

## Setting the Standard for Comparison: Census Accuracy from 1940 to 2000

John F. Long, J. Gregory Robinson, Campbell Gibson, Population Division, U. S. Census Bureau

Key Words: Census Accuracy, Coverage Evaluation, Item Nonresponse

**Abstract:** After almost every recent census, there have been claims that it was the best census ever (often accompanied by counter claims of its deficiencies). This paper attempts to look at the historical record to see how good or bad recent historical censuses have been. While the primary emphasis will be on the measurement of census coverage, the paper also addresses related issues of census quality such as the level of missing data (and consequently allocation) for recent censuses.

### INTRODUCTION

Was Census 2000 the “best census ever”? Almost every decade there have been claims that the most recent census was the best census ever. Have censuses been constantly improving? Are there some censuses that are particularly bad? This paper examines the quality of previous censuses in order to provide a standard from which historical census quality can be judged. There are many dimensions of census quality including coverage, accuracy, efficiency, and comprehensiveness.

Census coverage may be the most important factor since a census enumeration cannot be considered successful if it omits a substantial portion of the population. Moreover, differential coverage, where a particular segment of society has a significantly lower coverage rate than the general population, is a major detriment to census quality. Census coverage has been measured through demographic means for over 50 years and by statistical means for the last three censuses. The next section of this paper reviews those historical trends in coverage.

Census accuracy is perhaps even harder to measure since one never knows the exact underlying true values of the characteristics that the census measures. Even when a person has been counted correctly in the census, are his or her underlying characteristics correctly measured? There are many ways to address this question including whether the initial questions are posed correctly, whether the responses are accurate, and whether the results are correctly edited and tabulated. Measures of reliability from reinterview surveys, of the degree of item non-response that has to be allocated, and of the extent of question editing can all give some indication of accuracy. The second section of this paper discusses census accuracy with special emphasis on the level of allocation of responses.

Although this paper concentrates on the coverage and accuracy of census results, there are other dimensions of census quality. The comprehensiveness of a census is important. Did it measure the important characteristics that should be determined for small areas of geography? Does it provide benchmarks for all or most of the important characteristics for which society needs small area data?

Similarly, was the census done efficiently with the most economical methods and procedures to still produce high quality results? These questions are important but are beyond the scope of this paper.

### TRENDS IN CENSUS COVERAGE

Almost everything known about the size of undercount and the differential undercount comes from the Census Bureau’s own program of coverage evaluation. The Census Bureau has historically used two approaches to measuring the undercount. One method uses births and death records, immigration records and previous censuses to estimate the true population. This estimate is compared to the census count to measure the difference. This method is called Demographic Analysis (see fuller description in Appendix A). In recent decades, the Census Bureau has also conducted special surveys to measure the undercount. Scientific samples of census blocks are interviewed independently of the census enumeration. The results of these interviews are checked against the census records on an individual basis to see who was missed and who was counted in error. The method the Census Bureau uses was called the Post-Enumeration Survey in 1990 and the Accuracy and Coverage Evaluation in 2000 (see fuller description in Appendix A).

In the next section, we use the results of the Demographic Analysis and the coverage measurement survey to review the historical record on the completeness of coverage in the decennial censuses.

#### Historical Trends in Net Coverage, 1940-2000

Table 1 presents historically-consistent estimates of percent net undercount for the decennial censuses from 1940 to 2000. Two significant observations stand out. First, the demographic estimates document the significant decline in net census undercount over the last 60 years. The net undercount in Census 2000 is estimated to have been only 0.1 percent, representing a culmination of a long-term decline from the estimated 5.4 percent in 1940. The estimated undercount has declined for both Blacks (from 8.4 in 1940 to 2.8 percent in 2000) and Nonblacks (5.0 to a 0.3 percent net overcount). The demographic estimates show one aberration in the long-term decline for all groups, the net undercount in 1990 was higher than in 1980, but below 1970 levels.

---

Note: This paper reports the results of research and analysis undertaken by the U.S. Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

The second observation is that despite the overall declines in net undercount, the undercount rate of Blacks has remained persistently higher than the rate of Nonblacks and the net undercounts of males have exceeded the undercounts of females in each census from 1940 and 2000. (The historical data on births, deaths, and immigration needed to calculate the demographic estimates are not available for detailed race groups or for Hispanics). The excess of the net undercount rate of Blacks relative to Nonblacks fluctuated in the range of 3.4 to 4.4 percentage points from 1940 and 1990; the differential dropped to 3.1 points in Census 2000. The gap in the undercount of males relative to females actually increased from 1940 to 1970, and fluctuated between 1.5 and 1.9 percentage points since then.

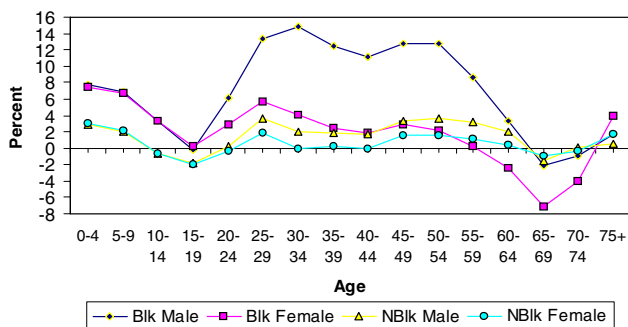
	1940	1950	1960	1970	1980	1990	2000
Total	5.4	4.1	3.1	2.7	1.2	1.6	0.1
Black	8.4	7.5	6.6	6.5	4.5	5.5	2.8
Nonblack	5.0	3.8	2.7	2.2	0.8	1.1	-0.3
Male	5.8	4.4	3.5	3.4	2.2	2.4	0.9
Female	5.0	3.8	2.7	2.0	0.3	0.9	-0.6
Diff. Black-NonBlack	3.4	3.6	3.9	4.3	3.7	4.4	3.1
Diff. Male-Female	0.8	0.6	0.8	1.5	1.9	1.5	1.5

**Age, Sex, and Race Patterns of Undercount**

Patterns of net undercount vary considerably within race and sex groups. Appendix Table 1 and Figure 1 and 2 display the detailed estimates of percent net undercount for race, sex, and 5-year age groups in the 1990 and 2000 censuses. In terms of level of percent undercount, the most notable pattern is the consistently high levels of undercount for Black men ages 20 to 64. The “extreme” differential undercount of adult Black men has been measured in every census since 1940. In 2000, the estimated net undercount ranged between 6 and 11 percent in these age groups. 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. In contrast, undercount estimates for Black adult women are relatively low. In fact, net “overcounts” for Black women aged 65-69 and 70-74 are relatively high. This reflects a case where net content error (persons of other ages reporting as ages 65-69 and 70-74) has as much an effect as coverage error on the observed net undercount

For Nonblack males and females, the net coverage patterns exhibit relatively low levels of net undercount. In fact, the percent net undercount estimates for Nonblack females straddled the zero undercount line for most age groups in 1990 and fell below zero in 2000. The one exception to these high coverage rates is the relative large net undercount of Nonblack children under age 5, especially in 2000.

**Figure 1 Estimated Percent Net Census Undercount by Race, Sex, and Age: 1990**



**Figure 2 Estimated Percent Net Census Undercount by Race, Sex, and Age: 2000**

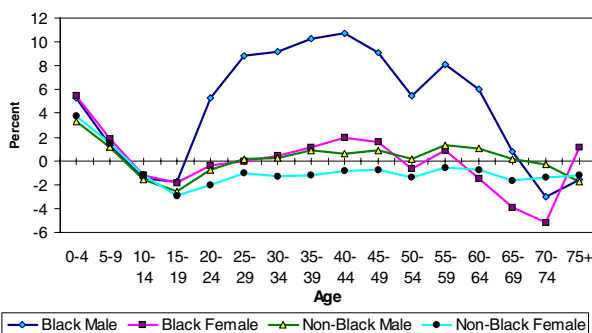
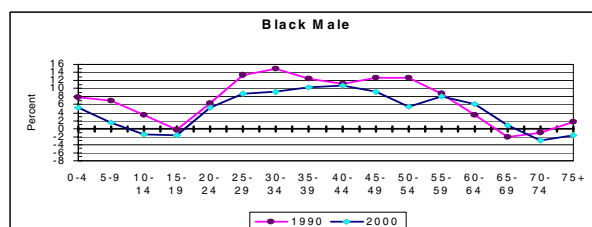
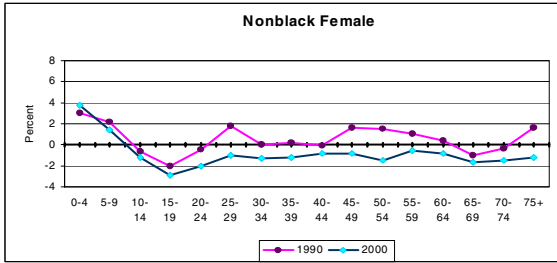
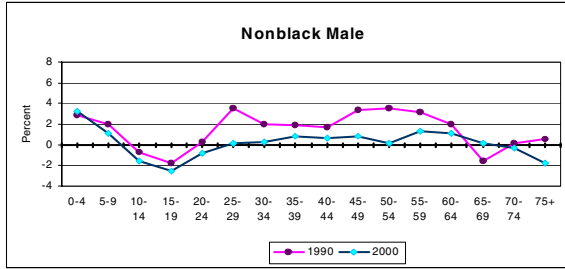
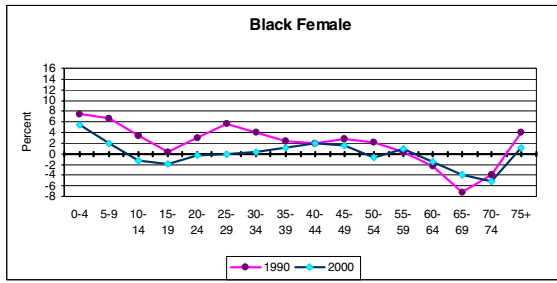


Figure 3 reconfigures the estimates into race-sex groups, where the change in net undercount from 1990 to 2000 is more visible. Compared to 1990 estimates, the DA estimates for 2000 show a broad decline in net census undercount rates for almost all race-sex-age categories. The net undercount rates for adult Black men aged 20-64 remained high in 2000 compared to other groups, though the rates were somewhat reduced from the even higher levels in 1990. The undercount rates for Black male and Black female children fell from 1990 to 2000. For Nonblacks, the net undercount rates were reduced from 1990 with one notable exception—children under age 5. For Nonblack females, ages 0-4 and 5-9 are the only age groups for which net census undercounts were observed; net census overcounts were measured for all age groups 10 and over. For Nonblack males, ages 0-4 is the only age group for which net undercounts of over 1.5 percent were measured in 2000 (in 1990, net undercount rates of 1.5 percent or more were estimated for ten 5-year age groups).

**Figure 3. Revised Demographic Analysis Estimates of Percent Net Census Undercount: 1990 and 2000**





**Table 2. Demographic Analysis Estimates of Percent Net Undercount, by Race, Sex, and Age: 1960-2000**

Race, sex, age	1960	1970	1980	1990	2000
Total	3.1	2.7	1.2	1.6	0.1
Males 20-64	4.4	4.4	3.4	3.4	1.5
Ages 0-9	2.4	3.4	1.7	3.3	2.6
Other	2.7	1.7	0.0	0.3	-1.2
Black	6.6	6.5	4.5	5.5	2.8
Black males 20-64	13.1	13.1	11.3	11.3	8.4
Black 0-9 (male & female)	5.4	8.1	7.0	7.2	3.4
Other Black	3.9	2.9	0.4	1.9	-0.4
NonBlack	2.7	2.2	0.8	1.1	-0.3
NonBlack males 20-64	3.4	3.3	2.4	2.3	0.5
NonBlack 0-9 (male & female)	1.9	2.6	0.7	2.5	2.4
Other NonBlack	2.6	1.6	0.0	0.1	-1.3

These dichotomous patterns of relatively “high” and “low” undercounts among Black subgroups in 1990 repeated the pattern from 1960-1980. Indeed, crude historical estimates of percent net undercount that extend back to 1880 demonstrate the intractable nature of the undercount for certain groups, especially adult Black men.

The net undercount of Blacks dropped overall by almost one-half from 1990 to 2000—from 5.5 percent to 2.8 percent. The undercount rates fell for all groups; the undercount rates of Black men (8.4 percent) and Black children (3.4 percent) remained high relative to overall levels (0.1 percent), but the differentials were reduced from the historically high rates for these groups.

Compared to the estimated net undercount rates for Blacks, the rates for Nonblacks have not been relatively high for any broad age group over the 1960 to 2000 period. The net undercount rates of Nonblack groups differed by less than 1.5 percentage points from the overall rate in each census with one exception (the net undercount rate of 2.4 percent for Nonblack children aged 0-9 in 2000).

**Findings from Coverage Measurement Surveys**

The persistence of the differential net undercounts documented by Demographic Analysis estimates is also conveyed by results from coverage measurement survey-based estimates, including the 1990 Post-Enumeration Survey (PES) and 2000 Accuracy and Coverage Evaluation (A.C.E.). In addition to the historical undercounts of Blacks, the 1990 survey results also reveal disproportionately high net undercount for groups such as Hispanics, American Indians on reservations, and persons living in renter-occupied housing. With the survey estimates, the “differential” undercount can be refined to indicate the excess undercount of these groups compared to Non-Hispanic Whites. Comparison of the 1990 PES and 2000 A.C.E. results provide important information on the reduction of differentials in Census 2000.

**A Closer Look at the Patterns of Differential Undercounts**

With the demographic analysis estimates, we will take a closer look at patterns of differential undercounts over the last several decades. For the 1990 census and Census 2000 we can use the census coverage measurement surveys to examine the patterns of differential undercount for more detailed groupings than with the DA estimates.

**Findings from Demographic Analysis**

Table 2 gives a perspective of how the relatively high net undercounts of Black adult men and Black children contribute disproportionately to the well-known differential undercount of Blacks. In 1990, the Black undercount of 5.5 percent was really a weighted average of two very different net undercount levels: (1) the estimated 11.3 percent net undercount of Black men ages 20-64 and 7.2 percent undercount of Black children under 10, and (2) the estimated net undercount of 1.9 percent for other Black groups (males 10-19, 65+, females 10+). Contrasted to the total undercount rate of 1.6 percent in 1990, the undercount of Black adult men and Black children was disproportionately high, while the net undercount of other Blacks did not exhibit any appreciable differential.

As shown in Table 3 (and illustrated in Figures 1 and 2), Census 2000 markedly reduced the differential net undercount. In 1990, the PES estimated relatively large net undercounts of most groups other than Non-Hispanic Whites. The net undercount of all other groups as a whole was 4.51 percent at the national level, significantly higher than the 0.68 percent for Non-Hispanic Whites. Similarly, the net undercount of persons living in renter-occupied housing (4.51 percent) was disproportionately high relative to the overall negligible undercount in owner-occupied units (0.04 percent). The existence of these differential net undercounts in the 1990 census was a major factor in the decision to adjust national population controls for surveys, such as the Current Population Survey.

The pattern of net undercount rates in Census 2000 as measured by the A.C.E. Revision II estimates is different from that in 1990. Not only was the overall estimated net undercount reduced markedly (from a 1.61 percent net undercount in 1990 to 0.49 percent net overcount in 2000), but the differentials between groups was narrowed considerably. For example, while the net coverage for Non-Hispanic Whites dropped from a net undercount of 0.68 percent in 1990 to a net overcount of 1.13 percent in 2000 (a change of 1.81 percentage points), the net undercount for all other groups fell from 4.51 percent in 1990 to 1.01 percent in 2000, a change of 3.50 percentage points. The reduction in the differential undercount of renters compared to owners is also notable. In fact, none of the individual groups in Table 3 shows estimated net undercounts of more than 2 percent in 2000, while all had net undercounts of over 2 percent in 1990 (exclusive of Non-Hispanic Whites and owners.)

Characteristic	Census 2000 Population	Percent Net Undercount (standard error in parenthesis)		
		1990 PES	2000 A.C.E. Revision II	Change: 1990 to 2000
Total	273,587	1.61 (0.20)	-0.49 (0.20)	2.10
Race/Hispanic Origin				
Non-Hispanic White	192,924	0.68 (0.22)	-1.13 (0.20)	1.81
All other groups	80663	4.51 (0.46)	1.01 (0.32)	3.50
Non-Hispanic Black	33,470	4.57 (0.55)	1.84 (0.43)	2.73
Hispanic	34,538	4.99 (0.82)	0.71 (0.44)	4.28
Non-Hispanic Asian	9,960	2.36 (1.39)	-0.75 (0.68)	3.11
AI on Reservation	540	12.22 (5.29)	-0.88 (1.53)	13.10
Tenure:				
Renter occupied	85,662	4.51 (0.43)	1.14 (0.36)	3.37
Owner occupied	187,925	0.04 (0.21)	-1.25 (0.20)	1.29

**Net Coverage Error and Its Underlying Components**

The estimates of coverage error discussed here measure net undercount in the census. They do not tell us about the separate components of net coverage error (omissions, erroneous inclusions and duplicates) or net content error. This distinction is important, because the amount of net undercount could be smaller than either the omissions or duplicates, and the trends in the omission or duplication rate could differ from the trend in the net undercount rate from one census to the next. For example, the net undercount of 338,000 or 0.1 percent in Census 2000 based on Demographic Analysis masks omissions and duplicates that number in the millions. Limited information on the magnitude of omissions or erroneous enumerations/duplicates can be indirectly obtained from coverage measurement surveys (such as the PES in 1990 or A.C.E. in 2000). Demographic Analysis provides only estimates of net undercount.

Results from research documents that support the A.C.E. Revision II estimates provide new information on the magnitude of omission or erroneous enumeration in Census 2000. One study estimated 5.8 million duplicated persons in Census 2000, and there is evidence that the number of

erroneous enumerations was much higher than that estimate (see U.S. Bureau of the Census, 2003 and Mule, 2002). Given the low level of net undercount, inferences can be made about the magnitude of omissions of persons in the census. Since a consistently defined set of omission or erroneous enumeration for 1990 and 2000 are not available, we can not make statements about the effect of changes in the levels of these offsetting components on the change in net undercount from one census to the next.

**ALLOCATION FOR ITEM NONRESPONSE**

**Historical background, 1940 and 1950**

While population coverage has been a major issue throughout much of the 210-year history of the U.S. decennial census of population, nonresponse on population characteristics for the enumerated population has been a major issue for only the last few censuses. Item nonresponse rates on population characteristics were relatively low through 1950, reflecting the use of direct interviews to enumerate each household, and rose significantly starting in 1960. Beginning in 1960, sample data were collected using a mail-back procedure (after enumerators left the sample questionnaire at each selected household). A mail-out/mail-back procedure now accounts for most data

collection in the decennial census. In this process, many questionnaires are returned with some questions not answered (item nonresponse).

It should be noted that item nonresponse and allocation for item nonresponse both refer to nonresponse remaining after the completion of field office editing. This operation, which may include both telephone and personal visit follow up, reduces the level of item nonresponse from that existing after census questionnaires are completed by enumerators or respondents.<sup>1</sup>

The 1940 census, frequently noted for the first use of sampling in the decennial census, was also the first in which item nonresponse was allocated for any characteristic. For example, allocation was used for age in 1940 and 1950, even though the nonresponse rates were extremely low: 0.16 percent and 0.19 percent, respectively (U.S. Bureau of the Census, 1953, p. 36). When age was not reported, it was allocated based on “other available information such as marital status, school attendance, employment status, age of other members of the family, and type of household.” The term “allocation” first appeared in 1960 census reports. Census reports for 1940 and 1950 used terms such as “assigned” or “estimated” for what is now called allocation.

Item nonresponse was handled in three basic ways in the 1950 census: (1) allocated for very few characteristics (perhaps just age and marital status); (2) allocated to a category for some characteristics (e.g., nonresponses on place of birth were assumed to be native, nonresponses on school enrollment in general were assumed to be enrolled if in the age group 5 to 17 and not enrolled otherwise, and nonresponses on labor force status were assumed to be not in the labor force); and (3) left in a separate “not reported” category for some characteristics (examples shown below). For some characteristics, such as sex and race, there apparently was no nonresponse when data collection was by direct enumeration.

While 1950 census reports did not include tables on item nonresponse rates, their relatively low levels are evidenced by a perusal of tables in the U.S. Summary Report for 1950. A few examples are listed below:

- Country of birth of the foreign-born population – 0.8 percent
- Citizenship status of the foreign-born population 21 years and over – 0.7 percent
- Years of school completed for the population 25 years and over – 2.7 percent
- Residence in 1949 for the population 1 year and over – 1.6 percent
- Occupation of the employed population – 1.3 percent
- Industry of the employed population – 1.5 percent
- Family income – 4.7 percent

<sup>1</sup> The data in this paper on allocation for item nonresponse exclude substitution, which refers to “the assignment of a full set of 100-percent population characteristics for all persons in a household.

### First extensive allocation, 1960

With the shift in the procedure for data collection from direct enumeration in 1950 to self-enumeration in 1960 (see next paragraph), item nonresponse rates rose, with a commensurate shift toward allocating for nonresponse in editing data (i.e., after field-office editing, as described earlier). The 1960 census reports were the first to include tables showing data on nonresponse and allocation. Nonresponse was defined as “the absence of an entry on the schedule or an entry that was poorly marked or was otherwise not readable by the mechanical equipment, or an entry that was inconsistent with another entry.” Allocation was defined as assigning a characteristic “during tabulation because of nonresponse” (U.S. Census Bureau, 1964, p. 343).

In 1960, allocation rates were quite low on the five population items asked on a 100-percent basis, reflecting the fact that while an Advance Census Report (ACR) was mailed to every household, enumerators visited each household, reviewed the ACR, and asked any questions necessary to complete the ACR before leaving a mail-back questionnaire for households in the 25-percent sample. Values for one or more 100-percent items were allocated for 3.0 percent of the population, and allocation rates for the five individual items ranged from 0.4 percent to 1.7 percent (Table 4).<sup>2</sup>

For sample items, which represented the vast majority of items in the census starting in 1960 (U.S. Census Bureau, 2002b, pp. 122-125), allocation rates were significantly higher. Values for one or more sample items were allocated for 18.9 percent of the population. Allocation rates for the six sample items shown for 1960 in Table 4 ranged from 2.5 percent to 10.6 percent. There are several items for which nonresponse rates in 1960 are not available because nonresponse was not allocated (e.g., means of transportation to work) or only partly allocated (e.g., residence 5 years ago), or because nonresponse was allocated but the rate was not published in general allocation tables (e.g., weeks worked in the preceding calendar year).

### Trends in allocation rates, 1960 to 1990

From 1960 to 1990, nonresponse was allocated for all of the 100-percent population items (five or six in each census), and data on allocation are generally comparable for this period. However, there was an increase in the number of sample population items for which nonresponse was allocated, from 6 to 14 among items included in Table 4 (see footnote 3). This increase presumably reflects the general increase in nonresponse (which probably increases bias) and a corresponding willingness to allocate among a large number of qualitative categories (e.g., state of residence 5 years ago) as well as quantitative categories (e.g., age).

<sup>2</sup>

The selected characteristics included in Table 4 in general are those for which information is available on allocation rates in the sources for Table 4 for at least three of the five censuses in the 1960-2000 period. Two exceptions reflect cases where there was a change in the category for which the allocation rate is available (nativity status changed to place of birth) or in the question itself (years of school completed changed to highest degree or level).

**Table 4. Allocation Rates for Nonresponse on Selected Population Characteristics in the Decennial Census of Population of the United States: 1960 to 2000**

(Percent of individuals, except for family income and household income. For information on confidentiality protection, coverage, sampling error, and nonsampling error, see source)

Characteristic (age group and other information)	Includes assignment <sup>1</sup>				2000 (excludes assignment) <sup>1</sup>
	1960	1970	1980	1990	
<b>100-PERCENT CHARACTERISTICS</b>					
<b>One or more items allocated (reflects changes in items, 1960-1990) .....</b>	<b>3.0</b>	<b>7.2</b>	<b>10.2</b>	<b>16.2</b>	<sup>2</sup> <b>9.7</b>
Sex (all ages) ..... ( *)	0.4	0.9	0.8	1.2	<sup>2</sup> 1.1
Age (all ages) ..... ( *)	1.7	2.7	2.9	2.4	3.7
Race (all ages) ..... ( *)	0.4	1.5	1.5	2.0	<sup>2</sup> 4.1
Origin, Hispanic type or not Hispanic (all ages) .....	(X)	(X)	4.2	10.0	<sup>2</sup> 4.4
Relationship (all ages) ..... ( *)	0.5	2.1	2.1	2.6	<sup>2</sup> 2.1
Marital status (14+ in 1960-1970, 15+ in 1980-1990) ..... ( *)	0.9	1.8	1.3	2.0	(X)
<b>100-percent characteristics shown for all years, 1960-1990 ..... ( *)</b>					
Number of characteristics .....	5.0	5.0	5.0	5.0	(X)
Median allocation rate .....	0.5	1.8	1.5	2.0	(X)
Average allocation rate .....	0.8	1.8	1.7	2.0	(X)
<b>SAMPLE CHARACTERISTICS .....</b>					
<b>One or more items allocated (reflects changes in items, 1960-1990) .....</b>	<b>18.9</b>	<b>34.7</b>	<b>44.9</b>	<b>53.3</b>	<sup>2</sup> <b>49.1</b>
Marital status (15+ in 2000) .....	(X)	(X)	(X)	(X)	<sup>2</sup> 2.8
Nativity status (all ages) .....	2.5	4.8	(NA)	(NA)	(NA)
Place of birth (all ages) .....	(NA)	(NA)	4.9	5.4	10.1
Language spoken at home (5+) .....	(X)	(X)	8.2	6.9	<sup>2</sup> 7.3
Residence 5 years ago (5+) .....	(NA)	(NA)	8.0	4.0	<sup>2</sup> 5.1
School enrollment (5-34 in 1960, 3-34 in 1970, 3+ in 1980-2000) ..... ( *)	8.3	5.8	4.5	4.5	<sup>2</sup> 6.9
Educational attainment, years of school completed (25+) .....	4.9	7.7	9.3	(X)	(X)
Educational attainment, highest degree or level (25+) .....	(X)	(X)	(X)	4.6	7.2
Means of transportation to work (workers 16+) .....	(NA)	(NA)	5.4	4.7	<sup>2</sup> 7.6
Labor force status (14+ in 1960-1970, 16+ in 1980-2000) ..... ( *)	3.1	5.1	3.9	3.9	<sup>2</sup> 10.9
Occupation (employed 16+, civilian only in 1990 and 2000) .....	(NA)	7.7	6.7	7.1	<sup>2</sup> 11.3
Industry (employed 16+, civilian only in 1990 and 2000) .....	(NA)	6.9	6.7	5.9	<sup>2</sup> 11.3
Weeks worked in preceding calendar year (workers 16+) .....	(NA)	7.3	9.4	14.9	<sup>2</sup> 19.8
Individual income (14+ in 1960-1970, 15+ in 1980-2000) ..... ( *)	6.2	12.5	11.5	14.2	30.8
Family income (all families) ..... ( *)	<sup>3</sup> 10.6	20.7	17.3	20.3	39.2
Household income (all households, family and nonfamily) .....	(NA)	(NA)	16.8	18.9	37.7
<b>Sample characteristics shown for all years, 1960-1990 ..... ( *)</b>					
Number of characteristics .....	4.0	4.0	4.0	4.0	(X)
Median allocation rate .....	7.3	9.2	8.0	9.4	(X)
Average allocation rate .....	7.1	11.0	9.3	10.7	(X)

(X) Not applicable. Question not asked, changed from 100-percent to sample (marital status), or not comparable to data for years prior to 2000. (NA) Not available. Nonresponse not allocated or partially allocated, or allocation rate not shown in source. \* Allocation rate shown for all years, 1960 to 1990.

<sup>1</sup> "Assignment" means that the missing or inconsistent answer was replaced with an answer based on related information from the same person. In 2000, the term "allocation" was limited to replacement answers based on information from a different person (in the same housing unit or in a different housing unit).

<sup>2</sup> Not comparable with allocation rate for previous years due to "assignment." See footnote 1.

<sup>3</sup> Revised from 11.7 in source table. See U.S. Bureau of the Census, 1964, p. LXXXVII.

Source: U.S. Bureau of the Census, 1964, 1972a, 1972b, 1983a, 1983b, 1992, 1993, and U. S. Census Bureau, 2001 and 2002a.

For the 100-percent population items, the frequency with which one or more were allocated rose from 3.0 percent of the population in 1960 to 16.2 percent in 1990, although it should be noted that there was some change in the 100-percent population items during this period. In other words, the proportion of individuals for whom the reported 100-percent data were completely acceptable dropped from 97.0 percent in 1960 to 83.8 percent in 1990. For the five 100-percent items included in each census in the 1960 to 1990 period, the median allocation rate increased from 0.5 percent to 2.0 percent, and the average allocation rate rose from 0.8 percent to 2.0 percent.

In general, allocation rates for 100-percent items either increased somewhat or remained about the same from one census to the next in the period from 1960 to 1990. The major exception to this generalization is the allocation rate for origin (Hispanic type or not Hispanic) for which the allocation rate jumped from 4.2 percent in 1980 to 10.0 percent in 1990.

The primary reason for the increase in the allocation rate for origin from 1980 to 1990 appears to have been a change in the procedure for failed edit follow-up. In 1980, an attempt was made to contact (by telephone or personal visit) respondents on all forms with incomplete responses on 100-percent items (U.S. Bureau of the Census, 1986, p. 5-27); however, in 1990 this procedure was performed for all incomplete long forms, but for only about one-tenth of incomplete short forms (U.S. Bureau of the Census, 1995, p. 6-29).

The relatively high allocation rates for origin (among 100-percent items) in 1980 and 1990 appear to be due in part to the order of the 100-percent questions on race and origin. In both of these censuses, the question on race (#4) preceded the question on origin (#7). After answering the race question, some non-Hispanics then did not answer the question on origin. In addition, some individuals of Hispanic origin indicated their Hispanic origin in answering the race question (reflecting ambiguity about the concept of race) and left the origin question blank.

For the sample items, the frequency with which one or more were allocated rose from 18.9 percent of the population in 1960 to 53.3 percent in 1990, although it should be noted that there was some change in the sample items during this period. In other words, the proportion of individuals for whom the reported census sample data were completely acceptable dropped from 81.1 percent in 1960 to only 46.7 percent in 1990. For the four sample items included in each census in the 1960 to 1990 period – admittedly a small proportion of the sample items – the median allocation rate increased from 7.3 percent to 9.4 percent, and the average allocation rate increased from 7.1 percent to 10.7 percent.

### Trends in allocation rates, 1990 to 2000

Published data on allocation for 2000 are not directly comparable to data on allocation for the 1960 to 1990 period. What constituted an allocation from 1960 to 1990 was divided into two categories for 2000: assignment and allocation (Love, 2003). Assignment means that the answer to the item was missing or inconsistent and related information from the same person or housing unit response record was used to determine the answer.

Under the new Census 2000 edit rules, the term allocation applies only to situations in which the answer to the item was missing or inconsistent and was determined from information on a different person response record in the same household or from information from a different housing unit record. Thus for all characteristics for which assignment was used in 2000, the change in allocation rates from 1990 to 2000, as shown in Table 4, understates the real change.<sup>3</sup>

The memorandum cited on assignment and allocation (Love, 2003) provides detailed estimates of comparable allocation rates for 1990 and 2000 for sample data on the household population. However, for the more general purposes of this paper, where data on allocation for the 1960 to 1990 period are for 100-percent data and sample data for the resident population (household population and group quarters population combined), allocation rates for nonresponse for 2000 are shown in Table 4 as defined for 2000.

For those items for which assignment was used in 2000 (as shown in Table 4 with a footnote), the allocation rates for 2000 understate the change from 1990 (i.e., understate the increase, or indicate a decrease that may not be real). However, there are several items for which assignment was not used in 2000 and for which allocation rates for 1990 and 2000 thus are generally comparable.

For all six items in Table 4 for which allocation rates for 1990 and 2000 are generally comparable (age, place of birth, educational attainment, individual income, family income, and household income), the allocation rate increased between 1990 and 2000.<sup>4</sup> For example, the allocation rate rose from 2.4 percent to 3.7 percent for age and from 20.3 percent to 39.2 percent for family income. In addition, the memorandum cited above shows that for sample data for the household population, the allocation rate (adjusted for assignment in 2000) increased from 1990 to 2000 for the vast majority of population (and housing) items for which the census questions were comparable (Love, 2003).

As discussed earlier, the nonresponse rate was particularly high for the question on Hispanic origin in 1990. For Census 2000, the order of the race and origin questions was reversed (origin #5 and race #6), and an explicit instruction was included to answer both of these questions. These changes account for at least part of the decline in the nonresponse rate (and thus in the allocation rate for nonresponse) for the origin item from 1990 to 2000 (U.S. Census Bureau, 2001, Technical Documentation, Appendix B).

<sup>3</sup> For the 100-percent and sample items, the 1990 rates are 16.2 percent and 53.3 percent, respectively. The published, noncomparable rates for 2000 are 9.7 percent and 49.1 percent, respectively. The question is how much higher will comparable rates be? The guess is that the 2000 rate for 100-percent items will not be as high as 16.2 percent, in part because of fewer 100-percent items in 2000 and in part because the 1990 rate was abnormally high due to the high rate of nonresponse (and allocation) for Hispanic origin, as discussed in the paper. For sample items, the guess is that the comparable rate for 2000 will be higher than 53.3 percent.

<sup>4</sup> Marital status is not included here because it was a 100-percent item in 1990 and a sample item in 2000.

The published allocation rates for 2000 for one or more items allocated for nonresponse are 9.7 percent for 100-percent items and 49.1 percent for sample items, as shown in Table 4. As discussed previously, these rates are not comparable to rates for 1990 and overstate the decline from 1990 to 2000. When these rates for 2000 are adjusted to be consistent with rates for 1990, they increase to 11.0 percent and 52.7 percent, respectively. Thus, on a comparable basis, the allocation rate for nonresponse for one or more items declined from 16.2 percent to 11.0 percent for 100-percent items. The corresponding rate for sample items was essentially unchanged: 53.3 percent in 1990 and 52.7 percent in 2000. The sizable decline in the rate for 100-percent items reflects the decline between 1990 and 2000 in the nonresponse rate to the question on origin, as discussed previously.

## CONCLUSIONS

So was Census 2000 the best census ever? In terms of estimated net coverage and differential net coverage, it certainly appears to be so. There is still the remaining question of the amount of gross coverage error (omissions and duplicates) and how that has changed over time.

The trends in census accuracy for characteristic data are not as positive. Allocation rates for item nonresponse have generally increased, meaning that more of the results rely upon the appropriateness of edit and allocation rules. Plans to replace the long-form questionnaire in the 2010 census with the American Community Survey (ACS) may reduce item nonresponse rates (and thus the need to allocate for nonresponse) because the ACS includes a nonresponse follow-up operation, unlike Census 2000. In addition, the fact that the ACS is ongoing means that the enumerators will be more experienced than were enumerators in Census 2000, and this may help to reduce item response rates in comparison to those in Census 2000.

## REFERENCES

- Das Gupta, Prithwis. 1991. "Models for Assessing Errors in Undercount Rates Based on Demographic Analysis," Preliminary Research and Evaluation Memorandum No. 84, U.S. Bureau of the Census.
- Hogan, Howard. 1993. "The 1990 Post-Enumeration Survey: Operations and Results," *Journal of the American Statistical Association*, Vol. 88, No. 423, pp. 1047-1060.
- Kostanich, Donna 2002. "A.C.E. Revision II: Design and Methodology," DSSD A.C.E. Revision II Memorandum Series #PP-35, U.S. Bureau of the Census.
- Love, Susan. 2003. "Comparison of Item Imputation Levels by Mode: Census 2000, Census 2000 Supplementary Survey, and the 1990 Census," Internal U.S. Census Bureau memorandum, April 24.
- Mule, Thomas. 2002. "Revised A.C.E. Results: Further Study of Person Duplication," DSSD A.C.E. Revision II Memorandum Series #PP-51, U.S. Bureau of the Census.
- Robinson, J. Gregory, Arjun Adlakha, Kirsten K. West. 2002. "Coverage of Population in Census 2000: Results from Demographic Analysis", Paper presented at the Annual Meeting of the Population Association of America, Atlanta, Georgia, May 8-11.
- U.S. Bureau of the Census. 1953. *Census of Population: 1950, Vol. II, Characteristics of the Population, Part 1, United States Summary*.
- \_\_\_\_\_. 1964. *Census of Population: 1960, Vol. I, Characteristics of the Population, Part 1, United States Summary*.
- \_\_\_\_\_. 1972a. *Census of Population: 1970, General Population Characteristics, Final Report PC(1)-B1, United States Summary*.
- \_\_\_\_\_. 1972b. *Census of Population: 1970, General Social and Economic Characteristics, Final Report PC(1)-C1, United States Summary*.
- \_\_\_\_\_. 1983a. *Census of Population: 1980, General Population Characteristics, Part 1, United States Summary, PC80-1-B1*.
- \_\_\_\_\_. 1983b. *Census of Population: 1980, General Social and Economic Characteristics, Part 1, United States Summary, PC80-1-C1*.
- \_\_\_\_\_. 1986. *Census of Population and Housing: 1980, History, PHC80-R-2*.
- \_\_\_\_\_. 1992. *Census of Population: 1990, General Population Characteristics, United States, 1990 CP-1-1*.
- \_\_\_\_\_. 1993. *Census of Population: 1990, Social and Economic Characteristics, United States, 1990 CP-2-1*.
- \_\_\_\_\_. 1995. *Census of Population and Housing: 1990, History, 1990 CPH-R-2*.
- U.S. Census Bureau. 2001. *Census 2000, Summary File 1, on the Census Bureau Internet site (www.census.gov)*.
- \_\_\_\_\_. 2002a. *Census 2000, Summary File 3, on the Census Bureau Internet site (www.census.gov)*.
- \_\_\_\_\_. 2002b. *Measuring America: The Decennial Censuses From 1790 to 2000*, by Jason G. Gauthier.
- \_\_\_\_\_. 2003. "Technical Assessment of A.C.E. Revision II, March 12, 2003.

## APPENDIX A: METHODS OF MEASURING

### COMPLETENESS OF COVERAGE IN THE CENSUS 1. DEMOGRAPHIC ANALYSIS

#### Description of the Method

In general, the demographic method of estimating coverage involves developing estimates for the population at the census date by the analysis of various types of demographic data essentially independent of the census, such as birth, death, and immigration statistics, as well as emigration estimates, and Medicare data. The difference between the estimated population (P) and the census count (C) measures the net census undercount, u, and net undercount rate, r (see equations 1 and 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 sample of the population.

#### Component Age Groups

The particular analytic procedure used to estimate coverage nationally for the various demographic subgroups depends primarily on the nature and availability of the required demographic data. Two principle demographic techniques were used to produce the demographic analysis estimates for 1990, one for the population under age 65 and another for the population 65 and over.

(1) Ages under 65. The demographic analysis estimates for the population below age 65 in 1990 are based on the compilation of historical estimates of the components of population change: births (B), deaths (D), immigration (I), and emigration (E). Presuming that the components are accurately measured, the population estimates (P) are derived by the basic demographic accounting equation applied to each cohort:

$$P_{0-64} = B - D + I - E \quad (1)$$

The actual calculations are carried out for single-year age cohorts. 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.

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

$$P_{65+} = M + m, \quad (2)$$

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.

#### Development of Historical Estimates for Multiple Censuses

The foundation of the demographic method is the logical consistency and relation of the underlying demographic data. With the use of components of change (births, deaths, net immigration) the estimated population for a birth cohort can be carried forward through time to derive estimates of net undercount in a series of censuses as the cohort ages (e.g., age 0-4 in 1950, age 10-14 in 1960, 20-24 in 1970, 30-34 in 1980, and 40-44 in 1990). Similarly, an older cohort in 1990 based on Medicare data can be carried backward in time to derive estimates for the cohort at younger ages (e.g., 65-69 in 1990, 55-59 in 1980, 45-49 in 1970, 35-39 in 1960, etc.). In this way, consistent estimates of net undercount for 1940 to 1990 based on demographic analysis are produced.

These multiple series of net undercount estimates for cohorts across censuses are linked through the components of population change. This linkage of the estimates provides a consistent basis to judge changes in patterns of coverage over time and to assess the plausibility of the demographic estimates themselves (see Passel, 1991).

#### Limitations of the Demographic Estimates

The aggregate administrative data and estimates that are incorporated in equations 1 and 2 are corrected for various types of errors. Many assumptions go into this estimation process, some of which can be validated and some which cannot.

The overall accuracy of the demographic estimates depends on the quality of the demographic data and corrections. Research has been conducted in the past few years to develop methods for assessing the uncertainty of the demographic coverage estimates (see Das Gupta, 1991). This work demonstrates that 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 undercount according to age, sex, and race and measuring changes in net undercount between censuses. This greater confidence in statements describing differences in undercount between groups is important to make, because coverage differentials are the focus of this paper.

Finally, it should be noted that the principal demographic estimates for race, sex, and age groups measure net undercount in the census. They don't tell us about the separate effects of net coverage error (omissions, erroneous inclusions) or net content error.

For more information of the development of the demographic analysis estimates, see Robinson et al (2002).

**2. COVERAGE MEASUREMENT SURVEYS**

The second coverage measurement method uses a sample survey and the dual system estimation methodology to estimate coverage error to measure net undercounts. This approach involves case-by-case matching of persons in an independent survey with persons in the census to determine who was missed or counted in error. The survey-based coverage measurement program associated with Census 2000 is known as the Accuracy and Coverage Evaluation (A.C.E), in the 1990 Census it was called the Post-Enumeration Survey (PES); for the 1980 Census was called the Post-Enumeration Program (PEP). While the 2000 and 1990 coverage surveys produced one “preferred” set of coverage estimates, twelve different estimates were generated for the 1980 PEP program with none being deemed “preferred”. Therefore, the 1980 estimates are not included in this paper. The 1990 PES and 2000 A.C.E. PES consisted of two parts. The first part was a

sample of the population, known as the P sample. The proportion of the P sample that was included in the census is an estimate of the proportion of the total population that was included in the census. The second part consisted of a sample of the census enumerations used to estimate the proportion of erroneous census enumerations. This sample is known as the E sample. These enumerations were checked against the census itself to determine the extent of duplication. They were also checked in the field to determine the extent of fictitious enumerations, inclusions by the census of people born after the census reference day, and the extent to which people were counted in the wrong location. Several other steps are involved in the generation of the coverage measurement estimates, including the poststratification of the population into geographic and demographic groups (such as age, sex, race, origin, tenure). For a description of the 2000 A.C.E. estimates and results, see U.S. Bureau of the Census (2002). For a description of the 1990 PES estimates and results, see Hogan (1993).

**Appendix Table 1. Demographic Analysis Estimates of Percent Net Undercount by Age, Sex, and Race: 1990 and 2000**

Age groups	Black Male		Black Female		Nonblack Male		Nonblack Female	
	1990	2000	1990	2000	1990	2000	1990	2000
All ages	8.1	5.1	3.1	0.5	1.6	0.2	0.6	-0.8
0-4	7.8	5.3	7.4	5.4	2.9	3.3	3.0	3.8
5-9	6.9	1.4	6.7	1.9	2.0	1.1	2.2	1.5
10-14	3.4	-1.5	3.4	-1.2	-0.7	-1.6	-0.7	-1.2
15-19	-0.2	-1.7	0.3	-1.8	-1.8	-2.6	-2.0	-2.9
20-24	6.2	5.3	2.9	-0.4	0.2	-0.8	-0.4	-2.0
25-29	13.4	8.8	5.7	0.0	3.6	0.2	1.8	-1.0
30-34	14.8	9.2	4.1	0.4	2.0	0.2	0.0	-1.3
35-39	12.5	10.3	2.4	1.2	1.9	0.8	0.2	-1.2
40-44	11.2	10.8	1.9	2.0	1.7	0.6	-0.1	-0.9
45-49	12.8	9.1	2.9	1.6	3.3	0.8	1.6	-0.8
50-54	12.7	5.5	2.2	-0.7	3.6	0.2	1.5	-1.4
55-59	8.7	8.1	0.3	0.9	3.2	1.4	1.1	-0.5
60-64	3.4	6.0	-2.4	-1.5	2.0	1.1	0.4	-0.8
65-69	-2.1	0.8	-7.1	-3.9	-1.5	0.2	-1.0	-1.7
70-74	-0.9	-3.0	-4.0	-5.2	0.1	-0.3	-0.3	-1.4
75+	1.7	-1.6	4.0	1.1	0.6	-1.7	1.7	-1.2