

## DISCUSSION

Richard Ruggles, Yale University

This session on the statistical uses of administrative records is of central importance to the development of the statistical system, for a variety of reasons. First, the investigation of what administrative records are available, makes explicit the enormous potentiality of this important source for statistical purposes. Second, these investigations make increasingly evident the further need for coordination between administrative data and statistical data. Third, all of the papers presented at this session underscore the fragmentation, decentralization, and lack of coordination of the present state of the Federal statistical system. Finally, it is apparent that this lack of a formal statistical system is a serious impediment to further development, and that if there is to be satisfactory development of Federal statistics over the next several decades more attention needs to be given to fitting the various pieces together.

The paper by Joseph Knott on the documentation and potential uses of major administrative record files is an extremely useful report on the activity of the Subcommittee on Statistical Uses of Administrative Data. This compilation of the administrative files and their content and coverage, and the survey of their uses represents a major step forward. Knott quite correctly points out that the increased capability of the computer has made administrative files very much more valuable, since the basic data are in a form that can now be processed economically. He also emphasizes that it has been the greater need for geographic detail brought about by such governmental programs as revenue sharing that has confirmed the importance of certain administrative files such as those of the Internal Revenue Service and the Social Security System.

I found it somewhat surprising, however, that the paper contained no mention of the large number of business-oriented files such as those collected by the Securities and Exchange Commission, the Federal Trade Commission, the Federal Reserve Board, and the regulatory commissions and financial institutions. I also missed discussion of the relation of administrative files to one another and to existing statistical files. Each of the administrative files was considered in isolation, but in fact they are often closely related to major statistical files. For example, the BLS producer price file could be related by BLS to the administrative data collected by the Council on Wage and Price Stability. Employment, hours and earnings data could be related to the Social Security files. It is this type of inter-relationship among the various kinds of information which over the longer run will prove most fruitful in the enhancement of the statistical system. Finally, it would have been most useful to view the question of the use of administrative files not only from the point of view of the information they contain but in terms of the information needs of society. This was specifically done in the case of revenue sharing programs. But it would also be useful to ask what already

existing administrative data could be brought to bear upon such questions as the inflationary process, productivity change, and capital formation.

The paper by David Cartwright and Paul Armknecht on the statistical uses of administrative records reminds us that the use of administrative records is not a new phenomenon. The Bureau of Economic Analysis and its predecessors have always used IRS, Social Security, and State employment data for the national income accounts. The authors also note that almost all administrative agencies publish tabulations of their administrative data and that these tabulations have traditionally found wide use in the statistical system. They recognize, of course, that there are serious problems in using such information. The lack of consistent definitions between sources, uncoordinated formats, and different coverage all pose serious difficulties. The authors emphasize that the microdata files of the Statistics of Income and the Social Security Work History files are valuable sources that are increasingly being used. Their case study of the unemployment insurance system is very encouraging with respect to the potential of such information, but it is extremely discouraging in terms of the problems of coordination. Here I would have hoped for a somewhat more positive approach. Traditionally, the design and content of administrative records has been considered to be a matter to be completely determined by the administrative agency itself. The fact that the administrative records may be poorly designed even for the administrative agency's own use, capricious, and failing to take other uses into account has generally been considered lamentable but unavoidable. Certainly the studies of reporting burden have indicated that it is the administrative and regulatory records that are responsible for the major reporting burden on respondents, and that a broader point of view needs to be taken if reporting burden is to be reduced. It is reasonable to suggest that, just as administrative agencies are subjected to budgetary control over their administrative expenses, so also should they have to justify the expenditure of real resources in terms of reporting burden on respondents, in the context of the sum total of information being gathered by the Federal government.

The paper by David Hirschberg and Vernon Renshaw on access to administrative records on establishments and individuals for public policy analysis provides an excellent overview of the problem. They emphasize the importance of longitudinal microdata records, an emphasis which is very much needed if we are to understand the processes of change. Aggregate data reported by the statistical system measures the average change taking place over time, but this average change is a composite of the behavioral change of individual units and the shift in the composition or structure of the individual units. Much more can be done within the present administrative and statistical system to create longitudinal records. Thus for example it would be possible to create

longitudinal records for the individual establishments included in the Annual Survey of Manufactures and the Census of Manufactures. These records could, furthermore, be grouped by company and related to the IRS tax data for such companies. It would have been extremely useful for the authors to devote more attention to the relation of the proposed Survey of Income and Program Participation to their concerns, discussing the explicit coordination of statistical surveys with existing administrative files.

The paper by Lois Alexander on legal issues with respect to administrative records explores the nature of the confidentiality problem. This problem has of course been central to the utilization of administrative records by anyone outside the administrative agency collecting the data. In all fairness it should be pointed out that problems of breach of confidentiality have never been laid at the door of the statistical agencies. Rather, the real confidentiality problem is one of other regulatory or administrative agencies getting access to information on individuals or businesses which is then used to the disadvantage of such individuals or businesses. Unfortunately in many of these areas the Congress has facilitated the use of such administrative records. Thus IRS records are used by security agencies, or distributed to State and local government agencies. Employment records are used by welfare agencies. What it comes down to is that these are matters of national policy, and it is for Congress to decide what constitutes proper or improper disclosure of administrative records. There is of course general agreement that records collected for statistical use alone should never be made available to regulatory, administrative or security agencies.

The Bonnen committee in its proposed confidentiality legislation formally recognized this dichotomy and set up a mechanism for the uniform protection of statistical files.

One of the interesting questions not covered in the Alexander paper is whether certain information should always be considered to be in the public domain irrespective of whether it is contained in administrative or statistical files. Certain administrative files are in the public domain. Thus the Securities and Exchange Commission Quarterly Financial Reports are by law made available to the public. More controversial, however, is the question of whether the Standard Statistical Establishment List (SSEL), giving the name, address, company affiliation, industry and a general employment size indicator, should or should not be information in the public domain. A number of States publish such directories. However, some members of the business community consider this information to be highly confidential. I personally would feel that for enterprises such listings should be in the public domain. On the other hand, I do not feel that information on individuals, even their names and addresses, should be considered to be public information.

One technique of confidentiality protection which was not mentioned is the running of computer programs on confidential files held by agencies to produce output or analyses which themselves would contain no disclosure. Although this

device has been used to a limited extent, its potentiality is becoming very much greater as the power of the computer increases. It is extremely useful where microdata processing is required. No sanitizing of the original tapes would need to be done, since only the results of the analysis would be disclosed.

The final paper by Daniel Garnick and Maria Gonzalez on where do we go from here does, I believe, catch the essence of the major problems. This paper recognizes the widespread discontent with the fragmentation of information, which results in duplication of data and inconsistencies between similar bodies of data collected by different agencies. Serious questions are raised about the quality of data which results from the present situation. On the other hand the paper also recognizes the importance of administrative record systems for providing cost effective intercensal estimates of important socio-economic variables needed at regional and local levels. It notes that in relating data sources to one another regression and matching techniques are quite promising. In this connection much more work is needed in terms of testing and developing robust methods. I wish more emphasis had been given to the linking and matching work (both exact and statistical) that has already been pioneered by Fritz Scheuren and Daniel Radner and which is being used by BEA in its income distribution work. Although the authors emphasize the need for greater consistency in classification and coding, this point also needs greater amplification. We all speak about the Federal statistical system, but like the weather no one really does much about it. Thus, for example, for many years the BLS producer price index has used its own 8-digit WPI industrial code, which is not the same as the Standard Industrial Code (SIC), but the price data are nevertheless used by Census and BEA to deflate value data classified according to the SIC.

What is needed is a more comprehensive framework that can be used to coordinate all kinds of economic and social data. The national income accounts currently serve this function for a large portion of economic data; but the accounts were designed 32 years ago, before the computer came into important use, and they are not even comprehensive in terms of economic data since they do not cover such things as flow of funds and national balance sheets. Furthermore, they do not fit well with microdata, and this is of considerable consequence when it comes to fitting social and demographic information such as that contained in household surveys with the national aggregates of economic data. The Framework for Planning U.S. Federal Statistics for the 1980's, published by the Office of Federal Statistical Policy and Standards, represented a valiant attempt to focus on this problem. It is a valuable and useful effort. My main objection to it is that since it represents the contributions of a large number of interagency committees and ad hoc task forces it does as an inevitable consequence resemble the proverbial camel. What is needed at this juncture is more attention devoted to the analytic formulation of an overall framework that will be capable of providing consistent and operational guidelines for both administrative and statistical data.