


United States Nuclear Regulatory Commission Official Hearing Exhibit	
In the Matter of:	Entergy Nuclear Operations, Inc. (Indian Point Nuclear Generating Units 2 and 3)
	ASLBP #: 07-858-03-LR-BD01
	Docket #: 05000247   05000286
	Exhibit #: NYS000212-00-BD01
	Admitted: 10/15/2012
	Rejected:
Other:	Identified: 10/15/2012
	Withdrawn:
	Stricken:

NYS000212  
Submitted: December 16, 2011

**Estimation of population coverage in the 1990 United States**  
*Journal of the American Statistical Association*; Sep 1993; 88, 423; ProQuest  
pg. 1061

# Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis

J. GREGORY ROBINSON, BASHIR AHMED, PRITHWIS DAS GUPTA, and KAREN A. WOODROW\*

This article presents estimates of net coverage of the national population in the 1990 census, based on the method of demographic analysis. The general techniques of demographic analysis as an analytic tool for coverage measurement are discussed, including use of the demographic accounting equation, data components, and strengths and limitations of the method. Patterns of coverage displayed by the 1990 estimates are described, along with similarities or differences from comparable demographic estimates for previous censuses. The estimated undercount in the 1990 census was 4.7 million, or 1.85%. The undercount of males (2.8%) was higher than for females (.9%), and the undercount of Blacks (5.7%) exceeded the undercount of Non-Blacks (1.3%). Black adult males were estimated to have the highest rate of undercounting of all groups. Race-sex-age patterns of net coverage in the 1990 census were broadly similar to patterns in the 1980 and 1970 censuses. A final section presents the results of the first statistical assessment of the uncertainty in the demographic coverage estimates for 1990.

KEY WORDS: Coverage error; Demographic analysis; Undercount.

## 1. INTRODUCTION

The general method of demographic analysis as a tool for coverage evaluation is well developed and has been actively used at the Census Bureau to assess the completeness of coverage in every census since 1960. (See Siegel and Zelnik 1966; U.S. Bureau of the Census 1974; and U.S. Bureau of the Census 1988 for the basic demographic evaluations of the 1960, 1970, and 1980 censuses.) Demographic analysis estimates of coverage have become the benchmark by which national differences in coverage for age, sex, and race groups and changes in coverage over time are measured.

The purpose of the demographic analysis evaluation program for 1990 has been twofold: (1) to evaluate the completeness of coverage of population in the 1990 census based on demographic analysis, and (2) to develop a statistically based assessment of the accuracy of those demographic estimates of net coverage. This article reports the results of the demographic estimates of coverage for 1990 and the assessment of the accuracy of the estimates. An important by-product of the demographic program is the historical estimates of coverage provided for every census since 1940. The demographic estimates of net coverage for 1990 were also used to evaluate the overall quality of the national estimates of net coverage based on the 1990 Post-Enumeration Survey (PES). (See Hogan 1992 for a description of the PES.)

Section 2 describes the methodology of the demographic estimates. Section 3 describes the estimates of coverage in the 1990 census based on demographic analysis and compares the estimates with those for previous censuses. Section 4 presents the results of the first-time assessment of uncertainty in the demographic coverage estimates for 1990. Section 5 presents our conclusions and plans for future research.

## 2. THE GENERAL METHOD OF DEMOGRAPHIC ANALYSIS

Estimation of census coverage based on demographic analysis involves developing demographic estimates of the resident population in various categories, such as age-sex-race groups, by combining various sources of administrative and demographic data. The independent population estimates ( $P$ ) are then compared with the corresponding census counts ( $C$ ) to yield an estimate of the net census undercount,  $u$ , and net undercount rate,  $r$ :

$$u = P - C \quad (1)$$

and

$$r = (u/P) * 100. \quad (2)$$

Demographic analysis represents a macro-level approach to measuring coverage, where analytic estimates of net undercount are derived by comparing aggregate sets of data or counts. This approach differs fundamentally from the PES, which represents a micro-level approach where estimates of coverage are based on case-by-case matching with census records for a sample of the population.

The particular analytic procedure used to estimate coverage nationally in 1990 for the various demographic subgroups depends primarily on the nature and availability of the required demographic data. Different demographic techniques were used for the populations under age 55, 55-64, and 65 and over; the total population is the sum of these subgroups. Figure 1 summarizes the cohort estimation procedure for each group.

### 2.1 Estimation of Subgroups

**2.1.1 Age under 55.** The demographic analysis estimates for the population below age 55 in 1990 are based on the compilation of historical estimates of the components

\* J. Gregory Robinson is Chief, Bashir Ahmed is Demographic Statistician, and Prithwis Das Gupta is Mathematical Statistician, Population Analysis and Evaluation Staff, Population Division, U.S. Bureau of the Census, Washington, DC 20233. Karen A. Woodrow is Adjunct Research Associate, Center for Social and Demographic Analysis, State University of New York, Albany, New York 12222. The authors thank the referees and Special Section Editor for their helpful comments and suggestions.

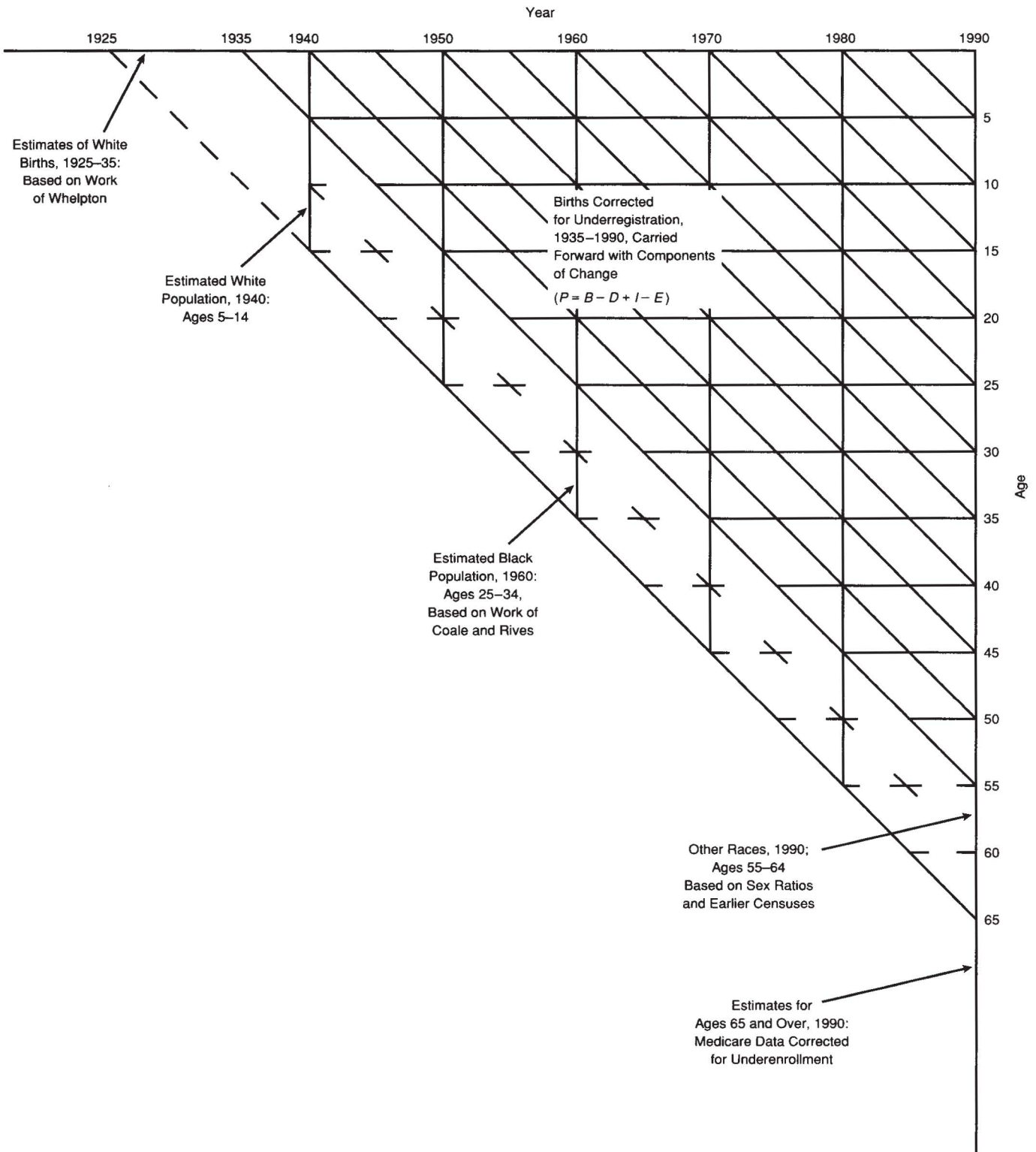


Figure 1. Lexis Diagram Showing Construction of Demographic Analysis Estimates of Census Coverage for 1990.

of population change: births ( $B$ ), deaths ( $D$ ), immigration ( $I$ ), and emigration ( $E$ ). Presuming that the components are accurately measured, the population estimates ( $P_1$ ) are derived by the basic demographic accounting equation applied to each cohort:

$$P_1 = B - D + I - E. \quad (3)$$

For example, the estimate of the population age 40 on April 1, 1990 is based on births from April 1949 to March 1950 (adjusted for underregistration), reduced by deaths to the cohort in each year between 1950 and 1990, and incremented by estimated immigration and emigration of the cohort over the 40-year period. (Follow the diagonal lines in Fig. 1.) All other single-year age cohorts are estimated in this manner.

It should be noted that the population under age 55 comprised 79% of the total population in 1990, so the population estimates based on Equation (3) have the greatest impact on the overall undercount estimates.

2.1.2 Age 55–64. For the population age 55–64 in 1990, the absence of national data on registered births and underregistration factors for this group (i.e., births from 1925 to 1935) necessitates using other data sources and methods. Hence we used different analytic techniques to develop the demographic estimates ( $P_2$ ) for this age group,

$$P_2 = T - D + I - E, \tag{4}$$

where  $T$  is the estimate in a previous time period (1925–1935 births for Whites, 1960 population for Blacks, 1990 population for Other Races ) and  $D$ ,  $I$  and  $E$  are as used in Equation (3).

For the White population, estimates for births for 1925–1935 developed by Whelpton (1950) are carried forward to 1940 with lifetable survival rates and to 1990 with components of change to estimate the population age 55–64 (see Fig. 1). For Blacks, revised population estimates developed by Coale and Rives (1973) were carried forward to 1990 with components of change. Estimates of the Other Races population age 55–64 were derived from assumptions about the consistency of age patterns of coverage in earlier censuses and the use of expected sex ratios. The demographic estimates of this age group are considered the weakest of the three broad age groups, but age 55–64 comprises only 8.5% of the total population.

2.1.3 Age 65 and Over. Administrative data on aggregate Medicare enrollments were used to estimate the population age 65 and over ( $P_3$ ) in 1990,

$$P_3 = M + m, \tag{5}$$

where  $M$  is the aggregate Medicare enrollment and  $m$  is the estimate of underenrollment. Although Medicare enrollment is generally presumed to be quite complete, adjustments to the basic data must be used to account for groups known or

suspected to be omitted (i.e., persons eligible for Medicare coverage but not enrolled, aliens resident in the country for less than 5 years, certain Federal employees and annuitants). The population age 65 and over represents 12.5% of the total population.

2.1.4 Total Population. The estimated total population in 1990 based on demographic analysis ( $P$ ) represents the sum of the individual estimates for ages under 55, 55–64, and 65 and over:

$$P = P_1 + P_2 + P_3. \tag{6}$$

As shown in Table 1, an estimate is first actually developed for the total legally resident population (row 1). Then the estimate of the number of undocumented residents (row 2) is added to produce the demographic estimate of the total resident population of 253.4 million in 1990 (row 3).

## 2.2 Estimation of Components

2.2.1 Estimating Births. The historical data on births,  $B$  in Equation (3), came from the vital registration system and were available at the national level only since about 1935. Births are by far the largest component of population change involved in the demographic analysis system; thus even relatively small errors in the estimates of births can have significant effects on the demographic estimates of coverage. Tests of birth registration completeness conducted for 1940, 1950, and 1964–1968 had provided correction factors for those years; factors for other years were obtained by interpolation and extrapolation. The effect of errors in these estimates (as well as the other components) is addressed in Section 4.

One important strength of the demographic analysis method is the ability to examine the internal consistency of the estimates in terms of their conformance with expected age, period, and cohort patterns of coverage; anomalous patterns for specific birth cohorts and components of change can be identified and possibly corrected. As a case example, the unrevised series of undercount estimates for Black males revealed an odd “cohort effect” whereby the birth cohort of

Table 1. Demographic Analysis Estimates of Population and Percent Net Undercount in the 1990 Census by Race and Sex

Estimates	Total			Black			Non-Black		
	Both sexes (1)	Male (2)	Female (3)	Both sexes (4)	Male (5)	Female (6)	Both sexes (7)	Male (8)	Female (9)
1. Revised legally resident population, 4-1-1990	250,061	122,928	127,133	32,039	15,625	16,414	218,022	107,303	110,719
2. Undocumented residents, 4-1-1990	3,333	1,792	1,541	281	134	147	3,052	1,658	1,394
3. Total resident population, 4-1-1990 (3 = 1 + 2)	253,394	124,720	128,674	32,320	15,759	16,561	221,074	108,961	112,113
4. Census count, 4-1-1990	248,710	121,239	127,470	29,986	14,170	15,816	218,724	107,069	111,655
5. Race modification	0	0	0	497	250	247	-497	-250	-247
6. Modified census count 4-1-1990 (6 = 4 + 5)	248,710	121,239	127,471	30,483	14,420	16,063	218,227	106,819	111,408
7. Net undercount, 4-1-1990 (7 = 3 - 6)	4,684	3,480	1,204	1,836	1,338	498	2,848	2,142	706
8. Percent net undercount, 4-1-1990 (8 = 7/3 × 100)	1.85	2.79	.94	5.68	8.49	3.01	1.29	1.97	.63

NOTE: These demographic estimates are subject to revision as new research becomes available; population is in thousands.

1935–1945 displayed unusually high undercount rates in each census as it aged (e.g., at ages 15–24 in 1960, 25–34 in 1970, 35–44 in 1980, and 45–54 in 1990). An investigation demonstrated that the underlying cause was that the Birth Registration Test of 1940 significantly overstated the “true” level of birth underregistration for Blacks, thereby leading to overcorrection of births and overstatement of the demographic estimates of Black population that are tied to the 1940 test results (Passel 1991; Robinson, Das Gupta, and Ahmed 1990). Based on this analysis, the sizes of the Black cohorts born between 1935 and 1945 were lowered.

The consistency in classification of births by race is important for making comparisons between the demographic analysis estimates and the census. Until 1990, the National Center for Health Statistics (NCHS) assigned rules for the classification of births by race, which favored race categories other than White. In 1990 NCHS adopted a new rule, called the “mother rule,” which assigns the mother’s race to the child regardless of the father’s race. Alternatively, one can develop the “father rule” where the child is assigned the father’s race. Work by Passel (1990) and U.S. Bureau of the Census (1991i) shows that the “father rule” tended to match more closely with the response patterns in the census. Estimates of births consistent with the father rule were used to develop the population estimates in Equation (3).

**2.2.2 Estimating Deaths.** The component of deaths ( $D$ ) is based on administrative records believed to be relatively complete. There is little information available on which to quantify empirically the possible extent of underregistration of deaths (U.S. Bureau of the Census 1991g). Therefore, unlike births for which the probable magnitude of underregistration can be empirically quantified, the magnitude of underregistration of deaths must be based on speculation. For infant deaths before 1960, it was assumed that deaths were underregistered at one-half the rate of underregistration of births; no adjustment for infant deaths was made for years after 1960. In addition to actual deaths, life table survival rates are used to carry forward the older cohorts, to estimate sex ratios, and to estimate migration.

**2.2.3 Estimating Immigration.** The immigration component ( $I$ ) in the demographic accounting equation has two major parts: legal and undocumented. The legal part consists of legal alien immigration, net migration from Puerto Rico, net arrival of civilians living abroad, net arrival of foreign students, net movement of military personnel, and refugees and parolees adjusting status. For the 1990 estimates, the legal part also includes immigrants legalized under the Immigration Reform and Control Act (IRCA) of 1986. Of these elements, legal alien immigration and undocumented immigration are the largest.

Data on legal alien immigration and adjustees are based on administrative records from the Immigration and Naturalization Service (INS). The INS data are believed to be quite complete, are timely, and require little estimation in comparison to other immigration components. The race of alien immigrants has to be estimated, based on the race of recent immigrants by country of origin as reported in the

most recent census. Revisions to the estimates for 1990—reflecting changes not in totals but in the race distribution of immigrants—cannot be accomplished until 1990 sample data from the census are available (in late 1993).

The number of undocumented immigrants is extrapolated from analyses of data on the foreign-born population obtained from the censuses and from periodic supplements to the Current Population Surveys (CPS). These analyses involve a residual estimation technique in which an estimate of the legally resident foreign-born population from a census is carried forward to the survey date and compared with the foreign-born population in the survey. The difference represents the number of undocumented immigrants included in the survey (Woodrow 1992). The figure is the net of entries, exits, and legal status changes of undocumented immigrants. The undocumented immigrant component may also include some unknown number of legal residents of the Special Agricultural Worker program. Revisions to the estimates of undocumented aliens may be made when detailed sample data from the 1990 census are analyzed.

**2.2.4 Estimating Emigration.** The emigration component ( $E$ ) represents emigration of legal residents only. The volume of emigration for the 1980s is based on simple extrapolations of emigrant levels during the 1960s and 1970s; the estimates are subject to greater error because no standard technique using current empirical data has yet been developed for them. Measurement of emigration based on the multiplicity sampling technique—where information is collected from residents about their immediate relatives who have ever lived in the United States and who are now living elsewhere—is a promising new data source for the 1990s (Woodrow 1990).

## 2.3 Limitations in the Scope of the Demographic Estimates

Because demographic analysis works with aggregate data from both the census and independent data sources, the estimates provide measurements of *net* census error; they do not identify the separate effects of omissions, duplications or erroneous inclusions, and reporting errors (age, sex, race) in the census. Further, because few of the independent sources contain adequate or consistent historical data on the Hispanic population or specific race groups other than White or Black (e.g., American Indian, Chinese, Japanese), demographic analysis cannot develop reliable coverage estimates for these groups. Finally, demographic estimates of coverage have been successfully produced only for the national population. In 1990, the currently available coverage estimates for Hispanics, Asians, and subnational areas are from the PES.

Another limitation of the demographic estimates concerns the comparability with the racial tabulations of the census. Before the demographic estimates of population for race groups are compared to the census to calculate the net undercount, the race categories of the census counts must be “modified” so that they are consistent with the race categories of the historical demographic estimates. Specifically, 9.8

million persons in the 1990 census (mostly of Hispanic origin) reported their race in the "Other Race-Not Specified" category, a category not included in the demographic estimates. This modification added 497,000 persons to the census count for Blacks (see row 5 of Table 1).

## 2.4 Internal Consistency of the Demographic Estimates: A Major Strength

Before discussing the results, it is important to emphasize how the logical consistency and interrelationships of the underlying demographic variables and the data used to measure them are a strength of the demographic method in several respects.

First, the internal consistency of the demographic estimates and components of change over the period 1940–1990 allow the production of undercount estimates for each census (1940, 1950, 1960, 1970, 1980, and 1990). For example, the net undercount of the 1949–1950 birth cohort (noted in Sec. 2.1) can be measured at age 0 in 1950, 10 in 1960, 20 in 1970, 30 in 1980, and 40 in 1990. Likewise, the population age 65 and over in 1990 based on Medicare data can be carried backward in time to estimate the population age 55 and over in 1980, 45 and over in 1970, and so on. The historical estimates of coverage for all ages that can be developed with demographic analysis are important for assessing trends in coverage.

Second, with multiple observations of net undercount rates across several censuses, it becomes possible to judge the quality of the demographic estimates and possibly identify and correct for anomalous patterns for specific cohorts and components of change. A prime example of this evaluative mechanism is the revisions to the demographic estimates for Blacks incorporated in the 1990 estimates, as described previously.

Finally, the particular methods, data, and assumptions used in demographic analysis are not fixed. As analytic methods evolve over time and new data or information become available, the estimates are subject to change, with the constant goal of improving the estimates. Because the demographic estimates are additive and internally consistent, revisions to the estimates (such as the change to the Black cohorts just noted) can be easily made and automatically lead to revision of the demographic estimates of population

and coverage in previous censuses. The robustness of the estimates to changing data and assumptions can be easily assessed. The historical experience has shown that the demographic estimates are quite robust with regard to measurement of broad patterns of coverage. Despite the many changes and improvements incorporated into the demographic methodology since 1960, the relative levels and age-sex-race patterns of net coverage for each census have not been appreciably affected by subsequent revisions. We do not anticipate that the first set of demographic estimates for 1990 will change much based on later revisions.

## 3. COVERAGE OF POPULATION IN THE 1990 CENSUS AND COMPARISON WITH ESTIMATES FOR PREVIOUS CENSUSES

### 3.1 Net Coverage Levels for 1990

The estimated net undercount in the 1990 Census was 4.68 million, or 1.85% (see Table 1). The estimated undercount for males (3.48 million, or 2.79%) exceeded the estimate for females (1.20 million, or .94%). In other words, about three-fourths of the net omissions in 1990 were males. The estimated net undercount rate of Blacks (5.68%, or 1.84 million) was higher than that of Non-Blacks (1.29%, or 2.85 million). These estimates imply that about two-fifths of net omissions in 1990 were Blacks. The race-sex details in Table 1 show that for both Blacks and Non-Blacks the undercount of males was greater than for females. Black males were estimated to have been missed at the highest rate of all race-sex groups—8.49%.

### 3.2 Comparison of Coverage in 1990 with Estimates for Previous Censuses

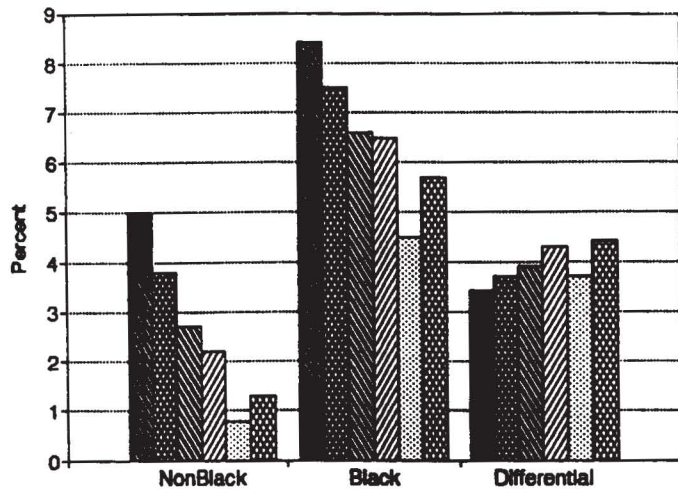
The historical estimates of net undercount for the 1940–1990 censuses presented in Table 2 show that the percent net undercount of the 1980 and 1990 censuses is lower than that measured for the censuses of 1970 and earlier. The long-term decline in undercount rates from 1940 to 1990 is clear. Also apparent in Table 2 is the sole exception to this trend: the lower rates of *net* undercount in the 1980 census relative to the 1970 and 1990 censuses. (Another view of the discontinuity is that the 1990 net undercount rate is higher than the 1980 rate.)

When comparing the change in coverage between 1980 and 1990 with the change between 1970 and 1980, it is important to remember that the demographic estimates provide only measures of *net* undercount. We do not know as much about the trends in the underlying—and offsetting—components of net undercount, which are the omission of persons from the census (undercounts) and the duplication of persons in the census (overcounts). A full assessment of the change in net coverage over the 1970, 1980, and 1990 censuses must consider the impact of both components.

The estimates of percent net undercount for sex and race groups in Table 2 show that the differences in coverage by sex and race in 1990 were similar to those of recent censuses. The greater undercount of males than females is common to all censuses since 1940; a finding for 1990 is that the grow-

Table 2. Historical Demographic Analysis Estimates of Percent Net Undercount and Differences by Race and Sex: 1940–1990

Race and sex	1990	1980	1970	1960	1950	1940
Total	1.8	1.2	2.7	3.1	4.1	5.4
Male	2.8	2.2	3.4	3.5	4.4	5.8
Female	.9	.3	2.0	2.7	3.8	5.0
Black	5.7	4.5	6.5	6.6	7.5	8.4
Male	8.5	7.5	9.1	8.8	9.7	10.9
Female	3.0	1.7	4.0	4.4	5.4	6.0
Non-Black	1.3	.8	2.2	2.7	3.8	5.0
Male	2.0	1.5	2.7	2.9	3.8	5.2
Female	.6	.1	1.7	2.4	3.7	4.9
Difference						
Male:female	1.8	1.9	1.5	.8	.6	.8
Black:Non-Black	4.4	3.7	4.3	3.9	3.6	3.4



Source: Table 1.

Figure 2. Percent Net Undercount by Race, 1940–1990. ■ = 1940; ■ = 1950; ■ = 1960; ■ = 1970; ■ = 1980; ■ = 1990.

ing gap in the undercount rates of men and women observed over the 1940 to 1980 censuses did not continue to grow for 1990 (see the next to the last row of Table 2).

Similarly, a common observation for all censuses since 1940 is that the estimated undercount rate of Blacks has remained persistently higher than the undercount rate of Non-Blacks. In fact, the excess of the net undercount rate of Blacks has hovered in the range of 3.4 to 4.4 percentage points over the last six censuses (see Fig. 2 and the last row of Table 2).

### 3.3 Age, Race, and Sex Differences in Census Coverage

Table 3 and Figure 3 display more detailed estimates of percent net undercount for race, sex, and age groups in the 1990 census. In terms of level of percent undercount, the

most notable pattern is the high levels of undercount for Black men age 25–64, where the estimated net undercount has ranged between 10 and 14%. These levels contrast sharply with the relatively low measured *net* undercount rates for Black males age 15–19 and 65 and over. For both Black males and females, the undercount rates for ages 0–4 and 5–9 are relatively high (though not as high as the rates for Black adult men). Also in contrast to the high net undercount estimates for Black adult men are the relative low undercount rates of Black adult women. For Non-Black males and females, the estimates of net undercount for 1990 are quite low across all ages. In fact, the net undercount estimates for Non-Black females straddle the zero undercount line for most age groups.

Figure 4 compares the 1990 coverage estimates by age, race, and sex with estimates for 1980 and 1970. Most notable is the general consistency of the net coverage patterns across censuses for each race-sex group. Thus the coverage patterns described previously are not unique to the 1990 census. These demographic estimates indicate that the net undercounts of Black men and Black children have been relatively high, and the net undercount of Non-Blacks relatively low, in each of the last three censuses.

In interpreting the estimates of net coverage for race, sex, and age groups that have been discussed, attention must be given to the fact that the demographic estimates are really approximations of the *exact* level of net undercount for a given group. As indicated by the ranges in Table 4 (and discussed later), there is considerable uncertainty in the detailed estimates of coverage by race, sex, and age. For Black age groups in particular, there is a wide range within which the “true” net undercount rate may fall. Nonetheless, it is clear from the alternative range of estimates for 1990 that the demographic estimates of percent undercount for Black adult males remain relatively high under any reasonable “uncertainty” assumption. The “lowest” alternative estimate for

Table 3. Amount and Percent Net Undercount by Age, Sex, and Race: 1990

Age (years)	Black				Non-Black			
	Male		Female		Male		Female	
	Amount	Percent	Amount	Percent	Amount	Percent	Amount	Percent
All ages	1,338	8.5	498	3.0	2,142	2.0	706	.6
0–4	140	8.6	129	8.2	224	2.7	224	2.8
5–9	114	7.7	108	7.5	216	2.7	218	2.8
10–14	57	4.1	55	4.0	39	.5	53	.7
15–19	–2	–.2	6	.4	–176	–2.3	–120	–1.7
20–24	78	5.7	34	2.5	–66	–.8	–50	–.6
25–29	192	12.7	75	4.9	444	4.5	199	2.1
30–34	207	14.0	52	3.5	380	3.8	67	.7
35–39	148	11.9	29	2.2	231	2.6	18	.2
40–44	103	10.6	15	1.5	113	1.4	–54	–.7
45–49	87	11.9	19	2.4	169	2.7	40	.6
50–54	72	12.0	10	1.6	143	2.8	22	.4
55–59	63	12.1	0	0	140	3.0	17	.4
60–64	48	10.3	–15	–2.9	120	2.6	4	.1
65–69	8	2.1	–36	–7.7	95	2.2	–48	–.9
70–74	11	4.1	–14	–3.8	60	1.9	–4	–.1
75+	11	3.2	30	4.4	7	.2	120	1.5

NOTE: Numbers are in thousands. The base of percents is the estimated population. A minus sign denotes a net overcount.

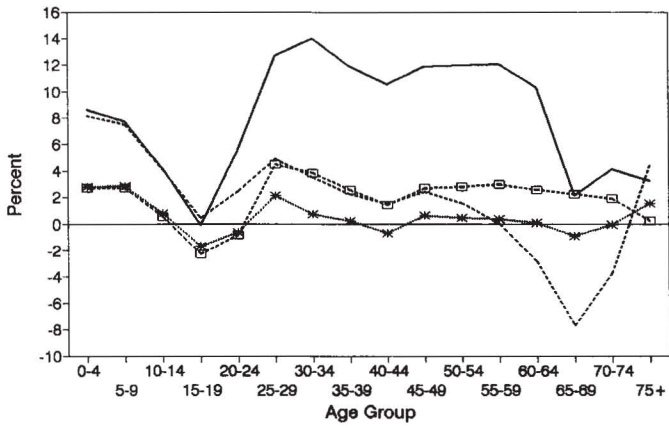


Figure 3. Percent Net Undercount by Race, Sex, and Age: 1990. — Black male; --- Black female; -□- Non-Black male; ...\* Non-Black female.

Black males is above 6% for each broad age group between 20 and 64. With the exception of Black males and females under age 10, net undercounts consistently above even 2% are not found for any other race-sex-age group when the total range of uncertainty in the estimates is taken into account.

### 3.4 Expected Sex Ratios

The last evaluation tool of demographic analysis to be considered is the comparison of sex ratios (males per 100 females) of the demographic estimates and the census. The sex ratio from demographic analysis is generally considered to be one of the method's most robust measures. As illustrated in Figure 5, the sex ratios of the census counts for Blacks fall well below the "expected" demographic ratios between ages about 25 and 65. These sex ratio patterns imply that the net undercount rates of Black men in the 1990 census are high *relative* to the rates of Black women. The "gap" for Non-Blacks is much smaller in comparison, denoting smaller differences in the undercount rates of Non-Black men and women. These observations are consistent with the direct estimates of percent net undercount for race-sex-age groups shown in Table 3.

## 4. EVALUATION OF THE DEMOGRAPHIC ANALYSIS ESTIMATES

### 4.1 The Eleven Demographic Analysis Evaluation Projects

Eleven demographic evaluation projects (D1–D11) were carried out to provide the first comprehensive assessment of

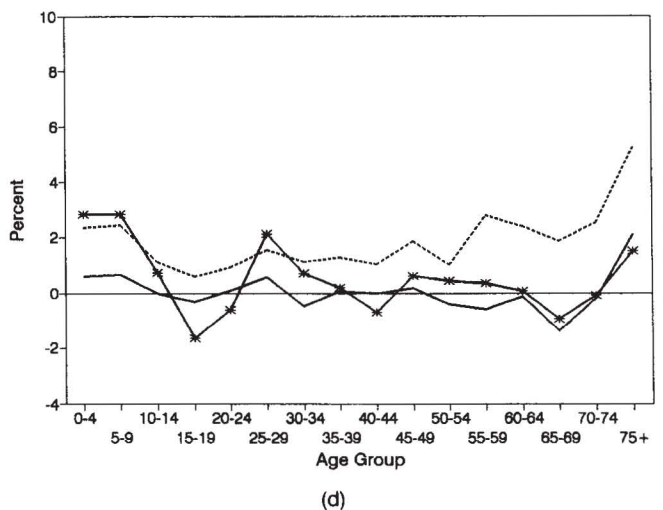
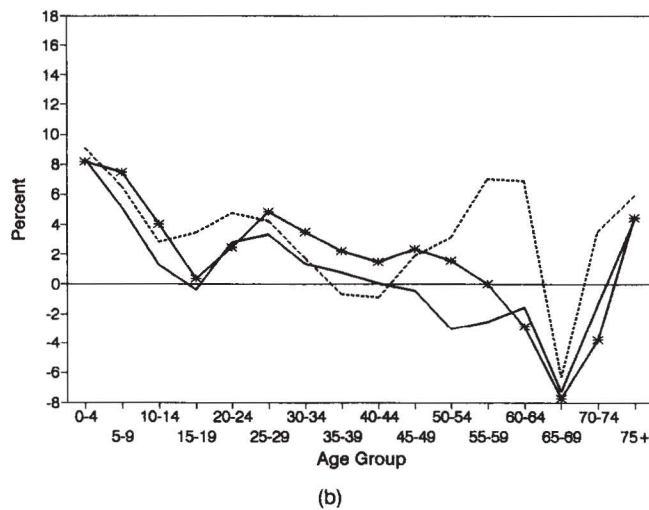
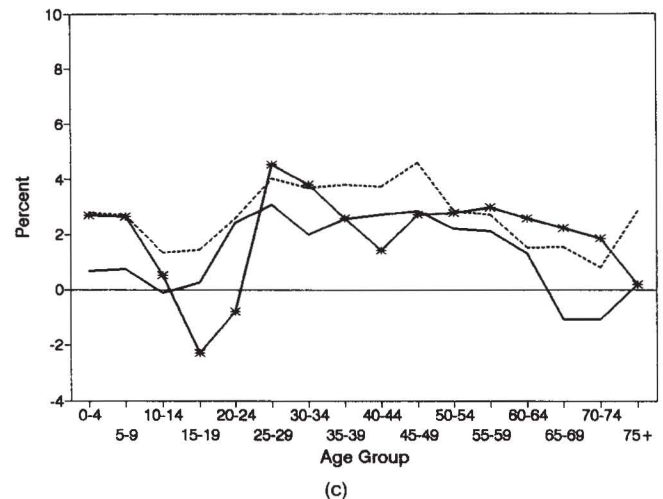
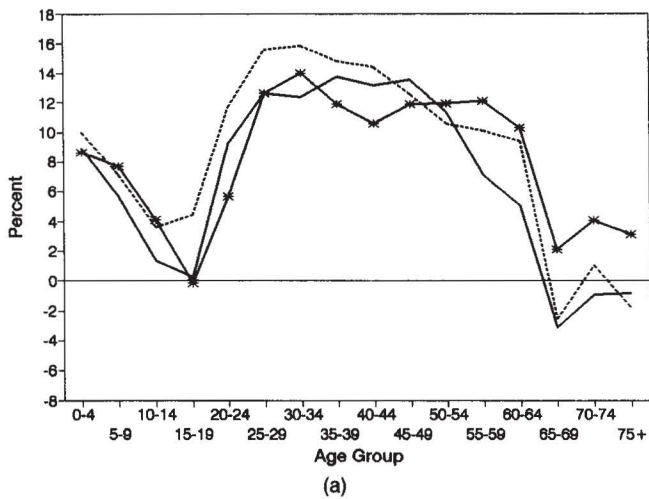


Figure 4. Percent Net Undercounts 1970–1990: (a) Black Males; (b) Black Females; (c) Non-Black Males; (d) Non-Black Females. —\*— 1990; — 1980; --- 1970.

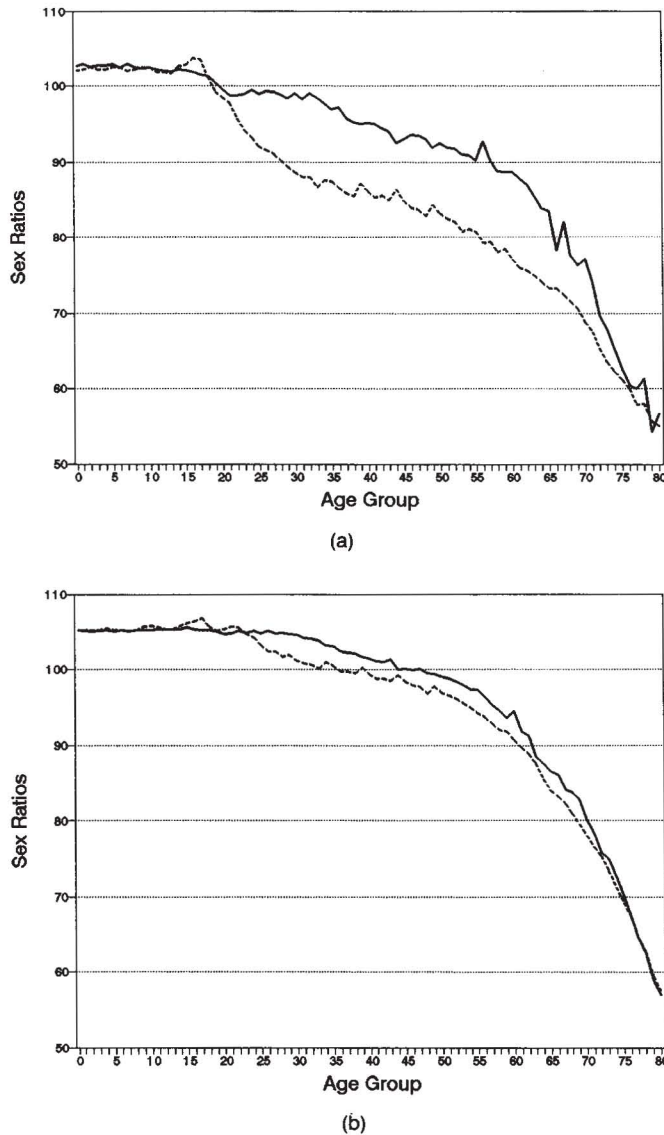


Figure 5. Census Sex Ratios (· · · ·) and Expected Sex Ratios (—) for Blacks and Non-Blacks by Age: 1990. (a) Blacks; (b) Non-Blacks.

the accuracy of the individual components of the demographic estimates of population and coverage. These evaluation projects are listed in the references. Demographic Evaluation Project D11 provides a detailed description of the total error interval (or uncertainty) model developed over the last several years. The other 10 demographic analysis evaluation projects address specific sources of uncertainty in the demographic estimates of population and coverage (e.g., births, undocumented residents, emigrants, Medicare data, inconsistency in racial classifications). These evaluations are important in providing the low and high multipliers that are critical ingredients to the total uncertainty model.

#### 4.2 The Total Error or Uncertainty Model

The purpose of the model we have developed is to provide a statistically based assessment of the uncertainty in the demographic estimates of coverage. In the model the demographic analysis estimate  $p$  for the total population or for a

subpopulation by race, sex, and age is regarded as a random variable given by

$$p = P_1x_1 + P_2x_2 + \dots + P_nx_n, \tag{7}$$

where  $P_i$ 's are observed demographic estimates (the so-called "point estimates") of components that are treated as constants (e.g., births, deaths, legal immigration, legal emigration, undocumented residents), and the random variables  $x_i$ 's are multipliers that fluctuate around 1. When all  $x_i$ 's are 1,  $p$  becomes identical to the observed total demographic estimate  $P$ .

The undercount rate  $u$  in the census is a random variable given by

$$u = \frac{p - c}{p} \times 100, \tag{8}$$

where  $c$  is the population enumerated in the census.

From Equation (7), we have

$$E(p) = \sum_{i=1}^n P_i \mu_i \tag{9}$$

and

$$V(p) = \sum_{i,j=1}^n P_i P_j \rho_{ij} \sigma_i \sigma_j, \tag{10}$$

where  $\mu_i = E(x_i)$ ,  $\sigma_i^2 = V(x_i)$ , and  $\rho_{ij}$  = correlation coefficient between  $x_i$  and  $x_j$  ( $\rho_{ii} = 1$ ,  $\rho_{ij} = \rho_{ji}$ ).

Assuming no errors in the census number  $c$  in (8), we obtain from the large-sample formulas

$$E(u) = \frac{E(p) - c}{E(p)} \times 100 \tag{11}$$

and

$$V(u) = \frac{(100)^2 c^2 V(p)}{[E(p)]^4}. \tag{12}$$

We assume each component  $x_i$  in (7) to be normally distributed so that  $p$  is also normally distributed. The  $\alpha$ -percent error ("confidence") interval for  $p$  is, therefore,  $(P_L, P_H)$ , given by

$$E(p) \pm \tau_\alpha \sqrt{V(p)}, \tag{13}$$

where  $\tau_\alpha$  is the value of the corresponding normal deviate and  $L$  and  $H$  stand for low and high.  $P_L$  and  $P_H$  can be substituted for  $p$  on the right side of (8) to obtain the  $\alpha$ -percent error interval  $(U_L, U_H)$  for the undercount rate  $u$ .

Alternatively, although  $u$  is not normally distributed, if we obtain the error interval for  $u$  directly from

$$E(u) \pm \tau_\alpha \sqrt{V(u)} \tag{14}$$

by using (11) and (12), then the results are almost identical. For our convenience, we take this approach in (14) to obtain the error interval for  $u$ . We have considered two values for  $\alpha$  in (14), 95% and 99%, with the value of  $\tau_\alpha$  corresponding to these  $\alpha$ 's being 1.96 and 2.58.

Strictly speaking, because of the modifications of the census race data mentioned earlier,  $c$  in (8) is not a constant.



For Blacks, we have assumed a uniform distribution between the unmodified number ( $c_l$ ) and the modified number ( $c_h$ ), so that

$$E(c) = \frac{c_l + c_h}{2} \quad (15)$$

and

$$V(c) = \frac{(c_h - c_l)^2}{12}. \quad (16)$$

The distribution of  $c$  for Non-Blacks follows from the fact that the total for Blacks and Non-Blacks is kept unchanged. From the expression for  $u$  in (8), using large-sample formulas, and assuming no correlations between  $c$  and  $p$ , we revised  $E(u)$  and  $V(u)$  in (11) and (12). We then used these revised expressions to obtain the error intervals for  $u$  from (14).

#### 4.3 Estimation of $\mu_i$ , $\sigma_i$ , and $\rho_{ij}$

For each multiplier  $x_i$  in (7), we arrived at two certainty limits:  $l_i$  and  $h_i$ . Initially, these limits had been obtained by a judgmental consensus of the Census Bureau experts knowledgeable about estimation methodology and possible errors in the components of change. Subsequently, these limits were revised a number of times based on more elaborate studies of the components (described in Demographic Analysis Evaluation Projects D1–D10). Because certainty limits are more difficult to justify than probability limits, for the purpose of developing our model we have treated the derived limits as being  $\beta$ -percent probability limits. We have assumed two different values for  $\beta$  for the same set of the error intervals: 99.9% and 99%. Obviously, the latter assumption gives us wider error intervals for the undercount rates.

Computation of  $\mu_i$  depends on our assumption about the relationship between  $\mu_i$  and the multiplier 1 corresponding to the demographic estimate. If the true value of the multiplier lies between, say, .95 and 1.15 with certainty, then these asymmetrical values around 1 imply that the multiplier has a distribution for which the mean is greater than 1. But that can happen in various ways. For example, (a) the multiplier has a normal distribution with mean, median, and mode equal to 1.05 (i.e., the average of .95 and 1.15); (b) the multiplier has a positively skewed gamma distribution with a median of 1, a mode less than 1, and a mean greater than 1; or (c) the multiplier has a positively skewed gamma distribution with a mode of 1, a median greater than 1, and a mean even greater than the median.

We developed several models based on different interpretations of the demographic estimates in relation to the measures of location and also on our assumptions about the probability distributions of the multipliers. The results from these models are not much different, as far as the lengths of the error intervals are concerned. We finally chose model (a)—a normal distribution for a multiplier with the mean as the average of the high and low values—because this model is the simplest to understand and also easiest to interpret in terms of the normal curve depicting accidental errors. This model implies that multiplier 1 is the mean of the distribution of  $x_i$  with a constant bias, with the bias measured by the

difference between 1 and the mean of  $l_i$  and  $h_i$ . In other words, we compute the mean  $\mu_i$  by

$$\mu_i = \frac{l_i + h_i}{2}. \quad (17)$$

Because  $x_i$  is assumed to be normally distributed, we can estimate  $\sigma_i$  from

$$\sigma_i = \frac{h_i - \mu_i}{\tau_\beta} = \frac{\mu_i - l_i}{\tau_\beta}, \quad (18)$$

where  $\tau_\beta$  is the value of the normal deviate corresponding to our assumption about  $\beta$ . Because we have assumed for  $\beta$  the probability limits of 99.9% and 99%, the corresponding values for  $\tau_\beta$  are 3.29 and 2.58.

Finally, we have assumed values for possible correlations between pairs of components, because it is impossible to estimate them empirically. We have assumed most of these correlations to be 0.

#### 4.4 Contributions of Demographic Components to Overall Variance

Most of the uncertainty in the demographic estimates for 1990 is contributed by four specific components: births, undocumented immigration, emigration, and legal immigration. These four components account for about 83% of the overall variance for the total demographic estimate (39.2% for births, 16.5% for undocumented aliens, 15.2% for emigration, and 12.0% for legal immigration). For Black males and Black females, these components (mostly births) contribute about 88% of the total variance. For Non-Black males and females, the four components account for about 78% of the total variance.

The other components—including deaths, Medicare data, population age 55–64 in 1990, miscellaneous migration components, Armed Forces personnel stationed overseas—are all subject to error but do not contribute to the overall variance to the extent as that estimated for the four major components.

#### 4.5 Uncertainty Intervals for Estimates of Percent Net Undercount

Table 4 shows the uncertainty intervals corresponding to two combinations of  $(\alpha, \beta)$ : 95% error intervals, 99.9% multiplier probability limits and 99% error intervals, 99% multiplier probability limits. For obvious reasons, the latter combination gives wider ranges of uncertainty intervals.

In terms of the 95% error intervals in Table 4 for the total population, the interval for the net percent undercount is 1.73 (from 1.63 to 3.36—see the last row of Table 4). The intervals are wider for Blacks (4.25 for males and 4.18 for females) than for Non-Blacks (2.04 for males and 2.03 for females).

In general, the uncertainty intervals are wider for individual age groups than for the total population of any race-sex group. This is particularly true for the age groups 45–64 and 65 and over (and especially for Blacks, where the 95% error intervals are wider than 6 percentage points).

Higher undercount rates do not necessarily imply wider error intervals. A high undercount rate may indicate a problem with the census enumeration, whereas a wide error interval indicates a problem with the demographic analysis

Table 4. Alternative Uncertainty Intervals for the Demographic Analysis Estimates of Percent Net Undercount by Race, Sex, and Age: 1990

Race, Sex, Age (years)	Percent undercount		95% Intervals			99% Intervals		
	Observed	Mean	Lower	Upper	Length	Lower	Upper	Length
<b>Black male</b>								
0-9	8.07	8.59	5.96	11.22	5.26	4.34	12.84	8.51
10-19	1.95	2.51	.36	4.65	4.30	-.88	5.89	6.77
20-29	9.09	10.08	8.35	11.82	3.47	7.41	12.76	5.35
30-44	12.50	13.55	11.63	15.47	3.83	10.53	16.57	6.03
45-64	11.87	13.44	9.15	17.74	8.59	6.32	20.56	14.24
65+	3.00	2.34	-1.44	6.13	7.56	-3.88	8.57	12.44
Total	8.47	9.31	7.18	11.44	4.25	5.92	12.70	6.78
<b>Black female</b>								
0-9	7.75	8.21	5.63	10.79	5.16	4.00	12.41	8.41
10-19	2.13	2.62	.56	4.68	4.12	-.66	5.89	6.54
20-29	3.47	4.39	2.68	6.11	3.43	1.73	7.06	5.33
30-44	2.55	3.63	1.60	5.66	4.06	.41	6.85	6.44
45-64	.61	2.29	-2.07	6.64	8.72	-4.94	9.51	14.46
65+	-.95	1.58	-1.60	4.76	6.36	-3.64	6.80	10.44
Total	2.97	4.03	1.94	6.12	4.18	.69	7.37	6.69
<b>Non-Black male</b>								
0-9	2.63	3.19	2.34	4.03	1.69	1.79	4.59	2.80
10-19	-.89	-.16	-1.11	.79	1.90	-1.74	1.42	3.16
20-29	1.70	2.68	1.47	3.90	2.42	.66	4.71	4.05
30-44	2.89	3.85	2.70	5.00	2.30	1.92	5.78	3.85
45-64	2.73	2.93	.87	4.99	4.12	-.52	6.39	6.92
65+	1.42	.84	-1.14	2.83	3.97	-2.49	4.17	6.66
Total	1.94	2.51	1.49	3.52	2.04	.81	4.21	3.40
<b>Non-Black female</b>								
0-9	2.76	3.33	2.49	4.16	1.67	1.94	4.71	2.77
10-19	-.53	.17	-.73	1.07	1.80	-1.32	1.66	2.99
20-29	.63	1.47	.42	2.52	2.10	-.28	3.22	3.50
30-44	.22	1.14	-.09	2.36	2.45	-.91	3.19	4.10
45-64	.44	.70	-1.45	2.84	4.29	-2.90	4.29	7.20
65+	.40	1.24	-.43	2.92	3.35	-1.57	4.05	5.62
Total	.61	1.30	.29	2.31	2.03	-.39	2.99	3.39
<b>Total population</b>								
0-9	3.53	4.08	3.08	5.08	2.00	2.40	5.76	3.35
10-19	-.28	.40	-.55	1.35	1.90	-1.19	1.99	3.18
20-29	1.90	2.81	1.65	3.97	2.33	.86	4.76	3.90
30-44	2.30	3.25	2.14	4.37	2.23	1.38	5.12	3.74
45-64	2.02	2.40	.67	4.13	3.45	-.50	5.30	5.79
65+	.79	1.14	-.68	2.97	3.66	-1.92	4.21	6.14
Total	1.83	2.49	1.63	3.36	1.73	1.04	3.95	2.90

NOTE: The 95% uncertainty intervals represent an error model with a 95% uncertainty interval and multiplier limits defined as 99.9% certain. The 99% uncertainty intervals represent an error model with a broader 99% uncertainty interval and multiplier limits defined as 99% certain.

data. Thus, although the undercount rates for Black males and Black females for 1990 are widely different (8.47 and 2.97), the corresponding lengths of the error intervals are about the same (4.25 and 4.18).

The means of the percent net undercount in Table 4 clearly indicate that the demographic net undercount estimates are biased in that they may underestimate the "true" net undercount (compare the estimates in column 1 and column 2). In fact, for the younger age groups of Non-Blacks, the undercount estimates fall close to the lower bounds of the 95% error intervals. For example, the demographic "point" estimate of .63% for Non-Black females age 20-29 is near the lower bound estimate of .42, the estimated mean of 1.47% being more than double the point estimate.

#### 4.6 Limitations of the Demographic Uncertainty Estimates

The systematic and detailed evaluation of the quality of the demographic coverage estimates reported here represents an evaluation program *new* for the 1990 census. The as-

sessments conducted in the 11 evaluation projects are subject to change and improvement over time just as the basic demographic estimates have been. But we feel that the models, assumptions, and analysis of the available information for the evaluation projects provide a reasonable assessment of the overall uncertainty in the demographic estimates of population and coverage for the 1990 census.

## 5. CONCLUSIONS

The technique of demographic analysis is a powerful tool for measuring net undercount in a census. The 1990 demographic analysis program provided not only the completeness of census coverage based on demographic analysis but also an assessment of the quality of these coverage estimates.

The estimates of net undercount for particular race, sex, or age groups based on demographic analysis may be subject to considerable uncertainty for measuring the exact *levels*. But they are subject to *less* variability in terms of measuring *differences* in coverage according to age, sex, and race and

measuring *changes* in net coverage between censuses. Thus the demographic coverage estimates allow us to make statements about differences and patterns of coverage, such as the net undercount of males is greater than that for females, the net undercount of Blacks is greater than that for non-Blacks, and the net undercount is especially high for certain age groups (e.g., Black males age 25–54 in 1990). The range of uncertainty measures developed for race-sex-age groups can be compared to assess the veracity of these statements.

This property of the demographic estimates—that they provide better measures of coverage differences rather than absolute coverage levels—is attributable to the fact that many of the errors in the estimates are believed to be consistent across sex, race, and time and hence tend to “cancel” in comparisons. The consistency of errors in the birth component provides a case example of how estimation errors tend to be eliminated when coverage estimates are being compared.

Finally, the internal consistency of the demographic estimates of population (and presumed consistency of errors) allows us to assess *changes* in net coverage patterns and levels over time with more confidence than the exact level of net coverage in any given census. Historical demographic estimates of net coverage show that the patterns of differential coverage for race-sex-age groups in 1990 is generally consistent with the pattern of recent censuses, and that overall net undercount rates have generally declined over the last several censuses.

The demographic estimates of coverage presented here are based on the best available data and are subject to change to incorporate new data and research findings (e.g., continued research may lead to revisions in our estimates of emigration and undocumented immigration). As in the past, these revisions are likely to affect the estimated levels of net coverage more than measured age, sex, and race differences in coverage. Finally, a new avenue of research is the application of demographic techniques that may provide indicators of coverage patterns for *subnational* areas. A principal project is the development of demographic estimates of coverage for regions and states for 1990, using data on state of birth from the 1990 sample data and Medicare data for the population age 65 and over. Also, we are investigating techniques for developing demographic indicators of coverage for selected age groups of the Asian and Hispanic populations. We are developing analytic indicators of coverage at the county level for very limited age groups (e.g., under age 10 and 65 and over).

[Received January 1992. Revised December 1992.]

## REFERENCES

- Clogg, C. C., Himes, C. L., and DeJani, A. N. (1991), “An Evaluation of Demographic Analysis as a Method for Estimating Population by Age, Sex, and Race,” Joint Statistical Agreement No. 76-5370-259 between the U.S. Bureau of the Census and the Pennsylvania State University.
- Coale, A. J., and Rives, N. W. (1973), “A Statistical Reconstruction of the Black Population of the United States, 1880–1970: Estimates of True Numbers by Age and Sex, Birth Rates, and Total Fertility,” *Population Index*, 39, 3–36.
- Hogan, H. (1992), “The 1990 Post-Enumeration Survey: Operations and Results,” in *Proceedings of the Survey Research Methods Section, American Statistical Association*.
- Passel, J. S. (1990), *Demographic Analysis: A Report on Its Utility for Adjusting the 1990 Census*, Washington, DC: The Urban Institute.
- (1991), “Age-Period-Cohort Analysis of Census Undercount Rates for Race-Sex Groups, 1940–1980: Implications for the Method of Demographic Analysis,” *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 326–331.
- Robinson, J. G., Das Gupta, P., and Ahmed, B. (1990), “A Case Study in the Investigation of Errors in Estimates of Coverage Based on Demographic Analysis: Black Adults Aged 35 to 54 in 1980,” *Proceedings of The Social Statistics Section, American Statistical Association*, pp. 187–192.
- Robinson, J. G., Ahmed, B., Das Gupta, P., and Woodrow, K. A. (1991), “Estimating Coverage of the 1990 United States Census: Demographic Analysis,” in *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 11–20.
- Siegel, J. S., and Zelnik, M. (1966), “An Evaluation of Coverage in the 1960 Census of Population by Techniques of Demographic Analysis and by Composite Methods,” in *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 71–85.
- U.S. Bureau of the Census (1974), “Estimates of Coverage of Population by Sex, Race, and Age: Demographic Analysis,” by Jacob S. Siegel, *Evaluation and Research Program*, PHC(E)-4.
- (1988), “The Coverage of Population in the 1980 Census,” by Robert Fay, Jeffrey S. Passel, and J. Gregory Robinson. *Evaluation and Research Reports*, PHC80-E4.
- (1991a), Demographic Analysis Evaluation Project D1: “Error in the Birth Registration Completeness Estimates” *Preliminary Research and Evaluation Memorandum No. 74 (PREM)*, by J. Gregory Robinson.
- (1991b), DA Evaluation Project D2: “Preliminary Estimates of Undocumented Residents in 1990” *PREM No. 75*, by Karen A. Woodrow.
- (1991c), DA Evaluation Project D3: “Uncertainty Intervals for Estimated White Births, 1915 to 1934” *PREM No. 76*, by J. Gregory Robinson.
- (1991d), DA Evaluation Project D4: “Uncertainty Intervals for Estimated Black Births, 1915–1934” *PREM No. 77*, by J. Gregory Robinson.
- (1991e), DA Evaluation Project D5: “Preliminary Estimates of Emigration Component” *PREM No. 78*, by Karen A. Woodrow.
- (1991f), DA Evaluation Project D6: “Robustness of the Estimates of the Population Aged 65 and Over” *PREM No. 79*, by J. Gregory Robinson.
- (1991g), DA Evaluation Project D7: “Uncertainty Measure for Other Components” *PREM No. 80*, by J. Gregory Robinson, Karen A. Woodrow, and Bashir Ahmed.
- (1991h), DA Evaluation Project D8: “Uncertainty for Models to Translate 1990 Census Concepts Into Historical Racial Classifications” *PREM No. 81*, by J. Gregory Robinson, David L. Word, and Gregory S. Spencer.
- (1991i), DA Evaluation Project D9: “Inconsistencies in Race Classifications of the Demographic Estimates and the Census” *PREM No. 82*, by J. Gregory Robinson and Susan Lapham.
- (1991j), DA Evaluation Project D10: “Differences Between Preliminary and Final Estimates of Percent Net Undercount” *PREM No. 83*, by Bashir Ahmed and J. Gregory Robinson.
- (1991k), DA Evaluation Project D11: “Models for Assessing Errors in Undercount Rates Based on Demographic Analysis” *PREM No. 84*, by Prithwis Das Gupta.
- Whelpton, P. (1950), “Birth and Birth Rates in the Entire United States, 1909 to 1948,” *Vital Statistics Special Reports*, 33, 137–162.
- Woodrow, K. A. (1990), “Emigration from the United States Using Multiplicity Surveys,” paper presented at the annual meeting of the Population Association of America, Toronto, Canada.
- (1992), “A Consideration of the Effect of Immigration Reform on the Number of Undocumented Residents in the United States,” *Population Research and Policy Review*, 11, 117–144.
- Woodrow, K. A., and Passel, J. S. (1990), “Post-IRCA Undocumented Immigration to the United States: Assessment Based on the June 1988 CPS” in *Undocumented Migration to the United States: IRCA and the Experience of the 1980s*, eds. F. D. Bean, B. Edmonston, and J. S. Passel, Washington, DC: The Urban Institute, pp. 33–75.