

**EXPANSION OF ADMINISTRATIVE RECORDS USES  
AT THE CENSUS BUREAU: A LONG-RANGE  
RESEARCH PLAN**

*By*

*Ron Prevost and Charlene Leggieri  
Administrative Records Research Staff  
Planning Research and Evaluation Division  
U.S. Bureau of the Census  
Washington D.C. 20233*

Presented at the November 1999  
Meeting of the Federal Committee on Statistical Methodology  
Washington D.C.

**Expansion of Administrative Records Uses at the Census Bureau:  
A Long-Range Research Plan**

*By*

*Ron Prevost and Charlene Leggieri  
Administrative Records Research Staff  
Planning Research and Evaluation Division  
U.S. Bureau of the Census  
Washington D.C. 20233*

**ABSTRACT**

A vision and strategy are presented for the process through which the Census Bureau could reinvent and revolutionize data collection and processing operations in the next millennium. This aggressive strategy is designed to draw strength from current operations through integration, research, and maximizing the utility of administrative records. This strategy will provide the ability to expand annual intercensal products, and alternative approaches for statistically representing the United States with cost reductions in 2010. A major component of this strategy is a complete assessment of how effectively we are addressing our customers' and stakeholders' needs. This needs assessment is encompassed in a long-term strategy with three basic goals: (1) Reduced direct data collection cost; (2) Increased data quality; and (3) Reduced respondent burden.

**I. A VISION FOR ADMINISTRATIVE RECORDS USES AT THE CENSUS BUREAU**

The environment in which the Census Bureau must carry out its mission is changing dramatically. Costs for traditional data collection methods are increasing at the same time that federal budgets are shrinking and public cooperation is declining. Concurrently, the Census Bureau is faced with increasing demands for high quality statistics that are more current and that provide information for small geographic areas. Administrative records offer a solution for generating timely statistics at much lower costs while reducing respondent burden. Profound advances in computing technology and record linkage accuracy have significantly increased the feasibility of expanded uses of administrative records for statistical purposes that avoids repetitive and burdensome inquiries of the public. These advances coupled with spiraling costs of, and public resistance to, traditional data collection have increased the opportunity for significant benefits through using administrative records in data collection, estimation, and evaluation systems.

The administrative records vision presented at the 1996 COPAFS meeting (Prevost, 1996) remains unchanged: ***"to develop a corporate administrative records infrastructure that will maximize the effectiveness of each survey, census, and estimation operation by 2010."*** The interrelation of administrative records with the Master Address File (MAF)/Tiger System, the Standard Statistical Establishment List (SSEL), and current periodic surveys will provide a cost-effective means for creating accurate current statistics for the United States and its environs. Most major programs currently conducted by the Census Bureau are free standing operations. When integrated, these activities will borrow strength from each other through quasi-independent measures of key statistics such as address and person characteristics. This integration will provide the basis upon which policy relevant statistics and analysis can be conducted. The partnership between and across production and research areas will develop a joint data repository that is not just a snapshot of past statistics but is a dynamic tool and statistical system that can reduce costs and increase data accuracy. Individual responses from major surveys are blended with administrative data, thereby increasing their value and providing statistics on an annual basis. For years administrative records have been key elements in the production of intercensal statistics through synthetic estimation and survey controls. We believe they may also be one of the keys to facilitate integration of Census Bureau products.

Our vision assumes that the immediate users of products from a centralized administrative records effort are the Census Bureau's Demographic and Economic programs, with longer term benefits likely for the 2010 decennial census. Some of the program benefits include:

- *Population Estimates:* An expansion of the intercensal estimates program using administrative records can produce timely, small area estimates of population totals and characteristics. Research on the development of small area estimates (census blocks and tracts) will provide the ability to construct intercensal tabulations for Congressional Districts, urbanized areas, school districts, and other user defined areas. The expansion of administrative records capabilities also supports the Small Area Income and Poverty Estimates program (SAIPE), and provides a basis for expanding the Census Bureau's partnerships with local governments for the purpose of creating and reviewing statistical products.
- *Demographic Surveys:* In addition to improved survey controls, an expanded use of administrative records will result in improved survey estimates of population and housing characteristics. Administrative records offer a potential means for: evaluating the quality of survey responses; providing sample selection or stratification information, and; developing edit and imputation schemes.
- *Economic Programs:* Emerging interest in the linking of employer-employee data sets is supported by a centralized administrative record research effort.

- *MAF/TIGER System:* Addresses from administrative records can be used cost effectively as a direct source of updates and as a means to target areas for improvement of the MAF.
- *Decennial Censuses:* Administrative records can be used to target special enumeration methods, improve coverage (Non-Response Follow-Up operations - NRFU), and enhance imputation for missing responses. If developed to the fullest, and with broad acceptance, administrative records could offer alternatives to traditional census taking methods that provide huge cost savings.

The long term beneficiaries of administrative records applications to official statistics are the public (as taxpayers and respondents), other government agencies (who use Census Bureau data), the Congress, and data users looking for timely, relevant statistics at reasonable costs.

## II. CREATING AN ADMINISTRATIVE RECORDS RESEARCH AGENDA

In late 1994 , the Census Bureau formed a team of researchers (Team for Administrative Records Planning) to determine how enhanced use of administrative records might be beneficial to the Census Bureau (Gates and Long, 1995). As a result of this team's recommendations and those of the National Academy of Science (1995), the Census Bureau created the Administrative Records Research Staff in 1996; a corporate staff to conduct research on expanding the uses of administrative records.

### **April 1996 - December 1998: The Infrastructure Development Phase**

Data acquisition, file research, processing, and formal evaluations based on the 1995 and 1996 decennial test censuses, and the 1998 Dress Rehearsal have provided us with a greater understanding of administrative records strengths and weaknesses. During the past three years we have:

- Established an Administrative Records Steering Committee comprised of Division Chiefs from program, research and policy areas of the Census Bureau;
- Assessed computer processing requirements and purchased computers with large processing and storage capacity;
- Developed, expanded, and co-located a corporate research staff within a new division, Planning Research and Evaluation Division (PRED);
- Instituted a Census Bureau-wide administrative record security and data access policy (Clark and Gates, 1999);

- Established a review process for new external and internal administrative records projects;
- Established the Assistant Division Chief for Administrative Records Research in PRED as the central point of contact within the Census Bureau for demographic administrative records data acquisition;
- Compiled inventories of administrative records projects and administrative record files at the Census Bureau;
- Explored new sources of data to rectify known problems in administrative records coverage and developed memoranda of understanding with data suppliers;
- Explored methods of sanitizing, standardizing, and matching records;
- Completed formal evaluations of decennial uses of administrative records (Wurdeman & Pistiner, 1997; Sweet, 1997; Buser, Kim, Huang & Marquis, 1998; Larsen, 1999);
- Initiated a Research Memoranda Series to share findings; and
- Developed a budget initiative to secure funding for administrative records research over the next five years.

**December 1998 - August 1999: The Long-Range Planning Phase** - With our infrastructure in place we began development of a long-range agenda of research geared toward Census Bureau program needs. In the Spring of 1999 we presented to our Administrative Records Steering Committee, an initial vision of how administrative records might be implemented for the next ten years. We received their comments and suggestions, and met with high-level visionaries and executives around the Census Bureau to determine the scope and priority of projects that administrative data might address. Furthermore, we contracted with Fritz Scheuren to develop a report reviewing the federal perspective of administrative records applications and providing recommendations on how we should proceed (Scheuren, 1999).

Our customers and contacts have assured us that continued research on the implementation of administrative records is both necessary and desirable. We distilled all suggested applications to the following most frequently referenced objectives.

1. Administrative records should be used to improve or target improvements to the MAF/TIGER System, and to classify addresses by their commercial and residential uses.

2. Focus on research projects that will yield short-term successes demonstrating tangible benefit; build a history of evidence with iterative successes (e.g., reduced cost, response burden, program and process improvement).
3. Determine the coverage of administrative records for decennial non-respondents (and for the full universe). Determine the accuracy of content (race, ethnicity, gender, and age characteristics of the population).
4. Provide research support for expansion of small area estimates to smaller geographic units including improving input data and geographical coding.
5. Develop administrative records to the point where they can be used to support survey sample design/stratification, sample selection and screening.
6. Use administrative records to link business and demographic data and conduct longitudinal studies -- support the Longitudinal Employer - Household Dynamics (LEHD) Project and review SSEL addresses.
7. Measure the accuracy, availability, operational feasibility of utilizing administrative records to measure response error/bias and replace questions on periodic surveys that typically have high nonresponse rates (or imputation of missing data).
8. Compare administrative records with Census 2000 results. Start with aggregate comparisons, although ultimately we may have to demonstrate micro-level comparability to gain full support.
9. Consider uses of aggregate administrative records data (like food stamps, and Internal Revenue Service income data for SAIPE) for modeling and estimates (e.g., unemployment data, health insurance, school lunch program).
10. Research should support “expanded” uses of administrative records that are consistent with Census Bureau mission and data suppliers’ approved data uses.
11. Conduct decennial testing in 2001-2003, to prepare the agency for a major test in 2005.
12. Research the definitional differences between demographic variables and their administrative record counterparts; determine if we can bridge these differences or if we must consider revising operationally defined demographic variables.
13. Develop appropriate messages about statistical uses of administrative records.

14. Explore how administrative records could be used to enhance American Community Survey (ACS) data.

### **III. IDENTIFYING THE RIGHT ISSUES AND OVERCOMING PIVOTAL CHALLENGES**

In laying out a long-range research agenda, it is important to focus on those aspects of the field of study that will make significant contributions and move the discipline forward. A corollary focus is on the issues that present fatal flaws or are “show-stoppers” if not overcome. For statistical uses of administrative records, these potential “show-stoppers” might be classified into three broad issues.

#### ***File Access/Use***

Title 13 of the U.S. Code explicitly directs the Census Bureau to use administrative records to the maximum extent possible in lieu of conducting direct inquiries. While this legislation provides the appropriate authorization to acquire administrative records from other agencies, parallel legislation governing administrative agencies often exists that restricts access to these data. These legal barriers are in the form of federal laws that specify who can access data and for what purpose (e.g. IRS); federal laws that permit access to other agencies only for purposes of program administration (e.g. Food Stamps); and the Federal Privacy Act which permits access for routine use only, with authorized exemptions for statistical purposes (e.g. Social Security).

The Census Bureau has a 30 year history of receiving IRS files to support the Intercensal Estimates Program as well as our economic statistics programs. Even with this long history, the IRS regulations limit Census Bureau access to data for non-reimbursable programs and only to the extent necessary in order to carry out Title 13 program activities. In other cases, where the Privacy Act allows access for statistical purposes, administrative agencies are cautious, and rightfully so, about privacy implications and public perception issues associated with sharing of confidential data. Adding to the difficulties in obtaining administrative data is the trade imbalance that occurs as a result of agreements that create a data flow in one direction only (e.g. data collected for statistical purposes cannot be shared with administrative agencies). These imbalances can be (and have been) overcome through research collaboration and reciprocal services that do not involve Title 13 data (e.g. geographical coding). Ultimately, authorizing legislation may be the best solution to protect the continued access to, and use of administrative data that are important for national statistical products.

### ***Methodological Challenges***

In addition to data access and use, numerous methodological challenges are inherent in using administrative data for statistical purposes. Administrative data, by definition, are collected for purposes other than statistical and are not subject to the same controlled collection procedures as survey or census data. These issues have been enumerated in many previous discussions of this topic and fall generally into four categories: coverage, content, spatial assignment, and record linkage. By coverage, we mean total coverage of the population of interest, coverage of specific population groups and geographic areas within the universe, and the methods employed to measure coverage since no statistical technique is perfect. Content issues include the availability of content items for the universe, the consistency of variable definitions across files and in comparison to census/survey definitions, the accuracy of data items, and the time reference of records. Some of the most difficult problems to solve relate to the geo-spatial characteristics of administrative records: what is the accuracy and currency of address information on administrative records and how well can the addresses be geo-located within units of statistical measurement (e.g. blocks, census tracts, school districts, political jurisdictions, etc.). Finally, the accurate linkage of records across different data sets is a significant challenge, especially when data are inconsistent, missing or erroneous. The use of Social Security Numbers facilitates record linkage but is not present nor accurate for all records.

### ***Privacy/Acceptability***

The existence of enormous amounts of personal information in electronic form and the Internet revolution both have added to heightened public awareness of, and concern about, information privacy. Information privacy has been defined as the individual's right to control the terms under which personal information is acquired, used and disclosed. The news is full of stories about actual and potential abuses of personal information, especially in the private sector but also, less frequently, in the public sector. But there is another dimension to privacy. More than a century ago, U.S. Supreme Court Justice Louis Brandeis described privacy as "the right to be left alone". Statistical agencies like the Census Bureau infringe on personal privacy each time a survey or census questionnaire is sent to a household or an interviewer knocks on a door. This situation occurs even though the data collection may be required by law and is conducted in the national interest to profile our citizenry and provide critical data for public policy and commerce.

These two dimensions of privacy present conflicting positions when discussing statistical uses of administrative records. The use of existing records presents an opportunity to reduce the response burden and cost of surveys and censuses, but on the other hand, employs information for a different purpose than originally collected. How do we balance these two conflicting interests and gain acceptance by the public, Congress and our data suppliers? If there is resistance to these ideas, are we prepared to make a convincing case that statistical uses of administrative records cause no personal harm, are used only to produce aggregate



numbers, provide cost and response burden reductions, and create an opportunity to generate timely new statistical products?

The Census Bureau is engaged in an active program of privacy policy and research to address some of these difficult questions. One focus of this program is to continually gauge public opinion on privacy through meetings with privacy advocates, national attitude surveys, and experiments in Census 2000. Another focus of this program is to provide and enforce stringent safeguards that protect the confidentiality of administrative data used at the Census Bureau. If statistical uses of administrative records are to expand, we must be diligent in considering and addressing privacy issues.

#### IV. STRATEGIC DEVELOPMENT OF A COMPREHENSIVE RESEARCH PLAN

Given the priorities and variety of activities proposed by our steering committee and internal customers, we have developed a research agenda that crosscuts multiple uses and intersects with our customers' program research schedules. Furthermore, we have chosen to conduct activities that increase in complexity and sensitivity over time. The research agenda supports Census Bureau strategic planning goals and nests within broad Census Bureau objectives to "***Meet emerging customer product and service needs***", "***Better meet customer and sponsor cost objectives***" and to "***Have continuous program innovation.***" In 1998, the Census Bureau established the following tactical goal to meet these objectives: "***Increase the use of administrative records for data collection, processing and evaluation to reduce respondent burden and costs and develop innovative, useful products.***"

We developed our research agenda on the following tenets:

- Focus on research goals that crosscut, especially across program directorates. In this way, the agency can leverage finite resources for research and development activities.
- Begin with a solid assessment of possible options based on research completed. Develop a research plan that progresses in a stepwise fashion. Each new effort builds on a previous activity or activities. This approach and the interdependence of our activities provide a continuous measure of quality, assist in maintaining a scientific approach, and ensure that we do not stray from our ultimate vision; iterative successes.
- Develop an incremental approach that broaches privacy issues gradually. This strategy allows acceptance building, time to measure public climate, and public discourse that has a foundation in weighing of benefits and costs, as well as feasibility.

- Plan research in a time-frame to meet program decision points. In particular, conduct research in time to support decisions about post-2000 programs to update the MAF, expansion of intercensal estimates, and design for the 2010 census.
- The final scope of the research agenda is subject to resource availability pending the acceptance of FY2001 budget initiatives.

**V. ADMINISTRATIVE RECORDS LONG RANGE RESEARCH PLAN  
FY2000 - FY2007**

Based on recommendations of the National Academy of Science and discussions with our internal customers we have developed the following long range research plan for the expanded use of administrative records at the Census Bureau. There are multiple reasons why the Census Bureau has embarked on the path of administrative record research. Corporate administrative records research can support several major operations that the Census Bureau is undertaking: the development of the Master Address File; the Intercensal Population Estimates Program; the Small Area Income and Poverty Estimates Program; Demographic Surveys; and the 2010 Decennial Census. Therefore, we have integrated our research in an attempt to provide a service to these programs. The success of our proposed research program has the potential to provide benefit to some or all of these applications. The goals of our research program follow and a discussion of the research activities supporting each goal are found Section VI.

**RESEARCH GOALS**

**Goal 1. Complete Operational Requirements**

To meet subsequent goals, our first priority is to continue to acquire, process, document, and evaluate national administrative record systems. Our research requires the acquisition, processing, storage, and manipulation of massive national administrative records data files. Our first attempt at building a national portrait of housing units, and population by age, sex, race, and ethnicity will be in FY2000 to support the decennial census experiment (AREX 2000). This national prototype will be based on administrative record systems that provide as broad and unbiased a picture of our nation as possible. Research we have conducted over the past several years has led us to believe the following files will begin to address that requirement:

|   |                       |
|---|-----------------------|
| IRS's Individual Master File-1040 Returns           | (117 million records) |
| IRS's Information Returns Master File               | (750 million records) |
| HCFA's Medicare Enrollment Database                 | (55 million records)  |
| SSA's Numident File                                 | (750 million records) |
| HUD's Tenant Rental Assistance Certification System | (3.3 million records) |

|  |                       |
|--|-----------------------|
| Selective Service System's Registrant File | (13 million records)  |
| Indian Health Service's Registration File  | (2.6 million records) |

In FY2001, we plan to expand our FY2000 population and housing research by creating household statistics and exploring the following national administrative records systems for their ability to provide enhancements to our operations:

- HUD's Computerized Homes Underwriting Systems (*FHA loan application file*)
- USPS's NCOA/LACS (*National Change of Address and Local Address Conversion System*)
- Education's FAFSA- Free Application for Federal Student Aid (*Student loan and grant application databases*)

***Proposed Data Contents of our Annual Research File***

Administrative record files will annually be processed, edited, combined and modeled to create an annual research data file of Census 2000 short form items and selected long form items such as units in structure, year housing unit built, presence of a business, marital status, income, citizenship, and state or country of birth. This file will provide a snapshot of information for a given year. Data items maintained in this file will require further processing, namely geographic aggregation prior to use in Census Bureau programs. We plan to develop, and improve the variables over the next several years beginning with Census 2000 short form items. Based on our file acquisition, research and evaluation the contents of this file may change over time.

***Activities Supporting Goal:***

- Administrative Records Acquisition and Processing:
  - S Annual acquisition of national administrative record files; maintain relationships with data providers by informing them of our needs and planned uses for their data
  - S Annual preprocessing, unduplication, and geocoding of national administrative record files
  - S Maintain controlled access and safeguards for administrative records
  - S Social Security Number validation capabilities
- National Population and Housing Characteristics Tally Evaluation
- Maintain a research and policy program to determine public acceptance of using administrative records for creating statistics.

***Benchmarks:***

Our first measure of success is the acquisition, processing, and unduplication of national administrative record files in FY1999 and FY2000 to support both the AREX tests and the

National Population and Housing Characteristics Tally Evaluation. We expect these tests to assist us in focusing our efforts on improving both the coverage and operations to complete a fully developed prototype for file acquisition and evaluation by FY2002.

## **Goal 2. Improve the Accuracy of the Master Address File/TIGER System**

---

---

The Census Bureau has been using administrative records from the United States Postal System since 1995 to provide new addresses for the Master Address File (MAF). The USPS Delivery Sequence File provides addresses that the Post Office considers in its mailing universe. In city-style address areas, addresses from administrative records provide a complementary universe of information; where individuals want their mail directed. We believe that the combination of these information sources will provide an accurate up-to-date record of city-style addresses in the United States. Rural addresses present more of a challenge. We plan to test administrative records in the Census Bureau's Community Address Update System as a targeting mechanism for directing field operations to collect address updates in rural areas. An accurate up-to-date MAF/TIGER System will improve operations for the survey frames of all the Census Bureau's periodic demographic surveys as well as the decennial census. In addition, geocoding capabilities of a continually enhanced MAF/TIGER System will be needed to geo-locate administrative records for program applications (see Goals 3 and 4).

### ***Research Activities Supporting Goal:***

- AREX Site Specific Tests (2000/2001)
- Baseline National Master Address File Evaluation (2000/2001)
- CAUS Dress Rehearsal (2001)
- Evaluation of Address Updating Procedures (2005)

### ***Benchmarks:***

Our ability to accurately implement these activities is key to all operations that might use administrative records. Key measurement points occur in: FY2001 with the completion of the baseline National Master Address File Evaluation; and in FY2001 with the dress rehearsal of Community Address Updating System (CAUS) procedures. If these tests prove to be successful, a final Evaluation of Address Updating Procedures in conjunction with the CAUS will be required in FY2005 for potential use in the 2010 Census.

## **Goal 3. Explore the Ability to Expand the Intercensal Estimates Program**

---

---

Aggregate administrative records have been employed in the Intercensal Estimates Program for decades as a source of information for the creation of annual statistics of our population. These

estimates are employed by federal agencies to distribute resources from federal grant-in-aid programs. These estimates are also used by periodic demographic surveys as controls to reduce variance. Improved aggregate data will increase the accuracy of these estimates and thereby improve the accuracy of both federal resource distribution and the Census Bureau's current demographic surveys.

The goal is to conduct research that will lead to increased data quality, coverage and geographic detail for the Intercensal Estimates Program. These improvements combined with increased geographical coding ability will provide the basis for creating new intercensal methodologies and statistics for census tracts, congressional districts, school districts, urbanized areas and user defined areas. Note: This goal builds on Goals 1, 2.

***Research Activities Supporting Goal:***

- ARES Site Specific Tests (2000/2001)
- National Population and Housing Characteristics Tally Evaluation (2001)
- Revised Household Income and Household Size Tally Evaluation (2001)
- Baseline National Master Address File Evaluation (2000/2001)
- CAUS Dress Rehearsal (2001)
- Expansion of Race Coding to Revised OMB Directive 15 (2002)
- National Population and Housing Characteristics Micro-data Evaluation (2003)

***Benchmarks:***

This goal is dependent on the same benchmarks as Goal 2. At a minimum, this goal requires address coding precision for county and governmental units. However, the ultimate expansion of the Intercensal Estimates Program requires accurate address coding for census blocks. We will conduct evaluations of person, housing and household data (with their characteristics) from administrative records against the results of Census 2000. By FY2003 we anticipate the ability to recommend whether administrative records can provide inputs to expand the Intercensal Estimates Program.

**Goal 4. Explore the Ability to Improve Estimates from the Small Area Income and Poverty Estimates Program**

---

---

In 1998, the Census Bureau released its first set of county and school district estimates of persons in poverty and income statistics from the SAIPE program. These estimates were based on models that utilized IRS and food stamp program administrative data. The goal is to conduct research and explore data files that provide a better means to predict household size and household income

necessary for these estimates. This may result in identification of administrative record files or combinations of files for use in statistical models. This goal builds on Goals 1, 2, and 3.

***Research Activities Supporting Goal:***

- National Population and Housing Characteristics Tally Evaluation (2001)
- Revised Household Income and Household Size Tally Evaluation (2001)
- Baseline National Master Address File Evaluation (2000/2001)
- CAUS Dress Rehearsal (2001)
- National Population and Housing Characteristics Micro-data Evaluation(2003)
- Evaluation of Address Updating Procedures (2005)

***Benchmarks:***

There are two measures of success for this goal. The first is the ability to improve the tabulation of household size and household income measures. We will have the capacity to recommend the use of administrative records to enhance these inputs into the SAIPE program by FY2001. The second measure of success is related to our ability to accurately geographically code information at the census block level, thereby providing improved inputs for school district estimates. We anticipate having the ability to make a recommendation on the use of administrative records for this purpose by FY2003.

**Goal 5. Provide a Cost-Effective Means for Improving Periodic Surveys**

---

---

Throughout a decade, administrative records can provide information on housing units and their occupants that is more up-to-date than the previous decennial census. The Census Bureau conducts a multitude of periodic surveys for a variety of clients. Often our clients are attempting to develop information for subsets of the population and housing universe. Administrative records can provide cost-effective measures to stratify and select sample cases, thus reducing the cost to the consumer. In addition, administrative records have the potential to improve final survey products by providing up-to-date information for item and household non-response including editing, substitution and allocation schemes, and weighting. The goal is to conduct research that leads to improvement in periodic survey sampling and item non-response methods.

**Research Activities Supporting Goal:**

- Revised Household Income and Household Size Tally Evaluation (2001)
- National Population and Housing Characteristics Tally Evaluation (2001)
- Baseline National Master Address File Evaluation (2000/2001)
- CAUS Dress Rehearsal (2001)

- National Population and Housing Characteristics Micro-data Evaluation(2003)
- ACS Micro-data Evaluation (2004)
- Evaluation of Address Updating Procedures (2005)

***Benchmarks:***

This goal is dependent on the same benchmarks as Goal 2. In addition, we will conduct evaluations of person and household level data from administrative records against the results of Census 2000 and the American Community Survey. By FY2004 we anticipate that we will have enough information to determine if administrative records can accurately portray information required to stratify survey samples.

**Goal 6. Explore Our Ability to Use Administrative Records in the 2010 Decennial Census**

Numerous questions must be answered before administrative records could be employed as a source of information to replace or supplement traditional census taking methods. The success of Goals 1 and 2 will provide for a limited use of administrative records in the 2010 Decennial Census for address database building. However, we must answer the following questions posed by the U.S. Census Monitoring Board (1999) before we can proceed with more extensive uses. Will data from multiple administrative record systems provide:

1. The coverage necessary to conduct a population census?
2. The data items necessary to conduct a population census?
3. Timely information?
4. Information that does not duplicate individuals?
5. Verifiable information?
6. Information in a manner that does not infringe on individual privacy?

We have identified five potential 2010 Decennial Census uses of administrative records. These uses are listed from least to most extensive.

1. The use of administrative records (and current estimates) to target local areas for special field or other procedures to reduce the differential undercount.
2. The use of administrative records to edit incoming census data and provide a statistical procedure to accommodate item non-response. This would replace current data allocation and substitution schemes that are operationally driven (based on information from another house in the neighborhood).
3. The use of administrative records to improve coverage of people missed by traditional methods through matches of administrative records to census enumerations.

4. The use of administrative records to provide a statistical procedure to accommodate whole household non-response. This activity would replace costly field follow-up activities and data allocation and substitution methods.
5. The use of administrative records to provide a core of information to replace traditional census taking procedures. Field procedures would be conducted only on those housing units for which no census information was available.

**Research Activities Supporting Goal:**

- AREX Site Specific Tests (2000/2001)
- National Population and Housing Characteristics Tally Evaluation (2001)
- Baseline National Master Address File Evaluation (2000/2001)
- CAUS Dress Rehearsal (2001)
- National Population and Housing Characteristics Micro-data Evaluation(2003)
- ACS Micro-data Evaluation (2004)
- Administrative Records Census Experiment (2003)
- Cost/Benefit Analysis of Administrative Records Census (2004)
- Field Test of Likely 2010 Census Designs (2005)

***Benchmarks:***

The benchmarks for this goal build on all other goals previously mentioned. Tests of our ability to measure the population and public acceptance will be conducted throughout the early part of the decade. We anticipate a final decision on the use of administrative records in the 2010 Decennial Census by FY2006, after final address tests and 2005 tests.

**VI. BRIEF DESCRIPTION OF PROPOSED ACTIVITIES**

**1. AREX Site Specific Tests**

The Census Bureau is planning an experiment to conduct an Administrative Records Census, called AREX 2000, in parallel with Census 2000. This experiment will be conducted in two sites within two states, Colorado and Maryland. A number of national files are assembled, unduplicated and geocoded. The files we plan to use are the IRS 1040 and 1099 files, Medicare files, Indian Health Service files, Selective Service files and Housing and Urban Development files. We hope to supplement our national administrative records' database with state Medicaid Enrollment records for the two test sites to eliminate a suspected bias of our current data files: persons in poverty. Records for the two test sites are extracted from the national database and some additional procedures are performed, such as clerical geocoding resolution, address verification and special handling of PO Box addresses on the



administrative files. Results will be aggregated and compared to Census 2000 results for various geographic levels and population groups.

The experiment will be repeated in 2001 for the same sites but with a set of refreshed files. For AREX2001, we will conduct an independent Coverage Measurement Survey to estimate the coverage of the administrative records.

## **2. National Population and Housing Characteristics Tally Evaluation**

After completion of the national databases in FY2000 and FY2001, we will tally housing and person data at the census block level and compare with the results of Census 2000, accounting for the absence of special field, clerical and other procedures implemented for the specific site tests as part of the AREX2000 and AREX2001. This evaluation will determine if geographic biases are present in administrative files or data processing activities. It will also focus on the age, race/ethnicity, and gender distributions. This nationwide database evaluation will be used in conjunction with the AREX2000 and AREX2001 site specific tests to assist in ascertaining the operational feasibility of our national processing system.

## **3. Revised Household Income and Household Size Tally Evaluation**

In FY1998, we prototyped the aggregation of household characteristics from multiple IRS income tax forms for use in post-stratification estimation research for the American Community Survey. This effort was conducted in conjunction with the Small Area Income and Poverty Estimates Program. This process included combining information from married persons filing separately, zero exemption tax returns, and other multiple returns filed from the same residence. Outputs from our processing included revised persons-per-household measurement, income measurement, and assignment of poverty status. We also included a geographic component that exempted multiple returns from a business address that was defined as a tax preparation operation. Preliminary research has shown that this technique worked well in our test counties. We will replicate this process for the national file. This activity will also include a nationwide evaluation of county-level tallies of household size, income, and poverty status from our revised procedures against Census 2000 measures of these same tallies.

## **4. Baseline National Master Address File Evaluation**

Following Census 2000, the Census Bureau will measure housing unit coverage using data from its Accuracy and Coverage Evaluation. The objective is to measure the effectiveness of the MAF in accurately reflecting all housing units existing in the nation on April 1, 2000. In addition, evaluations will chronicle the overall Master Address File building process and review the incremental impact of each of the many operations that went into building and updating the MAF. This will serve as a set of baseline statistics as we move forward in developing post-2000 updating methodologies, including the use of administrative records.

**5. Community Address Updating System (CAUS) Dress Rehearsal**

After field tests in 1999 and system tests in 2000, the Census Bureau plans to “dress rehearse” procedures for updating the Master Address File using community contacts and administrative records. This activity is an evaluation of our ability to employ administrative lists and our final address procedures to update the MAF.

**6. National Population and Housing Characteristics Micro-data Evaluation**

We plan to conduct a nationwide match of administrative records to person, housing unit, and household data collected from Census 2000. This activity will measure our ability to replicate census content and characteristics. We also plan to explore whether the U.S. Postal Service National Change of Address File information can be used to update current addresses of people.

**7. 2003 Administrative Records Census Experiment**

We plan to conduct a pilot test of an administrative records census with a focus on operational timing to determine if receipt of files and processing requirements will support delivery of census counts by December 31, as required by law. Experiments to this point will not have addressed this issue. This test will also be designed to measure public acceptance of administrative records use. A potential design is to conduct a mailout to a national sample of households notifying them of a test census that will use existing administrative records to reduce the high cost of traditional census taking. The notification will include a short form questionnaire for those who prefer that we not use their administrative records, including a request for SSN which will allow accurate substitution of their administrative record with the questionnaire information.

**8. Evaluation of Address Updating Procedures**

This activity is a final field test of our ability to update the MAF in FY2005.

**9. Cost-Benefit Analysis of Administrative Records Census**

We plan to analyze the costs of conducting an administrative records census or a census with significant administrative records input. A comprehensive analysis should include an assessment of the cost savings, potential reduction in data content, as well as any social costs in terms of privacy concerns.

**10. ACS Micro-data Evaluation**

Conduct a nationwide match of administrative records to person, housing unit, and household data collected from the American Community Survey. Measure our ability to replicate survey content and characteristics and infer our ability to stratify samples and develop imputation techniques for survey item non-response. Determine if differential coverage

exists and determine how well the two methods of measurement interface as potential 2010 Decennial Census components.

#### **11. Test of Likely 2010 Census Designs**

As part of a major field test conducted in 2005 in order to select the optimal design alternative for the 2010 Census, the most promising uses of administrative records will be tested.

#### **12. Expansion of Race Coding to Revised OMB Directive 15 (October 30, 1997)**

Under revised OMB Directive 15, federal data collection of race and ethnicity must provide for multiple race as a choice of race designation. Currently none of the key national administrative files collect race according to the new directive. Following Census 2000 when we will have actual data collected under the new directive, we will need to expand the capabilities of our current modeling process to support these new requirements. We plan to develop modeling techniques that link parental and child data so that mixed race coding can be assigned.

#### **Time Schedule**

The following table displays our proposed schedule from FY2000 through FY2007 for completing these research activities in relation to other planned Census Bureau operations.

## Timetable for Research Activities and Other Related Program Activities

| Start Date | End Date | Activity   | Goal(s)   | Notes                     |
|------------|----------|--|-----------|---------------------------|
| FY1999     | FY2000   | Administrative Records Acquisition & Processing (1999)                         | 1         |                           |
|            |          | AREX2000 Site Specific Tests   | 2,3,6     |                           |
| FY2000     | FY2001   | Administrative Records Acquisition & Processing (2000)                         | 1         |                           |
|            |          | AREX2001 Site Specific Tests   | 2,3,6     |                           |
|            |          | National Population and Housing Characteristics Tally Evaluation               | 1,3,4,5,6 |                           |
|            |          | Baseline National Master Address File Evaluation                               | 2,3,4,6   | CAUS/StARS                |
|            |          | Revised Household Income and Household Size Tally Evaluation                   | 3,4,5     |                           |
| FY2001     | FY2001   | CAUS Dress Rehearsal   | 2,3,4,6   | CAUS/StARS                |
| FY2001     | FY2002   | Administrative Records Acquisition & Processing (2001)                         | 1         |                           |
| FY2002     | FY2002   | Expansion of Race Coding to New OMB Directive                                  | 3         |                           |
| FY2001     | FY2003   | National Population and Housing Characteristics Micro-data Evaluation          | 3,4,5,6   |                           |
| FY2002     | FY2003   | Administrative Records Acquisition & Processing (2002)                         | 1         |                           |
| FY2002     | FY2003   | 2003 Administrative Records Census Experiment                                  | 6         | 2010 Census Activity      |
| FY2003     | FY2003   | Nationwide Implementation of Community Address Update System                   |           | CAUS                      |
| FY2003     | FY2003   | Full Implementation of the American Community Survey                           |           | ACS                       |
| FY2003     | FY2003   | Selection of Most Likely 2010 Design Alternatives                              |           | 2010 Census Activity      |
| FY2003     | FY2004   | Administrative Records Acquisition & Processing (2003)                         | 1         |                           |
| FY2003     | FY2004   | ACS Micro-data Evaluation  | 5,6       |                           |
| FY2004     | FY2004   | Design of 2010 Implementation Tests for FY2005                                 | 6         | 2010 Census Activity      |
| FY2004     | FY2004   | Cost Benefit Analysis of Administrative Records Census                         | 6         | 2010 Census Activity      |
| FY2004     | FY2005   | Administrative Records Acquisition & Processing (2004)                         | 1         |                           |
| FY2004     | FY2005   | Preparation of Administrative Records for Census Field Test Sites              | 6         | 2010 Census Activity      |
| FY2005     | FY2005   | Census Field Tests of Likely 2010 Designs                                      | 6         | 2010 Census Activity      |
| FY2005     | FY2005   | Evaluation of Address Updating procedures                                      | 2,3,4,6   | 2010 Census Activity/CAUS |
| FY2005     | FY2006   | Administrative Records Acquisition & Processing (2005)                         | 1         |                           |
| FY2006     | FY2006   | Final 2010 Design Recommended  | 6         | 2010 Census Activity      |
| FY2006     | FY2007   | Administrative Records Acquisition & Processing (2006)                         | 1         |                           |
| FY2006     | FY2007   | Preparation of Administrative Records for Dress Rehearsal Sites (if necessary) | 6         | 2010 Census Activity      |
| FY2007     | FY2007   | 2010 Dress Rehearsal   | 6         | 2010 Census Activity      |

## VII. SUMMARY

In 1994, the National Academy of Sciences Panel to Evaluate Alternative Census Methods said: "*The future holds attractive prospects for using administrative records as the keystone in developing a greatly improved small-area demographic data system that can provide data more frequently at no increase and possibly a significant reduction in costs over the decade....*" We have presented an aggressive strategy for expanding the use of administrative records at the Census Bureau in the next decade. By testing and exploiting administrative records in MAF operations, intercensal estimates, and other demographic programs early in the decade, the Census Bureau will gain the experience needed to assess the potential of these records for application in the 2010 decennial census and beyond.

The Administrative Records Research Staff has learned a great deal about federal administrative data files over the past three years. Significant, but not insurmountable, technical and methodological challenges will have to be addressed and overcome in the next several years as we test the limits of using non-statistical data for statistical purposes. Our past investment in this research and the massive investment the American public is making in Census 2000 should be employed as the cornerstone on which we build our program. Our goal is to continue to reap benefits from these investments and develop a new paradigm for collecting, preparing, producing timely statistics that are the basis on which the American democracy and economy thrives.

## VIII. SUGGESTED READINGS

Baucom, Sharon "*What We Know About the TY93 Information Return Master File.*" Administrative Records Research Memorandum #12. U.S. Bureau of the Census. 1997.

Bye, Barry. "*Race and Ethnicity Modeling with SSA Numident Data: Individual-Level Regression Model-Level-2.*" Westat Report, Bethesda Maryland. 1998.

----- "*Administrative Record Census for 2010.*" Westat Report, Bethesda Maryland. 1997.

Buser, Pascal, J. Kim, E. Huang and K. Marquis,. "*Administrative Records File Evaluation.*" 1996 Community Census Results Memorandum #17:, U.S. Bureau of the Census. February 1998.

Clark, Cynthia, and Gerald Gates. "*Policy for Access to, and Uses of Systems of Administrative Records.*" Internal Census Bureau Memorandum. June 1999.

Czajka, J., L. Moreno, and A. Shirm. "*On the Feasibility of Using Internal Revenue Service Records to Count the U.S. Population.*" Mathematica Policy Research Report, Washington D.C.. 1997.

Dumais, Jean, S. Eghbal, M. Isnard, M. Jacod, and F. Vinot. "*An Alternative to Traditional Census Taking: Plans for France.*" Paper presented at the Population Estimates Methods Conference, Washington D.C.. June 1999.

Executive Office of the President, Office of Management and Budget (OMB), Office of Information and Regulatory Affairs. "*Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity.*" Federal Register, Vol 62, Number 210. October 30, 1997.

Gates, Gerald and John Long. "*Summary of Current Administrative Records Issues and Activities in the Census Bureau; Team for Administrative Records Planning Memorandum 95-7.*" Internal Census Bureau Memorandum. September 1995.

Larsen, Michael. "*Predicting the Residency Status of Administrative Records that do not Match Census Records.*" 1996 Community Census Results Memorandum #25, U.S. Bureau of the Census. February 1999.

Leggieri, Charlene. "*Uses of Administrative Records in United States Census 2000.*" Paper Presented at the ECEUN/Eurostat Work Session on Registers and Administrative Records for Social Statistics, Geneva Switzerland. March 1999.

National Academy of Sciences. "*Measuring a Changing Nation: Modern Methods for the 2000 Census.*" Report of the Panel on Alternative Census Methodologies. National Academy Press, Washington D.C., 1999

----- "*Massive Data Sets.*" Report of the Committee on Applied and Theoretical Statistics. National Academy Press, Washington D.C. 1996.

----- "*Modernizing the U.S. Census.*" Report of the Panel on Census Requirements in the Year 2000 and Beyond. National Academy Press, Washington D.C., 1995

----- "*Counting People in the Information Age.*" Report of the Panel to Evaluate Alternative Census Methods; Committee on National Statistics. National Academy Press, Washington D.C., 1994

----- "*A Census that Mirrors America.*" Report of the Panel to Evaluate Alternative Census Methods; Committee on National Statistics. Washington D.C. 1993.

Prevost, Ronald C. "*The Usefulness of IRS Information Returns in the Development of a National Administrative Records Database.*" Administrative Records Research Memorandum #12. U.S. Bureau of the Census. 1997(a).

----- "*COMPASS (Coordinated Operations and Methods to Produce Annual Social Statistics), A Direction for the Future.*" 2010 Planning Offsite Minutes, U.S. Bureau of the Census. September 1997(b).

----- "*Development of a National Housing Estimation System: Accomplishments and Future Directions.*" Paper presented at the Annual Meeting of the Southern Demographic Association in Orlando, Florida. September 1997(c).

----- *"How Can our Research on Administrative Records in Census 2000 Lead to Greater Use in 2010?"* Paper presented at the Council of Professional Associations of Federal Statistics meeting in Washington, D.C.. November 1996.

----- *"Administrative Records and a New Statistical Era."* Paper presented at the Annual Meeting of the Population Association of America in New Orleans, Louisiana. May 1996.

----- *"Assessing the Accuracy and Impact of Population Statistics Through Distributional Analysis."* Paper presented at the Annual Meeting of the Population Association of America in Denver Colorado. April 1992.

----- *"Evaluating the Utility of Population Estimates as the Basis for Distributing Federal Revenues."* Doctoral Dissertation, Bowling Green State University, Bowling Green Ohio. 1991.

Prevost, Ronald C., and Michael Batutis. *"Database Structure and Population Estimates Methodology: Directions for the Coming Decade."* Paper presented at the Annual Meeting of the Population Association of America in Washington, D.C.. 1991.

Prevost, Ronald C., and Annetta Clark. *"Continuous Measurement Procedures and the Statistical Representation of Persons Residing in Group Quarters."* Paper presented at the Annual Meeting of the Population Association of America in San Francisco, California. 1995.

Sailer, P., M. Weber, and E. Yau. *"How Well Can the IRS Count the Population?"* Proceedings, Section on Survey Research Methods, American Statistical Association. 1993.

Scheuren, F. *"Administrative Records and Census Taking."* Report to the U.S. Bureau of the Census. May 1999.

----- *"Integrating Federal Surveys and Administrative Records."* Paper presented at the American Statistical Association, Section on Survey Research Methods. 1992.

Scheuren, F., T. Petska, and R. Wilson. *"Turning Administrative Records Into Information Systems."* Journal of Statistics. 1993.

Sweet, E. *"Using Administrative Record Persons in the 1996 Community Census."* Proceedings, Section on Survey Methods Research, American Statistical Association. 1997.

Urban Institute. *"Democratizing Information: First Year Report of the National Neighborhood Indicators Partnership."* March 1996.

U.S. Census Bureau Monitoring Board. *"Presidential Members Report to Congress."* February 1, 1999.

Winkler, W. E., *"Matching and Record Linkage,"* in BG Cox (ed.) Business Survey Methods, J. Wiley, New York, pp. 355-384. 1995.

----- *"Advanced Methods for Record Linkage."* Proceedings, Section on Survey Methods Research, American Statistical Association. 1994.

Wurdeman, K. and Pistiner, A. *"1995 Administrative Records Evaluation-Phase II."* 1995 Census Test Results Memorandum #54, U.S. Bureau of the Census. 1997.