

ADMINISTRATIVE STATISTICS: A BLS PERSPECTIVE

Janet L. Norwood, Bureau of Labor Statistics

The issue before the panel today is a very broad one. Tom Jabine and Fritz Scheuren have done a good deal of work in this area, and we are all very much indebted to them for it.

In commenting on these issues today, I would like to make a few general observations based on our experience with administrative data at the Bureau of Labor Statistics (BLS). That experience began very early. Almost a century ago, BLS did a study of marriage and divorce in which all of the data came from the records from the thousands of courts throughout the country.

Let me start with a few general points and then describe some special uses of administrative data at the BLS.

1. First, we must remember that we need to look at a very broad array of data when we talk about administrative records, not just Internal Revenue Service (IRS) and Social Security data.

Without administrative records, we would not have much of a program at the BLS. We rely on administrative data for much that we do -- especially as the basis of almost all of our sampling frames. It would be hard to develop some of our economic data without the access we have to administrative records. I will return to this point in a moment.

2. We have to remember, however, that administrative records pose considerable difficulties for statistical agencies trying to use them, difficulties which some of us tend to overlook. Administrative records are seldom designed for statistical purposes. They usually are available only with a considerable time lag, and the statistical properties of the data as well as the reliability of their processing frequently leaves a great deal to be desired.
3. Administrative records have special value in certain circumstances where survey data are just not practical -- for example, for local area data, for use in benchmarking surveys (when the administrative data set represents the universe), and for research purposes.
4. And now let us look at the matter of the standard list. The complexity of the problems in establishing a single list for use in sample development by all or even by a few statistical agencies is far greater than many people realize.

There are, of course, legal and policy issues -- most of which, in my view, are far easier to deal with than the bureaucratic ones. I am pleased to hear that the Census Bureau appears to have changed its position. But I hope you will forgive me if I say that I have heard this before. I am waiting for action.

But there is something more here that we should not forget. We must remember that the quality of statistical series may be affected adversely if we insist on the use of a single list -- without taking into

account the strengths and weaknesses of that list, and without taking into account the purpose and the concept which the survey is designed to measure. Probability sampling needs to be related to the concept that is being measured. Most BLS programs, but not all, should have samples developed with probabilities proportional to employment. Probabilities proportional to product sales makes more sense for our price programs, for example, than probabilities proportional to employment.

Indeed, there is no perfect single list. The Census list -- because of the Census Survey of corporate organizations -- probably has better breakouts of multi-establishment firms. The BLS list, now that unemployment compensation has been fully extended to all employers, we believe, has better coverage of small establishments, a matter of no small importance since more than 1/2 of all workers work for relatively small establishments.

What we need is access to all the lists and research to evaluate the strengths and weaknesses of them. We at the BLS are getting ready for possible Census action to make their list available. We now have commitments from all States to include Employer Identification numbers in addition to the Unemployment Insurance (UI) number on the 1984 UI address file. This will permit research to match and compare the lists more easily. Indeed, we have currently in process a study comparing the UI list with the Social Security list in one State (Texas).

But let there be no misunderstanding of our position. The Bureau of Labor Statistics will oppose any legislation which forces use of the Census list. We believe that the Standard Statistical Establishment List (SSEL) should be made available; we believe that research should be undertaken to evaluate it. And only then can we all decide which list, or combination of lists, would be best for our programs and for the statistical system.

5. And finally, I don't think that anyone has yet understood the tremendous problems that occur in Standard Industrial Classification (SIC) coding in our decentralized statistical system. The lack of consistency between agencies in the way establishments are coded; the erroneous view that an SIC code once assigned can be dependable for a number of years; the mistaken view that our classification system itself is rational and carefully thought through -- these are only some of the issues involved. And yet, we have not really begun to think very much about them.

I can tell you that at BLS, we have begun to do so. In a number of our programs, we have introduced more detailed probability

sampling -- not just to the level of the establishment but also down to the selection of the individual product to be priced or the particular occupation for which wages will be collected.

What we have found out in this process is that the SIC codes on the establishment lists are often wrong. Establishments selected for particular industry price indexes, for example, frequently do not produce any products in the particular SIC at all. As a result, at BLS we have instituted a whole new refiling system to keep our coding up to date.

And now, let me tell you a bit about some specific work we are doing at the BLS to use administrative data for program purposes.

First I will mention the UI system and the Quarterly Report on Employment and Wages, the ES 202. Technical responsibility for the ES 202 was transferred to the BLS some years ago because of the importance of the 202 data for the BLS monthly business survey and because the data from the 202 were important input to the national accounts. Our mandate was to improve the quality of the data and to speed up its timing.

That improvement required the cooperation of the Research Divisions of each State Employment Security Agency, as well as the Unemployment Insurance Division. We were able to get the cooperation we needed because we found that both groups needed good data. When we were able to show the UI people how important good 202 data were to check the tax reports that employers filed with them, we were able to get the support of this group, which is responsible for administering the records. The State research people, who needed the data for the BLS Federal/State cooperative business survey, improved the coding, editing and summarization of these records. The UI tax people realized that the research staff was indeed doing them a service because the data were reviewed and edited, inconsistencies were flagged, and those responsible for program administration were able to identify problem cases more easily.

The challenge is often to find ways to make administrators, who use the data to implement programs, recognize the benefit to them of work done in statistical units. By demonstrating how our goals can coincide, we can get better data more easily.

Second, we have in the 202 data a powerful tool for research. The cooperative nature of the Federal/State programs puts constraints on sampling and estimation design. The 202 records

contain, in most cases, (albeit with a lag of 6 to 9 months) a complete census of all employer records. So, we can experiment with alternative designs and estimation strategies by making use of the 202. And, in addition, we have in hand data that can be used to benchmark our sample survey. This is especially important for the BLS business survey, which is so important in the development of public policy.

And, finally, administrative records are also used in the BLS Occupational Safety and Health Program. While the core program in this area is the BLS Annual Survey of Occupational Injuries and Illnesses, that survey cannot meet a number of important needs. The Annual Survey produces estimates of the number of cases and how frequently they occur, but it cannot provide information about the kinds of injuries which occurred and the kinds of accidents which produced them. Such information, which is important to accident prevention work, can be secured from Workers' Compensation reports of individual accidents. BLS has developed the Supplementary Data System (SDS) to obtain these data.

Distribution of characteristics of cases can be important in program planning. For example, information that nearly one-quarter of all injuries affect the back, or that half of all cases in a normal business year involved employees in their first year of service, has important planning implications. There are, however, serious problems in using worker compensation records. Since there are no national standards for workers' compensation, each State sets its own reporting requirements. In some States, all injuries must be reported; in others, reports are required only if some specified period of disability has elapsed. This reporting variability prevents making national estimates of the number of injuries.

We have discovered, however, that State files which include similar kinds of cases have similar characteristics. Although we cannot estimate the absolute number of cases occurring nationally from workers' compensation sources (we obtain those from our annual survey), we can say a good deal about the characteristics of work injuries and accidents.

The speakers on this panel have demonstrated that we have gone a long way in the use of administrative records, both as sampling frames for survey work and for the substantive data we can get from them. Our responsibility now is to move ahead in the future to assure that the data are reasonably fit for the purpose intended and that we do the best job we can to standardize and improve them.