

EMPLOYER REPORTING UNIT MATCH STUDY (ERUMS): A PROGRESS REPORT

Warren L. Buckler, Social Security Administration

BACKGROUND

Late in 1977, the Federal Committee on Statistical Methodology (FCSM), which was then in the Office of Federal Statistical Policy and Standards (OFSPS), Department of Commerce and is now a standing committee in the Office of Management and Budget (OMB), formed a subcommittee to inquire into the statistical uses of administrative records along the general lines of: 1) evaluation of the quality of administrative data and their suitability for statistical applications, and 2) assessment of the problems of access to administrative records for statistical purposes and of needs for improved coordination between statistical and administrative components.

The Subcommittee on Statistical Uses of Administrative Records (SUAR) was comprised of representatives from various federal statistical agencies and statistical components of program agencies. The subcommittee drew upon the expertise and experience of its members to concentrate its investigation on administrative programs that collect important social and economic information from individuals and businesses. The subcommittee, in its final report in 1980, listed 11 recommendations for dealing with the issues it studied. These fell into three main areas: 1) Identifying and formulating solutions for common problems related to statistical standards for administrative information programs. 2) Identifying and meeting various problems related to access to administrative record systems. 3) Identifying collection programs and research activities requiring government-wide coordination and support.

Though the Subcommittee on SUAR issued its final report, work in this area was by no means finished. Early in 1981, another subcommittee was formed with the charge of attempting to implement some of the recommendations made by the original group. This new group, now known as the Administrative Records Subcommittee, established several work groups to deal with specific recommendations which dealt with the manner in which employers file administrative reports for their establishments: Recommendation #1 - Common identifiers should be used whenever possible in collecting information pertaining to the same individuals or organizations, and Recommendation #3 - Consistent procedures should be used in administrative and statistical data collection efforts for defining reporting units, identifying and coding reporting unit characteristics and developing standards for data tabulation.

The principle reason for selecting these recommendations was that there was a direct tie-in to the underlying desirability of increasing the use of administrative records for statistical purposes. The employer reporting systems were seen as areas where relatively low cost modifications to existing systems could be made that would yield overall benefits to both administrative and statistical programs by reducing respondent burdens, data collection costs and data processing costs. The statistical applications to a coordinated employer reporting unit

system has the potential for providing for the creation of powerful data bases that can be used to measure economic activity and demographic changes for subnational areas.

To address the issues before the work group, three major administrative record systems were selected for study: 1) The Unemployment Insurance (UI) records collected by each State under rules and procedures established and coordinated by the Department of Labor, 2) The W-2/W-3 records submitted to and processed by the Social Security Administration (SSA) for both SSA and Internal Revenue Service (IRS) administrative purposes, and 3) The Census Bureau's Standard Statistical Establishment List (SSEL) records.

IRS tax return records for business were not selected because this information is collected on a company basis rather than an establishment and thus did not provide the breakdown of information needed for this study.

The work group identified three major tasks: 1) document the structural differences among the three systems, 2) study the factors contributing to statistical inconsistencies among the three systems, and 3) study the possibilities and problems involved in implementing the recommendations.

Task 1 was completed with little difficulty. The work group planned a micro study to shed more light on the extent and nature of the establishment reporting problem as part of task 2. Though much developmental work was done the micro study was not conducted, primarily due to confidentiality restrictions on access to data, limited resources and priority conflicts among the participating agencies. An implementation study to investigate the feasibility of converting SSA and BLS systems to the establishment units contained in Census's SSEL was planned for task 3. Here again, the problems that confronted the group were such that no formal proposals were developed.

The work of the Establishment Reporting Work Group is well documented in its final report which was completed in December, 1982. The fact that it stopped short of its original goals is evidence of the difficulties facing the statistical community in obtaining the needed improvements in administrative records. However, significant progress was made in identifying the issues that require attention and possibilities for further exploration. The report concluded with a recommendation for continued study of establishment reporting and completion of some of the unfinished business at hand. So, early in 1983 a new Work Group was formed within the Administrative Records Subcommittee to continue examining some of the issues identified by the previous group in a somewhat different direction and with a new focus and, at the same time, partially address a topic identified by the FCSM for study--the need for interagency sharing of statistical data/files.

The purpose of this paper is to report on what this group has done and identify the activities still remaining.

THE EMPLOYER REPORTING UNIT MATCH STUDY (ERUMS) PROJECT

The new Work Group was originally comprised of representatives from Internal Revenue Service (IRS), Bureau of Labor Statistics (BLS), Social Security Administration (SSA), Bureau of Economic Analysis (BEA) and Office of Management and Budget (OMB). Later a representative from the Committee on National Statistics and an observer from the Census Bureau joined the Work Group.

This group knew that an important factor contributing to its predecessor's inability to fully accomplish what it set out to was that many of the activities planned by the group were beyond the control of its members. In light of the difficulties encountered by the previous Work Group, the new group agreed that the objectives and tasks to be undertaken should be items that the members felt were achievable and could be controlled by the group.

Purpose

The group determined that its primary objective would be to conduct a pilot study designed to match information, at a micro-data level, from employer wage reporting and establishment reporting systems of BLS, SSA and IRS. (Thus the name Employer Reporting Unit Match Study (ERUMS) Work Group emerged.) This micro study would differ from that proposed by the previous work group in that it would focus on the reporting unit relationships between the BLS and SSA systems, supplemented with information from IRS at the employer level. This type of study would allow the group to carefully examine and gain insight into the differences and similarities of the three systems so that recommendations could be made regarding (1) the development of a system that uses common identifiers for collecting information pertaining to the same organization and (2) developing consistent procedures to be used in administrative and statistical data collection efforts for defining reporting units and identifying and coding reporting unit characteristics. In addition, conducting the micro-data match study would provide valuable experience in learning how to accomplish interagency exchanges of data and files within the framework of current regulatory constraints. Also, the group hoped to learn from the pilot study how a cooperative interagency data exchange can be used to identify and correct errors, deficiencies and shortcomings in the systems of the participating agencies.

Plans

In the first several meetings of the ERUMS Work Group, the members concentrated on outlining plans for the study. These early deliberations resulted in the following:

1. Scope of Study - Considering the resources available to the work group, it was decided that a sample of records should be selected from one State for the pilot study. This would make it possible to do a thorough review and analysis of matched and unmatched cases.

2. Data Access - It was clearly recognized from the beginning that, because of current restrictions on the release of identifiable information, careful consideration must be given to the steps to be taken in order for the group to gain access to and use the required microdata records. Instead of the approach outlined by the

previous group, this group felt that it would be necessary to conduct the study under inter-agency agreements between the participating agencies. Extensive discussion centered around what the terms of such agreements would be. It was very important that these agreements contain well-defined statements as to the purpose of the pilot study and assurances of the protection of confidentiality of identifiable information.

3. Data Sources - Each participating agency identified the files from their system that could be made available for the match study, given a satisfactory outcome of the access issue. The SSA would provide records from its master file of employers which contains information used to code geography and industry for workers in statistical files and from records of employer wage reports furnished on Form W-3 (Transmittal of Income and Tax Statements). The IRS would provide records of information from Forms 940 (Employer's Annual Federal Unemployment Tax Return), Form 941 (Employer's Quarterly Federal Tax Return) and Form W-2 (Wage and Tax Statement). BLS would furnish employer information from reports that States are required to file under the Unemployment Insurance program and summarized in their ES-202 report.

4. Data Processing - The BEA personnel offered their services in performing the computer processing required for the micro-data match. An appropriate sample of records from the BLS, SSA and IRS systems for the one State was to be selected and matched based on a specific set of variables. The group recognized that a substantial amount of manual data processing would be required after the electronic match was done. An examination and analysis of the matched and unmatched records would be a key part of the pilot study and should provide the work group with much of the information needed to meet its objectives.

The group then set out to define the specific tasks that needed to be done. These were:

- Develop a project description.
- Select a State and obtain their permission to use their records for the match study.
- Document the data files to be used in the match.
- Develop sampling criteria.
- Develop matching criteria.
- Draft interagency agreements to cover data exchanges and the work to be done.
- Develop a timetable for accomplishing specific objectives.

As the group got into these items, it became apparent that a number of obstacles were popping up that were inhibiting the smooth and orderly progress toward attaining the primary objective. The major problem area centered around confidentiality issues; such as, assurances that can be given to the State selected for the pilot study about protection of information from their records, decisions on who will have access to the identified micro data, and requirements for the protection of tax return information. There were also problems with the multi-agency type inter-agency agreement that had been drafted to cover the pilot study work. It was evident that modifications had to be made to the original plans.

The decision was made that BLS would perform the computerized match operations. It was felt that confidentiality problems could be substantially reduced if access to confidential records was limited to those with a "need to know", in this case, BLS and SSA personnel. The issue of expanding this "special" group would be addressed as the study proceeds and if it is determined to be desirable to do so. Proposals for the content and format of interagency agreements have been revised several times. The group finally decided that an agreement between IRS and BLS covering the use of tax information, the work to be performed and the products to be obtained would be the best way to proceed. An agreement covering the conditions of use of SSA data by BLS would be handled through a separate document. These agreements have not yet been finalized.

Description of Match

A draft was prepared which described the ERUMS project in terms of some of the specifics of the matching operations as well as a statement of the purpose. The current version of that project description is shown in Exhibit A. 1/ Currently, the Group is considering modifications to the section of the project description that defines the statistical products that will be produced.

The Group then accepted BLS' recommendation that records for the State of Texas be used for the micro-data match. The files to be used in the match were documented. Part of that documentation is shown in Exhibit B. 1/

Sample Design and Selection Procedure

To assist the Group in designing the sample of records to be used in the match, BLS obtained a set of universe counts from their UI Address File for the State of Texas. These counts provided the number of single and multi-unit records in the file in terms of both the UI Account Number and the Employer Identification Number (EIN). Exhibit C 1/ summarizes these counts.

Based on these counts, a sampling rate of 6 in 100 was selected for the initial stage sample. This is expected to yield about 200 multi-unit EINs (2,000 records) and a little over 16,000 single unit EINs. The selection will be based on the occurrences of six pairs of randomly selected digits in the EIN.

The EINs selected in the first stage will also be selected from SSA files. In addition, records with EINs not in the Texas UI file but having the same pattern of selection digits and at least one Texas establishment in SSA files will be selected. The records selected from the UI and SSA files will be matched and classified in broad terms as follows:

1. In both UI and SSA files, exact match
2. In both UI and SSA, partial match
3. In UI file only
4. In SSA file only

Counts are to be obtained based on the initial match status and the single or multi unit designations. From these counts the subsampling rate will be determined in order to yield a final sample of about 400 EINs.

Information from IRS records will be provided for the final sample of 400 EINs.

Accomplishments to Date

What has been presented so far, not necessarily in any strict chronological order, is a condensed version of some of the activities and events that have been completed in the ERUMS project to date. These can be summarized and restated in the following list of accomplishments:

1. Development of a formal statement regarding the purpose of the Work Group.
2. Documentation of the data files to be used in the match.
3. Preparation of a project description document.
4. Development of a set of criteria for performing the match operations.
5. Drafting of interagency agreements covering the match study work.
6. Selection of a State (Texas) for the match study.
7. Development of a sample design for the match study.
8. Development of specifications for obtaining universe counts of records in BLS files for the selected state.
9. Preparation of a set of counts related to item 8 above.
10. Development of specifications for the first stage sample selection rates.
11. Development of a preliminary timetable for remaining tasks.

Remaining Tasks

Much work remains to be done. In fact, the heart of the ERUMS project is still before the Group. The key tasks remaining include:

1. Preparation of an interagency agreement and conditions of use agreement that is appropriate and acceptable to all concerned parties, and obtaining agency approvals for these agreements.
2. Selection of the initial sample of records.
3. Performance of the computer matching operations.
4. Preparation of counts by match category and subsampling to obtain the final sample for the manual match.
5. Performance of the manual classification and matching operations.
6. Preparation of required statistical tabulations and other products.
7. Analysis of the results of the computer and manual matches.
8. Preparation of a written report on the project.

Exhibit D 1/ shows more detail of the tasks that remain, along with an estimate of when each will be completed. (Obviously these dates are subject to revision as work proceeds.) The work group must now get on with the job to be done!

1/ Copies of Exhibits can be obtained from the author.

COMMENTARY

I have avoided labeling this section as the "Conclusion" as would be customary because I don't believe there really is a conclusion in the normal sense of the word. Rather, I'd like to make some personal observations on the work being done and where it might lead.

Even at the successful completion of the ERUMS project (and I am convinced this will happen), there will not be conclusive evidence that will point emphatically in a particular direction on how to proceed with implementation of a system that will provide more consistent procedures for identifying and collecting information about the characteristics of employers and their establishments/reporting units. We will have gained more knowledge and experience in an area where we might have ideas about what things should look like but are not sure what actually exists. From this pilot we will probably determine the need for a full scale study to further support and extend our findings.

A lot of work has already gone into this project and much more remains to be done. The individuals who are involved in this current phase of investigation are committed to contributing to efforts to improve the effective and efficient use of administrative records for statistical purposes. Yet there is a paradox-- While on the one hand there is a dedication and spirit of cooperation that exists among members of the Work Group, on the other there are laws, rules and regulations, as well as possible conflicts with an agency's priorities, that hinder real progress in this area.

There is a need for the Federal statistical community to focus attention on ways to developing a better mechanism for drawing the issues to the attention of those who are in a position to and willing to do something beyond the study and

recommendation phase. I believe that significant progress in this area will depend on the combination of more collaboration between the custodians of administration and statistical record systems and some type of legislative actions. Perhaps legislation to mandate true government-wide establishment reporting is the only certain way to insure implementation of the recommendations. I raise this as a possibility; of course, I don't know for sure it will work.

REFERENCES

Statistical Policy Working Paper 6: Report on Statistical Uses of Administrative Records, U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, 1980.

Establishment Reporting in Major Administrative Record Systems, Establishment Reporting Work Group, SUAR Implementation Subcommittee, Federal Committee on Statistical Methodology, December, 1982.

An Update on Establishment Reporting Issues: Practical Considerations, David W. Cartwright, Bruce Levine and Warren L. Buckler -- paper presented at the annual meetings of the American Statistical Association, August, 1983.

Reports and notes from meetings of the Employer Reporting Unit Match Study Work Group, Administrative Records Subcommittee, Federal Committee on Statistical Methodology, February 1983-April 1985.

Statistical Uses of Administrative Records: Recent Research and Present Prospects, Department of the Treasury, Internal Revenue Service, Statistics of Income Division, July, 1984.