP	.0074	.0030	.0068	.0076
8.	9 - PI	.715	.918	.465	-
9.	10 - ϕ	1214	427	1238	1293

D. Rest Population

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of states where ARE(c_i) < ARE (e_i)	10	4	9	-
2.	2 - No. of states where ADP(c_i) < ADP(e_i)	25	19	29	-
3.	4 - MARE	.0054	.0041	.0054	.0123
4.	5 - Max ARE	.0210	.0193	.0271	.0367
5.	6 - Median ARE	.0042	.0028	.0046	.0111
6.	7 - α	6266	2926	6326	32814
7.	8 - SADP	.0045	.0029	.0045	.0045
8.	9 - PI	.477	.644	.430	-
9.	10 - ϕ	6266	2926	6255	6011

Table 3. Measures of Performance of Statistical Synthetic Estimators Compared to the Census at the State Level Using Artificial Population 2 by Total Population and Each of Three Race Groups

A. Total Population

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of states where ARE(c_i) < (e_i)	9	5	8	-
2.	2 - No. of states where ADP(c_i) < (e_i)	13	15	14	-
3.	3 - Apportionment	2	0	2	6
4.	4 - MARE	.0052	.0044	.0053	.0147
5.	5 - Max ARE	.0183	.0200	.0297	.0771
6.	6 - Median ARE	.0048	.0026	.0047	.0113
7.	7 - α	9074	6179	9925	77513
8.	8 - SADP	.0048	.0037	.0049	.0067
9.	9 - PI	.757	.703	.694	-
10.	10 - ϕ	9073	6179	9758	17368

B. Black Population

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of states where ARE(c_i) < (e_i)	10	4	9	-
2.	2 - No. of states where ADP(c_i) < (e_i)	19	22	18	-
3.	4 - MARE	.0225	.0199	.0218	.0524
4.	5 - Max ARE	.0680	.0606	.0610	.1183
5.	6 - Median ARE	.0197	.0154	.0190	.0502
6.	7 - α	14703	12783	15724	132871
7.	8 - SADP	.0184	.0167	.0189	.0188
8.	9 - PI	.494	.489	.457	-
9.	10 - ϕ	14700	12561	15617	14220

C. Hispanic Population

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of states where ARE(c_i) < ARE (e_i)	8	32	1	-
2.	2 - No. of states where ADP(c_i) < ADP(e_i)	14	32	11	-
3.	4 - MARE	.0053	.0141	.0088	.0107
4.	5 - Max ARE	.0308	.0394	.0466	.0486
5.	6 - Median ARE	.0031	.0104	.0064	.0083
6.	7 - α	575	1430	1935	3918
7.	8 - SADP	.0048	.0062	.0046	.0051
8.	9 - PI	.428	.146	.581	-
9.	10 - ϕ	559	1075	574	648

D. Rest Population

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of states where ARE(c_i) < ARE (e_i)	10	4	9	-
2.	2 - No. of states where ADP(c_i) < ADP(e_i)	25	18	25	-
3.	4 - MARE	.0041	.0031	.0041	.0093
4.	5 - Max ARE	.0164	.0157	.0205	.0293
5.	6 - Median ARE	.0035	.0021	.0035	.0082
6.	7 - α	3428	1606	3440	18198
7.	8 - SADP	.0034	.0021	.0033	.0034
8.	9 - PI	.468	.754	.485	-
9.	10 - ϕ	3428	1606	3440	3376

Table 4. Measures of Performance of Statistical Synthetic Estimators Compared to the Census at the State Level Using Artificial Population 3 by Total Population and Each of Three Race Groups

A. Total Population

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of states where ARE(c_i) < (e_i)	9	7	6	-
2.	2 - No. of states where ADP(c_i) < (e_i)	9	8	8	-
3.	3 - Apportionment	4	2	4	8
4.	4 - MARE	.0050	.0045	.0047	.0136
5.	5 - Max ARE	.0184	.0228	.0300	.0773
6.	6 - Median ARE	.0040	.0026	.0032	.0092
7.	7 - α	8979	5866	9344	82339
8.	8 - SADP	.0048	.0033	.0047	.0078
9.	9 - PI	.701	.872	.715	-
10.	10 - ϕ	8923	5810	9266	22048

B. Black Population

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of states where ARE(c_i) < ARE(e_i)	10	6	9	-
2.	2 - No. of states where ADP(c_i) < ADP(e_i)	19	16	18	-
3.	4 - MARE	.0225	.0172	.0218	.0524
4.	5 - Max ARE	.0680	.0484	.0610	.1183
5.	6 - Median ARE	.0197	.0167	.0190	.0502
6.	7 - α	14703	12096	15724	132871
7.	8 - SADP	.0184	.0160	.0189	.0188
8.	9 - PI	.494	.589	.457	-
9.	10 - ϕ	14700	12093	15617	14220

C. Hispanic Population

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of states where ARE(c_i) < ARE(e_i)	14	11	6	-
2.	2 - No. of states where ADP(c_i) < ADP(e_i)	15	14	15	-
3.	4 - MARE	.0266	.0195	.0204	.0422
4.	5 - Max ARE	.0947	.0877	.1240	.1599
5.	6 - Median ARE	.0220	.0128	.0145	.0327
6.	7 - α	8895	3890	9448	61741
7.	8 - SADP	.0194	.0083	.0190	.0193
8.	9 - PI	.542	.910	.433	-
9.	10 - ϕ	8893	3835	9031	8501

D. Rest Population

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of states where ARE(c_i) < ARE (e_i)	11	9	8	-
2.	2 - No. of states where ADP(c_i) < ADP(e_i)	23	21	23	-
3.	4 - MARE	.0026	.0023	.0024	.0055
4.	5 - Max ARE	.0105	.0104	.0139	.0195
5.	6 - Median ARE	.0021	.0016	.0020	.0049
6.	7 - α	1270	789	1187	6541
7.	8 - SADP	.0020	.0014	.0020	.0020
8.	9 - PI	.582	.708	.593	-
9.	10 - ϕ	1270	695	1187	1224

B.4. Discussion of State Results. The total populations of all three artificial populations were estimated by all three SSEs with smaller error than by the census according to the measures used. The change in apportionment results for AP2 and AP3 are striking. Syn 2 almost always indicated smaller error than syn 1 or syn DA. The measures of goodness for the three "race" groups exhibit larger differences and levels than those observed for the total population. Differences between the artificial populations affected the performance of the SSEs, especially syn 2. This is because syn 2 treated Blacks and Hispanics alike which favors its performance under AP3 but not AP2. For Blacks, syn 2 appears to perform better than the other SSEs. For Hispanics, syn 2 is again superior to the other SSEs for AP1 and AP3 but is inferior to the census as well as the other two SSEs under AP2. The summary measures for Rest showed the same patterns as total population except the level of error was lower. Syn 2 had a lower error than syn 1, syn DA and the census for all summary measures. Syn 1 and syn DA had lower errors than the census for absolute relative error and the α measure but showed no improvement over the census for the other measures.

We have not analyzed the results of the tables in detail. However, it has been pointed out that certain anomolous results exist in the tables with respect to measures 2, 9 and 10. These measures deal with estimated proportions. Consider Hispanics in Tables 3 and 4 under syn 1 and the census. In Table 3, measures 2 and 10 are in agreement but measure 9 is not. In Table 4, measures 2 and 9 are in agreement but measure 10 is not. Measures 2 and 9 are all or nothing type measures in that the same value for a state is assigned irrespective as to how close the adjusted proportion is to the standard. Measure 10, on the other hand, assigns values that are affected by the closeness of the proportions. All three measures are dominated by the states with large populations.

There are 15 states with Hispanic population exceeding 100,000. These 15 states account for 91.5% of the total Hispanic population. Using both AP2 and AP3, syn 1 adjustment performs better than the census in nine out of the fifteen. The particular composition of states differs however. So, while measure 2 remains the same (for these 15 states), measure 9, the PI measure differs (.428 for AP2 to .542 for AP3). The principal cause for the change in PI measure appears to be in the largest Hispanic state for which adjustment is inferior under AP2 and otherwise for AP3. The fourth and fifth largest Hispanic states affect measure 10 to the largest degree. Under AP2 measure 10 is 559 versus 648 for the census. It is the performance of adjustment for one of these states, and not the largest state, that affects measure 10. For this state, adjustment did much better than the census. Conversely, in AP3 adjustment performed much worse for the other of the two states causing the census measure 10 to be smaller than that for adjustment.

B.5. Description of County Results. In this section we present the summary measures for adjustment of total population of counties. The table below provides the census undercount rate for each artificial population by "race" group. Census undercount rates for each artificial population by state can be found in the Appendix.

Table 5. Census Undercount Rate for AP1, AP2 and AP3

	"Race"	AP1	AP2	AP3
1.	Total Population	.0142	.0163	.0163
2.	Black	.0263	.0652	.0652
3.	Hispanic	.0221	.0151	.0595
4.	Rest	.0119	.0089	.0054

For counties we used AP2 and AP3 for comparison purposes because both seemed likely to be closer to the 1980 census undercount rates by "race" than AP1. AP1 also lacked an age-sex differential that was observed via demographic analysis over several censuses. We omitted looking at population by "race" in Table 6 below because small population sizes tended to distort the summary measures.

Table 6. Measures of Performance of Statistical Synthetic Estimators Compared to the Census at the County Level (based on 3137 counties) Using Artificial Populations 2 and 3 for Total Population*

A. AP2

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of counties where ARE(c_i) < ARE(e_i)	1369	1219	1201	-
2.	2 - No. of counties where ADP(c_i) < ADP(e_i)	815	917	870	-
3.	4 - MARE	.0093	.0089	.0086	.0128
4.	5 - Max ARE	.2151	.2131	.2192	.2236
5.	6 - Median ARE	.0070	.0056	.0056	.0076
6.	7 - α	36842	31218	37825	115755
7.	8 - SADP	.0084	.0074	.0086	.0115
8.	9 - PI	.689	.736	.703	-
9.	10 - ϕ	36842	31218	37657	55525

B. AP3

	Measure No./Description	Syn 1	Syn 2	Syn DA	Census
1.	1 - No. of counties where ARE(c_i) < ARE(e_i)	1510	1325	1266	-
2.	2 - No. of counties where ADP(c_i) < ADP(e_i)	693	723	707	-
3.	4 - MARE	.0082	.0081	.0074	.0111
4.	5 - Max ARE	.2683	.2946	.2757	.3067
5.	6 - Median ARE	.0055	.0044	.0039	.0055
6.	7 - α	41773	34688	41508	134577
7.	8 - SADP	.0083	.0071	.0084	.0131
8.	9 - PI	.724	.783	.747	-
9.	10 - ϕ	41717	34633	41430	74347

*Measures utilizing proportions use the entire U.S. as a base as opposed to using relevant states as bases. Measures using states as bases will be computed in a separate report.

Unlike the state analysis we found syn DA to perform better than syn 1 and syn 2 for both AP2 and AP3. Syn DA had the smallest MARE and median ARE measures. While syn 2 had a smaller SADP measure the percent of counties in which the census was superior to syn 2 was not much different from syn 1 and syn DA. Likewise the PI measures were similar among all three SSEs.

The universe of counties were divided by population into three size groups 0 to 10,000; between 10,000 and 50,000; and those exceeding 50,000 with 25%, 50% and 25% of the counties. Each of the three groups were looked at separately. This analysis indicated that syn DA fared well for the smaller population size, syn 2 fared well for the larger population size and for the middle group various SSEs fared well on some of the measures.

While syn 2 was the better of the 3 SSEs for states and syn DA for counties, the observation that syn 2 was also superior for large counties suggests that there may not be a single statistical synthetic estimator satisfactory for all areas but that we may need to apply separate SSEs over portions of the universe of all areas. A second consideration is that of sampling error which is covered in section III that follows. Since in practice the adjustment factors need to be estimated the sampling error of the SSEs warrant consideration.

B.6 Description of Enumeration District (ED) Results. We were not able to complete an application of the measures of performance listed earlier on all of the approximately three hundred thousand EDs in the universe. We had intended, at a minimum, to investigate the adjustment of EDs by states for California, Mississippi and North Dakota (the three states containing 1986 and 1987 census test sites). Unfortunately, due to its size (25,000 EDs), California could not be used, as existing computer programs could not be modified in the time available to complete the report. Hence, to the extent that California would have provided EDs in large cities with large minority percentages, the results from the other two states provide an uneven picture of an ED level adjustment. The results for total population for syn 2 and syn DA are presented in Table 7 below.

Table 7. Measures of Performance of Statistical Synthetic Estimators Compared to the Census at the ED level for Mississippi (3595 EDs) and North Dakota (2536 EDs) Using Artificial Populations 2 and 3 for Total Population.

A.* AP2/Mississippi

	Measure No./Description	Syn 2	Syn DA	Census
1.	1 - No. of EDs where ARE(c_i) < ARE(e_i)	2215	1980	-
2.	2 - No. of EDs where ADP(c_i) < ADP(e_i)	1353	1389	-
3.	4 - MARE	.0245	.0261	.0173
4.	5 - Max ARE	1.0	1.0	1.0
5.	6 - Median ARE	.0142	.0134	.0005
6.	7 - α	5428	5968	5968
7.	8 - SADP	.0198	.0217	.0210
8.	9 - PI	.603	.562	-
9.	10 - ϕ	5320	5680	5249

B. AP3/Mississippi

	Measure No./Description	Syn 2	Syn DA	Census
1.	1 - No. of EDs where ARE(c_i) < ARE(e_i)	2189	1966	-
2.	2 - No. of EDs where ADP(c_i) < ADP(e_i)	1278	1384	-
3.	4 - MARE	.0219	.0240	.0151
4.	5 - Max ARE	1.0	1.0	1.0
5.	6 - Median ARE	.0113	.0119	0
6.	7 - α	4674	5136	5104
7.	8 - SADP	.0177	.0208	.0192
8.	9 - PI	.616	.560	-
9.	10 - ϕ	4566	4889	4530

C. AP2/North Dakota

	Measure No./Description	Syn 2	Syn DA	Census
1.	1 - No. of EDs where ARE(c_i) < ARE(e_i)	1190	778	-
2.	2 - No. of EDs where ADP(c_i) < ADP(e_i)	2180	1346	-
3.	4 - MARE	.0042	.0036	.0016
4.	5 - Max ARE	.4317	.4372	.4372
5.	6 - Median ARE	0	.0024	0
6.	7 - α	175	157	152
7.	8 - SADP	.0038	.0029	.0033
8.	9 - PI	.351	.19	-
9.	10 - ϕ	152	160	150

D. AP3/North Dakota

	Measure No./Description	Syn 2	Syn DA	Census
1.	1 - No. of EDs where ARE(c_i) < ARE(e_i)	805	610	-
2.	2 - No. of EDs where ADP(c_i) < ADP(e_i)	2245	2331	-
3.	4 - MARE	.0021	.0021	.0010
4.	5 - Max ARE	.3087	.3087	.3087
5.	6 - Median ARE	0	0	0
6.	7 - α	56	66	56
7.	8 - SADP	.0024	.0033	.0019
8.	9 - PI	.334	.201	-
9.	10 - ϕ	56	58	53

Our overall impression of adjustment of EDs for undercount is that for the two states considered, adjustment was inferior to the census. For both AP2 and AP3 and for almost all measures, the census performed better than the two adjustment methods. The undercount rates for Mississippi were .0169 and .0148 for AP2 and AP3, respectively. For North Dakota, the undercount rates were .0020 and .0012 for AP2 and AP3, respectively. A complete assessment of the ED level adjustments needs to be conducted. In this area, as in all other areas of small area research, we have attempted to provide as much illustrative material (estimates) as possible with the intention of conducting detailed analyses at a future time.

In section II we have presented model error. In section III we also include the sampling error effect.

III. Statistical Synthetic Estimation - Sampling Error

A. Background

Since syn 2 and syn DA were found to be superior to syn 1 for at least some collection of areas, we dropped syn 1 from further analysis. In order to examine the sampling error effect we devised a simple sample design using EDs as the sampling unit even though it is likely that a smaller unit such as a census block is likely to be used in 1990. The ED was the smallest geographical unit on our data file. For each ED, counts by race-age-sex were available for the 1980 census and our artificial population variables AP2 and AP3. Details concerning the simple sample design can be found in Huang (1986). We briefly summarize the general design and the simulation procedure.

The sample design was constructed to support estimation of the 96 adjustment factors of syn 2. In this respect, the universe of EDs was stratified along adjustment factor definitions. The sample number of EDs was set at 1440. This number was determined by assuming that an ED contained on average seven blocks and hence approximates a 10,000 block sample design that had been suggested as a rough sample size for a PES in 1990. Sample sizes of EDs were allocated proportionally to the population of the sampling stratum. The sampling weights were approximately 200. The EDs were assigned to sampling strata on the basis of geography and 1980 census percent minority category. Because of this, some sample estimated adjustment factors within census division are correlated but they are never correlated between divisions. Due to the limitations of the computer, 90 replicates were selected, each containing 1440 EDs. Each replicate represents a potential sample realization; the replicates were obtained via equal probability systematic sampling. From each replicate a set of 96 adjustment factors for syn 2 and a set of 30 adjustment factors for syn DA were computed. These were used to compute covariance matrices for each set of adjustment factors.

We chose one of the 90 replicates and computed the summary measures presented earlier on total population for AP2 and AP3 for both syn 2 and syn DA when states and counties are of interest. The results are presented in the tables that follow. The covariance matrix will be used to compute mean square errors of the SSEs as well as to study regression methods in small area estimation. Because syn 2 requires more parameters to be estimated its summary measures are expected to be affected to a larger degree than those for syn DA. When viewed in the 1990 context some inefficiency in the sample design used e.g., EDs versus blocks, is balanced by the fact that we used the current 1980 census data which for 1990 will not be available. The net effect of this balancing of conditions is not known.

B. Description of State Results

It is instructive to view the results of the summary measures for both syn 2 and syn DA used in state adjustments for a sample. In general, when compared to the previous tables, (3 and 4) syn 2's performance has diminished while syn DA's has remained about the same. Both remain superior to the census. On the basis of Table 8, one would be inclined to select syn DA. The results of a single sample only are provided in Table 8 because of prohibitive costs.

Table 8. Measures of Performance of Statistical Synthetic Estimators Compared to the Census at the State Level for Total Population Using Artificial Populations 2 and 3 for a Single Replicate

A. AP 2

	Measure No./Description	Syn 2	Syn DA	Census
1.	1 - No. of states where $ARE(c_i) < ARE(e_i)$	6	8	-
2.	2 - No. of states where $ADP(c_i) < ADP(e_i)$	20	13	-
3.	3 - Apportionment	2	2	6
4.	4 - MARE	.0060	.0053	.0147
5.	5 - Max ARE	.0218	.0288	.0771
6.	6 - Median ARE	.0039	.0048	.0113
7.	7 - α	12189	9282	77316
8.	8 - SADP	.0056	.0048	.0067
9.	9 - PI	.481	.757	-
10.	10 - ϕ	11985	9282	17391

B. AP 3

	Measure No./Description	Syn 2	Syn DA	Census
1.	1 - No. of states where $ARE(c_i) < ARE(e_i)$	8	8	-
2.	2 - No. of states where $ADP(c_i) < ADP(e_i)$	17	9	-
3.	3 - Apportionment	4	4	8
4.	4 - MARE	.0060	.0049	.0136
5.	5 - Max ARE	.0362	.0290	.0773
6.	6 - Median ARE	.0038	.0033	.0092
7.	7 - α	19227	9180	82365
8.	8 - SADP	.0068	.0046	.0078
9.	9 - PI	.635	.703	-
10.	10 - ϕ	18968	9129	22032

C. Description of County Results.

As in the case of states the county sample based adjustments were also summarized. The results are in Table 9. These results for counties are similar to those for states in that syn DA is better than syn 2 or the census. The differences of the performance measures are due to the effects of sampling. These effects can be minimized in a number of ways. Remaining within the constructed adjustment factor domains, one way is to use a more efficient sampling procedure. Another way is to increase the sample size while a third is to construct estimators of the adjustment factors with smaller variance. This latter possibility involves smoothing of the estimated adjustment factors by way of model assumptions.

Table 9. Measures of Performance of Statistical Synthetic Estimators Compared to the Census at the County Level (3137) for Total Population Using Artificial Populations 2 and 3 for a Single Replicate*

A. AP2

	Measure No./Description	Syn 2	Syn DA	Census
1.	1 - No. of counties where ARE(c_i) < ARE(e_i)	1104	1254	-
2.	2 - No. of counties where ADP(c_i) < ADP(e_i)	999	862	-
3.	4 - MARE	.0092	.0087	.0128
4.	5 - Max ARE	.2200	.2192	.2236
5.	6 - Median ARE	.0052	.0058	.0076
6.	7 - α	44859	36703	115755
7.	8 - SADP	.0093	.0085	.0115
8.	9 - PI	.625	.702	-
9.	10 - ϕ	44515	36703	55525

B. AP3

	Measure No./Description	Syn 2	Syn DA	Census
1.	1 - No. of counties where ARE(c_i) < ARE(e_i)	1122	1358	-
2.	2 - No. of counties where ADP(c_i) < ADP(e_i)	821	702	-
3.	4 - MARE	.0081	.0077	.0111
4.	5 - Max ARE	.3007	.2720	.3067
5.	6 - Median ARE	.0042	.0044	.0055
6.	7 - α	61485	41095	134577
7.	8 - SADP	.0098	.0084	.0131
8.	9 - PI	.680	.743	-
9.	10 - ϕ	61172	41045	74347

*Measures utilizing proportions use the entire U.S. as a base as opposed to using relevant states as bases. Measures using states as bases will be computed in a separate report.

IV. References

1. Huang, E.T. (1986). "Survey Based Estimates of Census Adjustment Factors and Their Variances and Covariances - A Monte Carlo Study," unpublished memorandum, Bureau of the Census.

2. Isaki, C.T., Schultz, L.K., Smith, P.J. and Diffendal, G.J. (1985). "Small Area Estimation Research for Census Undercount - Progress Report," paper presented at the International Symposium on Small Area Statistics, May 22-24, Ottawa, Canada, 29 pgs.

3.* Isaki, C.T., Diffendal, G.J. and Schultz, L.K. (1986). "Statistical Synthetic Estimates of Undercount for Small Areas," paper presented at the Bureau of the Census' Second Annual Research Conference, March 23-26, Reston, Virginia, 30 pgs.

4. National Research Council (1985). "The Bicentennial Census - New Directions for Methodology in 1990," report of the Panel on Decennial Census Methodology, Natinal Academy Press, Washington, D.C.

5. Schultz, L.K., Huang, E.T., Diffendal, G.J. and Isaki, C.T. (1986). "Some Effects of Statistical Synthetic Estimation on Census Undercount of Small Areas," paper presented at the 1986 Annual Meetings of the American Statistical Association, Chicago, IL.

6. Tukey, J.W. (1981). Discussion of "Issues in Adjusting the 1980 Census Undercount," by Barbara Bailar and Nathan Keyfitz, paper presented at the Annual Meeting of the American Statistical Association, Detroit, MI.

7. Tukey, J.W. (1984). "Points to be Made," presented to the Committee on National Statistics - Panel on Decennial Census Methodology, 6 pgs., Washington, D.C.

8. Tukey, J.W. (1986). "Discussant's Remarks at the Census Bureau's Second Annual Research Conference - session titled 'Deriving Appropriate Standards for Census Undercount Adjustment', 3 pgs. Reston VA.

V. Appendix

A. Adjustment Strata for Syn 2

In an attempt to evaluate the errors in the statistical synthetic method for the undercount, adjustment strata need to be defined. Because of limitations of conducting a coverage measurement survey, only 50-200 strata would be possible for 1990. Listed below are a set of 96 strata that will be used to evaluate the statistical synthetic estimation procedure termed syn 2.

The strata are based on 3 variables that are believed to highly influence the undercount: geography, size of place, and race. Geography used here is nine groupings of states that follow the nine census divisions of the country. Size of place has been shown to differentiate the undercount. Large central cities and rural areas generally have high undercounts and suburban areas have low undercounts. Race has been the most analyzed undercount characteristic due to the ease of obtaining data for undercount estimation. Blacks have a very high undercount and whites have a low undercount. The data on hispanics is much less than for blacks and whites, but is believed to be about equal to the blacks. White is defined as nonblack, nonhispanic and nonwhite is defined as blacks and hispanics.

Each race category listed on a separate line is a strata. For example for the New England strata, the first strata is whites living in the central city of 50,000 people or more. The second strata is whites living in a SMSA but not in a central city. The total number of strata is listed for each geography grouping.

Tabulation Strata

1. New England - MA, ME, VT, NH, CT, RI

   Total number of strata = 6

   a) Central Cities 50,000 +
      White

   b) In SMSA, not in Central City
      White

   c) a and b
      Nonwhite

   d) Cities 10,000 - 50,000
      White

   e) Rural 0 - 10,000
      White

   f) d and e
      Nonwhite

2. NY, NJ, PA

Total number of strata = 15

a) New York City
   Black
   Hispanic
   Nonblack, Nonhispanic

b) Central Cities 250,000 +
   Black
   Nonblack, Nonhispanic

c) Central Cities 50,000 - 250,000
   Black
   Nonblack, Nonhispanic

d) b and c
   Hispanic

e) In NY City SMSA, not in Central City
   Nonblack, Nonhispanic

f) In SMSA, not in Central City 250,000 + (except NY SMSA)
   Nonblack, Nonhispanic

g) e and f
   Black and Hispanic

h) In SMSA, Not in Central City 50,000 - 250,000
   Nonblack, Nonhispanic

i) Cities 10,000 - 50,000
   Nonblack, Nonhispanic

j) Rural 0 - 10,000
   Nonblack,Nonhispanic

k) h, i, and j
   Black and Hispanic

3. South - WV, VA, NC, SC, GA, FL, MD, DE, DC

Total number of strata = 15

a) Central Cities 250,000 +
   Black
   Nonblack, Nonhispanic

b) Central Cities 50,000 - 250,000
   Black
   Nonblack, Nonhispanic

c) a and b
Hispanic

d) In SMSA, not in Central City 250,000 +
Black
Nonblack, Nonhispanic

e) In SMSA, not in Central City 50,000 to 250,000
Black
Nonblack, Nonhispanic

f) d and e
Hispanic

g) Cities 10,000 - 50,000
Black
Nonblack, Nonhispanic

h) Rural 0 - 10,000
Black
Nonblack, Nonhispanic

i) g and h
Hispanic

4. KY, TN, AL, MS

Total number of strata = 7

a) Cental Cities 250,000 +
White

b) Central Cities 50,000 - 250,000
White

c) a and b
Nonwhite

d) In SMSA, not in Central City
White

e) Cities 10,000 - 50,000
White

f) Rural 0 - 10,000
White

g) d, e and f
Nonwhite

5. MI, OH, IN, IL

Total number of Strata = 12

a) Chicago and Detroit
Nonblack, Nonhispanic
Black

b) Central Cities 250,000 +
Nonblack, Nonhispanic
Black

c) a and b
Hispanic

d) Central Cities 50,000 - 250,000
White

e) In SMSA, not in Central City 250,000 +
White

f) In SMSA, not in Central City 50,000 - 250,000
White

g) d, e, and f
Nonwhite

h) Cities 10,000 - 50,000
White

i) Rural 0 - 10,000
White

j) h and i
Nonwhite

6. MN, WI, IA, MO, KS, NB, SD, ND

Total number of strata = 9

a) Central Cities 250,000 +
White
Nonwhite

b) Central Cities 50,000 - 250,000
White

c) In SMSA, not in Central City 250,000 +
White

d) b and c
Nonwhite

e) In SMSA, not in Central City 50,000 - 250,000
White

f) Cities 10,000 - 50,000
White

g) Rural 0 - 10,000
White

h) e, f and g,
Nonwhite

7. TX, OK, AR, LA

Total number of strata = 11

a) Houston and Dallas
Black
Hispanic
Nonblack, Nonhispanic

b) Central Cities 250,000 +
Nonblack, Nonhispanic

c) Central Cities 50,000 - 250,000
Nonblack, Nonhispanic

d) b and c
Black
Hispanic

e) In SMSA, not in Central City
Nonblack, Nonhispanic

f) Cities 10,000 - 50,000
Nonblack, Nonhispanic

g) Rural 0 - 10,000
Nonblack, Nonhispanic

h) e, f, and g
Black and Hispanic

8. NM, CO, WY, MT, ID, UT, AZ, NV

Total number of strata = 7

a) Central Cities 250,000 +
White

b) Central Cities 50,000 +
White

c) a and b
Nonwhite

d) In SMSA, not in Central City
White

e) City 10,000 - 50,000
White

f) d and e
Nonwhite

g) Rural 0 - 10,000
All races

9. CA, OR, WA, AK, HI

Total number of strata = 14

a) Los Angeles
Black
Hispanic
Nonblack, Nonhispanic

b) Central Cities 250,000 +
Nonblack, Nonhispanic

c) Central Cities 50,000 - 250,000
Nonblack, Nonhispanic

d) b and c
Black
Hispanic

e) In SMSA, not in Central City 250,000 +
Nonblack, Nonhispanic

f) In SMSA, not in Central City 50,000 - 250,000
Nonblack, Nonhispanic

g) e and f
Black
Hispanic

h) Cities 10,000 - 50,000
Nonblack, Nonhispanic

i) Rural 0 - 10,000
Nonblack, Nonhispanic

j) h and i
Black and Hispanic

B. Adjustment Strata for Syn 1

The rationale for stratum formation for Syn 1 was to separate population groups by age-race-sex and by geography. Emphasis was placed on grouping geographic areas regardless of political boundaries such as states and cities. The basic unit to be adjusted is the enumeration district (ED) which is a contiguous collection of blocks usually less than 1600 persons. Each ED was assigned to one of the 90 strata to be described below.

The 90 strata under Syn 1 were formed by constructing five areal groups that cover the entire U.S. Each areal group's population is further broken down by sex by three race groups (Black, Non-Black Hispanic, Rest) and by three age groups (0 to 14, 15-44, 45 plus).

The five areal groups were defined on the basis of ED's in district offices (DO's). The DO's were coded in the census as centralized (essentially covering cities), decentralized (remaining mail collection areas) and conventional (personal enumeration). For our purposes, the ED's were classified as urban (if its population was 70 percent urban or higher) and non-urban otherwise.

The first of the five areal groups consisted of all urban ED's in DO's listed under group 1. These DO's are associated with the 35 of 49 largest SMSA's (with respect to population) with percent minority (Black, non-Hispanic) population exceeding 25. The DO's used were essentially centralized with a few decentralized DO's also included. The DO's in the remaining 14 SMSA's were used in constructing a separate areal group.

The second areal group consisted of urban ED's of decentralized DO's surrounding the DO's listed in group 1 and urban ED's of centralized DO's not assigned to group 1 and not located in the 14 SMSA's previously mentioned. Group 2 was constructed to include the suburban areas of the 35 SMSA's. The DO's in this group are listed under group 2.

The third areal group consisted of non-urban ED's in DO's in group 1, non-urban ED's in the 14 SMSA's previously mentioned and the non-urban ED's of DO's in group 2. The fourth areal group consisted of urban ED's of DO's in the 14 SMSA's and the urban ED's of DO's in all remaining decentralized DO's. This group consists of areas of mail coverage but not associated with large metropolitan areas of high minority percent. The DO's are listed under group 4.

The fifth group consisted of non-urban ED's of remaining decentralized DO's and all ED's in conventional DO's. Within each areal group, 18 age-race-sex factors are used in the statistical synthetic 1 estimation method.

Five Groups of Areas for Syn 1

Group 1 Urbanized ED's in 35 of 49 largest SMSA's with % minority greater than or equal to 25. Definition of areas by ED's located in DO's listed.

a. New York
2240-2256

b. Los Angeles
3240-3244

c. Chicago
2540-2549
2551

d. Philadelphia
2340-2344
2346

e. Detroit
2440-2442

f. San Fran-Oakland
3245-3248

g. DC-MD-VA
2841-2842

h. Dallas-Ft. Worth
3040

i. Houston
3041

j. Boston
2140-2142

k. St. Louis
2550
2641

l. Baltimore
2348-2349

m. Atlanta
2940

n. Newark
2257-2260

o. Cleveland
2444-2445

p. Miami
2942

q. Denver
3140

r. Pittsburgh
2345

s. Cincinnati
2448

t. Milwaukee
2642

u. Kansas City
2640

v. San Jose
3221

w. Buffalo
2147

x. New Orleans
3042-3043

y. San Antonio
3013

z. Ft. Lauderdale
2919

a1. Sacramento
3227

a2. Rochester
2120

a3. Memphis
2941

a4. Louisville
2553

a5. Birmingham
2925

a6. Dayton
2447

a7. Greensboro
2805

a8. Norfolk
2815
2840

Group 2 Urbanized ED's of decentralized DO's surrounding group 1 areas

a. New York
2201
2202
2203

b. Los Angeles
3201
3202
3203
3204
3205
3206
3207
3208

c. Chicago
2502
2503
2506

d. Philadelphia
2302
2303
2316
2318

e. Detroit
2401
2402
2403
2404
2405

f. San Fran-Oakland
3222
3223
3225

g. DC-MD-VA
2821
2822
2325
2326

h. Dallas-Ft. Worth
3001-3003

i. Houston
3015-3017

j. Boston
2101
2102
2105
2106

k. St. Louis
2510
2607
2608

l. Baltimore
2323
2324
2327

m. Atlanta
2901
2902

n. Newark
2212
2213

o. Cleveland
2415-2416
2418

p. Miami
2920-2921

q. Denver
3101-3103

r. Tampa-St. Pete
2915-2916

s. Pittsburgh
2309-2310

t. Cincinnati
2424
2522

u. Milwaukee
2624-2625

v. Kansas City
2601
2604

w. San Jose
3224
3220

x. Buffalo
2118-2119

y. New Orleans
3021
3022

z. San Antonio
3014

a1. Ft. Lauderdale
2918

a2. Sacramento
3228

a3. Rochester
2121

a4. Memphis
2934

a5. Louisville
2518-2519

a6. Dayton
2423

a7. Greensboro
2804
2806

a8. Norfolk
2816-2817

Additional centralized offices
2144-2146
2347
2443
3142

Group 3 Non-urbanized ED's of decentralized DO's surrounding group 1 areas and of Group 1 areas and of 14 of 49 largest SMSA's

Note: Group 1 areas are areas covered by the DO's listed in the group 1 definition.

The decentralized DO's surrounding group 1 areas are listed in the group 2 definition. (Use all DO's listed in the group 2 definition)

The area covered by the 14 of 49 largest SMSA's is provided in the group 4 definition (DO numbers listed).

<u>Group 4</u> Urbanized ED's of 14 of 49 largest SMSA's <u>and</u> in remaining decentralized DO's.

a. Minneapolis
2619-2620

b. San Diego
3212-3214

c. Seattle
2701-2704

d. Riverside
3215-3216

e. Phoenix
3141
3107-3108

f. Portland
2708-2709

g. Indianapolis
2552

h. Columbus
2446
2422

i. Salt Lake
2715

j. Providence
2143
2109

k. Nashville
2931

l. Albany
2114

m. Anaheim
3210

n. Oklahoma City
3114

Note: The urbanized ED's of remaining decentralized DO's refers to those decentralized DO's whose DO numbers do not appear in groups 1, 2 & 4.

Group 5 Non-urbanized ED's of remaining decentralized DO's and all ED's in conventional DO's.

Note: In group 5, all ED's in conventional DO's are used here. In addition, non-urbanized ED's in decentralized DO's not elsewhere specified are also used here.

C. The Statisticsl Synthetic Estimator

Let

$f_{i,\alpha}$ be the true number of category i persons in area α divided by the census count of the number of such persons in area α.

Let

$C_{i,\beta}$ be the census count of persons in population category i and geographic area β where area β is contained in the union of several α areas.

Then, a statistical synthetic estimator of total population for area β is given by Y_β, where

$$Y_\beta = \sum_{\substack{i,\alpha \\ \alpha \supset \beta}} f_{i,\alpha} \, C_{i,\beta}$$

and where the summation is over all categories i and areas α containing area β.

D. Census State Undercount Rates for Total Populations for AP1, AP2 and AP3

Table 10. State Census Undercount Rates for AP1, AP2 and AP3

State	AP1	AP2	AP3
Alabama	.0167	.0220	.0189
Alaska	.0177	.0158	.0107
Arizona	.0147	.0115	.0144
Arkansas	.0109	.0113	.0088
California	.0137	.0136	.0178
Colorado	.0137	.0119	.0121
Connecticut	.0075	.0078	.0072
Delaware	.0066	.0079	.0068
District of Columbia	.0398	.0771	.0773
Florida	.0304	.0344	.0352
Georgia	.0205	.0278	.0239
Hawaii	.0089	.0075	.0061
Idaho	.0044	.0034	.0027
Illinois	.0217	.0277	.0301
Indiana	.0124	.0111	.0079
Iowa	.0034	.0026	.0018
Kansas	.0121	.0104	.0078
Kentucky	.0124	.0103	.0067
Louisiana	.0243	.0302	.0257
Maine	.0067	.0052	.0035
Maryland	.0122	.0185	.0166
Massachusetts	.0104	.0103	.0092
Michigan	.0071	.0100	.0091
Minnesota	.0054	.0044	.0031
Mississippi	.0121	.0169	.0148
Missouri	.0118	.0133	.0108
Montana	.0122	.0093	.0058
Nebraska	.0040	.0034	.0024
Nevada	.0258	.0241	.0209
New Hampshire	.0112	.0086	.0054
New Jersey	.0117	.0170	.0179
New Mexico	.0301	.0229	.0453
New York	.0121	.0156	.0174
North Carolina	.0142	.0178	.0149
North Dakota	.0027	.0020	.0012
Ohio	.0078	.0093	.0077
Oklahoma	.0189	.0174	.0130
Oregon	.0087	.0069	.0048
Pennsylvania	.0112	.0129	.0109
Rhode Island	.0100	.0084	.0065
South Carolina	.0277	.0383	.0335
South Dakota	.0059	.0042	.0025
Tennessee	.0229	.0268	.0220
Texas	.0209	.0216	.0296
Utah	.0028	.0019	.0013
Vermont	.0057	.0043	.0028
Virginia	.0099	.0125	.0106
Washington	.0125	.0105	.0079
West Virginia	.0206	.0163	.0105
Wisconsin	.0063	.0062	.0050
Wyoming	.0112	.0093	.0071