

STATISTICAL USES OF ADMINISTRATIVE RECORDS: WHERE DO WE GO FROM HERE?

Daniel H. Garnick, Bureau of Economic Analysis
Maria Elena Gonzalez, Office of Federal Statistical Policy and Standards

"The 'Paper Blizzard' is the story of...Government-imposed paperwork that is burying the American businessman, or so the American businessman will tell you: 4,400 different Federal forms...if it weren't for the paperwork, the cost of their products to consumers would come down." (Mike Wallace, 60 MINUTES, CBS Television Network, Sunday, January 14, 1979.)

In 1977 the Federal Committee on Statistical Methodology formed a Subcommittee on the Statistical Uses of Administrative Records.* The papers at this session highlight some of the specific issues that the Subcommittee has addressed.

Among the central questions dealt with by the Subcommittee are--

Where do administrative records fit into the U.S. statistical system?

What research is needed for more effective statistical uses of administrative records?

The discussion in this paper begins with an overall response to these questions as background for providing our personal recommendations in answer to the question in the title of today's talk, "Where do we go from here?"

WHERE DO ADMINISTRATIVE RECORDS FIT INTO THE U.S. STATISTICAL SYSTEM?

Among professionals in the Federal statistical agencies and among informed users of the statistics, there is basic agreement that U.S. statistical programs are creditable and generally of high quality. The more usual expressions of concern by users center on their wish for more data series than are currently available and for better access to existing series.

But there is also widespread discontent: The Federal statistical system is fragmented among many agencies. On the one hand, there is a profusion of statistical series measuring similar entities, but, too often, incomparabilities among many of these series raise questions about the quality of the data and make it difficult to estimate statistical relationships. On the other hand, the fragmentation of the statistical system has resulted in a profusion of reporting forms entailing duplicated, wasteful reports and, perhaps, unnecessarily poor quality data.

All the costs involved in data collection and information management are not borne by the statistical agencies. And the absence of complete

assessment of the social costs for generating and managing information has permitted concealed inefficiencies in the allocation of resources in the Federal statistical system.

In 1965, the Committee on the Preservation and Use of Economic Data (the Ruggles Committee) issued its report [16] and took note of the underutilization of administrative records for general statistical purposes. Off and on since then there have been papers and memoranda comparing the advantages and disadvantages for general-purpose statistics of surveys and censuses over more intensive utilization of administrative records, with only casual reference to comparative costs. More recently, as a result of the report of the Paperwork Committee, President Carter specifically required the substitution of the use of administrative records for subnational surveys, and the Federal Statistical System Project (the Bonnen Committee), in its report [10] on issues and options, called for exploiting new data sources provided by administrative programs including, where necessary, influencing them to make some changes so as to increase their statistical utility.

Administrative record systems provide the only sources of information that hold promise for socioeconomic variables currently needed for the allocation formulae of many Federal grants to State and local governments, or for planning and evaluating programs likely to have differential geographic impacts. They represent the only basis for controlling Federal information collection costs and for reducing the reporting burden imposed on the public resulting from Government information requests.

In fact, as fragmented as the U.S. statistical system is, it is remarkable that there is as much utilization of these records as currently exists [5], particularly in light of--

1. the legal and institutional impediments to accessing the data among agencies [2], as well as
2. the incomparabilities and inconsistencies from one statistical series to another, from one time period to another, from one reporting unit to another, and from one geographic identification to another [11].

Current utilization can be viewed as a tribute to the individual agencies for the extent to which they have succeeded in managing information, in the face of the inaccessibilities and incomparabilities, in order to illuminate analytic, program, and policy issues cost-effectively within their areas of responsibility and authority, however suboptimal the utilization is in terms of the overall U.S. statistical system.

Many of the serious problems of low priorities and

*The members of the Subcommittee are listed in the acknowledgements at the end of this paper.

inadequate resources allocated to the development of general-purpose statistics by the administrative agencies, which have responsibility for the basic collection and processing of data associated with the programs they administer, result when the information has few administrative applications, even though they may have statistical applications of critical importance. The key to improved coordination of information collection for administrative and statistical purposes entails the cooperative design and implementation of information collection programs. The need for cooperation is recognized in the Bonnen Committee draft report as a basis for reducing reporting burdens. Perhaps even more important is the contribution that administrative procedures can make to more efficient and reliable collections of statistical information. In particular, an ongoing, smoothly functioning administrative program establishes communication links that can greatly facilitate information in general. Moreover, administrative systems often provide incentives for respondents to maintain records for reporting purposes that they would be unlikely to maintain for purposes of responding to statistical surveys. Thus, the statistical system could generally expect to obtain more reliable information if it is careful to coordinate its information collection with administrative collection systems. And where administrative systems require special adjustments or added features to be consistent with statistical needs, careful coordination is required to ensure a system design that meets both statistical and administrative needs adequately as well as to ensure equitable sharing of the costs of system maintenance by the statistical agencies together with the administrative agencies.

Administrative programs which have only limited use for detailed geographic or industrial reports by multiestablishment companies, for example, have no incentive for requiring the systematic reporting of such detail. Some programs have only voluntary plans for collecting information by industry and county rather than mandatory establishment reporting which could provide invaluable information for regional analysis, fund allocation formulae, administrative mechanisms and planning tools for State and local government jurisdictions, etc. The Social Security Administration (SSA) and the State Bureaus' Employment Security unemployment insurance (UI) programs, for example, use different reporting unit concepts than the Census Bureau uses for its economic census. These differences result in inconsistencies and incomparabilities from data series to data series. Reliable geographic and industrial information needed for statistical applications is also needed for administering nonfederal programs (e.g., State and local taxes on business). And when different agencies require different geographic breakdowns of the data, reporting burdens rapidly escalate. It would also be desirable that those nonfederal agencies demanding financial reports from business cooperate in a Federal-State-local program for designing reports so that business can supply the needed information without incurring ex-

cessive costs for duplicative and inconsistent reporting at all levels of Government.

However diligent individual agencies may be in pursuing cost-effective statistical programs in areas of limited application and authority, there remain the questions of cost-effective overall information collection and management. Make no mistake about it: whatever distortions entered Mike Wallace's description of the "Paper Blizzard," the statistical community cannot avoid facing up to the rising demands for information and the rising discontent with wasteful procedures and with the difficulties users face in accessing and relating statistical series currently produced. Changes on both the supply and the demand sides of the information process continue to occur since the report issued by the Ruggles Committee, and subsequent reports by Dunn [7], the Wallis Commission [9], and the Office of Federal Statistical Policy and Standards [1].

WHAT RESEARCH IS NEEDED FOR MORE EFFECTIVE STATISTICAL USES OF ADMINISTRATIVE RECORDS?

Administrative records, including a diversity of data elements for populations of individuals and of establishments, need to be further explored to determine the nature of the available data. Knott [15] reports on a survey covering selected files. This survey was done to determine availability and characteristics of administrative records. Important questions investigated were the units covered, the extent of coverage, and the smallest geographic areas for which the administrative record data are available.

An inventory of existing administrative record files and their availability for statistical uses is essential to determine the feasibility of proposed statistical uses of administrative records, especially those required to meet the escalating demand for statistical estimates for small areas at frequent time intervals. Three approaches to satisfy this demand in a cost-effective manner are discussed below. These are regression methodologies, matching applications and the system of national income and product accounts.

Regression Methodologies.--Statistics obtained through indirect methodologies, such as "synthetic" estimates, often make use of combinations of sample surveys and administrative record files. Such applications not only reduce the paperwork burden, they often reduce the costs of required statistics by orders of magnitude. Erickson [8], for example, proposed a regression method to estimate population changes of local areas (counties) where data on population growth from the Current Population Survey (CPS) was combined with ratios of births, deaths, and school enrollment from administrative record sets.

At present, the statistics for employment and unemployment for local areas are derived by a handbook method called the "70-step" method [17]. One of the recommendations of the National Commission on Employment and Unemployment

Statistics [6, p. 315] is to "improve the hand-book procedures by using more recent data, including new sources of data from administrative statistics systems and using regression methods to improve components of the system."

Synthetic and regression estimates of unemployment for small areas were also studied by Gonzalez and Hoza [12]. The 1970 Census estimates were used to derive alternative synthetic estimates of unemployment for counties, and a distribution of a relative method error for the synthetic estimates was estimated. In addition, regression estimates for CPS primary sampling units were derived using CPS unemployment data as the dependent variable and alternative synthetic and "70-step" estimates as independent estimates.

Matching Applications.--Matching applications, using different data sets, have been made to develop sampling frames, to develop estimates of coverage, and to improve the coverage of censuses or surveys. Such applications have become increasingly common in recent years. The files being matched may be administrative record files or survey data, or combinations of each.

To estimate the coverage of the 1980 Census of Population, for example, a Post Enumeration Survey (PES) will be matched to the Census questionnaires. Since it is expected that there will be a high correlation between the PES and the Census, an additional match is planned between Federal individual income tax and Medicare records [3].

Another 1980 Census coverage improvement project involves matching (administrative) lists of driver licenses with Census questionnaires, in selected urban areas, to check whether the persons on the driver license list were enumerated. The plans are to add to the Census count those persons that the Census Bureau is able to confirm were not counted [4]. Another administrative list being used for this purpose is the Immigration and Naturalization Service list of legal aliens.

A Linked Administrative Statistical Sample (LIASS) is being developed. The initial efforts have focused on the record systems of the IRS, the National Center for Health Statistics, and the Social Security Administration. LIASS is being proposed as a parallel effort to the household component of the Survey of Income and Program Participation [18]. Hirschberg and Renshaw [14] also describe greater longitudinal associations of individual and business establishment records, largely relying on SSA and IRS data. Were employer files retained, other associations could be made in which the dynamics of small area growth or decline could be traced over time to determine how business births, deaths, expansions, and contractions have affected overall activities in particular areas and how (various demographic subgroups of) workers respond to such changes in business activity.

National Income and Product Accounts.--The

national system of income and product accounts and related tables represents a massive and systematic application of synthetic estimation techniques and other statistical applications, intensively utilizing existing administrative records in combination with existing censuses and surveys. This system of accounts and tables represents, thus far, the single most comprehensive statistical use of business administrative records and the premier instrument for instituting coherence, with respect to economic statistics, in an otherwise fragmented statistical system. Thus, the recent Gross National Product Data Improvement Project Report is an example of the application of appropriate standards for expanding and improving already existing data series (in this case, principally, but not entirely, surveys and economic censuses) and for providing for well-defined data deficiencies that cannot otherwise be met through more intensive utilization of existing administrative records.

WHERE DO WE GO FROM HERE?

Administrative record systems already exist at all levels of Government, and in combination they represent established universal record systems. Their more effective utilization in statistical programs depends only in part on the research needs outlined in the preceding section in order to meet the growing demands for small area data at more frequent intervals. While a total redesign of the U.S. statistical system along the lines implied in the first section of this paper is not likely in the immediate future, legislative and partial redesign initiatives are needed to meet these rising demands for information cost-effectively. We already know much, if only in relatively broad outline, with respect to the needed improvements. Our specific recommendations are shown below.

Uniform coding. The Standard Statistical Establishment List (SSEL), developed by the Census Bureau, was intended to alleviate the problem of duplicate construction and maintenance of business lists. Pending legislation will provide Federal agencies access to the SSEL for statistical purposes. This legislation will remove the main obstacle to uniform industrial and geographic coding among statistical series generated by different statistical and administrative programs.

Sharing. By itself, the SSEL legislation will not reduce the paperwork burden or needless proliferation of record forms. Nor will it ensure uniform coding among the different statistical series. A number of information management legislative initiatives are pending. The Bonnen Committee recommends the sharing of data for statistical purposes. Cooperation among the agencies involved in collecting and using information will enhance the efforts to make more effective use of administrative records. In the absence of sharing records among agencies, incomparabilities among data series cannot readily be resolved.

Mandatory establishment reporting. At present, business must maintain records for different reporting units. Mandatory establishment reporting could serve to reduce the confusing proliferation of reporting units and the administrative records programs with each other and with information from the censuses (thereby improving the quality of information from the several collecting programs). The problem is that, while seeking to reduce reporting form duplication, we must avoid imposing new reporting burdens that are entirely out of consonance with the recordkeeping practices of business. The frequency of required reports is highly relevant to the question of mandating establishment reports. In general, the information obtained from the major business censuses at the establishment level is concerned with annual records, and most firms keep records of such information at that level insofar as it is required for their individual plant-level industrial location, expansion, and contraction decisions. Quarterly tax and monthly payroll and employment information, while necessary for business as well as Government economic monitoring, do not require that all the information requested for the business censuses be maintained more frequently than annually, however.

Municipal coding. Establishment records would be useful for urban analysis and planning, but whether or not establishment reports are mandated, there is a great need for timely statistics at the city level. At present, SSA and UI reporting units are identified only by county. This limits their usefulness to most cities which have very little information on their current industrial structures and payroll and employment conditions. Cities need this information for revenue forecasting and for planning educational and service functions as well as for economic development programs. Municipal coding should be added to present county coding practices. (Although the UI program does not currently call for municipal coding, some States have already begun to require added municipal coding to the reporting units in their refiling procedures.)

Trade-offs. Many anomalies exist in the collection and dissemination of business information by the Federal Government. The duplication of business reports results in wasted resources and confusion among data suppliers and users of information. Consideration should be given to greater centralization of information collection and tabulation in line with sharing the data among the affected agencies. From the point of view of the business community, it would be advantageous to have a single contact point in the Federal Government where a firm could elicit

all the answers it needs with respect to the information forms it is required to file as well as to participate in a program for reducing the number of forms and duplicated questions it must currently prepare. This not only presupposes sharing the information among affected Federal agencies, but the simplest, most straightforward approach to interagency information-sharing would be through the establishment of a central reporting and processing agency. Objections to such centralization have been made on the grounds that such a central agency would not have the most pressing needs for particular information and, this together with the routing of all information forms through a central processor, might needlessly delay the compilation of particular statistics. Other objections relate to the dangers in connection with centralized data banks containing more sensitive information about individuals and establishments than is actually required in order to meet any specific administrative or statistical need. Thus, technical problems of information management and possible confidentiality and privacy problems must be viewed against the issues of wasteful duplication of information collection and processing.

Need for resources. Whether the statistical system becomes more centralized or not, adequate resources, both in terms of trained personnel and in terms of budget, need to be assigned to research more effective statistical uses of administrative records, especially on improved statistical estimates for small areas. We have briefly discussed the potential for regression methodologies and for matching of records. Although both types of methodologies have been applied in selected instances, much potential for wider statistical uses of administrative records exists, but each application needs to be planned considering the specific circumstances involved. The matching applications can be used either to add information to the initial record sets or to validate the information in one of the record sets; the regression and other synthetic methodologies can be used to obtain estimates, not previously available, for small areas.

The large number of Federal programs allocating funds to State and local areas based on various formulae especially increases the need to improve the quality of local area data. The resources required to investigate the possibility of improving small area statistical data represent only a small fraction of these funds. And the potential cost savings in producing timely, accurate small area statistics by making more effective use of administrative records easily warrants further investigations and the required resources.

ACKNOWLEDGEMENTS

The authors would like to take this occasion to thank the other members of the Subcommittee on the Statistical Uses of Administrative Data for the insights they have shown in the almost two years the group has worked together. The present membership consists of--

Lois Alexander
Social Security Administration

Paul A. Armknecht
Bureau of Labor Statistics

Leroy Bailey
Bureau of the Census

David V. Bateman
Bureau of the Census

Robert E. Berney
Small Business Administration

Lawrence A. Blum
Bureau of the Census

J. Robert Brown
Bureau of Labor Statistics

Warren L. Buckler
Social Security Administration

David W. Cartwright
Bureau of Economic Analysis

John DiPaolo
Internal Revenue Service

Daniel H. Garnick
Bureau of Economic Analysis

Maria E. Gonzalez
Office of Federal Statistical
Policy and Standards

John A. Gorman
Bureau of Economic Analysis

David A. Hirschberg
Small Business Administration

Beth A. Kilss
Social Security Administration

Morris M. Kleiner
University of Kansas

Joseph J. Knott
Bureau of the Census

Lillian H. Madow
Bureau of Labor Statistics

Nash J. Monsour
Bureau of the Census

Allan Olson
Economic Development Administration

Elizabeth H. Queen
Bureau of Economic Analysis

Vernon Renshaw
Bureau of Economic Analysis

Frederick J. Scheuren
Social Security Administration

Max Shor
Bureau of the Census

Daniel F. Skelly
Internal Revenue Service

Hyman Steinberg
U.S. Postal Service

Martin Ziegler
Bureau of Labor Statistics

It should be noted that the recommendations made in this paper are the authors' and not those of the Subcommittee. Also, while the papers at this session grew out of the work of the Subcommittee, they represent only a portion of the material that will be included in the final report.

REFERENCES

- [1] A Framework for Planning U.S. Federal Statistics for the 1980's, Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, Washington, D.C., 1978.
- [2] Alexander, Lois, Statistical Progeny of Administrative Records: Some Legal Issues, American Statistical Association 1979 Proceedings, Section on Survey Research Methods.
- [3] Bateman, David and Charles Cowan, Plans for 1980 Census Coverage Evaluation, American Statistical Association 1979 Proceedings, Section on Survey Research Methods.
- [4] Bounpane, Peter and Clifton Jordan, Plans for Coverage Improvement in the 1980 Census, American Statistical Association 1978 Proceedings, Social Statistics Section, 1979, pp. 12-20.
- [5] Cartwright, David and Paul Armknecht, Statistical Uses of Administrative Records, American Statistical Association 1979 Proceedings, Section on Survey Research Methods.
- [6] Counting the Labor Force, National Commission on Employment and Unemployment, January 1979 (preliminary report).
- [7] Dunn, Edgar S., Jr., Review of Proposal for a National Data Center, Statistical Evaluation Report No. 6, December 1965.
- [8] Ericksen, Eugene P., A Regression Method

- for Reestimating Population Changes of Local Areas, Journal of the American Statistical Association, 1974, pp. 867-875.
- [9] Federal Statistics: Report of the President's Commission, Washington, D.C., 1971.
- [10] Federal Statistics System Project: Issues and Options, December 1978 (preliminary report).
- [11] Garnick, Daniel H., The Regional Statistics System, Conference on Modeling the Multi-Region Economic System, Perspectives for the Eighties, June 14-15, 1979 (to be published in the Conference Proceedings).
- [12] Gonzalez, Maria Elena and Christine Hoza, Small Area Estimation with Application to Unemployment and Housing Estimates, Journal of the American Statistical Association, 1978, pp. 7-15.
- [13] Gross National Product Data Improvement Project Report, Report of the Advisory Committee on Gross National Product Data Improvement, Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, Washington, D.C., 1977.
- [14] Hirschberg, David A. and Vernon Renshaw, Access to Administrative Records on Establishments and Individuals for Public Policy Analysis, American Statistical Association 1979 Proceedings, Section on Survey Research Methods.
- [15] Knott, Joseph, Major Administrative Record Files, American Statistical Association 1979 Proceedings, Section on Survey Research Methods.
- [16] Report of the Committee on the Preservation and Use of Economic Data, Social Science Research Council, Washington, D.C., 1965.