

THE FUTURE OF ADMINISTRATIVE RECORDS IN THE CENSUS BUREAU'S DEMOGRAPHIC ACTIVITIES

William P. Butz, U.S. Bureau of the Census

As the first panelist today, let me begin by congratulating the organizers of this session. They and their predecessors have sponsored sessions on various aspects of administrative records at these meetings since 1979. The discussions at these sessions and the written record in papers, proceedings, and collected volumes, form a detailed documentation of the uses and possibilities of administrative records in statistical systems. As a relative newcomer to the Federal statistical system, I have found this record of great use.

I will focus my remarks on demographic activities at the Census Bureau. This may be somewhat provincial, but since the Bureau is often considered a center of action--or inaction--in using administrative records, this focus may be useful. I will discuss goals 1 and 4 of the Jabine-Scheuren paper particularly. Their first goal calls for expanded use of administrative record systems in the conduct and evaluation of decennial population censuses and in current population estimates. The fourth goal promotes expanded use of administrative records in all phases of household surveys.

I begin here with a brief review of how the Census Bureau's demographic activities have benefitted from use of administrative records. Then I outline some difficulties that attend their continued or expanded use. I then describe some future uses of administrative records that we are now planning or considering. All this suggests some guidelines for future use.

Many important demographic activities at the Census Bureau have benefitted substantially from our statistical uses of administrative records. Here are a few recent examples.

In 1980 we used a sample of persons on individual tax records matched to census records to evaluate coverage of the decennial census. We also used administrative records to identify special target populations in the census for improving the coverage of the census. For the Income Survey Development Program (ISDP) we matched with a variety of administrative records. We also used Federal and state administrative record systems as sampling frames for pilot surveys associated with the ISDP. The Current Population Survey (CPS) and other current demographic surveys depend fundamentally on administrative records: local building permits are the basis for updating the CPS address sample for new construction between censuses. Finally, there is the Bureau's intercensal estimates program which produces estimates of the population of all states, counties and units of local governments. Federal, state and local administrative records of various sorts document the births, deaths and migrations that drive the changes in these intercensal estimates from year to year.

To understand the Bureau's plans for using administrative records in the demographic area in the future, it is useful to review briefly the problems in such use. There are technical and organizational problems, and there are perceptual problems. Among the technical problems we confronted in using such records in 1980 Census processes were lack of residential addresses,

geocoding problems in rural areas, lack of accurate apartment identification in multiunit structures, and matching difficulties in the absence of social security numbers in census records.

More generally, most problems in our use of administrative records arise because such records are collected and maintained for particular administrative purposes that rarely match our statistical purposes. Concepts and measurement periods are rarely the same. Quality control is sometimes insufficient. In addition, the coverage of administrative record systems is usually not coincident with the coverage of the statistical system with which they might be used.

Other problems arise because the program agencies that collect such records sometimes change them, or fail to change them without regard to our needs. One difficulty is communication. Sometimes the program agency does not inform us of such changes. Even when communications are good, the problems introduced by changes can be intractable. That is, it is sometimes impossible to distinguish data changes caused by changes to the administrative record system from real changes in the demographic or economic characteristics measured by the administrative record system.

These problems are no one's fault. They result when agencies with very different purposes collect data to meet their particular needs, then seek efficiencies by combining the data for statistical purposes. The more integrated are the uses of administrative records in statistical systems, the more troublesome these difficulties will become. Nonetheless, administrative records play a key role in the identification of errors, in imputation procedures, and in the extension of statistics to areas, groups, and time periods other than those for which the statistical data are directly collected. In these uses, administrative records are complementary to statistical data. When administrative records are substituted for statistical data, rather than complementing them, serious problems can and sometimes do arise. As a substitute, administrative records supplant statistical data. This can leave a statistical agency dangerously vulnerable to changes and failures outside its control, and expose users to potential loss of important data.

In either case, increasing use of administrative records for statistical purposes must, in my opinion, be accompanied by increasing influence of statistical interests upon the content, collection, processing, quality assurance and dissemination of administrative record data. This may occur in several ways. Statistical agencies now contract with program agencies for their data in many cases. The characteristics of these data and their delivery might be more explicit in these contracts and enforced in a more businesslike manner. Where the data are not thought adequate for the intended statistical purpose, perhaps the statistical agency must pay more money to buy the desired data

characteristics. This type of solution works through the market mechanism. Another type of solution would emphasize centralized coordination and, to some extent, control of characteristics of administrative record data through a centralized function in the federal government. In these or some other ways, I suspect that the independence of program agencies must give way to influence of statistical interests if statistical uses of administrative records are to increase markedly.

These considerations pertain to technical and organizational problems in using administrative record data for statistical purposes.

Of at least equal concern is the possibility that citizens and their elected representatives will come to perceive that statistical agencies are helping an administrative agency get its job done, where its job can directly influence the well-being of particular individuals and families. The recent German experience indicates that documented instances of such abuse are not necessary for an awakening of public interest. Critics argued there that an agency's identification of an illegal resident or a tax dodger can rarely be traced to a particular bureaucrat working with a particular data file. Hence, how can the people know whether statistical data given freely by them are being abused? German critics sought not just the illegality or impropriety of such misuse, but its impossibility.

I will return to this issue of perceptions. But first, what future demographic uses of administrative records are we planning? For one instance, we will use administrative records in upcoming pretests of the decennial census, to evaluate the quality of housing data and to improve it. We may even be able to substitute administrative data for questions heretofore asked on the census form. We are also seriously considering constructing matched administrative records files for evaluating coverage and content and for studying the practicability of adjusting the census results.

Next, the Bureau has developed an improved methodology for intercensal estimates based heavily on state and local administrative records. We are proposing to release our annual preliminary estimates, based on this method, more than a year earlier than estimates for some states have been released in the past.

Other projects are still in the dreaming stage, but we are thinking seriously about them. One is a longitudinal historical file on U.S. counties, including data from the Bureau and other sources. Another is a microdata file with merged records from the SIPP, establishment data from the economic area of the Bureau, and program records from elsewhere. The long-run goal would be to build these files for public release and statistical use.

Beside these projects planned and contemplated, it is useful also to list some demographic uses of administrative records we have recently decided not to pursue. One such decision was not to collect the FBI identification numbers of individual prisoners in our prison surveys and merge them with FBI arrest records to measure recidivism. Another was not to collaborate with

the Internal Revenue Service to collect home addresses and social security numbers of tax filers and their families for use in the 1985 census pretests.

In spite of these and other restrictive decisions, the Census Bureau does make extensive use of administrative records. Of the more than 70 papers presented at these meetings by Bureau staff, I count almost 40 percent that discuss direct uses of administrative records or plans for their use. When we move more slowly in this direction than some wish, it is because of the technical difficulties I briefly outlined above or the perceptual difficulties to which I now return.

The Jabine-Scheuren paper does not make light of the importance of public trust in the confidentiality of survey data. I want to underscore their concern. I think this issue is centrally important. That it is fuzzy, that it is badly measured and cannot perhaps be much better measured than it is, that the evidence is mostly anecdotal and from other countries, all make me uncomfortable. In our professional lives, many of us are unaccustomed to giving such matters strong consideration. But in this case I think we must.

I think our concern should extend beyond that expressed by the authors: "Without in any way denying that public attitudes and perceptions are important, we hope that starting in 1985 it will be possible to evaluate such proposals [for using administrative records] objectively on their merits." In my opinion, there are two realities of data confidentiality and use. First, the reality of the data abuses that are actually possible or occurring. Second, the reality of the perception of data abuses that people think are possible or occurring.

This second reality is no less real or less relevant than the first. I submit that consideration of the second reality is an important part of "evaluating...proposals objectively on their merits."

Two related arguments for greatly expanded statistical uses of administrative records are that such uses will increase efficiency of statistical systems and reduce costs. If the technical and organizational problems can be ameliorated, this may be true. Even then, though, we must not aim single-mindedly at efficiency in a society which still places a premium on individual privacy. If the demographic area of the Census Bureau heads toward more statistical use of linked administrative records, I want to be confident that we are following the leadership of public demands for efficiency, lower cost, and lower reporting requirements. If the public--the persons who are the source of all information as well as the ultimate consumers--want this efficiency and are not queasy about our producing it, then fine; we will follow and do our very best to produce both efficiency and safety in data gathering and use. But I will not be comfortable moving even more rapidly in this direction than we now are without confidence that the American people are leading us there.

An oft-stated principle is that data collected for administrative purposes should be available

for statistical purposes, but data collected for statistical purposes should not be used for any administrative purpose. Certainly this is true, but it may be too general to be useful. This principle states a distinction which may, in practice, escape the public and their elected representatives.

This danger is particularly clear when program agencies and private organizations can combine aggregated statistical data with their own administrative records to estimate characteristics of specific groups of individuals and to identify outlying individuals for review of compliance or for directed marketing attention. The distinctions of data confidentiality here are becoming quite difficult for the statistically unsophisticated to keep track of. Indeed, just what it is that actually is being kept confidential is a distinction that is becoming more difficult even for the statistically sophisticated.

In this environment I would offer four criteria for choosing among those expanded uses of administrative records that would arguably increase efficiency in demographic statistics:

1. Files should be kept separate and merged only on an ad hoc basis for specific purposes.
2. When merged, files should be narrowly restricted in content and use, rather than multipurpose.
3. Merged files should be used within a single agency rather than by multiple agencies for multiple purposes.
4. Administrative record use works better as a complement, rather than substitute for the functions of statistical data. Complementary

uses include benchmarking, error identification, and imputation. Substitution, on the other hand, can be dangerous for a statistical agency and its users. It is even dangerous for the program agency, because it can lead to sharply altered flexibility in collecting and reporting administrative records. Overall, substitution can reduce needed redundancy in a statistical system, thereby losing a principal source of error identification. Even more important, redundancy is insurance against unexpected failure in part of the system. Like insurance of all kinds, such redundancy in a statistical system has the characteristic of appearing wasteful and foolish during the long periods when it is not needed.

There is pressure now in the Federal statistical system to reduce costs and to reduce reporting burdens. We must all strive to do these, with or without the pressure. I think we are. But alongside these efforts, I hope we will not have to react by reducing the amount of productive, protective redundancy in the statistical system. I hope also that we will not react in ways that may harm the public's trust in the confidentiality and uses of the statistical data it offers. Instead I hope that decisionmakers will keep in mind that the part of reporting burden on the public which is associated with the collection of statistical data is already very small, and that the value of statistical information--though it is inestimable--is very large.