

RESEARCH ON THE USES OF ADMINISTRATIVE RECORDS FOR THE 1990 CENSUS

Rachel F. Brown, U.S. Bureau of the Census

I. RESEARCH AREA

The Bureau of the Census is examining methods for the 1990 census that will maintain or improve the quality of the data and coverage of the population and housing inventory, expedite the availability of census data and reduce the cost of taking a census. This research plan examines the potential uses of administrative lists for improving a complete count census or for taking the census in 1990.

II. DEFINITION AND BACKGROUND

Administrative records are data lists maintained by government agencies, private companies and organizations for one of many purposes, including record of membership enrollment or entitlement to program benefits, customer services, and taxation; thus, administrative records include different types of lists, such as telephone and utility company customer service records, drivers license lists, rosters of persons receiving unemployment benefits, college student rosters, union membership lists, state and Federal tax files, and files from past censuses and Bureau surveys. In form, they range from index card files to sophisticated computerized lists, and often contain some of the personal and housing data items requested in a census. It is theorized that most individuals and residential addresses are covered by at least one set of administrative records.

In some cases the information contained in administrative records is based upon observations or measurements made by a third party; in many instances the information has been provided by the person identified on the record. Administrative records can serve as an independent source for verifying the coverage and accuracy of the census, but this use of administrative records is predicated upon the assumption that the data in the records are more accurate than the census. If administrative records data are as accurate or more accurate than data obtained by a self-enumeration census, administrative records provide a potential means for conducting all or part of the census at a lower cost than a direct enumeration.

Administrative records were used by the Bureau in the 1980 census for a number of purposes, including: coverage and content evaluation, coverage checks for persons in minority areas who may have been missed in the enumeration, identification of target populations (such as American Indian reservations and special places) requiring unique enumeration procedures, and the enumeration of institutionalized persons who were unable or unavailable to respond for themselves during the census period. The use of administrative lists in 1980 was restricted by major technical, resource or policy limitations:

1. Current addresses: The benefits of administrative records were contingent upon the currentness of the residential addresses. Many files contained address information that was

out of date, as well as nonresidential addresses, such as lock boxes, businesses or banks.

2. Geocoding: Many lists included rural delivery addresses that did not include a house number and street name and were difficult to identify geographically, as well as addresses without apartment designations for multiunit structures. Persons or housing units with such addresses could not be identified with enough precision to enable matching in some programs, such as the non-household sources coverage improvement check and the coverage evaluation programs.

3. Matching: Since social security numbers (SSN) were not collected on the 1980 census questionnaire, they could not be used to match and unduplicate administrative record lists used for the non-household sources coverage improvement program or the coverage evaluation program. Matching was done manually, using name and address. This process was slow and required a large staff. Additionally, the use of different name configurations between different sets of files (nicknames; initials; different surnames, such as married women using maiden names) made the matching of many cases difficult to resolve without the SSN.

4. Computerization: Confined by limited funds and resources, the Bureau only used national and state files that were readily available on computer tape from the responsible agencies. This resulted in the exclusion of some states from the nonhousehold sources program and precluded consideration of incorporating files from local organizations into the program.

5. Confidentiality: A number of agencies are prohibited by Federal or state laws or regulations from releasing their records to the Census Bureau. It was for this reason the Bureau did not use welfare files for the nonhousehold program.

These problems must be resolved before the Bureau can consider expanding the uses of administrative records for other purposes within the decennial program, increasing the number of files to be processed by the Bureau for the 1990 census, or using them for taking the census. Automated techniques, such as the development of computer matching and unduplication, are essential if the Bureau proposes to implement programs that merge files from a number of sources.

III. PROPOSALS FOR 1990

Presuming the Bureau resolves the confidentiality, logistical and technical problems that restricted the use of administrative records in 1980, there are a number of areas for which the expansion or introduction of administrative records potentially can improve the accuracy and reduce the cost of the 1990 census. It must be noted, however, that the incorporation of administrative records into more aspects of the census

process may generate greater public concern about the development of Federal data banks. Public support is required if the Bureau proposes to substitute administrative records for the collection of data or verify and correct deficiencies in the collection of data. Any perception that the Bureau is building a data bank of dossiers will certainly reduce public participation and cooperation in the census.

A. Coverage Improvement

The main purpose of a complete count census is to enumerate every person and housing unit. This is always difficult, especially when some persons have no permanent address and others move during the census period or do not want to be enumerated. Other problems can occur when addresses are missed during the creation of the master address file. The omission of population and housing will be reflected in the census count as "under-coverage."

Coverage improvement techniques are designed to identify persons and/or housing units that might otherwise be missed in the census. These techniques generally involve the comparison of names and/or addresses contained on administrative lists with census address lists and/or questionnaires (list matching). Administrative records were used in the 1980 nonhousehold sources program, designed to reduce the differential coverage of minorities living in minority neighborhoods in urban centers; these procedures could be refined and/or new applications explored when developing coverage improvement plans for 1990. For example, the nonhousehold sources program might be expanded to include both urban and rural areas with high concentrations of minorities or modified to include nonminority areas. A disadvantage of programs designed to check within household coverage, such as the 1980 nonhousehold sources program, was the need to clerically match the independent lists to the census records and resolve nonmatches while the census enumeration was in progress. This clerical activity was costly, time-consuming and dependent upon the understanding of complicated matching rules. If the ability to data capture the household roster in automated form and execute a computer match with the independent administrative file can be developed, the processing of this check would be expedited and might permit implementation of the program in more areas without increasing the time resource or cost requirements.

During the enumeration of past censuses, the Bureau conducted a number of activities concurrently within a scheduled time period; many of them were checks to verify the coverage of the housing unit inventory. The accuracy of the precensus address list is perhaps the most important factor in accomplishing complete coverage of the population and housing inventory and expeditiously finishing the enumeration activities. Local administrative records such as property tax or utility customer service records may contain more complete information about addresses, such as apartment designation, or more addresses for the designated area than the lists used by the Bureau in past censuses. Therefore, they should be investigated and evaluated as potential sources for compiling or updating the 1990 census address list.

B. Coverage Evaluation

Administrative records have been used to evaluate coverage of past censuses and surveys by providing estimates of population and housing counts and characteristics. In 1980 this involved matching the census with national administrative record sets containing similar data, such as files from the Social Security Administration, Internal Revenue Service and the Current Population Survey. The matching and geocoding problems described in Section II handicapped this program; in addition, coverage by these files of certain population groups was not adequate to evaluate their coverage in the census.

Representatives from the statistical community, local officials and other data users have recommended expansion of the Bureau's coverage evaluation program for the 1990 census to provide undercount measures by demographic characteristics, such as race, sex, age and ethnicity for substate areas. National administrative files may not adequately represent certain population groups and/or geographic areas; therefore, state, county or local lists need to be examined to identify those with coverage of the target groups and/or areas that can be used to supplement the national files to develop any detailed coverage evaluation measures. [1]

C. Content Evaluation

The most common evaluation method is the comparison of data on administrative records with the census data file, using individual record-matching or aggregate analysis. This method was used in 1980 to evaluate the accuracy of respondent recall of expenditures such as utility bills.

Comparisons of data from administrative records to questionnaire content could aid in assessing the extent to which respondents report spurious data for selected questionnaire items, particularly other housing data, such as type of plumbing and value of property. Data from administrative records also can be used in a content evaluation for the purpose of detecting the extent of curbstoning, by collecting preselected data for an automated quality control and content evaluation of the enumeration data.

D. Content Improvement

The enumeration process obtains personal and housing characteristics principally by asking individuals to record the information directly on a census questionnaire. This method may fail to produce the desired information, due to refusal by some individuals to provide all or part of the requested information, the inability of the Bureau to contact a respondent living at an address, or the fabrication of data by enumerators for housing units in their assignments.

In the past, the Bureau has limited its efforts to obtain data missing from census questionnaires by telephone or personal visit follow-ups with a household member, solicit data for a housing unit from landlords or neighbors, or use reference materials developed from administrative lists to code incomplete responses. Numerous government agencies, private industry and organizations maintain files that contain data similar to selected items on census questionnaires. In many cases, such groups obtain their data di-

rectly from the individuals represented in their files. Conceivably, data from these records could be substituted for the missing census data files by matching the administrative record of an individual to his/her census questionnaire or using aggregated counts from the administrative records in an imputation model.

E. Content Collection

Administrative records have not been used for collecting characteristics for the general population in a census; however, some data usually collected by the census may be more accurate and easily obtained from administrative lists. For example, certain housing characteristics, such as property value, type of heating, and utility costs, are often available from utility service billings and/or local government tax or property assessment records. The use of administrative records as a source of some census data may reduce respondent burden and improve the quality of census data without incurring enumeration costs.

F. Special Place Enumeration

The census must count persons living in special situations such as colleges, jails, and military bases, as well as those in housing units. The Bureau devised special procedures and questionnaires to enumerate special places for past censuses. For example, when enumerating places such as mental institutions or hospitals, the enumerators found some residents were unable to complete a census form because of mental and/or physical disabilities; for such cases, the enumerators sometimes obtained the data by asking an employee of the facility to refer to the patients' records. The use of administrative lists can be expanded to enumerate all individuals who are residents of certain types of special places such as isolated military bases, jails, prisons, colleges, or institutions.

G. Administrative Records Census (ARC)

A large proportion of the population is already represented on one or more of the administrative files maintained by Federal or state agencies. These files could conceivably be merged to compile the census counts and data at a cost that is expected to be far lower than the traditional enumeration method. [2] The acceptability to data users of an ARC may be determined by the ability of this approach to replicate certain features of the traditional enumeration method:

1. Representation of the population and housing inventory as of April 1, 1990.
2. Compilation of the apportionment counts by December 31, 1990, and delivery of the redistricting counts by April 1, 1991.
3. Identification of the population in accordance with the census residence rules.
4. Provision of small area data, including race and ethnicity counts.
5. Provision of the detailed characteristics data required for the implementation of formula-grant programs and other data user needs.
6. Verification of the data reliability.

A number of countries already use administrative records for census purposes; however, the

statistical uses of census data by these countries may not be mandated by specific laws require the application of residence rules nor the precision in the accuracy and completeness of the coverage required by Title 13 and the Federal and state program uses.

IV. CRITICAL ISSUES

Before the Bureau can decide how administrative records will be used in the 1990 census, a number of technical, legal and procedural issues must be addressed and the feasibility of the proposed applications must be tested and evaluated. The research and analysis conducted over the next few years should be directed to resolving these issues for the determination of what files are available, the programs for which they are appropriate, the resources required to support the proposed programs and the development of the technology and methodologies required to facilitate the usage of administrative records in the 1990 program.

1. Legal and policy issues: The Bureau must obtain access to the administrative records proposed for use in the decennial operations before it can benefit from any potential savings their usage may provide. The Bureau has had difficulty in the past obtaining specified files for selected or all areas, as a result of local, state or Federal statutes and/or regulations. These restrictions still exist. It will be useful for the Bureau to begin negotiations now with the officials of the agencies and organizations maintaining the files to examine if and how the specific concerns can be resolved by the protection of Title 13, special processing and security procedures that restrict access to the files, or changes to the statutes or regulations. In particular, it is necessary to reconcile any problems concerning Federal and state laws before the Bureau can entertain proposals dependent upon the use of administrative records for major programs.

2. Public perception issues: There are some segments of the population that are reluctant to provide information to any government agency and already perceive the decennial census to be a component of a move to create a master data bank on individuals. A 1979 poll conducted by Louis Harris verified the public did not believe information provided to the Bureau was kept confidential. Increasing the use of administrative records for the 1990 census, particularly if it involves the linkage of records of specified persons or addresses to the census questionnaires, may be construed by the public as the development and maintenance of a dossier file. Concerns by the public about personal privacy and confidentiality of the data will not elicit the cooperation needed to conduct a successful census; at the very least, it will result in a deterioration of data quality and could terminate in large-scale resistance to the census, as happened in West Germany.

The Bureau must be sensitive to these concerns when analyzing the proposals for the 1990 census and develop techniques that will respond to

the concerns about the protection of personal privacy. The Bureau should measure the effect on public cooperation of introducing programs that utilize administrative records, particularly if their use is contingent upon the collection of a universal identifier, such as the social security number on the census form for linkage purposes. If public cooperation is measurably affected by the use of administrative records, it would be injurious to the success of the census program to include them in the plans for 1990.

3. List content and sources issues: Proposals to use administrative records for the decennial census are predicated upon the assumptions that the same type of file (i.e., property taxes or welfare recipients) maintained by state and local agencies and organizations contains the same data items, uses the same definitions or criteria for the file content items, contains residential addresses, and equitably (more or less) covers all areas. In fact, states use different categories for many items, such as race, and files may cover persons living in urban areas more completely than persons in rural areas. Files also are updated with varying frequency and, hence, differ in accuracy of the data.

The Bureau needs to investigate whether or not the file content and concepts are reasonably standardized and, if not, what must be done to standardize the files for use in the census, so the data quality and coverage is equitable for all geographic areas and population groups.

4. Methodological and technical issues: If programs are integrated into the design of the 1990 census that require multiple sets of administrative records, the Bureau must develop an efficient system for controlling, processing and matching the records. Such a system will require automated processing techniques and standardized specifications and formats for the input administrative record files provided by the responsible agencies and organizations, which probably restricts the Bureau to only accepting administrative records presently available on computer files. Most Federal and state government agencies and large national organizations currently maintain their records on computerized files, but this type of record system may not be so prevalent among local government agencies and organizations.

The logistics and resource requirements for controlling processing and matching a large number of files also must be determined before a final decision is made about the use of administrative records to ensure there are adequate resources to support the successful completion of the program. This is particularly critical even if only one proposal requires the use of county-level files (over 3,000) or files maintained by local governmental units (about 40,000). Additionally, if the programs require the matching of names and/or addresses to the census address control file or the census questionnaires, the matching process should be automated to attain a high level of accuracy and process the large number of cases generated for the various programs.

5. Procurement issues: The procurement of the files may require funds for the agencies and organizations from which the files will be requested because they may not be able to absorb the costs to produce the tapes needed by the Bureau. The pressure to reimburse the responsible agencies and organization will be even greater if they are asked to computerize their clerical files. While most Federal and state files considered by the Bureau for use in 1990 are already computerized, many counties and local jurisdictions may not maintain their records by a computerized system, and the Bureau may have to bear the conversion costs, in order to get the files it wants for the 1990 programs.

The Bureau will need to establish a system for handling the contacts with the responsible agencies and organizations. The selection of files from a limited number of Federal agencies and/or national organizations and, perhaps, one type of file from each state could be handled by the available Bureau staff. However, the selection of substate files or multiple sets of files from state agencies may require the designation of a representative for each state to coordinate and resolve any problems and negotiate the process of transferring the files to the Bureau.

6. Scheduling issues: Agencies and organizations vary on the frequency with which the files are updated, so the currency of the files differ with respect to Census Day (April 1, 1990) or any other specified reference date. Additionally, the time required by an agency or organization to produce and deliver a file reflecting the file status as of Census Day, particularly important for the coverage evaluation programs, may differ by as much as a year. Consequently, it is important that proposals to use of administrative records integrate the selection of the reference dates and techniques for file availability into the program designs.

All activities related to the decennial census must be carefully scheduled to allow sufficient lead time and the efficient use and availability of resources and technology. Some agencies and organizations will not require much advance notice of the Bureau's interest in their files, whereas, others will require extensive negotiations to obtain cooperation and delivery of the files. Thus, the process of negotiating for access of the files should be scheduled to allow time for the resolution of the legal, technical and logistical problems without jeopardizing the Bureau's deadlines for processing the files and incorporating them into the census operations as planned.

7. Issues related specifically to using administrative records for coverage improvement: The approach to census-taking for past censuses, and probably for 1990, is to make every effort to enumerate every housing unit and group quarters, and persons within. To accomplish this task successfully, the Bureau must develop a data collection methodology that works effectively for most of the country; and supplement it with techniques to handle special circumstances and check for deficiencies in coverage and correct them. Files on properties and individuals

maintained by state or local agencies probably provide the best potential for improving the coverage of the population and housing inventory obtained in the census; however, these files may not in fact improve the coverage for the areas with the undercount.

The Bureau must identify what types of housing units and persons are likely to be missed in the census, then determine if there are any administrative records that do successfully identify these housing units/persons and determine how the files can be integrated into the census program to improve the coverage before the enumeration is completed. Furthermore, a decision must be made as to what geographic areas these techniques must be implemented in, since the time and resources are not available to process the files for all areas of the country and match every address and name to the census forms.

8. Issues related specifically to using administrative records for content collection and/or improvement: Some proposals have been developed that recommend the substitution of administrative records for missing questionnaire items or even in place of collecting the data in the census. The substitution of data from administrative records can be accomplished by matching individual records to the census forms or developing dummy records in place of the missing forms. The data from administrative records can also be aggregated by geographic area and used in an imputation model to "correct" for deficient record entries.

The linkage of individual records is, perhaps, the most complex use of administrative records because of the matching process, as well as the sheer volume of the files to be processed. It is also the type of activity that may generate the most controversy, because the Bureau, Congress and the public must balance the potential cost benefits of this approach against the perception that the Bureau is building a master file on individuals.

9. Issues related specifically to special place enumeration: Ideally, the data collected in the census is reported by each individual for him or herself or by a person who is knowledgeable about the residents of the household. However, some persons live in group quarters, such as nursing homes or long-term care institutions, who are physically or mentally incapable of answering the questions. Alternative techniques must be used to collect the data for these persons. The patients records are a logical source for much of the needed information rather than the implementation of special procedures by census enumerators.

10. Issues related specifically to an administrative records census (ARC): The conduct of the census using administrative records in place of enumeration techniques has been advocated by some persons who contend that an administrative records census will result in comparable coverage of the population at a greatly reduced cost. Before this approach can be implemented, techniques must be developed for allocating persons to their place of residence; unduplicating the entries for persons recorded on more than one file; consolida-

ting the source files that reflect the population inventory as of April 1, 1990, and producing the apportionment and redistricting counts by the mandated deadlines; and supplementing the source files, to obtain the detailed characteristics data needed for the formula-grant programs and other data needs.

The logistical and technical problems that will determine the feasibility of using administrative records for selected census activities are even more prevalent in assessing the practicality of an ARC. There are other issues that must be addressed for an ARC. The preparatory work for any census, regardless of the collection method implemented is likely to include checks to ensure complete coverage. The design of an ARC may require the development of a precensus address list that has been verified by postal and/or field checks, for matching against the addresses identified on the administrative lists used for the ARC, as well as the direct enumeration of those addresses missing from the ARC file. If coverage or data deficiencies can be resolved only through the merge of a large number of different administrative record sets, there is a serious risk of overcount because of duplicate coverage between the sets. If any direct enumeration is required to correct the deficiencies, or obtain the detailed characteristics data, the cost benefit of an ARC decreases as the field work increases. A decision to implement an ARC is unlikely to result in the elimination of field work by the Bureau, particularly if the counts are not adjusted for deficiencies.

If the Bureau is satisfied that the technical and logistical problems can be resolved and an administrative records census can be designed to meet the mandated requirements of a decennial census and the needs of users, the Bureau will initiate a formal review of Title 13, to determine if any changes are required before the Bureau can legally implement this approach for the 1990 census.

V. BASIC ASSUMPTIONS

There are a number of basic assumptions that will be used in planning, analyzing and evaluating the proposals for using administrative records for the 1990 decennial census:

1. A complete count census will be conducted.
2. A precensus address list will be compiled and automated.
3. Administrative records will be used to improve the base address list.
4. All litigations regarding administrative record use, including the modification of Title 13, will be completed on a schedule compatible with decennial planning and testing.
5. The administrative records to be used for each program will be identified and procured within sufficient time.
6. Content of the 1990 questionnaire will be similar to the 1980 questionnaire.
7. Appropriate measures will be devised to protect the physical security of the tapes and record content.
8. Individuals must be enumerated at his/her usual place of residence and the criteria for determining a person's usual place of residence will be consistent with those implemented for previous censuses.

VI. POTENTIAL OBSTACLES

The successful examination and implementation of new approaches for the 1990 census is contingent upon the accessibility of the administrative records maintained by various agencies and organizations and the availability of the proper resources to develop and test the techniques using administrative records efficiently in the census operations. Nonetheless, circumstances may curtail or prohibit a thorough examination of administrative records:

1. Public desire to maintain full comparability with the procedures of past censuses and public resistance to new methods.
2. Legislation or policy directives may prohibit the use of some administrative records.
3. The inability to develop appropriate methodology/technology to:
 - a. match administrative records with census data;
 - b. geocode administrative records (particularly addresses in rural areas);
 - c. unduplicate record sets; and
 - d. equitably represent population groups and geographic areas.
4. Failure to obtain administrative records reflecting the status and characteristics of the population as of April 1, 1990.
5. Inadequate funding or experienced staff to fully develop and/or apply the techniques described in the plan.

VII. DECISIONS/ACTIONS

Proposals for the use of administrative records in the 1990 census must be examined and resolved for several critical decisions scheduled to be made by the Bureau in the next few years: conceptualize enumeration -- use of administrative records for census enumeration, September 1985; use of administrative records for updating the address control list, September 1986; determine coverage improvement techniques, including use of state and local resources, January 1987; select specialized enumeration methodology (special places, Indians, and so forth), January 1987; determine 1990 coverage evaluation methodology, including use of state and local resources, April 1987; inform Congress of proposed content subjects for census collection, April 1987; determine basic methodology for processing and production of data products, June 1987.

A number of planning committees have convened at the Bureau during the past two years to review

different approaches for conducting the 1990 census, including the use of administrative records. The reports of these committees, together with the evaluations of the 1980 operations that involved administrative records, are being reviewed to assess the proposals for 1990 provide the greatest potential for reducing the census program costs, improving coverage and data quality and expediting the availability of census data.

While the examination of the proposals for using administrative records in 1990 is still underway the Bureau is expected to focus the developmental testing and evaluation activities related to the use of administrative records upon the coverage evaluation and improvement programs, content collection and improvement programs, automated matching techniques and negotiations with Federal agencies and the attorney generals of the states to resolve legal and regulatory constraints of those files the Bureau identifies as the most inclusive for the proposed programs. The Bureau also will examine the feasibility of conducting an administrative record census.

Decisions about the design of the 1990 operations using administrative records must be made by mid-1987, although the actual selection of what sets of records will be used may occur somewhat later. Additionally, some administrative records may be incorporated into the design of the 1990 experimental programs.

Before any final decisions are made about the use of administrative records in 1990, the Bureau plans to review the results of the testing activities and the Bureau's recommendations with the advisory committees, statisticians and members of the data user community.

NOTES AND REFERENCES

[1] The use of administrative records for coverage evaluation is discussed in more detail in Hogan, Howard R., "Research Plans on Adjustment for the 1990 Decennial Census," 1984 American Statistical Association Proceedings, Social Statistics Section.

[2] Alvey, Wendy and Scheuren, Fritz, "Background for an Administrative Record Census," 1982 American Statistical Association Proceedings, Social Statistics Section.