

15881

114144

REPORT BY THE

Comptroller General

OF THE UNITED STATES

Procedures To Adjust 1980 Census Counts Have Limitations

The 1980 Decennial Census results will affect the distribution of seats in the Congress and many billions of dollars in Federal and State assistance. Incorrect counts could cause inequities.

The Census Bureau acknowledges that in each census there is a net undercount. It conducts evaluations to estimate the size of the undercount, but the evaluations have known limitations for adjusting census counts. Bureau research suggests statistical analysis might be helpful.



114144

In obtaining an estimate of the undercount, the Bureau deleted a major part of its planned evaluation. Time was saved but some precision was lost. Congress was not notified.

GAO is recommending steps to (1) formalize a statistical analysis research program and (2) require a report of plans to estimate census over or undercounts and an independent evaluation of the plans. GAO is also recommending legislation to keep the Congress informed of the plans.

The Bureau currently believes there is no statistically defensible way of adjusting the census counts. It plans to continue research.



014261

GGD-81-28
DECEMBER 24, 1980

Request for copies of GAO reports should be sent to:

**U.S. General Accounting Office
Document Handling and Information
Services Facility
P.O. Box 6015
Gaithersburg, Md. 20760**

Telephone (202) 275-6241

The first five copies of individual reports are free of charge. Additional copies of bound audit reports are \$3.25 each. Additional copies of unbound report (i.e., letter reports) and most other publications are \$1.00 each. There will be a 25% discount on all orders for 100 or more copies mailed to a single address. Sales orders must be prepaid on a cash, check, or money order basis. Check should be made out to the "Superintendent of Documents".



COMPTROLLER GENERAL OF THE UNITED STATES
WASHINGTON, D.C. 20548

B-201114

The Honorable Robert Garcia
Chairman, Subcommittee on
Census and Population
Committee on Post Office and
Civil Service
House of Representatives

Dear Mr. Chairman:

This report in response to your June 13, 1980, request addresses the (1) Bureau of the Census' ability to develop accurate undercount estimates for subnational levels, and (2) effect of dropping the planned post-census survey on the prospects for developing accurate undercount estimates.

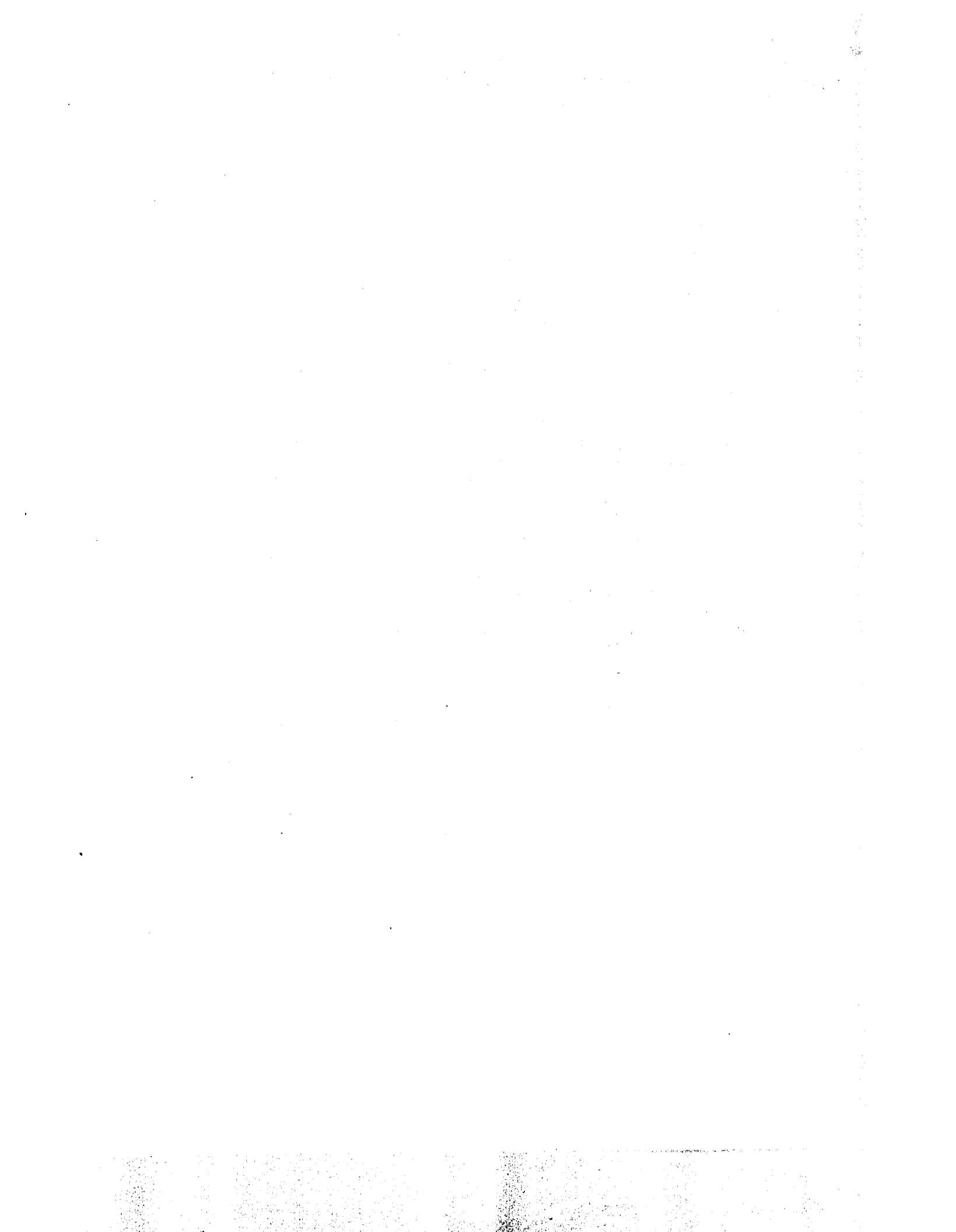
The findings of this report were provided to your staff during a briefing on October 1, 1980. The Bureau's official comments on our findings and recommendations are included in the report. As arranged with your office, we plan no further distribution of the report until 30 days from the date of the report unless you publicly announce its contents earlier. At that time, we will send copies to interested parties and make copies available to others upon request.

We are available to discuss our findings and to provide any further assistance you might need on the matters discussed in this report.

Sincerely yours,

A handwritten signature in black ink, appearing to read "Thomas B. Atchefs".

Comptroller General
of the United States



COMPTROLLER GENERAL'S
REPORT TO THE CHAIRMAN,
SUBCOMMITTEE ON CENSUS
AND POPULATION
COMMITTEE ON POST OFFICE
AND CIVIL SERVICE
HOUSE OF REPRESENTATIVES

PROCEDURES TO ADJUST
1980 CENSUS COUNTS
HAVE LIMITATIONS

D I G E S T

Each Decennial Census greatly influences political representation and government fund distribution for a decade. The number of congressional representatives a State receives as well as district boundaries are determined by decennial data. Past experience indicates the amount of Federal and State funds distributed annually, based at least in part on census results, could reach \$100 billion. A number of communities, contending that the 1980 census has missed large numbers of persons, and that this undercounting will result in their citizens being shortchanged in political representation and government fund distributions, have taken legal action to have the census count adjusted for the undercount. In response the Census Bureau has stated that a "statistically defensible" adjustment cannot be made.

Estimates of census errors at subnational levels are needed for correcting census data. The Census Bureau is aware that some persons will be missed and others double counted in any census. To estimate the level of coverage error for the 1980 census and as a possible springboard to adjustment the Bureau is using evaluation techniques, application of which includes known problems which are not likely to be overcome in the near future. Unfortunately, extending adjustment estimates becomes progressively more difficult and expensive at lower levels of government. Statistical analysis, used successfully in other projects, might help in estimating census error at subnational levels. To

date, however, the agency has not set up a formal program of applied research for this purpose. (See p. 13.)

The Bureau dropped the major improvement funded by Congress for estimating 1980 census errors at subnational levels, a postenumeration survey, mainly so that preliminary estimates would be available sooner. The Bureau estimates the precision lost will be greatest below the most populous State and city levels. Congress was not advised of this change. (See Chapter 3.)

GAO believes that the importance of these matters points up a need for the Bureau to inform Congress of its specific plans for dealing with census errors and to also advise the Congress of any changes in these plans.

GAO recommends that:

- The Secretary of Commerce require that the Bureau of the Census organize a program of experiments using statistical analysis to determine whether improved estimates can be developed of the true population at State and substate levels. (See p. 16.)
- The Secretary of Commerce require that for future censuses (1) the Census Bureau determine the feasibility of estimating and distributing population undercounts and describing in detail the methods to be used and (2) an independent assessment be made of the Bureau's proposed plans. (See p. 25.)
- The Congress enact legislation requiring that the Secretary of Commerce report to the appropriate congressional committees prior to each census on whether and how he intends to prepare credible estimates of census error at subnational levels. (See p. 25.)

Over the past three decades the Bureau has primarily used two coverage evaluation techniques, the demographic method and matching studies. These techniques as implemented by the Bureau do not provide estimates at geographical levels compatible with census data user needs.)

The demographic method combines records of births, deaths and net migrations to estimate an expected true population. It is the Bureau's primary technique for estimating national coverage. This method suffers from limited data, notably the absence of reliable estimates for the size of certain age groups and for the number of illegal aliens. (See p. 9.)

Matching studies is the other technique the Bureau has used for coverage evaluation. They can be useful in identifying subnational levels of census population error. The studies are of two types: reinterview studies and reverse record checks. Both involve the comparison of a list of sample persons or households to the census. In reinterview studies the list is developed by a pre-or post-census enumeration of a sample of the population. In reverse record checks the sample list is drawn from records independent of the census such as the medicare rolls or Internal Revenue Service records. (See p. 10.)

Both matching techniques have major weaknesses. People tend to respond to a reinterview study in the same way they do to census enumeration. Thus if a respondent in the census desires not to reveal the total number of persons in a household he will do the same in the reinterview. Persons missed by the census are therefore also likely to be missed by the reinterview study. In many cases, incomplete or invalid data prevent the Bureau from resolving whether a person has been omitted from the census. In addition,

matching is a time consuming process because it requires manual comparison case by case. (See pp. 10 through 13.)

To complete its evaluation of the 1980 census coverage at subnational levels, the Bureau planned a postenumeration survey (a matching study) of 250,000 households but dropped this program at the last minute, even though it was part of its funded program. According to the Bureau, findings from this survey would not have been developed in time to be included in the Bureau's preliminary estimates of census population error. By cancelling the survey, the Bureau should save time but by doing so it sacrifices precision in its estimates of coverage error. This is particularly true for areas with populations under one million. (See p. 19.)

This cancellation could also save the Bureau about \$8 million. However, the overall savings will be more than offset by cost overruns in other census activities. The Bureau did not inform the Congress of the cancellation and associated savings that in effect were reprogrammed when it requested additional funds to complete the census. (See p. 19.)

Evaluation of the accuracy of census coverage and data on population characteristics is needed at substate geographical levels for equitable distribution of government funds. For example, the revenue sharing program distributes money to about 39,500 communities. Neither demographic analysis nor matching studies as implemented by the Bureau are expected to provide unqualified evaluations of census coverage and population characteristics at the needed geographic levels. (See pp. 8 through 13.)

Although other statistical methods of analysis for estimating coverage are available, the Bureau has made little use of them. While they may not be a panacea, Bureau and private sector statisticians believe they

offer some hope either alone or in combination with presently used methodology. (See pp. 13 through 15.)

Many communities believe that an increase in their census count will proportionately increase their share of Federal funds. Because of the nature of Federal program fund formulas, however, this is not necessarily the case. These formulas include factors besides population, such as per capita income. In addition nondata elements have a significant bearing on the funds disbursed. For example many programs are governed by a fixed total amount of money. In such cases, an increase in funds to some communities reduces the amount available to other communities. (See Chapter 4.)

AGENCY COMMENTS

The Census Bureau agreed that procedures available for adjusting census counts have limitations. It believes that there is no statistically defensible way to accurately estimate the undercounts for subnational areas or perhaps even at the national level. The Bureau also agreed that research is necessary to develop undercount estimation techniques. (See p. 16.)

The Bureau believes that the precision lost by dropping the large-scale postenumeration survey is small. However, the Bureau points out that a greater precision for estimating undercount is needed as the undercount decreases. It believes that the undercount in 1980 is less than in previous censuses. (See p. 25.)



C o n t e n t s

	Page
DIGEST	i
CHAPTER	
1	INTRODUCTION 1
	Objectives, scope and methodology 2
	Proposed legislation 3
	Litigation on 1980 census 3
	Bureau's modifications of raw census data 5
2	SUBNATIONAL ESTIMATES OF TRUE POPULATIONS STILL HAVE LIMITATIONS 8
	Census evaluation programs do not meet user needs 8
	Conclusion 16
	Recommendation 16
	Agency comments and our evaluation 16
3	BUREAU REDUCES ITS EVALUATION PROGRAM TO PROVIDE MORE TIMELY ESTIMATES 18
	Plans altered to expedite the estimates 18
	Millions that may be saved were not mentioned in funding requests 19
	Precision lost by deleting the large- scale survey 19
	Refinements made in evaluation program but limitations still exist 21
	Conclusions 24
	Recommendations 25
	Agency comments and our evaluation 25
4	HOW WILL ADJUSTING CENSUS DATA AFFECT FUND DISTRIBUTIONS--WHO KNOWS? 27
	Use of census data for funds distribution is widespread 27
	Impact of adjusted population data on funds allocation is dampened by a host of variables 28
	Public officials utilize a simplistic approach in calculating "lost" govern- ment funding 30
	Conclusion 31

APPENDIX

I	Letter dated December 5, 1980, from Bureau of the Census	32
II	Text of draft legislation on census reporting requirement	35
III	Data elements used in distributing Federal funds for ten large programs	36

ABBREVIATIONS

CPS	Current Population Survey
GAO	General Accounting Office
SMSA	Standard Metropolitan Statistical Area

CHAPTER 1

INTRODUCTION

Results of the 1980 Decennial Census of Population and Housing will be of great importance throughout the next decade because they will greatly influence this country's political representation and governmental fund distributions. The number of seats in the House of Representatives allotted a State for a decade is based on the population figures provided from a decennial census. Moreover, the States use census data to distribute these congressional seats equitably among their populations. The States also use population data to determine State legislative districts and other local districts. Federal and State governments use population figures along with other data elements, such as per capita income, obtained from the census, to distribute many billions of dollars annually to political subdivisions. These funds could amount to about \$100 billion annually. In addition, governments and industry base many policy decisions on census data.

In theory, the census should count the true population, but authorities agree that no census in history has completely succeeded in doing so. Each census misses some people and double counts others. In U.S. censuses the result has been an estimated net national undercount. The Bureau of the Census has estimated for the past three censuses of 1950, 1960 and 1970 a net national undercount rate of 3.3, 2.7, and 2.5 percent, respectively. It has made these estimates by comparing reported official census results to an expected true population derived from various administrative records and other sources.

However, more important than the national undercount rate is how this undercount is distributed. A disproportionate undercount among geographic areas and population groups can create inequities. Bureau studies have shown that the rate of undercount differs for various segments of the population. In general, blacks are more undercounted than whites; males more than females. Limited Bureau evidence also indicates that the poor are more undercounted than the affluent and persons in the South more than those in the North.

Acutely aware of undercounting and particularly of the disproportionate undercount, the Census Bureau planned several programs to address this problem in the 1980 census. It included in its census enumeration process several coverage improvement procedures such as a 100 percent recheck of housing units initially reported as vacant. In addition the Bureau planned an expanded undercount evaluation estimate

program which included a large-scale postenumeration survey. This survey was designed to compare results from independent interviews of 250,000 households to results of the census. The purpose of the survey was to assist in the development of undercount estimates at various subnational levels. Such information would be useful in evaluating the quality of the census as well as in possibly adjusting the counts.

In prior GAO reports we have discussed undercounts and methods for distributing them. In our report on "Programs to Reduce the Decennial Census Undercount" (GGD-76-72, May 5, 1976), we recommended that the Bureau assess and if necessary increase its research so that it might be better prepared to attain its 1980 census goal of distributing the undercount and publishing corrected population figures at State and major metropolitan areas. In a November 9, 1978, report (GGD-79-7) we discussed the planned large-scale survey and noted that the survey would not be able to develop undercount estimates below State and major metropolitan areas.

OBJECTIVES, SCOPE AND METHODOLOGY

Because of concerns over the probability of undercounting and particularly of a disproportionate undercount in the 1980 census, and also because of information that the large-scale survey had been cancelled, the Chairman, House Subcommittee on Census and Population, Committee on Post Office and Civil Service in a June 13, 1980, letter asked us to review the Bureau's program for estimating and possibly adjusting counts at subnational levels. The Chairman specifically asked that we compare the Bureau's original and revised plans for estimating and possibly adjusting for undercounts at subnational levels, determine the basis for the change, the effect of the change, the cost and time required to complete the revised procedures and also determine whether the Bureau's plans had considered users' needs.

In our review we examined the particular procedures mentioned in the Chairman's letter (Chapter 3) as well as the relationship of these procedures to the Bureau's overall undercount estimation program for the 1980 and prior censuses (Chapter 2). In addition we examined the relationship of decennial census data to funds distributions and data users' needs for these distributions (Chapter 4).

To obtain needed information, we:

--Interviewed Bureau officials,

--Reviewed Bureau instructions, plans and correspondence on estimation procedures for the 1980 census.

- Examined methodologies and results of the Bureau's past census evaluation and undercount estimation projects.
- Reviewed proceedings and results of the Bureau's undercount conferences.
- Examined proposed legislation relating to adjusting census counts.
- Monitored legal actions taken by communities and groups on the 1980 census.

PROPOSED LEGISLATION

Several legislative proposals have been introduced to require adjustments to census counts. In the 96th Congress S.1606 was introduced. If enacted, this bill would require the Secretary of Commerce to adjust population figures of the 1980 census to correct for the undercount. The bill would also require every Federal official who administers a program under which money is distributed according to population data to use these corrected figures. This requirement would not extend to the use of adjusted data for apportioning seats in the House of Representatives. It was the bill sponsor's view that the population count that is used for apportionment ought to be one in which every person can be associated with a name and an address, not with a statistical estimate.

In addition, several bills were introduced in the 96th Congress to limit the count to citizens. One of these, House Joint Resolution 581, was a proposed amendment to the Constitution limiting the count for congressional apportionment to citizens. Another bill sought to accomplish the same objective by placing restrictions on the use of a fiscal year 1981 appropriation.

LITIGATION ON 1980 CENSUS

Numerous lawsuits have been filed by political jurisdictions and groups on the 1980 census. Almost all of this litigation is pending at the District Court level. As of December 19, 1980, one suit, City of Detroit v. Secretary of Commerce et. al., has been decided on the merits, adversely to the United States, by the U.S. District Court for the Eastern District of Michigan. This case has been appealed. The plaintiffs in one other case, Federation for American Immigration Reform (FAIR) v. Hodges, were not successful in their efforts to obtain an injunction directing census to differentiate between lawful and unlawful residents, and to exclude unlawful residents from the census for apportionment purposes.

A general contention by the plaintiffs in the pending lawsuits is that because of an undercount or improper count: (1) citizens will be deprived of their rights to equal representation in State and Federal legislatures and (2) communities will be deprived of Federal and State financial aid to which they are entitled. This latter contention refers to the fact that many Federal and State programs allocate program funds on the basis of population and other data collected by the census. In each pending case, the plaintiffs request relief in the form of an upward adjustment computed through statistical methods supplementing or augmenting the actual headcount. Four basic legal questions are involved in this litigation.

First, Article I, section 2 of the Constitution as well as the Fourteenth Amendment provide for the apportionment of the Federal legislature to be based on "the whole number of persons" residing in each State. Article I, section 2 goes on to provide that the determination of population for apportionment purposes shall be made by "actual enumeration." Presently in litigation is the question whether this constitutional provision requires an apportionment based only on an actual headcount, thus precluding an adjustment by any form of statistical methodology, sampling estimate, or projection.

Second, section 195 of title 13, U.S. Code, provides that "Except for the determination of population for apportionment... the Secretary shall, if he considers it feasible, authorize the use of the statistical method known as 'sampling' to carry out the provisions of this title." Some argue that even if the Constitution does not require an apportionment based solely on an actual headcount, section 195 does. At least to date, no dispute has developed on the point that for purposes other than apportionment--e.g., population counts for program funding determinations, the actual headcount can be adjusted by statistical methods.

A third question is whether an adjustment can be made within the timeframes for reporting the census results for apportionment purposes, or alternatively, whether the timeframes as applied to the 1980 census are unconstitutional. In this regard, section 141 of title 13, U.S. Code, requires that decennial population counts by State be reported to the President by January 1, 1981, and the counts at substate levels by April 1, 1981. The basis for this argument is that the Bureau lacks sufficient time to make a statistically defensible adjustment within the statutory reporting dates. But reporting unadjusted figures to comply with the timeframes would, in the opinion of the plaintiffs, cause apportionment inequities of a constitutional

dimension. To avoid this result, the clear implication of the relief the plaintiffs seek is an extension of the reporting dates to allow sufficient time for an adjustment.

A fourth and final question is whether an adjustment for any purpose, apportionment or otherwise, can be made in a statistically defensible manner. The issue here is not the constitutional or statutory propriety of an adjustment, but whether any currently available adjustment procedure would provide more accurate numbers than the actual count.

BUREAU'S MODIFICATIONS OF RAW CENSUS DATA

In past censuses, the Bureau has revised information obtained from questionnaires during the enumeration process. It revised upward the raw census counts in the 1970 census by about 4.9 million persons in arriving at the official reported population counts. This was accomplished based on evidence about the existence of these additional persons. These additions to the population raw counts are referred to by the Bureau as imputations ¹/ or substitutions. Also the Bureau adds characteristics (age, sex, race) of persons missed and persons who filled out the questionnaire improperly.

According to the Bureau, all data collected from surveys have some imperfections. There are several reasons that data may not represent the whole truth. These reasons include missed housing units, nonresponse or incomplete response, including the exclusion of persons residing in the household.

The Bureau considers that its responsibility is to turn raw data into published, coherent, consistent, tabulated information for users. The Bureau wants to make the data consistent from one time period to the next so that change due to different reporting practices of respondents will not be misinterpreted as a real change. However, the Bureau cautions that modifying unsatisfactory data at reasonable cost does run the risk of significantly changing collected data and introducing serious biases.

In the 1970 census, the Bureau imputed 4.9 million persons as a result of five operations: a recheck of housing units originally considered vacant; the use of surrogate sources to impute household populations; a check with post

¹/A statistical process of developing estimates for missing or inconsistent data in a survey.

office records after enumeration; the addition of persons not counted because of processing problems; and the inclusion of data from supplemental forms.

The recheck of vacant units not seasonally occupied was performed on a sample of 13,546 units. This recheck showed that some of these units were actually occupied. The Bureau inserted into the record of each housing unit converted to "occupied" the number of occupants in a preceding occupied housing unit. The sample results were then inflated to increase the national population by about 1.1 million persons.

From returned census questionnaires indicating that a housing unit was occupied but not listing any occupants, the Bureau selected a previously processed household as a surrogate. This process added about .9 million persons to the census.

About .5 million persons were added to the census count as a result of a check with post office records of the addresses of all housing units enumerated in the rural portions of 16 southern States. The Bureau believed the South to be the area of highest undercount. This postenumeration post office check generated over 700,000 potentially missed addresses. The Bureau used sampling in its record checking and interviews to estimate the number and location of the addresses actually missed in those States. The number of persons added were imputed from other households in the areas in which the census additions were made.

During the data processing phase in the 1970 census, mechanical failures occurred. Questionnaires were skipped or misaligned and thus not read in the microfilming process. In addition some questionnaires were inadvertently destroyed or lost. For the households dropped this way the Bureau made substitutions by replicating a neighboring household. The number of persons included in this fashion totaled about 2.2 million.

As part of its coverage improvement activities, the Bureau used Individual Census Reports for persons who were away from home during enumeration and a "Were You Counted?" publicity campaign directed towards persons believed to have been missed. About 156,000 cases from these programs were received too late for processing by district offices. Through a series of procedures including searching some enumeration records to determine duplicate counts, many on a sampling basis, the Bureau increased the counts by between 153,000 and 193,000.

Prior to the tabulation of data in 1970, a computer editing procedure modified census raw data by allocating responses to unanswered questions and unacceptable entries. The response of the last person processed having the same characteristics as the person being edited was substituted. For example, total money income was inserted for 20.7 percent of the families.

In the 1980 census, unlike the 1970 census, no imputation is planned to the raw census data without a 100 percent physical verification of the existence of household structures. However, in the 1980 census, information will be inserted from surrogate sources for physically existing units. For many households the Bureau is unable even after a number of attempts to determine whether the housing unit is vacant or occupied. Statistical procedures will be used to impute persons into some of these units. Also, for many questionnaires, data items besides population will be missing. Here again statistical methods will be used to insert missing data.

CHAPTER 2

SUBNATIONAL ESTIMATES OF TRUE

POPULATIONS STILL HAVE LIMITATIONS

The Bureau's present application of evaluation techniques has known major limitations which preclude estimating the accuracy and reliability of census data at the geographical levels needed by users. Aside from the types of modifications made to the 1970 raw census count, the counts can be adjusted further only if the Census Bureau can develop acceptable estimates of the true population at State and local levels for comparison with census counts.

Since 1950, the Bureau has used three techniques to identify those portions of the population that are not accurately enumerated in a decennial census: demographic analysis, matching studies, and statistical analysis. For 1980 the Bureau had hoped that statistical analysis, alone or in combination with other methods, could be developed to reliably determine the rate of census coverage at levels needed by users. However, the Bureau currently believes there is no statistically defensible way to accurately estimate the undercounts for sub-national areas or perhaps even at the national level.

CENSUS EVALUATION PROGRAMS DO NOT MEET USER NEEDS

The continuing objectives of the Bureau's evaluation programs since 1950 have been to develop improved census taking techniques and to provide users of census data with measures of the accuracy and reliability of census data to guide them in the use of that data. The Bureau has accomplished the first objective, but it has not satisfied the second at the level needed by users. The Constitution requires State counts for congressional apportionment. While the legality of adjusting census counts for apportionment is currently being argued in the courts, adjustment for other purposes is not. The results of a recent survey conducted by the Department of Commerce indicated that Federal agencies need data at the lowest possible geographic level. The Office of Revenue Sharing, for example, uses census data to distribute funds to about 39,500 communities. Following the 1950, 1960 and 1970 census, the Bureau considered its evaluations of the accuracy of census enumerations reasonable at the national level only. For the 1980 census, the Bureau planned to evaluate the accuracy of census counts for States and major metropolitan areas only.

Demographic analysis--"good" only at the national level

Demographic analysis has been accepted by the Bureau as providing the best estimate of the true population at the national level. However, it has certain limitations and as yet cannot be used to accurately estimate subnational populations. Based on demographic analysis the Bureau estimated that the 1960 and 1970 censuses resulted in net national undercounts of 5.1 and 5.3 million persons respectively.

Demographic analysis involves the compilation, combination, and manipulation of various types of demographic data that are largely independent of the current census so as to derive an expected "true population" by age, sex and race. In 1980 data will be drawn from birth registrations, death records, immigration and emigration estimates, medicare rolls, and previous censuses.

For the population under age 45 in 1980, estimates will be developed from birth records 1/, and from death and immigration statistics to which certain correction factors have been applied. For the population over age 65, aggregate medicare data, corrected for underenrollment by record matching, will provide the basis for the estimate. For ages 45 to 64 the coverage estimate will be extensions of the ages 35 to 54 in the 1970 census. Actual death statistics will be used to allow for mortality up to age 74. The Bureau planned to use official immigration statistics supplemented by estimates of undocumented immigration and emigration to obtain net immigration. However, currently the Bureau has no way of estimating the number of illegal aliens in the country.

While demographic analysis is quick, easy, and inexpensive in comparison with matching studies, it does have limitations. The estimates for the 45-64 year age group in 1980 will be based on data on the 35-54 age group in the 1970 census group. The 1970 estimate for this age group, in turn, was derived from analyses of previous censuses which themselves contained errors.

Demographic analysis is also limited in that to date the Bureau has not devised a way of estimating the number of illegal aliens. By their nature they are difficult to include in immigration statistics, a component in the demographic analysis. Yet, the estimate of the 1980 census over/undercount

1/The Bureau by its tests has found birth records since 1935 to be relatively complete.

may well depend on the number of illegal aliens the Bureau chooses to include in its official demographic estimate. According to preliminary figures released by the Director of the Bureau of the Census, the number of persons enumerated in the 1980 census will exceed the Bureau's current estimate of the population. That estimate does not include illegal aliens. According to the Bureau, there are currently no reliable estimates for illegal aliens. Recent speculation put their number at between 3 and 6 million. 1/

Because of the lack of quality data on internal migration and the weaknesses of State birth records before 1935, the Bureau has no accurate way of estimating State undercounts using its demographic methods. After the 1970 census, the Bureau did attempt to make demographic coverage estimates for States, but the results were considered developmental.

For 1980 the Bureau again plans demographic estimates of State coverage. The Bureau hopes that matching studies, if successful, will provide the data necessary to evaluate the movement of population among the States. The Bureau is also hopeful that research presently underway will extend the data on corrected births back to 1925 or 1915 and thereby overcome the weakness in pre1935 birth records.

Besides having these limitations, demographic analysis has not proven successful in developing coverage estimates for the Hispanic population. Public law 94-311 requires the Department of Commerce to study ways of determining credible undercount estimates of the Hispanic population. Demographic estimates of census coverage for any population group require data adequate to construct an expected population to which census figures can be compared. For the Hispanic population this data base simply does not exist. In 1980 the Bureau plans to develop estimates for the Hispanic population from matching studies.

Finally, a further limitation to demographic analysis is that no way exists to test the reliability of its estimates.

Matching studies--hampered by problems

Matching studies, which have been used since the 1950 census to evaluate the accuracy of census coverage and data,

1/"Preliminary Review of Existing Studies of the Number of Illegal Residents in the United States", prepared by Bureau staff members in January 1980 at the request of the Select Commission on Immigration and Refugee Policy.

have problems. In some cases they have been hindered by errors highly correlated with census error. All matching studies are hampered by a basic problem--the resolution of pseudo mismatches.

These studies involve matching a sample of persons or living quarters to census files on a case by case basis. They fall into two categories: reinterview studies and reverse record checks.

A reinterview study consists of reenumerating a probability sample of households and matching the addresses and names of persons that are listed with the census records unit by unit and name by name. From this procedure the components of coverage error can be derived. Underenumeration (misses), overenumeration (duplicates), as well as persons missed in enumerated housing and persons missed in missed housing units can be projected from the matching results.

According to the Bureau, the reinterview method has two major weaknesses. First, reinterview studies have coverage errors, and these errors are correlated with errors in the census; namely, persons who are missed in the census are also likely to be missed in a reinterview. Estimates of underenumeration based on the reinterview method are biased in the direction of understatement.

The second weakness in the reinterview technique is that it depends on the difficult matching process to determine whether or not a person or housing unit was enumerated in the census. The matching process is plagued by such problems as

- geographic coding of census questionnaires to an incorrect area,
- comparing a survey response to the census questionnaire of an imputed person,
- locating and matching persons who have moved since census enumeration,
- erroneous enumeration of persons in the census or survey,
- nonresponse to census questions resulting in insufficient information for matching,
- refusal to answer survey followup designed to reconcile inconsistencies.

Because of these problems, all cases cannot be classified as definitely in both the census and the sample, in the census only, or in the sample only. Many cases are unresolved. Since the estimate of coverage is based on the number of matches, unresolved cases can greatly increase the bias of the estimate.

To overcome matching difficulties, two Bureau statisticians in a paper presented to the American Statistical Association have suggested subtracting all unresolved cases from both sample and census counts. An estimate of coverage would be made on the residual cases. They believe that procedures which try to match on little or no information, perhaps by balancing erroneous matches with erroneous nonmatches, are bound to increase the variance of an estimate. Moreover, if this balancing fails, results would be biased. To date, the Bureau has reached no decision concerning acceptability of this proposal.

In addition, matching is very time consuming because it is done manually. For this reason the Bureau limited its planned postenumeration survey to a sample of 250,000 households. The Bureau recognized this sample would not permit undercount estimates below States and major metropolitan areas.

The reverse record check method consists of defining a specified list of names, or a combination of lists, as representing a population whose coverage in the census is to be checked. To preserve independence none of the lists may come from the census being evaluated, although they may be drawn from another census. A sample is drawn from the list and is compared with the current census records to measure underenumeration, overenumeration, and net coverage. Persons not found in the census rolls are contacted to obtain their census day address for further search. A Bureau report 1/ notes that this method is less subject to the correlation bias than reinterviews but that matching can be more difficult. In addition to sharing all of the matching problems of the reinterview, record checks suffer from address inconsistencies. The sample addresses come from lists that may identify housing units differently from the census. The lists may also be outdated or seriously incomplete.

1/Estimates of Coverage of Population by Sex, Race, and Age: Demographic Analysis, PHC (E)-4, February 1974.

Analysis of the 1960 and 1970 census matching studies, for which data were available indicates the magnitude of unresolved cases. The Bureau had problems in determining the enumeration status of 4 to 17 percent of the cases in the reinterview studies and 5 to 36 percent in the reverse record checks.

By use of matching studies the Bureau estimated the net undercount for the 1950 census at 2.1 million persons and for the 1960 census at 3.3 million persons. The Bureau considers these national estimates inferior to the demographic estimates which it believes to be more reasonable. However, according to the Bureau, the results of matching studies reflect subnational variations in undercount which the demographic analysis cannot.

In 1970 the Bureau attempted to use a Current Population Survey (CPS) 1/ to census match to evaluate the accuracy of coverage. The results were not included in evaluation publications. The Bureau indicated that the results were unreliable because of the small sample size (about 49,000 households). For the 1980 Census, a CPS/census match with a larger sample is a mainstay of the evaluation program. This will be discussed in greater detail in Chapter 3.

Statistical analysis--hope for the future?

Statistical analysis has not been thoroughly explored for use in identifying and measuring over/undercount at needed user levels. The Bureau has successfully used statistical analysis in other projects, but it has made little use of it in estimating the accuracy of census counts.

Synthetic adjustment and regression analysis are the two principal statistical techniques examined by the Bureau. Statisticians in the Bureau and in the private sector believe that either or both could be useful in distributing undercount below the national level and that regression could help in identifying the components of undercount. 2/ However, the Bureau has not conducted a formal program of research to develop a reliable application of these techniques.

1/The CPS is an ongoing study conducted by the Bureau to collect monthly data on the labor force.

2/Conference On Census Undercount, Proceedings of the 1980 Conference, U.S. Department of Commerce, Bureau of the Census, July 1980.

The synthetic adjustment involves applying specific rates of undercoverage for particular population groups at one geographical level to the same population groups at a subordinate geographical level. By this procedure the national rates of undercount by race, age and sex can be used to adjust the census counts by race, age, and sex for any and all parts of the land. For example, after the 1970 census, the Bureau estimated that black males aged 35 to 39 were undercounted nationally at a rate of 17.8 percent. By this simple synthetic method, the census count for this population component in any specific community would be adjusted upwards by 17.8 percent.

The use of the simple synthetic method of adjustment requires the acceptance of a maintained hypothesis and two assumptions. The maintained hypothesis is that for any population category there is no statistically significant difference in the rate of undercount across the universe. The simple synthetic method assumes the true population to be that derived by demographic analysis. This version of synthetic adjustment assumes race, or race, age, and sex to be the components of undercount.

If these assumptions and the hypothesis are accepted, the simple synthetic method provides an easy and understandable way to correct the census counts for all sub-national levels. However, all are subject to serious challenge.

There is ample evidence from past census evaluations that geographic variations in coverage exist and that the rate of census coverage is not uniform across the country. For example, in the relisting of 8,000 city blocks after the 1970 census, the Bureau found that of the 1,741 addresses missed in the census, 30 percent were concentrated in 8 blocks. Further, the demographic estimate of true population is untestable and is based on incomplete data. Also, a Bureau study suggests that age and sex contribute little to the variation in coverage among the States. Finally, some Bureau researchers feel that race may not be the major cause of undercount.

These researchers believe that, while undercount can be measured in terms of race, race itself may contribute far less to it than: the efficiency with which census district office personnel enumerate, geography, housing conditions, literacy of the population in an area, number of imputations to an area, and highly mobile or unattached persons. Other possible causes of undercount identified by Bureau personnel are urban-rural residence, income, employment, and migration. Some Bureau officials feel that the factors of undercount are not universally constant but rather vary from area to area.

A simple synthetic adjustment of undercount based on race, or on race, sex, and age could logically approach the truth more closely for the nation as a whole than not adjusting. But, given the uneven distribution of undercount, a risk exists that an adjustment not based on the real causes of undercount could actually adjust subnational counts further from the truth. The application of a national ratio of undercount to an area of little or no undercount could produce an overcount. And if illegal aliens are added to the demographic estimate of true population, according to a member of the Bureau's task force on undercount, a uniform distribution of them using the simple synthetic method would be contrary to evidence indicating their geographical concentration.

As a possible method of distributing any 1980 undercount and identifying its causes, the Bureau is examining regression analysis. Regression is a generally accepted statistical technique by which the relationship between a dependent variable (undercount) and a set of predictor or independent variables can be analyzed. In the case of undercount, tabulations or scores of factors believed to cause undercount (independent variables) could be weighted and totalled to obtain possible predictions and/or explanations. By regression analysis the Bureau can potentially identify the components of undercount.

Although there is evidence of some Bureau research using regression analysis, this research is largely at the conceptual level. The Bureau has yet to determine by experiment what census data to use as independent variables, how to measure them and the best geographic level of analysis. According to a member of the Bureau's Statistical Research Division, use of this technique to distribute 1980 undercount will depend on future experiments conducted with 1980 census data.

Discussion with key Bureau officials and review of Bureau records demonstrate that in recent years the Bureau has not performed a coordinated, vigorous program of applied research using statistical techniques to identify, measure, and distribute over/undercount.

In 1979 and 1980, the Bureau held several conferences on census undercount. For the February 1980 conference, the Bureau solicited papers from the statistical community. None of the participants reported findings from experiments that might be adopted to improve 1980 adjustment techniques. A principal Bureau conclusion from the conferences was that the only methodology readily available to the Bureau might be a combination of demographic analysis and the synthetic method.

Subsequent to the conferences, the Bureau accepted a proposal from a research firm to statistically analyze the 1970 CPS/census match data. Any techniques developed by this experiment could be extended to the 1980 census.

CONCLUSION

Over the last three decades, the Bureau's coverage evaluation program has concentrated on demographic analysis and the use of matching studies. Demographic analysis and matching techniques at this time offer, at most, hope of producing estimates at the State and major metropolitan area levels; however, estimates are needed at much lower substate levels.

The quality of demographic analysis has been improving with more refined data on births and deaths. However, estimating the number of illegal aliens does not seem practicable in the foreseeable future. Matching studies include the basic problem of assuring the validity of the match, as well as the other problems of correlation bias and the time needed for matching a large number of cases. These problems have existed for years and there is no evidence that the Bureau can resolve them in the near future.

While the use of the other technique, statistical analysis, does not guarantee success, it does, according to Bureau and private sector statisticians, offer promise. The Bureau has not thoroughly explored its use and applications. The Bureau has made some recent efforts consisting of soliciting ideas from the statistical community, and some preliminary research using 1970 census data. However, there is no clearly defined program with specific goals and milestones to perfect the use of statistical techniques in analyzing over/undercount errors.

RECOMMENDATION

We recommend that the Secretary of Commerce require that the Director, Bureau of the Census organize a program of experiments using statistical methods of analysis to determine whether those techniques improve estimates of the true population at the State and substate levels.

AGENCY COMMENTS AND OUR EVALUATION

The Census Bureau agreed that the methods available for estimating undercount have limitations. The Bureau currently believes there is no statistically defensible way to accurately estimate the undercounts for subnational areas or perhaps even at the national level. The Bureau stated that the 1980

census undercount using demographic analytical techniques and not considering illegal aliens is close to zero. The Bureau believes that the reduced level of undercount will make the task of estimating error in the counts more difficult. This occurs because the errors of measurement may be larger than the error being measured.

Regarding our recommendation, the Bureau plans to continue planning and research to develop undercount estimation techniques for use in the 1990 census. The Bureau commented that this effort will require substantial resource allocations and major development work over the next several years. This program will include experiments with statistical methods and specific goals and milestones as we recommended.

The Bureau did express concern that it appeared we were not recognizing that the Bureau has experimented and used statistical methods of analysis. As stated in the report the Bureau has successfully used statistical analysis in other projects. Statisticians in the Bureau and in the private sector believe that the application of statistical analysis should be explored for estimating the true population at sub-national levels.

CHAPTER 3

BUREAU REDUCES ITS EVALUATION

PROGRAM TO PROVIDE MORE TIMELY ESTIMATES

According to the Bureau, the large-scale postenumeration survey was deleted so that more timely preliminary estimates of undercounts at subnational levels could be provided. The survey would have included the interviewing of about 250,000 households and matching the results to the 1980 census. According to the Bureau, population estimates under the revised plan will lose precision. The lost precision is greater below the largest State and city levels. The practical significance of the lost precision will ultimately depend on the size and use made of the undercount estimates and users' acceptance of the estimates.

To compensate for benefits lost by dropping the survey and to improve upon previous evaluations, the Bureau added a survey that identifies erroneous enumeration and improves the CPS/census match originally included in its expanded evaluation program. The CPS/census match and the added survey (a reinterview) have limitations similar to other matching studies of their kind because of structure and correlation bias.

PLANS ALTERED TO EXPEDITE THE ESTIMATES

The Bureau dropped its planned large-scale postenumeration survey mainly because of time limitations. The Bureau expects to save at least six months by eliminating the large-scale survey. According to plans in February 1980, interviewing for the survey was to take place in October through December 1980 and matching to census records and tabulations would be completed around December 1981. Thus the results would not have been available for immediate comparison with demographic estimates scheduled to be available about six months earlier. Nor would the results have been available by July 1981, the time when data estimates will be needed for distributing revenue sharing funds.

Further compounding the problem was a delay in completing the maps needed for the survey. The Bureau thought this delay would stretch out the completion of the survey until mid-1982.

Consequently, in early April 1980, the Bureau decided to drop the survey. Thus, the Bureau limited its undercount evaluation program for subnational levels to its originally

planned CPS/census match, an added erroneous enumeration survey, and the originally planned match to administrative records.

MILLIONS THAT MAY BE SAVED WERE
NOT MENTIONED IN FUNDING REQUESTS

Although the Bureau believed it would save about \$8 million by eliminating the large-scale postenumeration survey, it did not mention these cost savings in its appropriations requests. Moreover, the overall cost of the census will not be affected because these savings will be needed to compensate for other census overruns.

In 1978, the Bureau estimated the cost of the survey at about \$16 million. Because of inflation the estimate was subsequently revised to about \$18 million. Money for most of the project was included in the Bureau's fiscal year 1980 budget. When the Bureau began to experience cost overruns in other 1980 census operations, it planned to postpone to fiscal year 1981 most of the operations of the large-scale survey.

However, even with the planned postponement of the survey and the reduction of other census operations, the Bureau still needed an additional \$50 million for fiscal years 1980 and 1981 to complete the census. In letters to the Appropriations Committees dated August 28, 1980, the Bureau alerted them to its financial problems. In these letters the Bureau identified its funding shortfalls and described its methods for reducing costs. The elimination of the large-scale survey with the associated cost savings of about \$8 million was not mentioned.

In mid-September 1980, the Congress passed an urgent supplemental appropriation which provided \$27 million for the census operations in fiscal year 1980. Also, the Bureau is requesting \$23 million in an amendment to its fiscal year 1981 appropriations.

PRECISION LOST BY DELETING
THE LARGE-SCALE SURVEY

Without the large-scale survey, time and money may be saved, but precision in the estimates will be lost. The original plan was to survey approximately 250,000 households. These households were to be drawn from a sample independent of the 1980 census. Geographic areas were to be selected and then a subsample of households was to be drawn. These households were to be interviewed by experienced interviewers. Once this was completed, the persons listed in the survey were

to be matched on a one-to-one basis with the 1980 census listing of names to provide estimates of the omission rate for the 1980 census. The primary objective was to provide estimates at the State level. As a secondary goal this survey would have provided estimates below this level. Specifically, the survey would have provided estimates for all States, 26 large cities--each with a population over 500,000--and their SMSA's ^{1/} and for six additional cities and their SMSA's which had a minority population of over 40 percent where "minority" is defined as black or Hispanic. This would also help the Bureau identify the miss rate for minorities in regional areas.

With the revised evaluation program the Bureau will be able to provide estimates for the large States and the largest cities with about the same level of precision estimated in the original program. However, its ability to provide estimates below that level will be changed as shown below.

Expected Loss in Precision Resulting from
Deleting the Large-Scale Postenumeration Survey

<u>Area</u>	<u>Bureau's estimated coefficients of variation a/ ranges</u>			<u>Range of precision lost (percentage)</u>
	<u>Original target</u>	<u>Original program</u>	<u>CPS alone</u>	
Large States--populations of at least 5 million	.17-.33	.3-.6	.3-.7	0-17
Other States	.34	.6-.7	.7-.8	14-17
Large cities--populations of at least 1 million	.24-.34	.5-.8	.5-1.0	0-25
Intermediate cities--populations between 500,000 and 1 million	.34	.8-1.1	1.0-1.8	25-64
Special cities (population between 250,000 and 500,000 and over 40 percent minority)	.35-.56	1.1-1.5	1.8-2.6	64-73

a/A measure of the variation around the average as a percentage of the average.

^{1/}A standard metropolitan statistical area (SMSA) is defined as a county containing at least one city with 50,000 inhabitants or more, or several economically and socially related contiguous counties with at least one city of 25,000 inhabitants or more. In the New England States, where SMSA's are comprised of cities and towns, the minimum population size is 75,000.

The effect of the lost precision will ultimately be determined by the rates of undercount. If the error in the estimates of undercount for adjustment exceeds the rate of undercount the value of the estimate will be lost.

Thus, for example, if one of the smaller cities with a large minority population had an enumerated population of 300,000 from the census and the evaluation studies identified an undercount of five percent, its estimated true population would be 315,000 persons. Under the original program this estimate would have a range, at the 95 percent level of confidence, of between 305,739 and 324,261 ($315,000 \pm 9,261$). Under the revised program considering a 73 percent loss in precision, the estimate would have a range of 298,947 to 331,053 ($315,000 \pm 16,053$). With the loss of 73 percent in precision, the estimate for this example is now such that one cannot determine if there was in fact an undercount because the lower limit of the range (298,947) is less than the population counted (300,000).

REFINEMENTS MADE IN EVALUATION PROGRAM
BUT LIMITATIONS STILL EXIST

Although the Bureau refined and modified its evaluation program to compensate for the elimination of the large-scale postenumeration survey, limitations in the program still exist.

CPS/census match
has limitations

With the elimination of the large-scale survey, the matching of results of the CPS to the census, which was a part of the original expanded undercount evaluation program, has now become a mainstay in the revised program. Refinements have been incorporated into this procedure. However, it contains certain limitations, including correlation bias and matching difficulties which are inherent to all matching studies.

The CPS is a monthly project whose purpose is to collect current data about the numbers of persons in the U.S. who are employed, unemployed, or not in the market for jobs. Approximately 82,500 households are selected for interviewing during the weeks containing the 19th of each month. Only about 66,000 are actually interviewed. This occurs because (1) many of the households are vacant or no longer exist and (2) about 3 to 5 percent of the families in occupied units do not respond. Since the CPS is required monthly, little time is allowed for following up nonrespondents. The base list of addresses is derived from the most recent decennial census, currently the 1970 census. In addition, area recanvasses, samples of newly

constructed households, and other supplements provide new addresses. However, the majority of addresses now used for the CPS is obtained from the 1970 census list.

The evaluation program will use the results of the April and August 1980 CPS/census matches to provide estimates of the omission rate for the 1980 census. Persons listed in the two monthly surveys will be matched on a one-to-one basis with the 1980 census listing of persons. However, due to the limited coverage of military and institutionalized persons in the CPS a refinement has been added. The Bureau has added a special place enumeration into the undercount evaluation program to include about 3,000 interviews in such places as nursing homes, hospitals, jails, and military barracks.

However, this revised CPS/census match, even with its refinements, has limitations. The hard-to-enumerate persons tend to be missed by both the census and CPS--persons who desire not to be counted, such as fugitives, illegal aliens, violators of building occupancy requirements, and drifters who are not associated with a specific household. This correlation bias in which the same people are missed in both the CPS and the census occurred in 1970. Males did not show a higher undercount rate than females, although that relationship had been consistently documented by demographic analyses.

This bias will be somewhat reduced because the 1980 evaluation has included an administrative records match. The administrative records match involves matching lists from at least two other sources which are independent of the decennial census to the census. The primary lists to be used are the Internal Revenue Service tax return file for 1980 and the medicare enrollment list for 1980. The tax return file will be used for people under 65 years old, and the medicare file primarily for those over 65. However, the results of this matching operation will probably not be available until early 1983.

Still another limitation to the revised program is its reduced coverage evaluation for households. Because the revised program starts with household addresses from the CPS instead of a geographic location as originally planned in the large-scale survey, certain estimates for a housing inventory will be impossible. There will be a less precise estimate of whole households missed because of the reduced sample size caused by dropping the large-scale survey. And there will be no estimate of the number of incorrect counts due to housing definitional differences. The Bureau had recognized these limitations in its evaluation program but as of September had not decided on any plan of action to resolve the problem. Some Bureau discussion has centered around a

further study to be conducted after the CPS/census match is completed; however, no plans for such a study on housing have been formalized to date.

E-sample generally compensates for erroneous enumerations

The large-scale survey would have identified such errors as households that were counted in the wrong block, individuals counted in the wrong households, children counted who were actually born after April 1, 1980, and any duplicate counts of people or households. The Bureau calls these errors "erroneous enumerations." Because the large-scale survey was deleted, an erroneous enumeration sample (E-sample) was incorporated into the revised evaluation program. When drawing the E-sample, a selection of areas from the 1980 census address registers is to be chosen first and then the serial numbers from the registers within that area are to be statistically picked. These serial numbers identify the households which will be interviewed from mid-November 1980 to early January 1981.

Movers further complicate record matching

The problem of resolving nonmatches discussed in Chapter 2 will also occur in the revised evaluation program. Movers present a special matching problem. Those persons who moved between census day and the time of the CPS interview can make matching almost impossible. The April CPS did not ask the respondent where he/she lived as of April 1. The people who had just moved to a new home are not identified in the April 1980 CPS and therefore will be enumerated at a different address in the CPS than in the 1980 census. The August CPS does ask people where they lived on April 1, but in the August interviews the problem of recalling the residency status of occupants as of April 1 can occur. As a result all movers may not be identified.

To help resolve these problems, several phases of review have been included in the matching process. A nonmatch will not be accepted without review by at least one other person. Also, the Bureau has included a followup interview phase to help reconcile these unresolved cases.

Delays in the revised program

This entire undercount evaluation program is dependent on the availability of material and data from the 1980 census. Any delays in the 1980 census will have a domino effect on the undercount evaluation program. Already delays of this sort have occurred. Since some 1980 census district offices

closed six weeks or more later than expected, the initial processing of the CPS supplements was changed from August 1 to September 1, 1980. The E-sample interviews were changed from mid-October to mid-November 1980. Thus, to date at least a 1-month delay has occurred.

Since the July 1981 targeted date for estimates did not allow for any slippages, this date is now doubtful. A more recent forecast includes a September 1981 date.

The sketch map process also has experienced delays. The purpose of the sketch maps was to enable geographers to spot the physical location of the CPS household on a 1980 census map and determine its precise 1980 area. CPS interviewers were given instructions to draw a map of the household's location. Later in the processing, this hand drawn map was to be assigned a geocode number, and this number would be used to facilitate matching the census records to the CPS form. Due to insufficient information, lack of zip code and address ranges, the geocoding of the sketch maps proved a difficult task and was more time consuming than originally planned. In addition, this geocoding slipped because too few clerks were assigned to the operation.

Other types of delays could cause problems in the future. During the holidays in November and December recruiting for interviewers may be difficult. Still another recruitment problem is the conflict with other surveys and their need for interviewers; during this same time period, the content reinterview for the decennial census and the Bureau's annual housing survey will be conducted. Also, during the winter months in some areas, weather conditions are not conducive to effective followup canvassing.

CONCLUSIONS

The Bureau has revised its plans for the evaluation program to allow for more timely preliminary undercount estimates at subnational levels. However, according to the Bureau, precision will be lost in providing estimates of population coverage. The amount of precision lost is greater below the larger State and largest city levels, and particularly for small cities with large Hispanic and black populations.

Whether the time and money saved outweighs the lost precision will ultimately depend on the size and use made of the undercount estimates and the users' acceptance of the estimates. However, given the national interest in population coverage estimates, we believe that the Census Bureau should

not have unilaterally made the decision to revise its evaluation program. The decision to not spend about \$8 million budgeted for the large-scale survey and to delete the survey was not communicated to the Congress when the Bureau requested additional funding to complete the census. We believe that this change in the undercount evaluation program should have been communicated to Congress, as it could affect the planned estimates of census coverage.

RECOMMENDATIONS

We recommend that the Secretary of Commerce require for future censuses (1) the Director, Bureau of the Census determine the feasibility of providing credible estimates of over/undercounts and distributing them to needed user levels, and if feasible prepare detailed information to describe the methods to be used, including the expected reliability of such estimates as well as the time and costs needed for these efforts and (2) an independent assessment of the Bureau's proposed methods be made.

We recommend that the Congress enact legislation requiring that the Secretary of Commerce submit a report to the legislative and appropriation committees having jurisdiction over the census 18 months prior to each census addressing whether and how he intends to prepare estimates of over/undercounts at subnational levels, including characteristics for the undercounted persons. The report should include the cost and time needed to make these estimates and an independent assessment of the Bureau's proposed methods. After the submission of the report and before completion of the census, if the Secretary finds new circumstances exist which necessitate a change in his plans, he should provide an updated report identifying the changes and their consequences. See appendix II for suggested text of the legislation.

AGENCY COMMENTS AND OUR EVALUATION

Regarding planning for the 1990 census, the Bureau said it will continue its research to develop undercount estimation techniques. Also, the Bureau agreed that an independent assessment be made of its proposed plans.

The Bureau commented that its decision to drop the large-scale postenumeration survey and rely on matching the CPS sample households to the census should be commended, not criticized. The Bureau said that the small losses in precision are more than offset by the gains in producing undercount data sooner, cost savings, reduced respondent burden, and the possibility of more accurate results.

By dropping the large-scale survey the Bureau lost some precision. The significance of this loss is unknown at this time. As the Bureau points out, currently the net undercount error is unknown. However, it believes that the error could be lower than in previous censuses. If true, the task of estimating the error becomes more difficult. As demonstrated in the report the errors of measurement may be larger than the error being measured and hence the need for a more precise measuring tool. Consequently, the precision lost in dropping the survey could become crucial in adjusting 1980 census counts at various subnational levels.

CHAPTER 4

HOW WILL ADJUSTING CENSUS DATA AFFECT

FUNDS DISTRIBUTION--WHO KNOWS?

Adjustments to 1980 decennial census figures would affect the amount of Federal and State program funding received by recipients. To predict the magnitude and direction of such changes, however, would be nearly impossible in a multiprogram situation. This unpredictability of results arises because of numerous dampening variables which inhibit and complicate the influence of fluctuations in formula data elements on funding allocations. These complicating factors have generally been overlooked by public officials when they have attempted to document "lost government funding" due to a census undercount.

USE OF CENSUS DATA FOR FUNDS DISTRIBUTION IS WIDESPREAD

Census data influences the annual distribution of billions of dollars in Federal and State funds. No precise estimate is available, but the combined fund distribution for fiscal year 1980 could have been in the range of \$100 billion. At this rate, census data could influence the distribution of \$1 trillion over the next decade. Federal grants-in-aid to State and local governments accounted for \$82.9 billion in assistance expenditures. Distribution of State revenues to local communities has never been accurately measured, but its magnitude is suggested by a 1977 survey of the States conducted by the Advisory Commission on Intergovernmental Relations. The 23 States which responded distributed about \$21.3 billion in fiscal year 1974.

Both Federal and State assistance programs rely heavily on formulas to allocate funds. The data for these formulas are principally derived from the census. For example, Federal programs which incorporate population and related characteristics in their funding formulas are generally required to use census data. This requirement affects a large number of Federal programs. A Congressional Research Service study completed in 1978 found that over 100 programs depend, at least partially, on population data.

The variety and importance of census data is confirmed by an analysis of ten large Federal allocation programs. These ten programs obligated about \$61 billion in fiscal year 1980. An analysis of these funds distributions is included as appendix III.

Recognizing the importance of census data in funds allocation programs, the Commerce Department initiated a survey of census data users to obtain their views on the impact of adjusting census counts. The response to this survey was only partial, and the quality of responses varied. Nonetheless, the responses reveal that respondents generally believe that if adjustments are made they should be made for all formula elements. The respondents emphasized those data elements which identify target recipient groups, such as handicapped individuals or school aged children. They also commented that data should be adjusted to as low a geographic level as possible. Bureau officials currently believe, however, that present adjustment methods may not be adequate to adjust census counts at any level in a statistically defensible manner.

This creates a problem for Federal assistance programs which distribute funds to as many as 39,500 separate government units. Among ten of the largest Federal allocation programs, eight direct at least a portion of their funds to local communities.

IMPACT OF ADJUSTED POPULATION DATA ON FUNDS ALLOCATION IS DAMPENED BY A HOST OF VARIABLES

The impact that adjusted census population data would have on recipient funding levels is difficult to predict because of a host of dampening factors which complicate the influence of changes in population data.

Population is one of several data elements

Population is often just one of several data elements in an allocation formula. Many of the formulas place great emphasis on other data elements. For example, per capita income is an important factor in such programs as general revenue sharing, medicaid, and aid for dependent children. Other formulas utilize data that relate directly to the funded activity, such as road mileage in highway assistance programs.

Finally, some data elements, such as population, can be used as either a direct or an inverse indicator of need. The community development block grant program, for example, views a population decrease as indicating a need for increased funding.

Fixed and open funding

"Fixed funding" refers to those programs that have a preestablished sum of money to distribute. These programs constitute the majority of Federal allocation programs. Since

most Federal assistance programs distribute a fixed sum of money, each community's or State's share of the pie is determined by "percent of the universe" calculations. An increase in funds to some communities reduces the amounts available to other communities. Consequently, the relative rate of population adjustment becomes more important than the absolute levels of adjustment.

The few programs which are open ended do represent a sizable dollar outlay; for example, both medicaid and aid for dependent children are open ended programs. "Open ended" programs have no fixed dollar amount to distribute; rather, the level of funding grows or diminishes based on the application of the funding formula. However, these programs are usually constrained by a matching grant arrangement.

Non-formula determinants of formula based grants

Certain dampening factors, while not directly related to the allocation formulas, have a major role in determining the funds actually received by local governments.

- In some programs, only a portion of funds is distributed by formula, with the balance allocated at the discretion of the sponsoring Federal agency. These discretionary allocations are often used to meet needs that are not reflected by the formula.
- "Hold harmless" provisions insure that program recipients will receive a percentage of prior year funding. For example, in the comprehensive employment and training program, the percentage of available funds going to an area in one year is 50 percent determined by its share in the previous year.
- Laws often mandate minimum and maximum allocations that can be made to individual States and U.S. territories. These minimums and maximums are specified in some programs as annual dollar amounts, in others as percentages of total available funds, and in some as both. Furthermore, some programs specify equal distributions among the States for at least a portion of the available funds.

Thresholds/eligibility requirements

While all the above factors tend to dampen or diminish the influence of single data elements, the thresholds/eligibility requirements increase the influence of certain data elements. Some programs have thresholds/eligibility requirements that are defined directly in terms of population. For example, the comprehensive employment and training program stipulates that recipient communities must have a total population of at least 100,000. More common are indirect thresholds, by which population change affects program eligibility by causing an area to be reclassified, for example, from rural to urban. Reclassification can cause communities to become eligible for some Federal programs and lose eligibility for others. However, the potential impact that thresholds may have is diminished when programs allow communities to combine in order to meet minimum eligibility requirements.

PUBLIC OFFICIALS UTILIZE A SIMPLISTIC APPROACH IN CALCULATING "LOST" GOVERNMENT FUNDING

Local officials have demonstrated the effect of a population undercount by oversimplifying the relationship of population to funds distribution. They do this by identifying the ratio of dollars received to population. Through this argument each additional person counted would increase that community's entitlement.

This approach suffers from three shortcomings in that it

- assumes that population is always the determining factor,
- assumes that program funding is "open ended" and therefore population undercount and levels of adjustment are relevant, and
- ignores the other dampening variables.

Predicting recipient funding levels on the assumption of a direct correlation between population and Federal assistance can lead to inaccurate funding estimates. In April 1978, because of its discontent with Bureau population estimates, the city of Mount Vernon, New York, conducted a sample survey of its population. Based on the results, the Bureau adjusted Mount Vernon's July 1, 1976, population estimate from 67,056 to 76,494. Agencies distributing Federal

funds were notified of this change in mid-1978. City administrators, found that 1979 funding allocations decreased for two major programs--general revenue sharing and the community development block grant program.

Because of the construction of the general revenue sharing formulas the population factor generally does not influence the distribution of funds to local governments. The formula for that purpose relies heavily on the per capita income factor and to a lesser extent on adjusted taxes. As a matter of record, Mount Vernon's subsequent entitlement dropped \$33,417 in funding from the prior year entitlement of \$729,667.

While population did not influence the allocation of Mount Vernon's general revenue sharing funds, it did have a direct impact on the allocations of community development block grant funds. These funds are allocated through formulas which use five data elements. One of the formula elements--population growth lag--rewards population decreases. In Mount Vernon's case its population increase resulted in a drop of \$308,000 (the 1979 allocation was \$1,991,000, compared to the previous year's allocation of \$2,299,000).

CONCLUSION

Individual program analysis serves to illustrate the intricacies of funding formulas and related issues. A comprehensive analysis, however, would require that all relevant Federal and State assistance programs be simultaneously evaluated. This would require an understanding of the funding formulas, access to formula data for all program recipients, and knowledge of the behavior of formula elements under the influence of dampening variables. The simplistic model is therefore not a good approach and could produce inaccurate results. Currently, no method exists for determining the effect on recipient funding levels of adjusting census population data in a multiprogram situation.



UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, D.C. 20233
OFFICE OF THE DIRECTOR

DEC 9 1980

William J. Anderson
Director
General Government Division
U.S. General Accounting Office
441 G Street, N.W.
Washington, D.C. 20548

Dear Mr. Anderson:

Thank you for the opportunity to review your draft report entitled "Procedures to Adjust 1980 Census Counts Have Limitations." We appreciate the effort expended by GAO in reviewing this portion of the Bureau's program and in preparing suggestions for future improvements. We should like to offer several general comments in the spirit of making this report potentially more useful for the purposes to which it will be put.

All of the methods available for estimating undercount are known to be imperfect tools. Each of the methods has its own limitations but each method is expected to yield important insights into the distribution of errors in the counts. These insights can be very useful for evaluation and planning purposes, but our concern is that these data may be too weak for adjusting census counts for official purposes. The current position of the Bureau is that there is no statistically defensible way to accurately estimate the undercounts for sub-national areas or perhaps even at the national level.

The Bureau's plans for evaluation, as well as GAO's review of those plans, were made without information on the completeness of the 1980 census counts. Early data now available from the 1980 census suggest that the undercount in this census may be substantially less than in previous censuses. In fact, the "measured" undercount estimated using demographic analytical techniques is close to zero. However, we believe the true undercount in the 1980 census is larger than the "measured" undercount because demographic analytic estimates do not include an allowance for illegal aliens residing in the U.S. On balance, the early data suggest we may have made some inroads in 1980 in reducing the level of undercount observed in previous censuses.

With this reduced level of undercount for the 1980 census, if realized, the task of estimating the error in the counts by the various methods becomes even more problematic. When measuring a small number with an imperfect measuring tool, the errors of measurement may be larger than the error being measured.

As the results of the various evaluation studies become available and we can analyze and quantify the errors in the undercount estimates, a better picture will emerge about the usefulness of those data for various purposes.

In summary, we agree with the overall thrust of the GAO report that the 1980 procedures available to adjust census counts have limitations. We are acutely aware of the needs to develop undercount estimates for sub-national areas and are continuing our planning and research to develop undercount estimation techniques for use in the 1990 census. This will require substantial resource allocations and major developmental work in the Bureau and in some of the other Federal statistical agencies over the next several years. Even though these efforts may not yield a fully satisfactory technique, we believe this issue merits high priority consideration.

Some specific comments are:

1. The report criticizes the Bureau for dropping the Post Enumeration Survey from the coverage evaluation plans and relying on matching the CPS sample households to the census. The Bureau's position is that this program change should be commended, not criticized. The change means that undercount data will be produced faster, will be less costly to produce, will reduce respondent burden, and more accurate results are likely to be obtained. (For example, some of the deficiencies cited in the report about the revised program may have been more severe in the Post Enumeration Survey.) The small losses in precision cited in the report are more than offset by the gains in timing, cost savings, etc.
2. The report recommends that there be a program of experiments using statistical analysis to determine whether improved estimates can be developed of the "true" population at State and sub-State levels. This recommendation is based on the statement that statistical methods of analysis for estimating coverage are available but the Bureau has done little in using them. The Bureau has had a variety of experiences in using regression analysis for making population and income estimates for State and sub-State areas. In addition, the Bureau developed a new method based on a combination of Bayesian analysis and regression methods that is being used to estimate income for sub-State levels. Thus, it is not a fact that the Bureau has done little in using these methods. The Bureau has used them, refined them, and developed them. However, for the 1980 undercount problem, statistical analysis methods will probably not be useful for distributing undercounts to sub-national levels because they have very little effect on correcting for the biases in the undercount estimates.
3. The report recommends that an independent review be made of the undercount estimation plans for future censuses. We have no objection to an independent review. In fact, the plans for estimating the 1980 undercount have had several independent reviews. First, these plans

were discussed in substantial detail with the nine advisory committees that the Bureau meets with regularly. In fact, some of the changes and modifications that have been made in these plans were a direct result of these reviews with the advisory committees. Second, these plans have been discussed extensively in various professional conferences. Third, the plans were reviewed by a panel established by the National Research Council in 1978 as a part of their appraisal of 1980 census plans. Finally, the Office of Federal Statistical Policy and Standards reviewed the plans and methodology as a part of their function in approving data collection forms.

Sincerely,



VINCENT P. BARABBA
Director
Bureau of the Census

TEXT OF DRAFT LEGISLATION ON
CENSUS REPORTING REQUIREMENT

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled, that this Act may be cited as the "Census Amendments Act of 1981."

Section 101. Title 13, United States Code, is amended by inserting the following new sections after subchapter I, section 15.

"16. (a). Reporting Requirement. Not later than the eighteenth month first preceding the decennial census date established by law, the Secretary of Commerce shall report to the Senate and the House of Representatives on the following, with respect to each such census: (1) the Department's intentions, if any, regarding preparation of over/undercount estimates at subnational levels, and ascertaining the affected population groups' characteristics; and (2) the methodology and procedures, and cost and time needed for making the estimates and determinations described in this section."

"(b). The Secretary shall be under a continuing obligation to advise the Congress as soon as practicable on material modifications or additions to the information reported under this section."

"17. The Secretary of Commerce shall ensure that the report required by section 16(a) of this title is accompanied at the time of its transmittal to the Congress by an independent assessment thereof."

Section 102. The judicial review provisions of the Administrative Procedures Act, 5 U.S.C. 551 et seq., shall not apply to the Secretary's reporting obligations under sections 16 and 17 of this title.

DATA ELEMENTS USED IN DISTRIBUTING FEDERAL
FUNDS FOR TEN LARGE PROGRAMS

<u>Program Name</u>	Fiscal Year 1980 Obligations (\$ billions)	Formula data elements	<u>Source of data</u>
Medicaid <u>1/</u>	14.8	State and national per capita income Administrative data	Decennial Census updated The States
Highway Research Planning and Construction	8.9	Road mileage State area State rural popula- tion Population in urban areas National Forest data Administrative data	Postal Service Decennial Census Decennial Census Decennial Census Department of Agriculture The States
Comprehensive Employ- ment and Training Act	7.2	Unemployment data Low income data	Bureau of Labor Statistics (States and Census Bureau)
Maintenance assistance <u>1/</u> (aid for dependent children)	7.1	State and national per capita income Administrative data	Decennial Census updated The States
General Revenue Sharing	6.8	Population Urbanized popula- tion Per capita income Relative income State income tax collection Tax effort	Decennial Census updated Decennial Census Decennial Census updated Decennial Census updated Census Bureau Survey Internal Revenue Service

1/Formula does not determine total allocation for funds, it only affects a limited Federal matching percentage scale.

<u>Program Name</u>	<u>Fiscal Year 1980 Obligations (\$ billions)</u>	<u>Formula data elements</u>	<u>Source of data</u>
Construction grants for Waste Water Treat- ment Works	4.5	State population Administrative data	Decennial Census updated Environmental Protection Agency and the States
Lower-Income Housing Assistance (section 8)	3.6	Poverty 1975 population Lacking plumbing facilities Overcrowded Vacancy deficit Renters with prob- lems	Decennial Census Decennial Census updated Decennial Census Decennial Census Decennial Census Decennial Census
Community Development Block Grants/Entitle- ment Grants	2.8	Population Population growth lag Housing overcrowd- ing Extent of poverty Age of housing	Decennial Census updated Decennial Census updated Decennial Census Decennial Census Decennial Census
Social Services for Low Income and Public Assistance Recipients (Title XX)	2.7	State and national population Administrative data	Decennial Census updated States
Educationally Deprived Children-Local Educa- tional Agencies	2.6	Number of children in poor families (1970) Children aged 5-17 from families be- low 50 percent of the national in- come for a 4 per- son family Administrative data	Decennial Census 1975 Survey of Income and Education (Census Bureau) National Center for Educational Statistics

(275154)





AN EQUAL OPPORTUNITY EMPLOYER

**UNITED STATES
GENERAL ACCOUNTING OFFICE
WASHINGTON, D.C. 20548**

**OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, \$300**

**POSTAGE AND FEES PAID
U. S. GENERAL ACCOUNTING OFFICE**



THIRD CLASS