IMPLICATIONS OF THE 1980 PEP FOR FUTURE CENSUS COVERAGE EVALUATION

Robert E. Fay, U.S. Bureau of the Census

1. INTRODUCTION

The undercount of the population censuses, a concern of the U.S. Bureau of the Census for many years, has become an issue of more intense public and professional interest and debate during the last decade. Much of the attention has focused on whether geographic differentials in the census undercount can be adequately estimated or measured, and whether the census counts or other characteristics should be adjusted to compensate for such differentials. The decision of the Census Bureau is that none of the evaluations of the undercount of the 1980 census are suitable for adjustment. At the same time, the bureau has undertaken a program of research to investigate the feasibility and implications of incorporating an adjustment into the counts of the 1990 census.

The complex issues of census adjustment have been addressed by a number of authors, including Bailar (1983) and Keyfitz (1979). A paper of Ericksen and Kadane (1985) states their own position on the feasibility of adjustment and includes accompanying discussion from a number of other points of view. Part of the debate has centered on the importance of measurement of undercount for geographic units through direct sample survey methods, and on the methodological difficulties and limitations of such an approach. In particular, the specific merits and deficiencies of the undercount study conducted for the 1980 census, the Post-Enumeration Program (PEP), has been part of this public discussion. The intent of this paper is to examine issues in the measurement of net census error by survey methods, as these issues are illustrated by specific problems of the PEP.

No conclusion with respect to census adjustment is offered here, although issues of measurement of net census error cannot be totally divorced from the issue of adjustment, since implicit in any decision to adjust the census is the assumption that a more suitable measurement of the total population may be obtained through means other than unadjusted census counts.

2. A THREE-CLASS MODEL FOR CENSUS OMISSIONS

Earlier studies of census undercount (e.g. U.S. Bureau of the Census 1960, Marks and Waksburg 1966) have remarked on the distinction between within-household and whole-household omissions. Within-household omission denotes failure to include a member or members of a household in which other members are properly included in the census. Obviously, whole-household omission refers to omission from the census of entire households.

Within- and whole-household omissions represent quite different circumstances of census error. In the case of within-household omissions, specific persons are left out by an otherwise cooperating household because of deliberate or accidental concealment, or misunderstanding of census definitions. Whole-household omissions represent failure to find, or to contact and correctly interview a household. Thus, this distinction offers potential insight into underlying causes of the census undercount.

The relative magnitude of these two sources of error is not entirely clear, however. Coverage studies based upon direct sample surveys have tended to point to whole-household omissions as the dominant factor. Demographic analysis of the U.S. population (e.g. Coale and Zelnik 1963, Coale and Rives 1973, Siegel 1974) suggests the possibility that sample survey evaluations have failed to capture large segments of the total undercount, much of which might be due to undetected within-household omissions. Demographic analysis employs statistics on births, deaths, immigration, emigration, Medicare registrations, and the counts of earlier censuses themselves to construct an estimate of the aggregate U.S. population.

Demographic analysis of the 1950, 1960, 1970, and 1980 censuses has consistently suggested higher omission of males than females. For example, in their demographic analysis of the net error of the 1980 census, Passel and Robinson (1985) estimated a net error in the national count of 1.0 percent for the legally resident population, which could be divided into an estimated 5.7 percent rate for Blacks and an estimated 0.4 percent for Whites-and-other races. Separately by sex, however, the estimated rates for Black males were 8.4 percent versus 3.0 percent for Black females, and 1.0 percent for White-and other races males versus -0.3 for White-and other races females. The last estimate denotes a small net overcount.

Table 1 shows detailed estimates of net census error in 1980 developed by Passel and Robinson. The differential in rates of undercount between males and females appears concentrated over the range from 20 to 64 years of age. For Black adults, the differences are particularly dramatic, since the undercount rate for males exceeds the corresponding rate for females by at least 10 percentage points over the range 30-54 years.

Table 1 Demographic Estimates of the Percent Net Undercount by Age, Sex, and Race: 1980

	Black		White-and-Other-Races	
Age	Male	Female	Male	Female
Under 5	9.5	8.9	0.3	0.3
5-9	6.0	5.5	0.3	0.3
10-14	1.7	1.6	-0.5	-0.4
15-19	-0.1	-0.4	-0.5	-0.8
20-24	8.5	2.5	1.1	-0.8
25-29	12.3	3.1	2.1	-0.2
30-34	13.9	3.0	1.3	-1.1
35-39	18.5	6.0	2.5	0.1
40-44	17.1	4.3	1.7	-1.0
45-49	17.7	4.9	2.5	-0.3
50-54	14.2	2.0	2.2	-0.5
55-59	8.6	1.8	2.3	-1.0
60-64	4.8	0.4	1.8	0.0
65-69	-4.0	-7.8	-0.4	-1.6
70-74	-2.5	-3.0	-0.4	-0.5
75+	0.1	6.1	0.7	2.4

Source: Passel and Robinson (1985) Table 3.

Although the evidence on this question is indirect or lacking, the differential undercount of adult males relative to females can be plausibly attributed primarily to within-household omissions from the census. In some cases, these omissions may represent deliberate concealment of persons by reporting census households when the

presence of adult males may affect eligibility for government benefits, or if the adult males wish to conceal their presence for other reasons. In other cases, respondents may simply misinterpret the intent of Census Bureau definitions of residence and omit persons marginally attached to the household, even if the omitted persons have no other address at which to be enumerated. For example, a parent of an adult child returning to the parent's home for an indefinite period of time may decide to exclude the child as not a permanent member of the household, even if the child has no other address at the time. Although the majority of the differential between males and females may be due to within-household omissions, part of the differential may arise from differences by sex of the homeless, who may in practice be difficult to enumerate in the census.

The age-specific undercount rates in Table 1 also suggest a within-household undercoverage of Black children under age 10, since the estimated net error for this age group far exceeds that for adult Black females. Some of this high net error may be attributed to misreporting of age in the census, but the PEP also found an apparent within-household undercount of Black children.

To suggest the potential impact of within-household underenumeration in the census, suppose that the effect of eliminating all within-household underenumeration would be to remove the observed differential between the sexes within each of the two race groups. Under this simplified assumption, underenumeration of females would be attributed solely to omission of whole households, while the underenumeration of males would be viewed as the sum of whole-household omissions occurring at the same rate as females, and within-household omissions accounting for the differential undercount of males. Obviously, this assumption oversimplifies the underlying situation and makes no allowance for the apparent within-household undercount of Black children. Nonetheless, the differential undercount for Blacks relative to Whites-and-other races would be reduced from the estimated values of 5.7 percent vs. 0.4 percent to 3.0 percent vs. -0.3 percent under this assumption. The difference in the estimated rates would drop from 5.3 percentage points to 3.3 percentage points. In other words, the differential would decline by roughly 35 percent. Consequently, the evidence from demographic analysis suggests that within-household omissions may be an important component of the total differential undercount by race.

What may be termed the "two-class" model of census undercoverage is based upon the distinction between within-household and whole-household omissions. Earlier matching studies employed this dichotomy and concluded on the basis of comparisons to demographic analysis that measurement of whole-household underenumeration appeared considerably easier than measurement of within-household underenumeration.

The three-class model to be suggested here is an outgrowth of the mail-out/mail-back census design, under which a mailing list is developed prior to census day, forms are mailed to households, most households respond by mail, and a census follow-up is conducted for housing units not returning forms in order to complete the enumeration. The 1980 census enumerated approximately 95 percent of the population through this means, a much larger proportion than in the 1970 census, which, in turn, was the only other census to incorporate this approach on a large scale. This design contrasts with the conventional census in which an enumerator is assigned to an enumeration district (ED) to canvass the assigned area, list the housing units, and interview the households.

In "mail areas," in which the mail-out/mail-back census is conducted, omissions may be categorized into three groups:

1. Class I - Omissions of persons within enumerated housing units;
2. Class II - Omissions of whole households that had received a census form in the mail but did not respond, or who did not receive a form but lived in a housing unit listed in the census address register for the ED in which they lived;
3. Class III - Omissions of whole households that never received a census form and whose housing units were not listed in the correct address register.

Class I simply represents the within-household omissions as studied earlier. Classes II and III divide whole-household omissions on the basis of their relationship to the mail census. A household in Class III is omitted largely through failure of census procedures. A household in Class II, however, often contributes to its own omission, particularly by not responding to the mail census.

Comparison of the PEP results to demographic analysis implies that Class I remains the most difficult class to measure effectively. Analysis of other evidence from PEP, however, suggests that Class II presents more severe methodological problems than does Class III. Depending upon the design and execution of the census, Class II may represent the largest class of omissions, as it appears to have been in 1980. Recognition of the specific methodological problems associated with Class II may thus prove of interest in the design of future studies of underenumeration. The next three sections of this paper details the findings from the PEP on these three classes.

3. OVERVIEW OF THE 1980 PEP

3.1 Balancing of the P- and E-Samples

The most obvious form of error affecting the population count is that of omission. To measure omissions, a sample as statistically independent of the census as feasible is required. In the terminology of the 1980 PEP, this sample was termed the "P-sample." For 1980, attempts were made to match each P-sample case to an enumeration in the census within a corresponding area of search. In most cases, the area of search was a single designated census ED thought to contain the sample address, but in cases of difficulty in determining the correct ED, the area of search was often expanded to a group of ED's.

The conceptual design underlying the P-sample design in the 1980 PEP was to attempt to classify all P-sample cases either as matched to a complete census enumeration within the corresponding area of search, or as not matched. Problems of missing data interfered with this simple conceptual design, and the models underlying the treatment of missing data were discussed by Fay and Cowan (1983).

The net error of the census is also affected by

problems of erroneous inclusions. This problem was recognized as early as 1950 in the design of the coverage evaluation for that census (U.S. Bureau of the Census 1960). For example, one form of erroneous enumeration is duplication, since multiple enumeration artificially inflates the count. A separate sample, termed the "E-sample" in the 1980 PEP, was selected from the census in order to measure duplications and other forms of erroneous enumeration.

Under the design of the 1980 PEP, restriction to a specified area of search in the P-sample also necessitated estimating the number of census enumerations where the housing unit was included in the census outside of the correct area of search (geocoding error) and the number of census enumerations at the wrong address (a component of definitional error). These two quantities, as well as other forms of definitional error, including fabrication of enumerations by census interviewers, were measured by the E-sample.

Cowan and Fay (1984) further discussed the design of the 1980 PEP and presented results specific to the coverage of the 1980 census. For completeness, the next two sections review aspects of the P- and E-sample design related to the question of the three-class description of census omissions.

3.2 Design of the P-Sample

The P-sample for the 1980 PEP was based upon the April and August, 1980 samples of the Current Population Survey (CPS), the monthly labor force survey conducted by the Census Bureau. Because interviews occur on approximately the third week of the respective months, some interviewed persons, termed "movers," had different April 1, 1980 (census day) addresses than their CPS addresses. Since effects of moves between census day and the CPS interview were substantial by August, a separate question was included in the CPS interview in August to collect census day address.

On the basis of available information, an initial attempt was made to assign P-sample cases to the census ED's in which they should have been enumerated, and to match P-sample cases to census enumerations. The initial effort achieved approximately an 85 percent match rate to the census. The cases unmatched at this point were generally assigned to one of the four pre-follow-up codes:
N1 - Person not matched to the census in a household with persons matching to the census;
N2 - Whole-household not matched to the census, but at a matching address;
N3 - Whole-household not matched to the census, with no address match;
PM - A probable match to a census enumeration.
There were a few additional codes related to the specific implementation of the PEP (Fay and Cowan 1983), but discussion of these will be omitted.

A similarity may be noted between these codes and the three classes defined in Section 2. The N1 cases here generally represent a preliminary classification as a within-household omission. The N2 cases denote that the address, but not necessarily the unit, was found in the census, and include instances in which a unit at a multi-unit address was completely omitted from the census, as well as those in which the unit was in the census but enumerated as vacant or occupied by another household. The N2 category in PEP also encompasses units counted in the census under census "close-out" procedures in which census follow-up interviewers obtained from neighbors minimal information about the status of the unit. The census close-out information was typically inadequate to form the basis of an unambiguous match to the P-sample, so that all such cases were counted as nonmatched in PEP. Cowan and Fay (1984) further explained the treatment of census close-out cases in the PEP estimation. Thus, the N2 category suggests a mixture of Class II and Class III. The code N3 implies a preliminary classification into Class III.

Most such cases, except those presenting special problems of missing data, were then assigned to a "PEP follow-up" in the winter and spring of 1981 to improve the geographic information available to assign the PEP case to the correct ED and to confirm the actual census day address. Incomplete follow-up interviews and cases where insufficient geographic information was obtained from the follow-up for assignment to a census ED represented two of the primary categories of missing data in the P-sample.

For completed cases, the PEP follow-up information became the basis for a final attempted match to the census to assign a final match code. Cases remaining nonmatched were given final match codes N1-N3 with the same definitions as the pre-follow-up codes. A special code, N5, was employed in a few instances when the PEP person appeared to correspond to a census enumeration of a person without the required number of items to be classified as a "data-defined" person, within an otherwise normally enumerated household. An imputation of a final match code was performed for incomplete P-sample cases.

3.3 Design of the E-Sample

The E-sample was sampled from the census to estimate geocoding errors, definitional errors, and duplications. By physically locating each sampled unit, E-sample interviewers spotted the sampled unit on a map to determine if the unit had been enumerated in the correct ED. If not, subsequent operations determined whether the unit was nonetheless in the probable area of search for the P-sample. If the case was outside the probable area of search, the case was classified as a geocoding error.

The measurement of definitional errors included determining whether the enumerated persons in fact existed and, if so, if the address at which they had been enumerated was correct by census definitions. A hierarchy of three sources was employed to assess whether the sampled census enumeration was correct: current occupants of the sampled housing unit, who were often the originally enumerated persons; neighbors or other knowledgeable persons, such as rental agents; and the Post Office. The third source was regarded as a last resort, and the E-sample questionnaire directed interviewers first to inquire whether a change-of-address form established the probable existence of the enumerated household at the address on census day. In the absense of a form, the postal carrier was asked whether mail had ever been delivered.

Duplications within-ED's were estimated by searching for duplicate enumerations for a 50-percent subsample of the E-sample. A later study provided an estimate of the number of between-ED duplications within the P-sample area of search.

4. RESULTS FROM 1980 SUGGESTING A THREE-CLASS MODEL FOR CENSUS OMISSIONS

4.1 P-Sample Estimates of Omissions

Several different assumptions concerning missing data, possible errors in subsets of the data, and response error were considered in the analysis of both the P- and E-samples of PEP. For purposes of discussion here, four sets of P-sample estimates will be examined, based upon the interviewed April CPS sample, the April sample supplemented by information from attempted matches of some of the CPS noninterviewed households, the August CPS, and the August CPS with movers (those with a different census day address from their August CPS address) deleted. Cowan and Fay (1984) further described these four sets.

Table 2 shows estimated nonmatch rates to the census for the four sets. The N2 category is consistently the largest single category of omissions estimated by the PEP. As noted earlier, this category contains most of the Class II omissions from the census, as well as some Class III omissions of individual units in multi-unit structures enumerated in the census.

Table 2 Estimated P-Sample Nonmatch Rates to the 1980 Census By Type

	April Interviews	April w/ Type A Nonint.	August	August Without Movers
All Races				
N1	1.1	1.1	1.5	1.3
N2	3.2	3.4	3.4	2.5
N3	1.0	1.0	1.2	0.9
N5	0.1	0.1	0.1	0.1
Black				
N1	2.8	3.0	3.2	3.0
N2	7.6	7.8	6.4	5.3
N3	1.2	1.2	1.3	0.9
N5	0.2	0.2	0.2	0.2
Non-Black Spanish				
N1	1.8	1.8	2.6	2.4
N2	7.0	7.2	7.9	5.9
N3	1.0	1.0	1.4	0.9
N5	0.2	0.2	0.1	0.1

The relatively low values for the N1's, representing within-household (Class I) omissions, suggest that PEP, like previous matching studies, failed to capture much of the within-household underenumeration implied by demographic analysis. Further examination of the estimates by sex and age confirm this, indicating, for example, that PEP found only a small fraction of the within-household underenumeration of Black males suggested in Table 1.

Variations in the overall nonmatch rate estimated by these four sets of estimates represent an important issue of interpretation and a limitation of the success of PEP. Table 2 also shows that, partly as a consequence of its status as the largest component, variations in the estimates for N2 are principally responsible for the variations in the overall rate among different sets. Thus, a consistent methodology to estimate Class II omissions appears an essential but difficult goal in the design of similar coverage studies.

4.2 Patterns of P-Sample Missing Data

Missing data represented a serious problem in the analysis of the PEP P-sample. The most critical category of missing data was incomplete PEP follow-up interviews, since this category was numerous and the assumptions of the imputation model (Fay and Cowan 1983) assigned a high proportion of these incomplete cases to a final status of "nonmatched."

The overall noninterview rate for follow-up interviews in the April sample was 23 percent, with higher rates for Blacks, 28 percent, and Spanish, 35 percent. Since follow-up was conducted in the winter and spring of 1981, the effect of moves from the original CPS address during the intervening period was a potential contributing factor to the high rates of nonresponse.

To investigate this hypothesis, a longitudinal match or "linkage" was performed between the April 1980 and April 1981 CPS samples for the overlapping half of CPS sample with the same designated housing units. Households linked between the two years may be viewed as probable non-movers during this period, while most nonlinking households can be considered to have moved.

Table 3 presents longitudinal nonlinkage rates of April 1980 PEP cases to the April 1981 CPS. The longitudinal nonlinkage rate for all cases slightly exceeds the typical 20 percent for annual mobility of the U.S. population, implying that most but not all of the longitudinal nonlinkages are true movers. Generally, the longitudinal nonlinkage rates for follow-up cases are larger, but especially so for N2 cases and for N1 cases who are not apparent college students. The longitudinal nonlinkage rates for N3 cases, on the other hand, appear only modestly higher than all cases. Thus, this table indicates that both Class I and Class II have a much higher propensity to move than the population as a whole.

Table 3 Percentages of April Cases Apparently Moving After One Year, By Pre-Follow-Up Status

	Black	Non-Black Spanish	Other
All cases	28	31	23
All follow-up	40	44	34
N1-student	34	37	32
N1-other	54	54	59
N2	42	50	42
N3	33	36	27
PM	30	30	37

Although Table 3 seems to suggest that Class III persons are only slightly more inclined to move than other persons, this interpretation is clouded by the large proportion, approximately 80 percent, of pre-follow-up N3 cases found to match to the census after the follow-up. Thus, it is still possible that the Class III persons included among the pre-follow-up N3's had a more pronounced propensity to move, but that the inclusion of a large number of cases eventually matching to the census among the pre-follow-up cases diluted this effect.

The high noninterview rates on PEP follow-up indeed appear to have been mostly the result of the high mobility of follow-up cases. Among those follow-up cases longitudinally linked to the April 1981 CPS, follow-up response rates were relatively satisfactory, in excess of 90 percent, even for Blacks and Spanish. For longitudinal nonlinkages to the April 1981 CPS, i.e. persons who had probably moved, the follow-up response rates were only

48 percent for Blacks, 40 percent for Spanish, and 61 percent for others. Thus, although the follow-up interview had included procedures under which attempts were made to reach sample persons who had moved at their new addresses, this aspect of follow-up was evidently of limited success.

Table 3 indicates the high propensity to move among follow-up cases over an interval of one year after the census date. Longitudinal linkage to the May, June, and July 1980 CPS affords additional information about the timing of these moves. Table 4 shows longitudinal linkage rates for the population as a whole, for follow-up cases, and for N2 follow-up cases. This table indicates that much of the differential longitudinal nonlinkage rate emerged during the 3 1/2 month period between census day and the July CPS, but also that a differential propensity to move continued even after July. The differential rates of move appear as pronounced during the period between the May and July CPS as during the period between census day and the May CPS.

Table 4 Longitudinal Linkage Rates for the April PEP Sample, As Percents

	May 1980	June 1980	July 1980	April 1981
All Cases	95	92	90	76
All Follow-up	90	84	82	64
Pre-Follow-up N2	88	80	75	58

5. THE EFFECT OF THE MAIL CENSUS

Table 5, presented earlier by Cowan and Fay (1984), displays E-sample results on components of erroneous enumeration for the 5 percent of the population enumerated by conventional census procedures, for households responding by mail in mail areas, and for households not responding by mail but enumerated instead by the census Followup. The conventional census produces a level of definitional error comparable with that for mail returns in mail areas, but levels of duplication and geocoding error are substantially lower than for the mail census. The error rates for nonmail households in mail areas, 8.2 percent, is drastically higher than for mail returns. Furthermore, the close agreement in the estimated number of within-ED duplicates between mail and non-mail returns suggests that most duplicates may have occurred as mail return/non-mail return pairs. Further discussion of this point was given by Cowan and Fay (1984). Thus, the error rate for the census Followup may have approached or even exceeded 10 percent.

Although the estimates in Table 5 pertain to errors of erroneous enumeration in the census, they portray the context in which Class II households are omitted from the census during Followup. Many duplicates during Followup may occur when other households are re-enumerated in place of Class II households. Class II households may have been missed as some census Followup interviewers fabricated fictitious interviews. A move by a Class II household between census day and completion of census Followup may have caused an incoming household to be enumerated instead, although enumeration of the incoming household at the new address was classified as a definitional error in the 1980 PEP. The census Followup nonresponse cases, enumerated by census close-out procedures, also belonged to the Class II category, although these households are excluded from Table 5.

6. THE DIFFICULTY OF DETERMINING NONMATCH RATES FOR MOVERS

The estimated nonmatch rate to the census for movers (specifically those with a different August CPS and April 1, 1980 address) from the August sample exceeded 20 percent for Blacks, Spanish, and others, separately. Although moving close to the time of census enumeration itself would probably affect the nonmatch rate to the census, a rate in excess 20 percent seems inflated.

To test this hypothesis, a longitudinal linkage of the August CPS to the May, June, and July CPS was performed for the overlapping part of the sample in each of those months. Thus, as in the analysis of the April sample, this longitudinal linkage information could be employed as a probable indication of the timing of move. Those who appeared to move during the period between April 1, 1980 and the May CPS did indeed have a somewhat higher nonmatch rate to the census than movers as a whole. Movers late in the summer, however, even those who moved between CPS week in July and in August, still had a very high nonmatch rate to the census, estimated at 21 percent.

An interpretation of these high nonmatch rates to the census, even for apparent movers during a period in which most of the census was already over, was that PEP matching was subject to considerably higher levels of matching error than for nonmovers. In fact, far more resources were available to determine the correct ED for CPS households with the same census day address. Matching for movers depended entirely upon the ability to determine the correct ED for a reported address, possibly subject to problems of recall or interviewer error.

A longitudinal analysis of the April sample further supports the conclusion that nonmatch rates to the census for movers were inflated in the August sample from matching difficulties. By using longitudinal linkage to reconstruct dates of move for the April sample during the period between the April and July CPS interviews, nonmatch rates to the census were estimated at 11.4 percent for movers between the May and June CPS interviews and 7.4 percent for movers between the June and July interviews. These rates, although higher than those for nonmovers, suggest serious over statement of the nonmatch rates for movers in the August sample.

7. CONCLUDING REMARKS

Demographic analysis indicates that Class I persons, i.e. within household omissions, continues to be the most elusive category for sample survey measurement of census undercoverage. The differential undercount of adult males relative to adult females may represent a significant proportion of the total undercount, yet one that is the most difficult to capture.

The mail census particulary serves to define a second category of omissions, Class II, that includes census noninterviews and other households whose addresses are listed in the census address register but who are not themselves enumerated, often because of nonresponse to the mail census. This group constitutes the largest class of omissions in the 1980 census according to the 1980 PEP. Thus, although the actual magnitude of

Class II estimated by PEP is plausible, variations in the estimates for this class over different PEP data sets accounted for the majority of the instability in the estimates.

The mail census measures the more cooperative and less mobile part of the population first. The enumeration of the less cooperative, more mobile group is deferred for almost two or more months. The greater mobility of this group interacted with the relatively high error rates in the 1980 census Followup. The resulting omissions posed difficult problems of measurement, particularly in terms of high nonresponse rates to PEP follow-up as a further consequence of high mobility.

From all appearances, measurement of Class III underenumerations presented less serious problems in the 1980 PEP. It is possible that CPS failed to cover some of the same units entirely omitted from the census, although no specific evidence is available on that question.

It is unlikely that evaluation of the 1990 census will assume the same form as the 1980 PEP. Nonetheless, recognition of the methodological challenges posed by the three classes should serve as a guide for interpretation of further research efforts.

Acknowledgements: The author wishes to thank John Bushery, Howard Hogan, Charles Jones, and Glenn Wolfgang for comments on an earlier version of this paper, and Alice Coburn for typing.

References

Bailar, B. (1983), "Counting or Estimation in a Census - A Difficult Decision," in Proc. Soc. Statist. Sec., Amer. Statist. Ass., 42-49.

Coale, A.J., and Rives, N.W. (1973), "A Statistical Reconstruction of the Black Population of the United States, 1880-1970: Estimates of the True Numbers by Age and Sex, Birth Rates, and Total Fertility," Population Index, 39, 3-36.

Coale, A.J., and Zelnik, M. (1963), New Estimates of Fertility and Population in the United States, Princeton, N.J.: Princeton U. Press.

Cowan, C., and Fay, R.E. (1984), "Estimates of Undercount in the 1980 Census," in Proc. Sec. Surv. Res. Meth., ASA, 566-571.

Ericksen, E.P., and Kadane, J.B. (1985), "Estimating the Population in a Census Year," with discussion, J. Amer. Statist. Ass., 80, 98-131.

Fay, R.E., and Cowan, C. (1983), "Missing Data Problems in Coverage Evaluation Studies," Pro. Sec. Surv. Res. Meth., ASA, 158-163.

Keyfitz, N. (1979), "Information and Allocation: Two Uses of the 1980 Census," The Amer. Statis. 33, 45-50.

Marks, E.S., and Waksburg, J. (1966), "Evaluation of Coverage in the 1960 Census of Population and Housing Through Case by Case Checking," in Proc. Soc. Statist. Sec., ASA.

Passel, J.S., and Robinson, J.G. (1985), "Revised Demographic Estimates of the Coverage of the Population by Age, Sex, and Race," memomorandum to Roger A. Herriot, U.S. Bureau of the Census, April 8, 1985.

Siegel, J. (1974), "Estimates of Coverage of the Population by Sex, Race, and Age in the 1970 Census," Demography, 11, 1-23.

U.S. Bureau of the Census (1960), The Post-Enumeration Survey: 1950, Technical Paper Number 4, Washington, D.C.

Table 5 Components of Estimated Census Overenumerations and Duplications from the 1980 PEP E-Sample (Weighted Estimates in Thousands)

	Total U.S.		Conventional		Mail Return in Mail Area		Non-Mail Return in Mail Area	
	Est	Pct	Est	Pct	Est	Pct	Est	Pct
Estimated complete enumerations (excluding census imputations for noninterviews)	220802.7	100.0	8791.0	100.0	173881.1	100.0	37992.7	100.0
Estimated erroneous enumerations								
Total	7581.8	3.4	145.2	1.7	4299.5	2.5	3121.3	8.2
Geocoding errors in census								
Total	2104.1	1.0	27.7	.3	1484.7	.9	583.0	1.5
Matching Duplicate in correct ED	618.0	.3	.0	.0	406.3	.2	211.7	.6
Other	1486.1	.7	27.7	.3	1078.4	.6	371.3	1.0
Definitionally Incorrect								
Total	3602.7	1.6	98.6	1.1	1841.5	1.1	1658.1	4.4
Based on Post Office response	982.0	.4	21.3	.2	346.9	.2	611.3	1.6
Other	2620.7	1.2	77.3	.9	1494.7	.9	1046.9	2.8
Duplication								
Total	1874.9	.8	18.9	.2	973.3	.6	880.2	2.3
Within the same ED	1551.1	.7	16.4	.2	761.0	.4	773.7	2.0
Between ED's in the same area of search	323.8	.1	2.6	.0	212.2	.1	106.4	.3