GAO

United States
General Accounting Office
Washington, D.C. 20548

152267

General Government Division

B-257820

August 2, 1994

The Honorable Thomas C. Sawyer
Chairman, Subcommittee on Census,
  Statistics and Postal Personnel
Committee on Post Office and
  Civil Service
House of Representatives

Dear Mr. Chairman:

This letter addresses your request for our assistance in
interpreting a Census Bureau memorandum entitled, "Could
the Census Bureau Reduce the Undercount by Not Using a
'Long Form'?" (see encl. I). You asked us to review the
effects on decennial census coverage (the accuracy of the
basic count of the population) of eliminating (1) the long
form used during the 1990 Census and (2) some of the
questions on the 1990 Census short form.

In its memorandum, the Bureau developed a statistical model
that could be used to estimate the overall rate at which
people were missed during the 1990 Census (the overall
omission rate). Using the model, the Bureau estimated the
possible effects on the omission rate of eliminating the
census long form (one of the variables used to estimate the
overall omission rate).

We examined the reasonableness of the assumptions used in
the Bureau's model, and we replicated calculations the
Bureau used to estimate the effects on census coverage of
eliminating the long form. We then used the Bureau's model
to examine the second question raised by your letter
concerning the effects on census coverage of eliminating
some of the questions on the short form. Our review of the
Bureau's memorandum relies on the results of Bureau
evaluations of the omissions rates in the 1990 Census and
the results of other Bureau evaluations done in 1992 and
1993 that tested the reaction of the public to different
decennial census questionnaires of varying lengths and
format. However, we did not independently validate any of

the Bureau's evaluation results. Our review also relies on our previous work.[1]

We did not analyze (1) the usefulness of the information collected in the 1990 Census, (2) the potential savings of eliminating the long form, or (3) other possible operational effects on the decennial census of reducing the number of questions. Therefore, our limited review of the Bureau memorandum does not by itself provide a basis for deciding whether to eliminate or keep any questions on the long or short form in the 2000 Census. We have obtained comments from the author of the Bureau memorandum on a draft of this letter and have responded to them where appropriate.

BACKGROUND

To collect data for the 1990 Census, the Bureau used a long form, which was sent to a sample of 16.7 percent of the housing units in the nation and contained 26 housing questions and 33 population questions.[2] It also used a short form, which was sent to all the other known housing units and contained seven of the 26 housing questions on the long form and seven of the 33 population questions. For most of the nation, the Bureau mailed the forms to the known housing units. It then sent Bureau staff (known as enumerators) door-to-door to follow up on those households that did not return the completed forms by mail to the Bureau. The Bureau memorandum examined the hypothesis that eliminating the long form and using only the 1990 Census short form would have improved the coverage of the 1990 Census (decreased the number of persons missed). The hypothesis is based on the established facts that (1) census forms returned by

---

[1]See Programs to Reduce the Decennial Census Undercount (GAO/GGD-76-72, May 5, 1976); On the Census Bureau's Preparations for the 1990 Decennial Census, statement by Gene L. Dodaro, before the House Subcommittee on Census and Population, Committee on Post Office and Civil Service (May 15, 1986); Decennial Census: 1990 Results Show Need for Fundamental Reform (GAO/GGD-92-94, June 9, 1992); Census Reform: Questionnaire Test Shows Simplification Holds Promise (GAO/T-GGD-92-59, July 1, 1992).

[2]In 1990, although the national sampling rate for the long form was 16.7 percent, the sampling rate varied with the population density of the cities and towns. In small areas (those with an estimated population of less than 2,500 individuals), the Bureau used a sampling rate of 50 percent. In densely populated urban areas, the sampling rate was 12.5 percent. All other areas had a sampling rate of 16.7 percent.

mail tended to miss fewer people than those filled out by Bureau staff going door-to-door to visit households that did not return forms by mail and (2) census long forms tended to require more door-to-door visits than did short forms.

In the Bureau's statistical model, the overall omission rate depends on (1) the omission rate of each type of form, which varied by whether a form was returned by mail or enumerator, (2) the mail return rate of each type of census form, and (3) the sampling rate used for the long form.

Specifically, according to the results of Bureau research, forms that respondents returned by mail missed fewer people than did forms returned by enumerators. The omission rate for mail-returned long forms was 1.8 percent, and for mail-returned short forms, was 1.9 percent. For enumerator-returned forms, the omission rate was 11.3 percent for the long form and 11.7 percent for the short form. The 1990 Census mail return rate of the long form was lower than that of the 1990 Census short form. The 1990 Census mail return rate for the short form was 74.9 percent, and for the long form, was 70.4 percent--a difference of 4.5 percentage points.

On the basis of the results of its statistical model, the Bureau memorandum concluded that any improvement in the coverage of the 1990 Census by eliminating the long form and using only the short form would have been trivial. According to the Bureau memorandum, the overall omission rate for both mail-returned and enumerator-returned forms would have decreased from 4.40 percent to an estimated 4.36 percent.

RESULTS IN BRIEF

Our review of the Bureau's memorandum confirms the Bureau's conclusion that simply eliminating the long form, assuming no other changes, would have had a trivial effect on the coverage of the 1990 Census. However, using the same model, we believe that substantially reducing the number of questions on the short form could have noticeably decreased the estimate of the overall omission rate of the 1990 Census. We base our belief on the results of Bureau tests conducted in 1992 and 1993 and our previous work. Specifically, the Bureau tests showed that an increase of 4.6 percentage points in the mail return rate of the short form would have been possible if 9 of the 14 questions on the short form were eliminated. Such an increase in the mail return rate of the short form could have decreased the estimated overall omission rate from 4.40 percent to 4.024 percent. This decrease in the estimated overall omission rate would have been 9 times that estimated by eliminating the long form.

The author of the Bureau memorandum has cautioned that our use of the model to examine coverage gains from reducing the length of the short form is inadvisable for several reasons, which we discuss and respond to later. Despite these cautions, we believe that a substantial reduction in the number of questions on a short form similar to the one tested by the Bureau could noticeably affect census coverage.

THE POSSIBLE EFFECTS ON COVERAGE OF ELIMINATING THE LONG FORM

The Bureau memorandum examined the hypothesis that fewer people would have been missed in the 1990 Census if the long form had not been used. The Bureau calculated the overall omission rate as the sum of the omission rate of each type of form (i.e., short and long)[3] weighted by the sampling rates of both forms. When the long form is included in the calculation (at a 16.7 percent sampling rate used in the 1990 Census and a mail return rate of 70.4 percent), the overall omission rate is estimated to be approximately 4.40 percent. If the long form is eliminated from the calculation and all forms are assumed to have been short forms, the overall omission rate estimate becomes 4.36 percent (a 0.04 percentage point improvement). On the basis of our review, we believe the statistical model used in the Bureau memorandum is appropriate to test the hypothesis that eliminating the long form would or would not have noticeably improved the overall omission rate. We have replicated the calculations made with the Bureau's model to test this hypothesis.

The 4.5 percentage point difference between the mail return rates in the 1990 Census (i.e., 74.87 percent for the short form and 70.35 percent for the long form) was the largest difference in any decennial census. For undetermined reasons, the difference between the mail return rates of the two forms increased from 1980 to 1990.

A special Bureau test in 1993, the Appeals and Long Form Experiment (ALFE), yielded an even larger difference between the mail return rates of long and short forms. In ALFE, short and long forms containing the same questions as those in the 1990 Census were redesigned to be easier for the respondent to comprehend and fill out (more respondent-friendly) by changing the instructions, the wording of the questions, and the spacing of the questions on the form. Two respondent-friendly versions of the long form were tested. The two versions yielded mail

---

[3]The derivation of the omission rate is explained in the Bureau memorandum. Note that a range of omission rates may be considered--e.g., from 4.3 percent to 2.0 percent. The Bureau's calculations and ours use the higher estimates.

return rates that were 11.3 and 12.8 percentage points lower than the mail return rate of the short form that was used as a control.[4] Assuming the largest difference between the long and short form mail return rates found in the ALFE test--12.8 percentage points--eliminating the long form would have resulted in a decrease in the estimated omission rate of 0.17 percentage points. This decrease is about 4 times the improvement calculated using the assumptions of the Bureau memorandum, which used mail return rate data from the 1990 Census. Assuming the ALFE mail return rate differential, the estimated omission rate would have improved by decreasing from 4.40 percent to 4.23 percent. The results of the ALFE test offer evidence that the trend toward a larger difference between the mail return rates of the long and short forms is likely to continue even if the forms are redesigned to be more respondent-friendly.

The Bureau memorandum examined the sensitivity of the model to changes in the difference between the mail return rates of the long and short forms. According to the Bureau's calculations, the difference in the mail return rates of the two forms would have to be at least 35 percentage points before the overall omission rate would have improved from 4.40 percent to 3.90 percent (i.e., one half a percentage point). We also have replicated these calculations.

THE POSSIBLE EFFECTS ON COVERAGE OF ELIMINATING SOME OF THE QUESTIONS ON THE SHORT FORM

At your request, we reviewed a second hypothesis (not discussed in the Bureau's memorandum) that eliminating a substantial number of questions from the short form would have noticeably improved coverage (i.e., decreased the omission rate). As a first step, using the same statistical model contained in the Bureau memorandum, we examined the rate of decrease in the overall omissions rate as the mail return rate of the short form increases. (Encl. II shows the results graphically.) With the long form included in the model, each percentage point increase in the short form mail return rate results in a decrease in the estimated overall omission rate of approximately 0.08 percentage points. When the long form is eliminated from the model, each percentage point increase in the short form mail return rate results in a decrease in the estimated overall omission rate of approximately 0.098 percentage points. (Encl. II shows the

---

[4]This test was designed to examine the effects on mail return rates of (1) two types of respondent-friendly long forms and (2) different "appeals" to the public about the benefits of the census (that a response is mandatory and that the confidentiality of responses is protected).

effects of both changes on the estimated overall omission rate.)
Thus, on the basis of our review of the model, we believe that a
modest increase (e.g., 5 percentage points) in the mail return
rate of the short form could noticeably improve census coverage
by about 0.40 percentage points instead of 0.04 percentage points
(which is the Bureau's estimate of the effect of eliminating the
long form).

The figure in enclosure II shows that modest improvements in the
short form mail return rate can affect the estimated overall
omission rate to a greater degree than the elimination of the
long form. The mail return rate of the short form would have a
much larger impact on the model's estimated overall omission rate
because the short form is sent to 5 times as many households as
is the long form (83.3 percent compared to 16.7 percent).

In 1992, the Bureau conducted the Simplified Questionnaire Test
(SQT) that tested several variables, including a proxy of the
mail return rates (referred to as completion rates) for a
redesigned short form (referred to as the "micro form") with only
five personal questions--name, age, gender, race, and ethnicity.
The SQT micro form had a higher return rate by mail[5] than did a
respondent-friendly version of the 1990 Census short form
containing 14 population and housing questions.[6] In the SQT,
71.4 percent of the micro forms were returned by mail, while the
respondent-friendly version of the 1990 Census short form had
66.8 percent returned by mail.[7] Using the Bureau memorandum's
model, we calculated the decrease in the estimated overall
omission rate, if the mail return rate of the short form

---

[5]The SQT used the completion rate as an approximation of the mail
return rate that the Bureau used to describe the 1990 Census
results. The SQT's completion rate eliminates from the
calculation of mail return rate forms returned by the postmaster
as undeliverable (because the address was nonexistent or vacant).
For this reason, the SQT and the actual 1990 short form mail
return rates are not directly comparable.

[6]We compared the micro form with a respondent-friendly version of
the actual 1990 Census short form that changed the instructions
on why the form should be returned, the wording of the questions,
and the spacing of the questions on the form.

[7]The elimination of questions designed to identify persons who
may have been left off the census form by the mail respondent and
those who may have been incorrectly included in the SQT micro
form may have other effects on coverage. The Bureau is
investigating the magnitude of the effects of such questions on
coverage in a special test called the Coverage Test.

increased by 4.6 percentage points, to be 0.376 percentage points. The resulting decrease from 4.4 percent to 4.024 percent is approximately 9 times as large as the 0.04 percentage point decrease that would result from eliminating the long form as estimated by the Bureau's model.[8]

We agree with the Bureau that the number of questions on the short form is only one of many variables that affect mail return rates. The Bureau correctly pointed out that the ALFE test (which was conducted after the SQT) showed several versions of the respondent-friendly short form achieving mail return rates that ranged from 76.4 percent to 78.1 percent (instead of 71.4 percent or 66.8 percent). In addition, the Bureau attributed the improvement to a notice (or instruction) on the envelope saying the return of the form was mandatory. The Bureau, however, does not dispute that substantial reductions in the number of questions on the short form can affect mail return to some degree.[9]

Concerns About Using the Model to Predict Possible Effects on Coverage of Eliminating Some of the Questions on the Short Form

The Bureau model assumed that all nonresponding households have equal potential for being missed. The author of the Bureau

---

[8]Although the micro form used in the SQT had a higher return rate (71.4 percent) than the respondent-friendly short form (66.8 percent), the Bureau pointed out that the micro form did not produce higher coverage--as shown by the commonly used proxy measure of coverage (i.e., average household size). However, this proxy measure of coverage cannot be used to evaluate the possible effects on coverage of a higher short form mail return rate (obtained by a micro form) because the Bureau did not follow up on those who did not respond by mail. Only if the Bureau had followed up on those who did not respond by mail would the effects on coverage be evident because of the large differential in the omission rate between forms returned by mail and those filled out by an enumerator.

[9]The ALFE test showed an 11.3 to 12.8 percentage point differences between respondent-friendly long and short forms, respectively. The test, however, did not include the micro form. Therefore, the Bureau could not compare the mail return rate of a micro form with the mail return rates of other forms used in the ALFE test. We do not know whether a micro form with a mandatory notice would have achieved a higher or lower mail return rate than the respondent-friendly forms.

memorandum has cautioned that this assumption may not be true. She also pointed out that the predicted changes in the omission rate that may result from changes in the mail return rate (of the short form) are likely to be overstated. Specifically, she warned that coverage error increases as the time to complete the enumeration process (door-to-door follow-up) increases, which indicates that there is or could be a group of potential respondents who will not cooperate (regardless of the kind of questionnaire used). Therefore, such respondents would be harder to enumerate. The author added that there is no evidence that this uncooperative group would be motivated to respond to a census by a reduction in the number of questions. Therefore, she concluded that the use of the model to predict the benefits of increased mail return rates without appropriate adjustment for differential error rates over time does not appear to be justified.

We acknowledge that the model is a simplification of reality and that the effects on omissions may not occur to the degree the model predicts. Nevertheless, on the basis of previous work,[10] we believe higher mail return rates (of the short form) will yield better coverage because there are explanations of why people are missed by the census in addition to the one that some people just do not want to cooperate in any way.

In a previous report on the results of the 1990 Census,[11] we discussed how the quality of data obtained from the enumeration process declined as data collection efforts were extended. In that report, we noted that field work completed more quickly provides the Bureau with time to review census counts and resolve apparent discrepancies. It also minimizes the problems associated with trying to count a mobile population that may move from one location to another. As more time passes after Census Day (the day the population count is made), more housing units can change status (e.g., from vacant or uninhabitable to occupied). Furthermore, enumerators may contact respondents who may provide inconsistent information or respondents who may forget who was living in a housing unit on Census Day. The Bureau also tries to complete the last portion of nonresponse cases by using special procedures that allow enumerators to contact neighbors of nonrespondents or persons who were not

---

[10]See GAO/GGD-76-72, May 5, 1976; On the Census Bureau's Preparations for the 1990 Decennial Census (May 15, 1986); GAO/GGD-92-94, June 9, 1992.

[11]See GAO/GGD-92-94, June 9, 1992.

B-257820

residents of the household on Census Day.[12] These procedures result in a disproportionately large number of omissions from the census count.

Although one should not attempt to use the model to predict a precise change in the estimated overall omission rate, one can use the model to get an idea of whether a change in the number of questions on the short form could noticeably affect census coverage. If every household received a short form and the short form mail return rate were increased by 4.6 percentage points, the overall omission rate would decrease by an estimated 0.45 percentage points, or a 10-percent reduction in the overall omission rate. This reduction approximately equals 1 million people, i.e., 10 percent of the 10 million people missed by the 1990 Census.[13] Even a smaller reduction, e.g., 5 percent, could add a significant number of people to the census count.

CONCLUSION

Only a limited sample of the population receives the long form. Thus, the elimination of it would not noticeably effect census coverage. However, we believe that a shorter version of the short form (which is sent to a much larger population) could have a noticeable effect on the short form mail return rate. A larger mail return rate reduces coverage errors (e.g., those caused by reducing the time needed to conduct follow up on nonresponse cases and allowing the Bureau more time to handle difficult or uncooperative cases). For these reasons we believe a substantial reduction in the number of questions on a short form similar to the one tested by the Bureau in the SQT could noticeably affect census coverage.
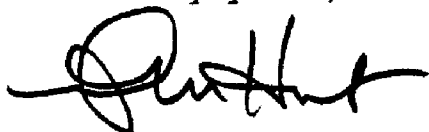
—   —   —   —   —

---

[12]In large city offices, an average of 20 percent of all nonresponding households were enumerated by these procedures.

[13]See 1990 Census: Reported Net Undercount Obscured Magnitude of Error (GAO/GGD-91-113, Aug. 22, 1991).

B-257820

We are sending the Acting Director of the Census Bureau a copy of this letter. If you have any questions regarding this issue or would like to discuss it further, please call me on (202) 512-8676.

Sincerely yours,

William M. Hunt
Director, Federal Management
   Issues

CENSUS BUREAU MEMORANDUM

September 10, 1993

COULD THE CENSUS BUREAU REDUCE THE UNDERCOUNT
BY NOT USING A "LONG FORM?"

Catherine Keeley

The 1990 census "long form" questionnaire was returned at a rate five percentage points lower than the "short form." This led to speculation that if the Census Bureau did not use a long form, the undercount rate could be reduced. This paper addresses the relationship between the 1990 census questionnaire return rates, length of the census questionnaire, and the undercount. We show that any increase in coverage if a long form had not been used would be trivial. These data show that the national mail return rate for the long form would have had to be as low as 40 percent before the use of the long form would even contribute one-half percentage point to the 1990 coverage error.

In the 1990 census, there were two principle types of data collection forms: a 100 percent questionnaire[1] (or "short form") and a sample questionnaire[2] (or "long form"). Each household received one of the two questionnaires. The short form contained 7 population questions and 7 housing questions. The long form had all the short form questions plus 19 housing questions[3] and 26 population questions[4].

The short form had a national mail return rate[4] of 75 percent and the long form rate was 70 percent; see [1]. This mail return rate varied for different geographic areas. However, returning the questionnaire by mail was not the only way households were counted in the census. For every housing unit on our address list for which we did not receive a questionnaire in the mail, we sent an enumerator to visit the unit and conduct an interview with the household.

---

[1] So called because everyone (or 100 percent of the population) is asked the questions contained on this questionnaire.

[2] So called because this form was sent to a statistical sample of approximately 1 in every 6 housing units. It contained all items asked of everyone as well as additional items.

[3] Some questions had multiple parts, but not every household had to answer every question.

[4] The 1990 mail return rate is the ratio of the number of households returning a census questionnaire by mail to the total number of occupied housing units that should have received a census questionnaire delivered by mail or by a census enumerator.

Evaluation of the 1990 census shows there is a higher coverage error rate (those missed in the census, also called census omissions)[5] for forms returned by an enumerator (11.6 percent) than for those returned by mail (1.8 percent) (Table 1). This is due to many reasons[6] but they are all related to the nature of the cases the enumerator is working, that is, the nonrespondents. Nonetheless, there is a 9.8 percentage point difference between the two estimates. However, an important factor to note is that, at the national level, short forms and long forms had about the same coverage error rates[7] among mail returns as well as among enumerator returns.

Table 1:   Estimated Coverage Error Rates
(Standard errors in parentheses)

Form Type	Source	
	Mail Returns	Enumerator Returns
Short Form	1.9% (0.1)	11.7% (0.5)
Long Form	1.8% (0.2)	11.3% (0.9)
Total	1.8% (0.1)	11.6% (0.4)

The observed lower mail response rate for long forms—meaning more cases becoming part of the nonresponse workload, and consequently completed by an enumerator—and the higher coverage error of enumerator returns, leads to questioning whether the use of long forms could be a significant contributor to the undercount.

QUESTION 1:

Would there have been a coverage gain if we had not used long forms in the 1990 census?

Besides the rates just discussed (mail return rates by short and long form, and coverage errors for mail and enumerator returns by short and long form), a critical factor in answering this question is the number of households receiving the sample form, or the sampling rate. At a national level, only about 17 percent of all housing units received a sample form. This rate is very important because 83 percent of the housing units received a short form—with an associated higher response rate—while only 17 percent received the long form. Since the long form went to such a small proportion of all households, the impact on coverage is minimized.

---

[5] See Appendix 1 for a discussion of coverage rates used in this paper.

[6] Discussed in Appendix 2.

[7] There is no statistical difference at the 90 percent level of confidence.

ANSWER TO QUESTION 1:

The coverage gain by not using the long form would have been 1/25 of 1 percentage point. A trivial difference.

The statistical model used to answer Question 1 is in Appendix 3.

QUESTION 2:

How low would the mail return rate for long forms have to be to see a substantial increase in coverage error due to using the long form?

ANSWER TO QUESTION 2:

When modeling lower return rates for the long form at a national level[1], we see that it takes about a 40 percent return rate for the long form before coverage error due to long form use is increased by 1/2 of 1 percentage point. The long form return rate must be 10 percent before the coverage error is increased by 1 percentage point[2].

Table 2:   Estimated Coverage Error Increases by Long Form Return Rate

Long Form Return Rate	Coverage Error Percentage Point Increases
60%	0.21
50%	0.37
40%	0.52
30%	0.68
20%	0.84
15%	0.92
10%	1.00

---

[1] Assumes the 1990 coverage error rates and the short form return rate.

[2] A modification of the model that centers on differences between rates, that is, the differences needed between the short form and long form rates before substantial coverage error rates are noted, supports these conclusions. The data presented here implies about a 35 percentage point difference in return rates for the two forms before the coverage error is increased by 1/2 of 1 percentage point (1990 return rate of 75 percent for the short form).

4

In conclusion, while the lower mail return rate of the long form does introduce some coverage error, the increase over the coverage error from the short form is trivial. For the long form to introduce substantial coverage error, the mail return rate would have to be at such a low level that the use of the long form would be suspended for other methodological reasons.

Future research will apply this model to coverage error rates for various population subgroups, such as Black and Hispanic, and geographic areas, such as urban and rural.

Appendix 1

Coverage Error Rates

The rates used in the tables are derived from combining the work on missed housing units from the Housing Unit Coverage Study and missed persons from the Post-Enumeration Survey (PES); see [2]. The coverage error rates shown in the table are census omissions. Using these calculations, the total omission rate is 4.3 percent. This differs from the well publicized, national undercount rate derived from PES of 1.6 percent. This is because the net rate from PES (1.6) is based on the rate of omissions minus erroneous enumerations. These rates are derived from specific matching rules unique to the PES and cannot be used separately because they were specifically designed to produce a net rate—not two independent rates. Therefore, to analyze only omissions, other rates, based on independent matching schemes, have been derived and can be considered proxies for missed persons.

These rates are given for mail-back areas that cover about 95 percent of the population. A "List/Enumerate" data collection procedure was used in sparsely settled areas of the country for the 1990 census. In these areas, enumerators picked up all census forms and administered all long forms. Data for List/Enumerate areas are not included.

The rates used in this paper may be considered the upper bound on omissions and, therefore, the most conservative approach. Another set of omission rates has been recently developed; see [3]. These rates show an overall omission rate of 2.0 percent and thus can be considered a probable lower bound on the omission rate. These rates have been used in the model with similar results.

The omission (or miss) rates used in this appendix are from the "lower bound" data (discussed in Appendix 1), the actual rates may be somewhat higher.

Census Bureau research over the years has found that forms returned by mail have higher quality data than those completed by a census enumerator[10]. These findings supported the institution of a mail census and its expanded use for the last three censuses.

Although it may be true that self response provides better quality data, we should not be quick to blame this on the enumerators or assume better training or procedures would remedy this problem. Enumerator coverage error rates increase over the duration of nonresponse follow-up from approximately 1.6 percent omissions to about 9.5 percent. In addition to these being more difficult cases initially (nonresponding cases), many other factors could contribute to this increase in omissions. Over time household composition may change, there is an increase in the impact of movers, and the hardest persons to enumerate are likely to be enumerated in the final stages of the census. Recall error is also a likely contributor.

Also, during the final stages of 1990 nonresponse follow-up the district offices implemented "last resort" and "closeout" procedures to enumerate the final set of unresolved cases. Last resort procedures allowed the enumerators to turn in questionnaires with most items unanswered. Closeout procedures allowed questionnaires to be accepted with even fewer data. Persons in households who were enumerated using either last resort or closeout procedures have a significantly higher rate of omissions than those enumerated under regular procedures (18.5 percent versus 4.1 percent).

Clearly, there are many problems inherent in collecting data from households that do not respond initially to the census. Only a few are discussed here. However, the data presented in this paper show that the form type is not the critical factor. Therefore, to make a substantial reduction in the coverage error from nonresponse cases (whether short form or long form), methodologies that concentrate on improving the coverage of the nonresponse universe, such as sampling and estimation, have the most potential.

---

[10] Intensive statistical research into 1950 and 1960 census data revealed that enumerators unintentionally inserted certain biases into the data they collected. Apparently, through their own idiosyncratic ways of asking questions or marking answers, enumerators suppressed natural variation that should have occurred in the data. In the mail-out/mail-back system, every questionnaire is filled out by a different person. So, there is no such bias and the full variability can emerge.

Appendix 3

Model of Effects of Return Rates and Form Length
on Census Undercount

The research question is, "Would there have been a coverage gain if the Census Bureau had not used long forms in the 1990 census?"

The key variables are (rates are at the national level):

o   Mail return rates for the short and long form:

   Short form = .7487 = SFMR
   Long form = .7035 = LFMR

o   Sampling rates for the short and long form:

   Short form = .8333 = SFSR
   Long form = .1667 = LFSR

o   Coverage error rates for the short and long form by mail and enumerator returned forms:

   Mail returned forms:

   Short form = 1.9 = SPCE;M
   Long form = 1.8 = LFCE;M

   Enumerator returned forms:

   Short form = 11.7 = SFCE;E
   Long form = 11.3 = LFCE;E

The following models these variables to analyze the relationship between response rates, form length, and coverage error.

o   The 1990 census can be modeled as:

   [SFMR(SFCE;M) + (1-SFMR)(SPCE;E)] SFSR +
   [LFMR(LFCE;M) + (1-LFMR)(LFCE;E)] LFSR

                        or

   [.7487(1.9) + .2513(11.7)] .8333 + [.7035(1.8) + .2965(11.3)] .1667 = 4.40 =
   short and long form coverage error rate

2

o   Assuming all short forms and thus a .7487 mail return rate overall, if all other error rates stayed the same:

.7487(1.9) + .2513(11.7) = 4.36 = short form only coverage error rate

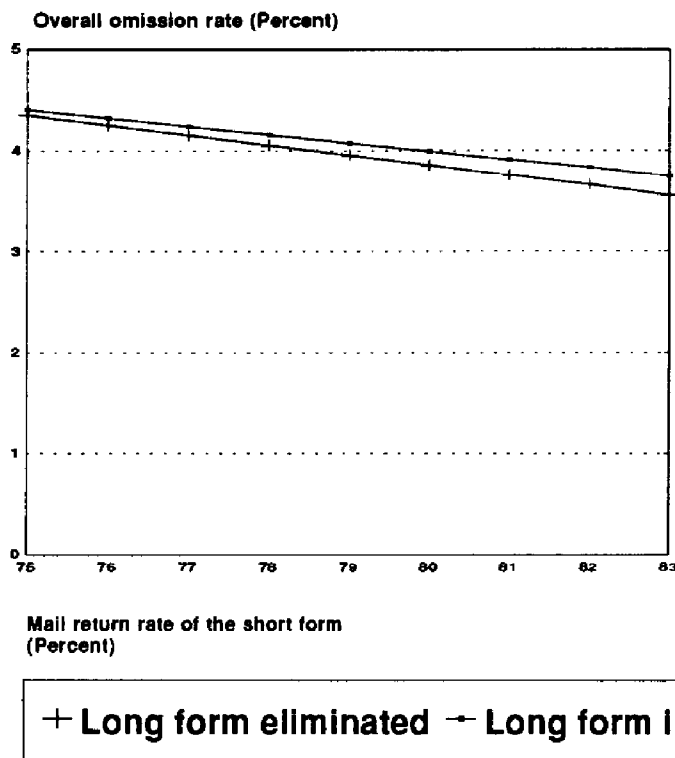o   Difference is 4.40 - 4.36 = .04 or 1/25 of 1 percentage point/coverage gain.

References

1. 1990 Census Mail Return Rates, Diane F. Barrett, October 15, 1992.

2. Unpublished tabulations, Danny R. Childers, August 1993.

3. Analysis of Census Omissions: Preliminary Results, Christopher L. Moriarity and Danny R. Childers, Draft, August 1, 1993.

PROJECTED EFFECTS OF MAIL RETURN RATES ON OMISSION RATES

Overall omission rate (Percent)

**Mail return rate of the short form
(Percent)**

┌───┐
│ **+ Long form eliminated ← Long form included** │
└───┘

Note: This figure shows the inverse relationship between the
overall omission rate and mail return rate of the short form
based on a statistical model using 1990 Census data. It is not
meant to be a prediction of the 2000 Census overall omission
rate, which may depend on other factors.

Source: GAO analysis of Census Bureau data.

(243052)