BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-87/03

REPORT ON DEMOGRAPHIC ANALYSIS SYNTHETIC ESTIMATION FOR SMALL AREAS

by

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

Recommended: Kirk M. Wolter

Report completed: January 29, 1987

Report issued: January 29, 1987

Report on Demographic Analysis Synthetic Estimation for Small Areas

by

C. Isaki, L. Schultz

I. Introduction and Executive Summary

A. This report summarizes the work of the Census Undercount Adjustment for Small Area group as it pertains to Demographic Analysis (DA) Synthetic estimation. The group presented and described its activity at several conferences and documented its work in several memoranda and in papers. The papers include comparisons of the performance of the Demographic Analysis synthetic estimator with other estimators. In the present report however, we restrict discussion to the DA synthetic estimator and comparisons to the Census using various "measures of improvement" and "truths". The "measures of improvement" include mean absolute relative error as well as other indicators of performance such as errors in apportionment when compared to the "truth". The sources used to represent "truth" were the results of the 1980 Post Enumeration Program (PEP) as well as constructed populations based on the 1980 Census. It was necessary to hypothesize what was meant by "truth" in order to obtain a standard for comparison.

B. In general a synthetic estimator, $\hat{N}_s$, of a population total, N_s, for area s as applied to Census adjustment for undercount is of the form

$$\hat{N}_s = \sum_{\alpha} F_{\alpha} C_{s\alpha} \qquad \text{where} \qquad (1)$$

α denotes categories of persons as available in the census

$C_{s\alpha}$ denotes the census count of persons in the α^{th} category in area s and

F_{α} denotes an adjustment factor that represents the ratio of actual persons to census counted persons in the α^{th} category.

The DA synthetic estimator, $\hat{N}_{DAS}$, for area s is distinguished from other synthetic estimators by the use of Demographic Analysis numbers in the numerator of F_{α} and by the restriction of the α categories to age-race-sex cells available from Demographic Analysis at the U.S. level. In some parts of the research, the definition of the DA synthetic estimator was slightly broadened to include an additional Hispanic "race" component to the usual Black/Nonblack race groups available from Demographic Analysis. The number of 5 year interval age groups is 18 and together with sex resulted in either 72 or 108 age-race-sex categories (depending on whether Hispanic is used).

C. Previous Work

Several papers on the DA synthetic estimator and its use exists in the literature. Hill (1980) applied the method to 1970 Census counts to obtain adjusted state data for total population and for Blacks. He introduced matching methods, demographic methods and imputations as other techniques for estimating undercount. On the basis of such criteria as internal consistency, simplicity, timeliness, flexibility, equity and reliability, he concluded that DA synthetic estimation at the state and local area level is a viable procedure. Schirm and Preston (1984) analyzed the DA synthetic estimator (using

two α categories, Black/Nonblack) of total population by state. They introduced several measures for comparing the DA synthetic estimator and the census with respect to truth (the actual population count). Unlike Hill, who had no standard of comparison, Schirm and Preston created their standard by modelling the population and focussed on estimation of proportions. They assumed a relationship between truth and the census count via a stochastic variable and looked at several situations. They also investigated three scenarios related to assumptions concerning the errors in the national Demographic Analysis estimates. In scenario I, they assumed that the DA national figures were correct. In scenario II they assumed that the DA national figures were measured with constant error while in scenario III the figures were assumed stochastic with mean error zero. As a result of their simulation work based on models, Schirm and Preston found that adjustment results were sensitive to the evaluation measures used. An interesting observation made was that synthetic estimation will probably overcorrect the population proportions in states where the heavily undercounted group is a large part of the states' population and undercorrect for states where the group is a small fraction of the states' population. They concluded that adjustment for census undercount by the synthetic method was expected to improve the estimated proportions for states. Robinson and Siegel (1979) applied DA synthetic estimation to 1970 Census data to examine its effects on revenue sharing results specifically among states and among substate units within each of two states, Maryland and New Jersey. They found that the adjusted population figures had less of an effect than income measurement on fund allocation.

Members of the Small Area Estimation for Census Adjustment Research group have completed several papers, reports and memoranda on DA synthetic estimation. While some documents pertained solely to DA synthetic estimation, others combine and compare several adjustment methods. The references due to the efforts of various members of the group are 1) "Examples of Some Adjustment Methodologies Applied to the 1980 Census" by Diffendal, Isaki and Malec (1982), 2) "Small Area Adjustment Methods for Census Undercount" by Diffendal, Isaki and Schultz (1984), 3) "Synthetic State Estimation Using Demographic Analysis" by Isaki and Schultz (1984), 4) "Demographic Analysis State Synthetic Estimates Using Census Substitution" by Isaki (1985), 5) "Small Area Estimation Research for Census Undercount - Progress Report" by Isaki, Schultz, Smith and Diffendal (1985), 6) "Statistical Synthetic Estimates of Undercount for Small Areas" by Isaki, Diffendal and Schultz (1986).

We briefly summarize each reference as it pertains to DA synthetic estimation and provide a detailed presentation of the results in the body of this report. In reference 1) the 1980 DA estimates (legal population) were supplemented with an estimate of the illegal population under three separate assumptions of the size of the illegal population. The combination was treated as 1980 DA estimates of the total U.S. population and used in DA synthetic estimation. The ratios of DA synthetic estimates of total population to Census total population for counties indicated a consistent pattern for each of the three illegal population sizes. That is, the ratios increased by 1980 population size of county and within size of population of

county by 1980 Census proportion Black. A similar observation holds for DA synthetic estimates of states. Reference 2) includes two DA synthetic estimates of county total non-institutional populations ratio adjusted to agree with PEP 3-8 state estimates. Depending on the estimator, counties in the south or containing high percentage Black or Hispanic in the west yielded higher undercount.

The focus of references 3) and 4) is the comparison of the performance of three versions of the DA synthetic estimator of total population at the state level. Using the directly computed PEP 3-8 estimates as a standard the results of the analysis indicated that the correlation between each of the three estimated undercount rates with that obtainable from PEP 3-8 was under .45 while correlations among the DA synthetic estimates of undercount were approximately .90. Plots of each of the synthetically derived estimates of undercount versus PEP derived undercount indicated a weak relationship. Removing the District of Columbia, an apparent influential point, reduced the various correlations to under .35. Hence, given that the PEP 3-8 estimates are correct, there does not appear to be a strong association between the census undercounts as measured by the synthetic estimators and that measured by the PEP. According to some measures of improvement however, a DA synthetic estimator was determined to be closer than the unadjusted census counts to the PEP 3-8 figures.

The previous DA synthetic estimators distribute undercount according to the census distribution of total persons. An alternative is to

distribute the undercount using some other variable felt to be related to undercount and available for all small areas. One variable of interest, mail return rates, was a likely candidate but not readily available in a suitable form (by age-race-sex) so we used census substitutions as an alternative. The results were negative in that the resulting DA synthetic state estimates did not perform as well as the unadjusted census counts using PEP 3-8 as the standard.

Small area DA synthetic estimates, at the district office (DO) level, were examined in reference 5) and compared with the census. The directly computed PEP 3-8 district office estimates were used as a standard. At the DO level, three versions of the DA synthetic estimator did better than the census according to the "measures of improvement" used. Finally in reference 6) several artificial populations were constructed and used as a standard with which to compare the performance of the DA synthetic estimator of total (and by race) population for states and counties. In general, over all three artificial populations the DA synthetic estimator was superior to the unadjusted census counts for states for total population and in almost all cases by race. The DA synthetic was also superior to the unadjusted census counts for total population at the county level. In this last reference, the artificial population counts by age-race-sex at the U.S. level were treated as coming from Demographic Analysis methods. Hence, the adjustment factors in the DA synthetic estimator were measured without error. This is not the case in practice for a number of reasons, the chief reason being the presence of illegal aliens in the population to be counted.

D. The Demographic Analysis estimates are derived using U.S. birth and death records and also involve estimation of immigrants and emmigrants. The universe of coverage is the legal population. In 1980, the size of the illegal U.S. population was the topic of considerable debate. Estimates of the number of illegal aliens in the 1980 Census have been made, and if accurate, provide a lower bound to the size of the illegal population. Should the illegal population remain a problem in 1990, unmodified use of DA estimates for adjustment purposes is untenable. Our goal is adjustment of the census counts to reflect coverage of the total population.

Other sources of problems with the components of the DA estimates is the accuracy of measurement of the components. Birth and death counts are subject to misreporting, a source of error whose measurement was suggested as a research project. Our understanding is that emigration estimates are based on a model and in 1980, the 45 to 64 year age group was estimated on a model basis as well. We mention these recognized sources of error with regard to the DA estimates because in the following discussion, the DA estimates are taken as given and any modifications to them are treated as a part of the small area estimation method. For example, use of an Hispanic category in constructing a DA synthetic estimator is treated as development of a separate DA synthetic estimator and not as a part of the DA estimation process.

For comparing the performance of the various types of DA synthetic estimators, we utilized several measures of performance outlined in

the next section. Such measures speak only to the numerical results and not to other considerations of small area adjustment such as timing, cost or implementation. Such considerations are beyond the scope of this report.

E. Recommendations

The comparisons of the performance of several versions of a DA synthetic estimator with that of the census depend on the main assumption that illegal aliens can be adequately measured by age-race-sex at the U.S. level. Apart from other deficiencies such as birth under-registration, emigration and the 45-64 cohort modelling, the illegal alien size and distribution is likely to be the biggest source of error in the Demographic Analysis estimates. The analysis in sections III and IV both assumed types of information on population at the U.S. level not normally considered a part of Demographic Analysis. In section III, both the census and the DA synthetic estimators were ratio adjusted to equal that of PEP 3-8 in total population. In doing this, differences in state adjusted figures among those from the census, PEP 3-8 and the DA synthetic estimators are due solely to the manner of estimation and are not affected by differential total population counts. In section IV when the artificial populations are used, the census was not adjusted, however, the DA synthetic estimator was constructed using the actual artificial population age-race-sex totals at the U.S. level. Consequently, the comparisons in section IV illustrate a favorable scenario for DA synthetic.

Using PEP 3-8 as a standard or using the artificial populations, the DA synthetic estimator performed better than the census for total population of states according to the measures of improvement used. However, using the artificial populations and again at the state level but by race group, the measures of improvement relating to proportions suggest that the census performs better than the synthetic estimator. Again by race group, the measures of improvement relating to absolute relative error indicate that the DA synthetic is superior to the census. In contrast to the proportion related measures, these latter measures are likely to be affected to a higher degree by knowing the total count by race at the U.S. level. Knowing the total count by race depends chiefly on knowing the illegal alien count. Consequently, we cannot recommend that DA synthetic be used to adjust the census unless it can be established that accurate estimation of the illegals is possible.

II. Measures of Improvement

A. Several measures of improvement were used in comparing the performances of several versions of the DA synthetic estimator with that of the census. Each measure requires a DA synthetic estimate, census and standard figure for each small area of interest. The choice of a standard figure is pivotal in the research results that follow. Consequently, in section III we use the PEP 3-8 estimates as a standard while in section IV we use the artificial population counts as standard figures.

B. The measures of improvement 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 serve as a rough reminder of the different types of measures.

Measures

1. MR = ME/MC where

   MC = count of the number of times the census total c lies in the interval $s \pm \text{Var}(s)^{1/2}$ where s denotes the standard and Var(s) denotes its sampling variance (applicable when the PEP 3-8 estimate is used as the standard)

   ME = Same as MC except replace census by the adjustment figure e.

2. MRP = ME′/MC′ where

   MC′ = Sum of s of areas where c lies within $s \pm \text{Var}(s)^{1/2}$

   ME′ = Sum of s of areas where e lies within $s \pm \text{Var}(s)^{1/2}$

3. Number of areas where ARE(c) < ARE(e)

   where

   $$\text{ARE}(c) = |(c-s)/s|$$

4. No. of areas where $| P^c - P^s |_i < | P^e - P^s |_i$

   where $P^c = c_i / \sum_{i}^{N} c_i$ for the i-th area

   $$\text{ADP}(c) = | P^c - P^s |$$

5. Number of states erroneously apportioned

6. $$\text{MARE} = \frac{1}{N} \sum_{i}^{N} \left| \frac{e_i - s_i}{s_i} \right|$$

7. Maximum ARE(e)

8. Median ARE(e)

9. Weighted squared relative error

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

10. $$\text{PRSAE} = \sum_{i}^{N} \left| P_i^c - P_i^s \right| / \sum_{i}^{N} \left| P_i^e - P_i^s \right|$$

where $$P_i^c = \frac{c_i}{\sum_{i}^{N} c_i}$$, etc.

11. $$\text{PRSSAE} = \sum_{i}^{N} (P_i^c - P_i^s)^2 / \sum_{i}^{N} (P_i^e - P_i^s)^2$$

12. $$\text{PI} = \sum_{i}^{N} \text{IMPV}_i / M$$

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

13. Weighted squared relative error differences

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

In the above listing of measures, the first five are of type 1, the next 4 are of type 2 and the last 4 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 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 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 13 measures of improvement 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.

Because of the evolutionary nature of small area adjustment research not all adjustment methods introduced in the next sections have been subjected to every measure of improvement. Some measures were suggested for use upon our completion of certain phases of the work. In addition, some measures such as apportionment are not relavent when race groups are of interest.

III. Using PEP 3-8 Estimates As A Standard

A. All of the measures of improvement presented in the previous chapter require knowledge of the true population parameter (or a consistent estimate), be it a total or a proportion. In this chapter, we present the results of our comparisons among various DA synthetic estimators and the census using PEP 3-8 estimates as the truth. In some instances the population of interest is restricted to the non-institutional population as defined in Cowan and Bettin (1982). The obvious weakness in the comparisons is the assumption that PEP 3-8 estimates are close to the truth. The accuracy of the various

versions of the PEP estimates is the subject of considerable debate. In this section of the report, we take the qualified position that the PEP 3-8 estimates are indeed the truth or consistent estimates of the truth. The PEP series of estimates provide the only source of directly estimated sub-U.S. level undercount estimates. Our choice of PEP 3-8 estimates is entirely historic and does not imply an endorsement of it over the other versions. Our initial introduction to small area estimation and the PEP involved a preliminary estimate termed PEP 1-7 (before PEP clean-up cases were processed in 1982). PEP 1-7 was discarded (the clean-up was completed) and PEP 3-8 was suggested for continued use because it was most similar to PEP 1-7.

B. We proceed to compare the performance of three DA synthetic state estimates of total population (See Isaki and Schultz (1984) for details). The PEP 3-8 noninstitutional state estimates were augmented with an estimate of the state institutional estimates using a raking procedure. The three DA synthetic estimates differ in the way the Hispanic category is treated. We termed the three DA synthetic estimates as adjustment method I, II and III. In adjustment I, only two race/ethnicity categories are used in defining adjustment factors, Black and Nonblack. In adjustment II, three race/ethnicity categories are used. For Hispanic, the Black adjustment factors are used and the adjustment factors for the remaining category, termed Rest, is derived so as to maintain the Nonblack adjustment factors used in adjustment I. In adjustment III, the Hispanic adjustment factors are taken from the PEP 3-8. The PEP

3-8 non-institutional estimated Hispanic adjustment factors are used in a similar manner as in adjustment II. In all three adjustment methods the Black adjustment factors are the same. Adjustment III is not a DA synthetic estimate in its entirety because it assumes knowledge of the Hispanic adjustment factors through an outside source.
The following observations are made concerning the computed adjustment results and the 1980 census.

1. While all of the state total population estimates (including the PEP 3-8) are highly correlated and the three undercount estimates are highly correlated among themselves, they are not highly correlated with the PEP 3-8 measured undercount of the census. None of the latter three correlations exceeded .45.

2. In almost all states (41 of 51) the undercount estimates for the three adjustments were of the order I > III > II. For most of the remaining cases (8 of 11) the reverse order II > III > I occurred, possibly due their high percent Hispanic population together with lower adjustment factors used in adjustment III.

3. Applications of some of the measures presented in Section II are presented below for states and DOs. Note that each of the three adjustments and the census were ratio adjusted so that the total U.S. population was equal to that of PEP 3-8.

Table 1. Measures of Improvement of DA Synthetic State and DO Estimates of Total Population Using PEP 3-8 as a Standard

A. States

	Measure No.	Census	Adjustment I	Adjustment II	Adjustment III
1.	1 - MR[b]	-	1.000 (22)	1.250 (25)	1.250 (25)
2.	2 - MRP[b]	-	1.181	1.113	1.113
3.	6 - MARE[a]	.0124	.0119	.0110	.0112
4.	10 - PRSAE[b]	-	1.014	1.142	1.088
5.	11 - PRSSAE[b]	-	.955	1.520	1.285
6.	12 - PI[c]	-	.505	.707	.688

B. DO

	Measure No.	Census	Adjustment I	Adjustment II	Adjustment III
1.	1 - MR[b]	-	1.078 (236)	1.123 (246)	1.123 (246)
2.	2 - MRP[b]	-	1.083	1.137	1.139
3.	6 - MARE[a]	.0328	.0308	.0300	.0300
4.	10 - PRSAE[b]	-	1.050	1.068	1.067
5.	12 - PI[c]	-	.559	.573	.587

(a) The smallest number is considered best.

(b) Numbers greater than one indicate that the adjusted data are better.

(c) Numbers greater than .5 indicate that the adjusted data are better.

Figures in parentheses indicate the number of times the interval contains the synthetic adjustment figure.

A special re-tabulation of the PEP 3-8 non-institutional data was run at the district office (DO) level (Isaki, et. al. (1985)). The three DA synthetic estimates at the DO level were also computed. The results are presented in Table 1 for 414 of 422 DOs. The remaining eight DOs were omitted due to small PEP sample size.

Given the above results we conclude that adjustment II performs best among the three adjustment methods and is superior to the census in

the sense that the adjustments are closer to PEP 3-8 according to the measures of improvement. While the DO measures are not as impressive as that for states, the overall impression is that adjustments II or III are superior to the census.

IV. Using Artificial Populations as Standards

A. In this section we overcome the lack of a standard at very small levels of geography by constructing three artificial populations at the enumeration district level and compare the performance of a single DA synthetic estimator and the "census" at the state and county level (Isaki, et. al. (1986)). The data detail on the file limited the "DA" synthetic estimator to 5 rather than 18 age groups. The "DA" estimator also assumed the existence of Hispanic Demographic Analysis data rather than creating hypothesized ones as detailed in Section III. The results that follow necessarily assume the existence of DA U.S. level age-sex estimates for Hispanics. As in the case of the measures of improvement where their construction and application were chronological, the research on DA synthetic as applied to the artificial populations is not complete. If direct estimates of Hispanic by age-sex at the U.S. level can be obtained by, say a post enumeration survey (PES), then the results and discussion pertaining to DA synthetic are reasonable for artificial populations 1 and 2. If the undercount rates for Hispanics are like the Blacks then the results and discussion concerning artificial population 3 is reasonable. If the illegal population is basically Hispanic in nature, and the PES is accurate and timely, then the

results concerning the artificial populations in regard to DA synthetic estimation are relevant.

One adjustment (most similar to method III) was used for each artificial population. Quotes were used on DA to alert the reader that a simulated DA data set is being used. We omit the quotes in the following discussion. Assuming that the U.S. level age-race-sex data are a proxy for what would be expected via demographic analysis the comparisons between the DA synthetic estimate and the census are relative to the constructed artificial populations. The key variable used to construct the artificial populations is census substitutions. Census substitutions are the result of imputing people into housing units. For example, people were substituted into the census for closeout cases (no form was completed, but people may have lived in the housing unit), for machine failure (questionnaire destroyed or misread) and when field counts for the area (usually a block or an enumeration district) were larger than the processed counts. Preliminary analysis using 1980 PEP data at the state level 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, we focussed on census substitutions as a proxy for undercount.

The three artificial populations (denoted AP1, AP2, AP3) constructed by age-race-sex at the enumeration district (ED) level are:

i) AP1 = (census - substitutions) + substitutions

ii) AP2 = Census + F_{DA1} * substitutions

iii) AP3 = Census + F_{DA2} * substitutions

where F_{DA1} and F_{DA2} are defined below and only the non-institutional population is used in subsequent analysis.

AP1 treats the term 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 for the non-institutional population). The factor, F_{DA}, is the ratio of the difference between the demographic analysis derived total and the comparable census figure to the U.S. total of substitutions (by age-race-sex; F_{DA} = (DA-CEN/SUB)). Since demographic analysis estimates do not provide for an Hispanic category, AP2 and AP3 differ on the basis of how the Hispanic artificial population data are derived. For AP2 we assumed Hispanics are like the Nonblack population implying that the F_{DA}'s are the same for both groups, likewise for AP3 we assume Hispanics are similar to Blacks.

Focusing first on states as the small area of interest (and counties later on) we list the results of applying several measures on the DA synthetic estimator and the census using each of the artificial populations as the standard. Table 2 contains the results for total

population, Blacks, Nonblack Hispanic and the remaining race group termed Rest for AP1. Tables 3 and 4 contain the results for AP2 and AP3. The DA synthetic estimator used here assumes that U.S. level age-race-sex estimates are available for use in adjustment (race includes a separate Hispanic estimate). The U.S. level estimates used are exactly those artificial population totals previously described. Hence, the DA synthetic estimator used in our study assumes away two deficient properties of the 1980 DA estimates 1) availability of an Hispanic category and 2) the illegal alien population.

Looking at the three tables by characteristic we observe that the DA synthetic estimator is superior to the census for total population of states for all ten measures considered. Comparison of the DA synthetic estimator with that of the census by race groups provide some conflicting results. For Blacks in all three tables, the DA synthetic exhibited better measures than the census except for measures 12 (PI) and 13 (ϕ) and 10 (PRSAE), the first two measures, PI and ϕ, weight the performance by the true population of the state while all three deal with estimation of proportions. The Hispanic and Rest are also estimated better by DA synthetic except for measures 12, 13 and 4 where the census is sometimes superior. In general, it would appear that DA synthetic is superior to the census, at least for these three artificial populations at the state total population level.

Table 2. Measures of Improvement of DA Synthetic and the Census at the State Level for Total Population, Black, Hispanic and Rest for AP1

	Measure No./Description	Total Population DA	Total Population Census	Black DA	Black Census
1.	3 - No. of states where $ARE(C_i) < ARE(DA_i)$	7	-	5	-
2.	4 - No. of states where $ADP(C_i) < ADP(DA_i)$	13	-	20	-
3.	5 - Apportionment	2	2	-	-
4.	6 - MARE	.0052	.0134	.0083	.0208
5.	7 - Max ARE	.0190	.0398	.0267	.0501
6.	8 - Median ARE	.0048	.0121	.0078	.0197
7.	9 - α	8533	55221	2686	20506
8.	10 - PRSAE	1.092	-	1.006	-
9.	12 - PI	.654	-	.362	-
10.	13 - ϕ	8211	9735	2494	2470

	Measure No./Description	Hispanic DA	Hispanic Census	Rest DA	Rest Census
1.	3 - No. of states where $ARE(C_i) < ARE(DA_i)$	2	-	9	-
2.	4 - No. of states where $ADP(C_i) < ADP(DA_i)$	16	-	29	-
3.	5 - Apportionment				
4.	6 - MARE	.0098	.0158	.0054	.0123
5.	7 - Max ARE	.0628	.0668	.0271	.0367
6.	8 - Median ARE	.0072	.0125	.0046	.0111
7.	9 - α	1722	8217	6326	32814
8.	10 - PRSAE	1.114	-	.991	-
9.	12 - PI	.465	-	.430	-
10.	13 - ϕ	1238	1293	6255	6011

Table 3. Measures of Improvement of DA Synthetic and the Census at the State Level for Total Population, Black, Hispanic and Rest for AP2

		Total Population		Black	
	Measure No./Description	DA	Census	DA	Census
1.	3 - No. of states where $ARE(C_i) < ARE(DA_i)$	8	-	9	-
2.	4 - No. of states where $ADP(C_i) < ADP(DA_i)$	14	-	18	-
3.	5 - Apportionment	2	6	-	-
4.	6 - MARE	.0053	.0147	.0218	.0524
5.	7 - Max ARE	.0297	.0771	.0610	.1183
6.	8 - Median ARE	.0047	.0113	.0190	.0502
7.	9 - α	9925	77313	15724	132871
8.	10 - PRSAE	1.360	-	.995	-
9.	12 - PI	.694	-	.457	-
10.	13 - ϕ	9758	17368	15617	14220

		Hispanic		Rest	
	Measure No./Description	DA	Census	DA	Census
1.	3 - No. of states where $ARE(C_i) < ARE(DA_i)$	1	-	9	-
2.	4 - No. of states where $ADP(C_i) < ADP(DA_i)$	11	-	25	-
3.	5 - Apportionment	-	-	-	-
4.	6 - MARE	.0088	.0107	.0041	.0093
5.	7 - Max ARE	.0466	.0486	.0205	.0293
6.	8 - Median AE	.0064	.0083	.0035	.0082
7.	9 - α	1935	3918	3440	18198
8.	10 - PRSAE	1.112	-	1.012	-
9.	12 - PI	.581	-	.485	-
10.	13 - ϕ	574	648	3440	3376

Table 4. Measures of Improvement of DA Synthetic and the Census at the State Level for Total Population, Black, Hispanic and Rest for AP3.

		Total Population		Black	
	Measure No./Description	DA	Census	DA	Census
1.	3 - No. of states where $ARE(C_i) < ARE(DA_i)$	6	-	9	-
2.	4 - No. of states where $ADP(C_i) < ADP(DA_i)$	8	-	18	-
3.	5 - Apportionment	4	8	-	-
4.	6 - MARE	.0047	.0136	.0218	.0524
5.	7 - Max ARE	.0300	.0773	.0610	.1183
6.	8 - Median ARE	.0032	.0092	.0190	.0502
7.	9 - α	9344	82339	15724	132871
8.	10 - PRSAE	1.643	-	.995	-
9.	12 - PI	.715	-	.457	-
10.	13 - ϕ	9266	22048	15617	14220

		Hispanic		Rest	
	Measure No./Description	DA	Census	DA	Census
1.	3 - No. of states where $ARE(C_i) < (DA_i)$	6	-	8	-
2.	4 - No. of states where $ADP(C_i) < ADP(DA_i)$	15	-	23	-
3.	5 - Apportionment	-	-	-	-
4.	6 - MARE	.0204	.0422	.0024	.0055
5.	7 - Max ARE	.1240	.1599	.0139	.0195
6.	8 - MEDIAN ARE	.0145	.0327	.0020	.0049
7.	9 - α	9448	61741	1187	6541
8.	10 - PRSAE	1.014	-	1.01	-
9.	12 - PI	.433	-	.593	-
10.	13 - ϕ	9031	8501	1187	1224

Turning to estimation at the county level we present the results for the same DA synthetic estimator applied toward estimating total population for counties using artificial populations 2 and 3.

Table 5. Measures of Improvement of DA Synthetic and the Census at the County Level (3137 Counties) Using Artificial Populations 2 and 3 for Total Population

		AP2		AP3	
	Measure No./Description	DA	Census	DA	Census
1.	3 - No. of counties where $ARE(C_i) < ARE(DA_i)$	1201	-	1266	-
2.	4 - No. of counties where $ADP(C_i) < ADP(DA_i)$	870	-	707	-
3.	6 - MARE	.0086	.0128	.0074	.0111
4.	7 - Max ARE	.2192	.2236	.2757	.3067
5.	8 - Median ARE	.0056	.0076	.0039	.0055
6.	10 - PRSAE	1.326	-	1.550	-
7.	12 - PI	.703	-	.747	-

The results in Table 5 favor DA synthetic over the census in all respects.

B. In summary, the DA synthetic estimator assuming an Hispanic component performed better than the census at the state and county level for total population for all measures considered. At the state level but by race groups, the results were mixed. In all cases, the results for race groups indicated that DA synthetic was superior to the census for absolute relative error type measures. However, in almost all cases for measures dealing with proportions, the reverse was true. In the case of the Rest group this could possibly be explained by noting that for this group small adjustments for undercount are required. Hence adjusted proportions differ little from the census proportions. In comparing proportion estimation of the Rest group

with that of total population, the differing results are likely to be due to the race distributions among states.

While it is recognized that no Hispanic DA estimates are available, Table 4 for AP3 illustrates that if Black and Hispanic undercount rates are approximately equal, considerable reduction in absolute relative error (by at least one half) is possible. In the current Demographic Analysis estimate context, this assumes that almost all illegals are Hispanics and possess undercount rates equal to that of Blacks. If these assumptions are true, no speculation on the size of the illegal population is needed.

References

1. American Statistical Association (1984). "Report of the ASA Technical Panel on the Census Undercount," American Statistician, 38(4); 252-256.

2. Diffendal, Gregg J., Isaki, Cary T and Malec, Donald J. (1982). "Examples of Some Adjustment Methodologies Applied to the 1980 Census," technical report, U.S. Bureau of the Census, Washington, D.C.

3. Diffendal, Gregg J., Isaki, Cary T. and Malec, Donald J. (1983). "Some Small Area Adjustment Methodologies Applied to the 1980 Census," Proceedings of the Section on Survey Research Methods, Toronto, Canada: American Statistical Association, 164-7.

4. Diffendal, Gregg J., Isaki, Cary T. and Schultz, Linda K. (1984). "Small Area Adjustment Methods for Census Undercount," invited papers to the Data Users Conference on Small Area Statistics, U.S. Department of Human Services, Washington, D.C. 52-56.

5. Hill, Robert (1980). "The Synthetic Method: Its Feasibility for Deriving the Census Undercount for States and Local Areas," Conference on Census Undercount, U.S. Government Printing Office, Washington, D.C., 129-141.

6. Isaki, Cary T. and Schultz, Linda K. (1984). "Synthetic State Estimation Using Demographic Analysis," unpublished report, 23 pgs.

7. Isaki, Cary T. (1985). "Demographic Analysis State Synthetic Estimates Using Census Substitutions," unpublished report, 12 pgs.

8. Isaki, Cary T., Schultz, Linda K., Smith, Philip J. and Diffendal, Gregg 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.

9. Isaki, Cary T., Diffendal, Gregg J. and Schultz, Linda 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, Virginai, 30 pgs.

10. National Research Council (1980). "Estimating Population and Income of Small Areas," report of the Panel on Small-Area Estimates of Population and Income, National Academy Press, Washington, D.C.

11. National Research Council (1985). "The Bicentennial Census - New Directions for Methodology in 1980," report of the Panel on Decennial Census Methodology, Natinal Academy Press, Washington, D.C.

12. Passel, Jeffrey S. and Robinson, Gregory J. (1984). "Revised Estimates of the Coverage of the Population in the 1980 Census Based on Demographic Analysis: A Report on Work in Progress," paper presented at the Meetings of the American Statistical Association.

13. Passel, Jeffrey S. and Robinson, Gregory J. (1984). "Unpublished Tabulation, Population Division, U.S. Bureau of the Census.

14. Robinson, Gregory J. and Siegel, Jacob S. (1979). "Illustrative Assessment of the Impact of Census Underenumeration and Income Underreporting on Revenue Sharing Allocations at the Local Level," paper presented at the Meetings of the American Statistical Association.

15. Schirm, Allen L. and Preston, Samuel H. (1984). "Census Undercount Adjustment and the Quality of Geographic Population Distributions," technical report, University of Pennsylvania.

16. Tukey, John 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.

17. U.S. Bureau of the Census (1980). Proceedings of the Conference on Census Undercount, U.S. Government Printing Office, Washington, D.C.

18. U.S. Bureau of the Census (1982). "Coverage of the National Population in the 1980 Census by Age, Race, Sex," Current Population Reports, Series P-23, No. 115, U.S. Government Printing Office, Washington, D.C.

19. Warren, Robert (1981). "Estimation of the Size of the Illegal Alien Population in the United States," Agenda Item B of the meeting of the American Statistical Association - Census Bureau Advisory Committee, November.