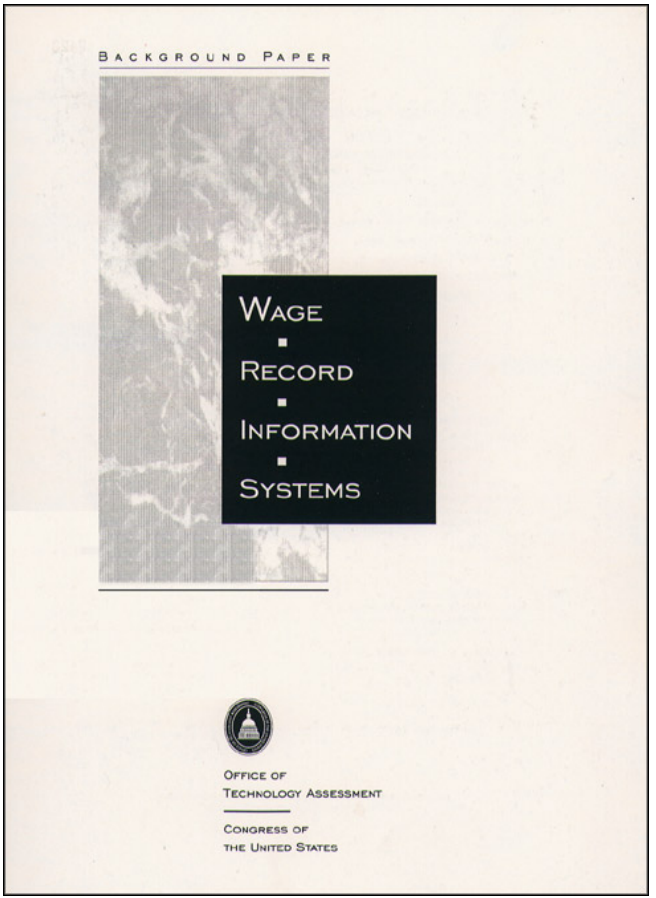
Wage Record Information Systems

May 1994

OTA-BP-EHR-127
NTIS order #PB94-170404



Recommended Citation:
U.S. Congress, Office of Technology Assessment, *Wage Record Information Systems*, OTA-BP-EHR-127 (Washington, DC, May 1994).

Project Staff

Clyde H. Behney
Assistant Director
Health, Education, and Environment

Nancy Carson
Program Director
Education and Human Resources

PRINCIPAL STAFF

John Wirt
Project Director

ADMINISTRATIVE STAFF

Cecile Parker
Office Administrator

Linda Rayford
P.C. Specialist

Tamara Kowalski
Research Secretary

R eviewers

Lorraine Amico
National Governors' Association

Virginia Chupp
U.S. Department of Labor

Paul T. Decker
Mathematical Policy Research

Paula Duggan
Northeast- Midwest Institute

Karen Greene
U.S. Department of Labor

Juliette M. Lester
National Occupational Information Coordinating Committee

Karen Levesque
MPR Associates

Brian MacDonald
U.S. Department of Labor

Jay Pfeiffer
Florida Department of Education

Alan Phelps
University of Wisconsin

Thomas Plewes
U.S. Department of Labor

David Stevens
University of Baltimore

Mary Sue Vickers
National Occupational Information Coordinating Committee

James Woods
National Occupational Information Coordinating Committee

NOTE: OTA appreciates and is grate :ful for the valuable assistance and thoughtful critiques provided by the reviewers. The reviewers do not, however, necessarily approve, disapprove, or endorse this report. OTA assumes full responsibility for the report and the accuracy of its contents,

Contents

Introduction	1
Purpose of this Paper	1
Background on Data Issues	2
Experience With Wage Record Information Systems	7
Quality of Studies Done With Wage Records	11
Possibilities for the Development of Wage Record Information Systems	15
Other References	19

WAGE RECORD INFORMATION SYSTEMS

INTRODUCTION

This background paper responds to section 408 of the 1990 amendments to the Perkins Act, which asks OTA to review activities to be undertaken by the National Occupational Information Coordinating Committee (NOICC) to encourage the use of wage records from state unemployment insurance systems for purposes of conducting policy studies or monitoring the outcomes of vocational education.

- The legislation asks NOICC to demonstrate the use of wage record information systems for these purposes, and develop "procedures" for establishing and maintaining a nationally accessible database of wage record information.

- It also asks OTA to evaluate the quality and usefulness of the demonstrations supported by NOICC and the technical problems involved in generating and using earnings record information.

After consulting with a number of federal agencies on these legislative requests, NOICC has:

. *Supported a study by the National Governors Association of the extent to which state vocational education agencies are using or experimenting with wage record information systems.*

The study shows that state vocational education agencies already have considerable experience in using wage records from the unemployment insurance system for different purposes.[1]

- *Sponsored the preparation of a comprehensive, "how to do it" guide for stales by MPR Associates on setting up* using it for purposes of examining the *outcomes* of vocational education and other training programs, conducting followup studies, or for other purposes.

MPR Associates will submit the guide to NOICC for publication in 1994.[2]

The guide will include an overview of wage record information systems, a comprehensive annotated bibliography of existing studies of wage record information systems, and a review of data quality issues.

PURPOSE OF THIS PAPER

The purposes of this background paper are:

- To briefly summarize what has been learned in vocational education and other program areas from studies and demonstrations of the use of wage records for purposes of program evaluation and policy research;

- To raise some issues about the quality of data resulting from wage record information systems that need to be considered; and

[1] Amico, Lorraine, 'Mate Capacity To Use Unemployment Insurance Wage Records: The Vocational Education Experience," Washington DC: National Governors' Association and the Office of Adult and Vocational Education, June 1993, pp. vii-viii.

[2] Levesque, Karen, and Karen Alt, "A Guide to Using Unemployment Insurance Data for Program Follow Up," prepared for the National Occupational Information Coordinating Committee, Berkeley, CA: MPR Associates and Institute for Family, Work, and Community, forthcoming.

. To identify steps that could be taken by the Congress to deal with these issues and thus encourage the sound development of this potentially important new source of information.

BACKGROUND ON DATA ISSUES

Purposes of Wage Record Information Systems. Wage record information systems can serve three main purposes of policy development and consumer information at different levels of government. These purposes are:

- ***To conduct policy-oriented research on trends in the employment and earnings of individuals in the labor market, and the employment effects of alternative strategies of income support, education and training, and social services.***

  One example of the kinds of analyses that might be done is to compare the employment outcomes of strategies of income support or other social services with strategies of education and training. Analyses might also be done to trace the labor market experience of individuals who are displaced from their jobs, and thus improve our understanding of what forms of support are most helpful in the long run. Studies like these could be carried out at the national, state, or local levels.

  The longitudinal natuure of the earnings data that can be assembled from wage record data and the commonality of the outcome information among individuals and across states are what make such studies possible. Most other federal surveys of income and employment, except for a few research data bases, are cross-sectional. This means that income data are available for only a single point in time from a sample of individuals in the whole population. This generally makes it difficult to analyze the effects of programs on earnings growth.

- ***To monitor and evaluate specific education and training or other social service programs or service providers on the basis of their effects on the economic welfare of individuals in order to make specific decisions about the improvement or funding of programs or institutions within a state, region, or at the national level.***

  This purpose involves using wage record information systems for managing governmental programs on the basis of their outcomes for individuals. These techniques of ***performance management*** may involve comparing specific providers to each other or against certain criteria or standards of performance. The comparisons could be made among all of the providers within a certain program, such as the Job Training Partnership Act (JTPA); all training programs of a certain type within a type of institution (such as all the business programs in community colleges); or particular categories of programs across several types of institutions (all of the community colleges, technical colleges, and proprietary schools within a state). With adequate means for describing participant populations, comparisons might also be made across widely different but nonetheless competing types of programs such as JTPA, the JOBS part of the federal welfare program, and the training components of many other federal programs. Eventually, it might even be possible to conduct such studies at the national level. However, in some program areas, and especially vocational education, this would require substantial capabilities for identifying program participants across the country in common terms. These capabilities do not exist.

- ***To provide the public with information about the outcomes of programs in their local areas and trends of employment in different industry, occupational and/or***

geographic areas that is useful to them in making career or consumer decisions.

These information services could provide people with data on: (a) job growth by industry and perhaps by occupation in a local area and other areas, (b) the earnings of graduates from different programs, and (c) information about program enrollments and trends.

The same kinds of information could be used by local institutional leaders for planning and evaluating their own strategies of improvement, adjustment, or expansion.

Clarity concerning these purposes is important, because the design of a wage record information system may need to be different depending upon its purposes. Some purposes may conflict with others in certain situations. Some states and federal agencies may place much higher priorities on some of the purposes than others. This means that it may not be possible to design one system meeting all purposes equally well within a state or at any other level.

The Unemployment Insurance System and Wage Records. The idea of wage record information systems is that the data for accomplishing these purposes already exist in the earnings reports that are collected from employers on a quarterly basis by ***State Employment Security Agencies SEASs).*** [3] This data is collected by these state agencies as part of their normal process of administering the nationwide system of unemployment compensation.

The employment and earnings data in these record systems have been estimated to include over 90 percent of the working population in the United States, or well over 140 million people.[4] Most state agencies accumulate this data over time, so that *longitudinal earnings histories* could be constructed. Not included are mainly self-employed individuals, some agricultural workers, people who are employed by the federal government or are serving in the military, independent contractors, and railroad workers. [5]

These earnings records, or wage record data, as they are usually called, are collected by SESAs in order to determine the tax liability of employers for unemployment compensation and verify the eligibility of applicants for unemployment compensation. **The key fact is that three of the data elements collected are common across the states.** [6] **These** three data elements are the Social Security numbers of all employees in the state who are covered by unemployment insurance, their quarterly earnings, and the standard industrial code (SIC) and/or the business name and address of the employer, [7] Some states also collect other data elements such as weeks worked.

This broad coverage and the possibility of assembling longitudinal histories of the labor market experience of individuals are the two main potential advantages of wage record information systems. The similarity of the earnings information among individuals and states means that a common set of outcomes might be developed for purposes of managing the performance of a wide range of educational, employment, income support, and social service programs, and conducting research on the labor market.

---

3 The exact organization and administration of state unemployment insurance systems differs among the states. See Levesque, op. cit.

[4] Baj, John, Charles E. Tort, and David Stevens, "A Feasibility Study of the Use of Unemployment Insurance Wage-Record Data as an Evacuation Tool for JTPA," Report on Project's Phase I Activities, Washington DC: National Commission for Employment Policy, January 1991, p. 12.

[5] bid. p. 8.

6 Private conversation with Brian MacDonald, Office of Employment and Unemployment Statistics, Bureau of Labor Statistics, Washington, DC.

[7] National Commission for Employment Policy, "Using Unemployment Insurance Wage-Record Data for JTPA Performance Management," Research Report 91-07, Washington, DC, June 1992, table 1.

Accomplishing these purposes requires knowing the Social Security numbers of all the individuals involved in the programs that are to be compared or analyzed. Once the Social Security numbers of the individuals involved are known, the data on their earnings in the files of the state's unemployment insurance system can be obtained.

The agencies who assemble databases for linking program participants with records of their earnings are *user agencies* in each of the states. These *user agencies* take lists of program participants to the SESA agency and receive in return the wage records for each person for whom there is earnings data in the files of the unemployment insurance system. The user agency then takes these *matched records* of program participation and earnings data from the SESA and assembles them into a *wage record information system.* This is done on an ongoing basis to track program participation and outcomes over time. These user agencies could be either special statewide offices created to assemble wage record information systems for several program areas and provide resulting analyses and reports to those areas, or offices dedicated to individual program areas. How these user agencies would be organized within a state would probably vary greatly for a variety of political, historical, and other reasons.

Obtaining Data From Other Sources. In most situations, achieving a sufficient level of accuracy in a wage record database may require obtaining employment and earnings data from the SESA agencies of neighboring states, other agencies of the state and in other states, and agencies at the national level, where earnings and other kinds of activity information for people who are not included in the unemployment insurance system of a state can be obtained. Program participants may be working for the federal government, or may have moved to another state and be working there, or may not be in the records of the unemployment insurance system because they have enlisted in the military or are engaged in some other kind of productive activity, such as enrollment in a community college. Without this additional information, there will be misleading gaps in the wage record information. Many people will appear to be doing nothing, when in fact they are doing something.

The danger is that without earnings and enrollment data for these others kinds of activities, analyses done to compare the outcomes of specific social service or education and training programs could be distorted. Analyses using only wage record information from one state's unemployment insurance system could, for example, show that certain programs in the state are much worse than other programs, when in fact people who participate in the second set of programs are much more likely to have moved out of state or to be enrolled in community or four-year college program than those who participate in the former group. Since these are all gainful activities, it would not be appropriate to penalize the second program without knowing about these other activities. Simply being located near the border of a state could be sufficient for the employment and earnings rates of programs to be seriously downrated in a scheme of performance management where no wage records from neighboring states have been obtained. Solving this general problem requires adding data on the employment and other activities of individuals from other agencies and states.

Data-Sharing Agreements. Obtaining this data from these other agencies and states generally requires reaching a **data-sharing agreement.** **These** data-sharing agreements prescribe the terms of data transfer. SESA agencies generally stipulate that the receiving agency may use wage record data for certain purposes and not any others. In most cases, ownership of the wage record data is strictly retained by the SESA, so that none of the information belongs to the recipient agency, and they are not allowed to share it in any way or release it in any form where the identity of an employer or an individual's earnings would be revealed. Strict regulation of the terms of data-sharing are required in order to maintain the "convenant" of confidentiality that exists

between SESA agencies and employers, which is a bedrock of the unemployment insurance system. Reaching these data sharing agreements generally involves considerable cost in time and effort.

National Wage Record Database (NWRD). Recognizing these costs, Congress asked the Bureau of Labor Statistics (BLS) to submit a plan for creating a NWRD in Section 405 of the Job Training Reform Amendments of 1992. Creation of the NWRD could provide an efficient way of pooling at least all of the wage records in the existing state unemployment insurance systems into a national database and in turn making all of this data available to all of the states. The language of the legislation states that:

> *. . the Commissioner of Labor Statistics, in cooperation with the states, shall determine the appropriate procedures for establishing a nationwide database.. in a longitudinal manner and for making such information available for policy research or program evaluation . . .*

Many alternatives are open to BLS in submitting the plan requested by Congress, which will be submitted in June 1994. Core data elements of the NWRD could reside in the states or be sent to a central data bank operated by the federal government. The database could consist of a national sample of individuals, samples within states, or the entire working population of the United States.

"One stop shopping" for wage record data is likely to be one of the main features of the database design proposed by BLS. State user agencies wanting wage record data would submit lists of social security numbers to the state office maintaining access to the central database, where they would be matched to wage records from all 50 states. [BLS may recommend that the state offices of *labor market information* should be the access point for the NWRD or they may recommend something else.]

With such a national database, the need for coordinating data sharing agreements for wage record data among states should decrease. If so, this could substantially lower the cost to state user agencies of developing wage record information systems. But the cost to BLS and SESA agencies of building and maintaining the NWRD could substantially offset these savings. A responsive data system containing the social security numbers and several years of earnings records for the complete working population of the United States (or well over 140 million people) would not be cheap.

Confidentiality Issues. The possibility of developing a national database of wage record information underscores serious issues of confidentiality and data privacy (which clearly also exist with existing arrangements for data-sharing). BLS is currently planning to propose to Congress statutory measures for dealing with these confidentiality and privacy issues. One possibility is that federal statutes might be written to define any data transferred to designated user databases as being **statistical information** rather than administrative records, and therefore not accessible with identifiers to anyone or susceptible to a supoena. [8]

Any wage record information systems developed will also need to be managed so as not to undermine public perceptions of the primary mission of state unemployment insurance systems and who the system is primarily intended to serve. Preservation of these perceptions is likely to require protecting the confidentiality of information in the systems. At the moment, most people are probably not aware that fairly complete histories of their earnings reside in the records of their state unemployment insurance systems and are updated on a regular basis with information from their employers.

A fourth general purpose of wage record information systems could be to employ them in enforcing compliance with policies of eligibility

[8] Private conversation with David Stevens, University of Baltimore, Baltimore, Maryland.

and payment in various program areas at the state and federal levels. Many state agencies are already obtaining wage record data for this purpose by going directly to SESA agencies. The areas of activity include the enforcement of income criteria for participation in a wide range of state and federal programs, past due child support payments that have been ordered by the courts, past due student loans, law enforcement, and many other possibilities. Several federal agencies, especially in the Department of Health and Human Services, are interested in wage record information systems precisely for such reasons.

The pressures from enforcement agencies to gain access to the wage record information systems of state user agencies are likely to grow as the systems grow. The pools of data available in the systems could become large and typically there would be many more data elements than are available from SESA agencies. Furthermore, many of these wage record information systems could turn out to be in the same departments or program areas where the agencies wanting the information for enforcement are located. Arguments for obtaining the data for enforcement purposes directly from the wage record information systems rather than going to the SESA agencies may be made in terms of the efficiencies to be gained from using the program-oriented systems for multiple purposes.

At the national level, it seems very unlikely that any such uses of a national database of wage records could be allowed without seriously undermining its use for purposes of performance management and research, because of the problems of confidentiality and perceptions of the purposes of the unemployment insurance system. BLS is clearly correct to carefully consider confidentiality issues as part of its planning effort.

Privacy laws are also a factor that needs to be considered in developing policy on wage record information systems. Requirements associated with the Buckley Amendments to the federal Privacy Act, the Social Security Act, and state versions of these laws often preclude being able to obtain participant records linked to the wage records. At the state level, fears regarding invasion of privacy have resulted in some specific curbs on programs for linking wage records to records of program participation. Concerns about the privacy issue could lead not only to the defeat of legislation to create wage record information systems, but to the enactment of laws making it illegal to collect social security numbers for purposes other than administering the Social Security Act. [9]

Cost Burdens on Unemployment Insurance Systems. In addition to these confidentially issues, SESA agencies should not be forced to bear any major new cost burdens in providing wage record data to other agencies that might hamper their ability to accomplish their primary mission of helping the unemployed. Large growth in the number of requests for wage record data from enforcement agencies, state user agencies, and the creation of a NWRD, could impose significant new cost burdens on SESAs. At a minimum, policies may need to firmly establish that SESA agencies are fully compensated for the costs of any services they provide to state user agencies or any national system of wage records.

Need for Congressional Review. The submission of BLS's plan for a NWRD will bring many of these issues to a head, and create a need for a careful review by Congress before embarking on any substantial expansion. The BLS plan is likely to focus mainly on issues concerning the organization, development, and costs of the NWRD. At least as much emphasis in any congressional review should be placed on issues of supporting the state use of data from wage record information systems and the quality of the studies that are being done. Issues of the quality of information and how this quality is related to the consequences of use have not received the systematic attention that they deserve. The potential advantages of wage record

[9] Private communication with Jay Pfeiffer, State Department of Education, Tallahassee, Florida.

information systems have been made clear enough, but the potential pitfalls have not. Sufficient experience with wage record information systems exists in the states and the research community to determine some of the critical issues of quality and their consequences for use, and how they might be dealt with.

EXPERIENCE WITH WAGE RECORD INFORMATION SYSTEMS

States and the federal government have acquired considerable experience with wage record information systems in the past five to ten years. One indication of this is the 50 percent growth that occurred in the total number of data sharing agreements at the state level between 1986 and 1991, as reported by the National Commission on Employment Policy. [10] By 1991, data sharing agreements had been reached in 43 states in the area of Child Support Enforcement programs, in 28 states for the JOBS portion of AFDC, and in 26 states in Food Stamps. The average number of data sharing agreements per state in 1991 was 17, and the range was between one and 117 per state.

One of the important reasons for this growth is undoubtedly the passage of the 1988 Amendments to the Deficit Reduction Act. Under these amendments, states were required for the first time to collect quarterly wage reports from all employers as part of a new Income Verification and Eligibility System.[11] Prior to this legislation, many states monitored eligibility for unemployment compensation using a "wage request" approach, whereby employers were only required to submit wage information when it was necessary to process a claim. [12]

Several persons interviewed in preparing this paper stated that state interest in using wage record data for purposes of performance management and policy research has continued to grow, and seems to have accelerated since 1991.

Use ***of Wage Records in JTPA. The*** program area where the most experience with wage records lies is the JTPA. In the NCEP study referred to above, 29 of the 47 states responding to a survey reported having at least one data-sharing agreement in JTPA. [13] In 25 of these states the data were being used for "program accountability purposes. " In the four other states, the data were being used for purposes of verifying eligibility for JTPA or the results of followup surveys. Sixteen state JTPA agencies are now conducting demonstration projects with support from the Office of Strategic Planning and Policy Development (OSPPD) in the Employment and Training Administration of the U.S. Department of Labor, to assess the likely effects on provider agencies and clients of switching from questionnaire surveys to wage records under JTPA performance standards. A report will be released in fall 1994.

A Leading State User Agency. The most advanced state-level user agency in the country is generally considered to be Florida's Education and Training Placement Information Program (FETPIP). This program started out as a project of the state legislature. It was originally located in the State Job Training Coordinating Council (SJTCC), but was funded from educational resources. Initially reports were issued on the employment and educational outcomes of vocational education programs. After about 4 years of development the project was moved to the Department of Education. In the program, earnings records obtained from the state's unemployment insurance system are combined

10 NCEP, op. cit., p. 58.

11 Jarosik, Daniel, and Alan Phelps, "Empowering Accountability for Vocational-Technical Education: The Analysis and Use of Wage Records," MDS-244, Berkeley, CA: National Center for Research in Vocational Education, November 1992, p. 1. Employers' reports are required to include gross earnings by Social Security number, industry of employment by Standard Industrial Classification code, and county of employment.

12 Baj, op. cit., p. i.

13 NCEP, op. cit., p. 59.

with data from the state's community college and university systems, the public schools, the state corrections system, and the federal government, including the Department of Defense. Employers are also surveyed to determine the occupation of all students tracked in the system. This provides a common set of educational enrollment and employment outcomes for measuring the results of education, training, and employment programs in the state. The outcomes include quarterly earnings, weeks worked, occupation, and further enrollment in either the adult education, community college, or higher education systems in the state. The employment and educational activities of a total of 1.8 million people are currently being tracked.

A number of statewide accountability reports required by the state legislature are regularly produced using data from FETPIP. One report shows the employment and educational outcomes of all secondary and postsecondary vocational programs in the state by program and institution. New performance measures for vocational education will include the placement rate, a productivity rate, and two earnings measures. Each measure will include rates for special populations. FETPIP has emphasized making these reports and files of all the outcome data in the system available to local institutions to use for their own planning and other purposes. Some institutions are beginning to ask for data from the system and request information that goes beyond what is required in the accountability reports, such as longitudinal data on the outcomes of programs. By law, all secondary and postsecondary vocational programs in the state must achieve a placement rate of 70 percent. Recently, a "hot ticket" bill has been passed by the state legislative to build a system of incentive bonus funds for completions and placements in certain occupational areas. Additional bonuses will be provided for enrolling, completing, and placing certain targeted groups of people. [14]

Over the years, FETPIP has worked to steadily broaden the number of program areas where participants are being tracked, and is becoming a factor in statewide processes of strategic planning and budgeting for education and training programs. These procedures require "conferences" between an educational sector (e.g., the community colleges) and the executive and legislative branches of government to agree on estimates of revenues and enrollments prior to the process of determining a budget for the sector in the state legislature. Recently, the purpose of these conferences has been expanded to include occupational forecasting. Data from FETPIP are combined with data from the state's Bureau of Labor Market Information to support the discussions of the conferees.

A second process of "occupational forecasting" for program planning has also been established in Florida, where the governor, the commissioner of education, the secretaries of labor and commerce, and the legislature must periodically come together and agree on the 30 or so occupations with the best "potential for employ merit." Data from FETPIP provides the basis for these discussions and agreement among the parties on the occupations. States funds are then directed to those occupational areas and away from others. This process will also be repeated in 28 regions of vocational education in the state.

FETPIP is now providing outcome information to the JOBS program and is working with the SJTCC to determine how followup could be provided for JTPA. JTPA is currently comparing the wage record data they are getting from FETPIP with their own information from questionnaire surveys as part of the OSPPD project mentioned above. In addition, employment and further education reports are provided to all the public schools in the state, the university system, the Adult Migrant Education Program, the General Educational Development (GED) program, an Exceptional Students Program, a study of Associate of Arts degree holders, and a study of training and recidivism for the prison system.

14 private communication with Jay Pfeiffer.

Other Experience With Wage Records in Vocational Education. There is other state experience with wage record information systems in vocational education, but generally not as much as in JTPA. According to a study conducted by the National Governors' Association (NGA) for NOICC, 20 state vocational education agencies use or have explored the use of wage record data for various purposes. [15] In 1992, there was "sustained" use of the wage record data in 12 of these states, "planned use" in two states, and a "short-term effort" in the remaining six states. A recent survey of state plans for implementing performance standards under the 1990 Amendments to the Perkins Act shows that only one state (Illinois) is planning to utilize wage record information. [16]

NGA found that in nearly all cases where wage record information is being used in vocational education, there is a history of prior use by a JTPA agency. [17]

One of the other states where wage records have been used in vocational education is Washington. Staff in the Washington State Board for Community College Education have combined wage record data from five states (Washington, Alaska, California, Idaho, and Oregon) with enrollment data from the state university system and employment information from the Department of Defense. Eighteen data elements describing the background characteristics (e.g., age and sex), vocational program completed, and employment histories of approximately 15,000 former students have been tracked for several years. [18]

One of the major problems in developing wage record information systems in vocational education is defining who is a vocational student. [19] Students may take one course or complete an entire program of studies. Programs of study differ among institutions even within a particular occupational area. Without consistency among vocational programs and institutions, there is no basis for making comparisons among vocational programs or between vocational education and other programs. Substantial efforts within states and among regions will be needed to reach agreements on common definitions of vocational programs before unemployment insurance wage record data can be very useful for evaluating vocational programs or conducting policy-oriented research. 20

Another serious problem in a substantial number of states is lack of access to the Social Security numbers of vocational students. Some states tie vocational education funding to the collection of Social Security numbers from students, while in others obtaining the social security numbers of students at the state level for any reason is considered an invasion of privacy. This is especially true at the secondary level. Only 15 of the 32 state vocational programs at the secondary level that responded to NGA's survey of wage record use in vocational education are currently collecting Social Security numbers from students. Seven of the 13 state postsecondary agencies that responded to the survey do so. [21] Overcoming this problem would be very difficult.

Costs of Development and Operation. States have also gained experience with the costs of developing and operating wage record information systems. Estimates for the annual

15 Amico, op. cit., table 2.

16 Rahn, Mikala, E. Gareth Hoachlander, and Karen Levesque, "State Systems for Accountability in Vocational Education," MDS-491, National Center for Research in Vocational Education, December 1992, appendix. Recently, a decision has also been made in Florida to utilize wage record information in the implementation of performance standards.

17 Amico, p. vii.

18 Profiles for Washington and seven additional states can be found in Jarosik, op. cit., pp. 22-26.

19 Stevens, David, Peggy Richmond, Joseph F. Haenn, and Joan S. Michie, "Measuring Employment Outcomes Using Unemployment Insurance Wage Records," prepared for the Office of Planning and Evaluation by Research and Evaluation Associates, Inc., Washington DC: U.S. Department of Education, December 1992, p. 100; and Jarosik, op. cit., p. 71.

20 Ibid., p. 83.

21 Amico, op. cit., table 5; and Jarosik, op. cit., p. 82.

costs of operation are in the range of one to ten dollars per person tracked, but there is a great deal of uncertainty in these numbers and a lack of documentation concerning what costs are included. [22] NCEP estimates the average operating cost in JTPA to be $1.75,23 while Stevens estimates it [24] to be $3.00 per person in vocational education. Sometimes only the costs of paying the SESA agency for computer runs to match a list of Social Security numbers with wage records are included, not the costs of staffing the information system, preparing lists of program participants, conducting analyses, and providing services. Startup costs are likely to be higher than long run average costs. One study of JTPA estimates startup costs of $20,000 per state and $2.00 per program participant for list preparation and paying the SESA agency for record matching. [25] Only $.20 of this was for the actual costs of matching. No estimates of how the annual costs of data collection compare with the total costs of the reporting system were provided.

On important issue is how the costs of obtaining earnings and employment data from wage records compares with collecting the same data using questionnaire survey or other methods. The NCEP study is one of the few sources. NCEP estimates that the annual costs of obtaining program outcomes information for JTPA performance standards using a telephone survey is $19.00 per completed interview. [26] Whether the questionnaire methods also provided data on outcomes other than quarterly earnings or other information was not revealed.

One good benchmark for the longer term total costs of operating a wage record information system is provided by FETPIP. FETPIP's total annual budget for collecting data, preparing reports, and providing user services is currently **$310,000** for the total of 1.8 million people being tracked, or about 16 cents per person. The main product of the program has been the annual report to the state legislature on the employment outcomes of vocational education, but now several other reports are produced for the new client agencies. The vocational education report covers about 200,000 secondary and postsecondary students[27] and was produced when the budget for the program was somewhat higher than it is now. The recent increase in the total number of program participants being tracked from 200,000 to 1.8 million was therefore accomplished with no increase in the total costs of the program. This indicates large economies of scale in the development and operation of wage record information systems. [28]

The cost advantages of wage record information systems compared to surveys may be especially great for time periods long past the end of participation in a program. These lower costs are inherent in the nearly universal coverage of the wage record data and the quarterly requirement for data submission by employers. In surveys, finding a high proportion of the original group of respondents usually becomes extremely difficult as the length of time since the last interview or after leaving a program increases. In experimenting with the collection of followup data 18 months after program completion, it has been found in JTPA that usable data cannot be obtained with surveys because of low response rates. [29] NCEP estimated the ongoing costs of a wage record information system for JTPA as one-fifth the cost using surveys. [30]

In conclusion, the costs of tracking the employment and earnings of program participants appear to be significantly lower using wage records than with survey or other methods; how-

22 Private conversation with staff of the National Governors' Association, Washington, DC.
23 NCEP, Op. cit., p. 65.
24 Stevens, Op. cit, p. vii.
25 NCEP, p. 65.
26 NCEP, op. cit., p. 2.
27 Jarosik, op. cit., table 1.
28 See also Baj, op. cit., figure 5-1.
29 Private conversation with Karen Greene, Office of Strategic Planning and Policy Development, U.S. Department of Labor, Washington, DC.
30 NCEP, op. cit., p. 43.

ever, the size of the cost advantage is very uncertain because of the lack of documentation in existing studies. The greatest cost advantages are likely to be in situations where long term followups or longitudinal outcome data are needed.

QUALITY OF STUDIES DONE WITH WAGE RECORDS

While the examples above indicate the potential value and operating costs of wage record information systems, they do not show the factors affecting the quality of studies. The fact that wage record data come from administrative sources can make them much more difficult to use and interpret than the results of surveys, where the categories of response are under the control of the analyst and samples of respondents can be randomly selected.

General Problems With Data From Administrative Systems. Data from administrative sources can be subject to several kinds of systematic errors. One is when incentives may be operating that cause respondents to report differently to the administrative system than they otherwise would, in order to minimize or avoid certain consequences. In the case of wage record systems, these incentives exist because of the tax liabilities of employers involved in unemployment compensation. For example, some employers may classify themselves in reporting as belonging to industries or jurisdictions with the lowest tax rates, irrespective of where the employees whose wages are being reported actually work. [31] Also, some groups of employers may not report non-wage payments to workers as accurately as they do wages and salaries. If so, the wage record data for people who go into industries or occupations where earnings are systematically under-reported will be biased downward.

A second reason is that policies may change in the administrative system so that what was originally included in reports may no longer be included, and vice versa. If analysts fail to keep up the changes, the comparison of outcomes at one point in time to another point in time using data could be too high or too low. In a highly decentralized system such as unemployment insurance, keeping up with all the changes that occur and figuring out whether they are important or not for the purpose of analysis could involve a great deal of effort.

A third problem can be that details of the reporting system that are not known to the analyst may affect the results of analyses. For example, in studies of displaced workers earnings estimated from wage record data can be distorted because employers typically report severance payments after the termination of employment. [32]

None of these problems are fatal, but they do indicate why caution is warranted and substantial efforts are needed to understand the specific factors affecting the quality of studies done with wage record data.

Overall Matching Rate. One of the most significant variables affecting the quality of studies is the number of sources from which data are obtained on the employment and other activities of individuals. The choices made about the number of sources can greatly affect the *overall matching rate* achieved. This matching rate is the percentage of all persons being tracked for whom some positive identification of employment, unemployment, or other activity appears in the outcome data. [33] This rate is analogous to the response rate in questionnaire surveys and is the linch pin of wage record information systems.

31 Jarosik, op. Cit., p. 82.

32 Decker, Paul T., "Systematic Bias in Earnings Data Derived From Unemployment Insurance Wage Records and Implications for Evaluation the Impact of Unemployment Insurance Policy on Earnings," Unpublished paper, Princeton, NJ: Mathematical Policy Research, April 1989, p. 5.

33 In general, the overall matching rate could be obtained by averaging the quarterly matching rates over the time individuals are traced in wage record information systems.

Generally, the higher the overall matching rate, the more confidence that can be placed in the results of studies. The administrative origins of the data and the non-random processes by which individuals are selected into programs are two reasons for this. For both reasons, the results of studies done with wage record data are vulnerable to being systematically biased *too high or too low* compared to the actual values. Short of having done the studies necessary to determine the consequences of low matching rates, the only way of guarding against such problems in wage record data is having positive identification of all the activities of the complete population of interest.

At present, there are no firm rules for deciding what the overall matching rate needs to be in what circumstances to have confidence in the results of analyses. Accumulating this knowledge will require: (a) time, (b) substantial efforts to summarize the experience of state agencies with different configurations of data and purposes, and (c) specific studies comparing the results of analyses using wage record information with data from other sources.

Adequate levels of the matching rate using *wage records alone* may be very different among population groups for several different reasons. Three of the most important are differences in the rates of unemployment among the groups, where they tend to be employed and how they earn their income, and their likelihood of engaging in other activities, such as education.

Rates of unemployment are important because positive identification of whether a person is actually not employed is generally not possible from the data that are available from state unemployment insurance systems. The reason for this is that only about one third of all people who are unemployed (according to the standard definitions of unemployment) receive compensation from the unemployment system.[34] People who receive compensation from the system can be positively identified, but those who do not cannot be.

Among JTPA participants, it has generally found that matching rates of 70 percent or so can be obtained using the wage records from only one state, while for vocational education graduates the rate is typically somewhat lower--about 60 percent. The primary reason for this is that more vocational education graduates continue on for additional education than do JTPA completers.[35] For populations of former prisoners, the matching rates in Florida have been in the neighborhood of 30 percent.

The basic way of increasing the matching rate for any population group is to increase the number of sources from which outcome data are obtained. In one study, Stevens has shown the possibility of increasing the matching rate for vocational program participants to 90 percent or more by obtaining wage records from four neighboring states, the home state's state university system, the home state's community college system, and the federal government.[36] Many analysts would argue that for most populations, overall matching rates of 90 percent or more are needed for studies done with wage record data.

The main criterion for deciding how high the matching rate should be is how it affects the results of studies. If the results change significantly when the matching rate is increased, then the first matching rate is too low. In one study of 14 community colleges in two states, the observed rates of employment changed by 20 percent by state depending upon whether the wage record data came from only the state where each of the colleges was located, or the state and three other neighboring states.[37] A performance funding formula based on a wage record data

[34] Private conversation with Paula Duggan, Northeast Midwest Institute, Washington, DC.; and Lynne Webb, unemployment Insurance Service, U.S. Department of Labor, Washington, DC.

[35] Private conversation with Karen Levesque of MPR Associates, Berkeley, California.

[36] Stevens, op. cit., chapter 6.

[37] Ibid., p. 103 and 106.

system using data from only one state could therefore be very unstable.

Stevens showed in another study that the employment rate for the former students of one community college was 64 percent compared to 42 percent for another community college when wage record data only from the state in which the colleges were located was utilized. On basis of this information, institution "b" was performing much more poorly than institution "a." However, when wage records were also obtained from adjacent states, as well from the federal government and the military, the positions of these two institutions changed completely. The employment rate for institution "b" increased to 76 percent, while the rate for institution "a" increased to only 66 percent. On this basis, institution "b" appears to be a considerably better place to go than institution "a." This provides a clear example of how the matching rate can be related to the accuracy of study results.

Variations like this are caused by the highly localized patterns of migration and labor market mobility that exist among localities, schools, industries, and occupations. These localized patterns can strongly affect the results of analyses when comparisons are made among specific institutions or programs rather than among statistically selected samples of individuals or institutions covering much larger areas. The analyses done by Stevens did not include any data on subsequent enrollments in postsecondary education or training programs of any kind, which could have changed the picture even more.[38] His comparisons also included no adjustment for economic conditions in the counties surrounding the two institutions, which could also have recast the relative performance of the two institutions.

Differences in Earnings Among Quarters. Another factor that affects the results of these studies is the quarter selected as the basis for computing the earnings of individuals. In order to simplify data processing, some states have elected to compute annualized earnings by multiplying the earnings from one quarter by four. Stevens has shown that annual earnings computed from earnings in the first quarter of the year will overstate actual earnings by an average of $2,000 per year, or 10 percent of total earnings.[39] Annualized earnings based on earnings in the fourth quarter are nearly correct. This means that in systems of performance management where earnings are estimated from one quarter's worth of earnings, data will tend to punish programs with larger proportions of their graduates in the fourth quarter compared to the first quarter. Estimating earnings from four quarters of data or compensating for the quarter used could result in a different picture.

Demographic Variables. Another potential problem in using wage record data is lack of means for identifying comparison groups and controlling for program effects due to differences in the background characteristics of individuals (e.g., age, sex, race, and social status). Wage record data by themselves do not contain information about gender, race, social background, age, or years of experience in the labor market, which are clearly related to employment and earnings. Identification of characteristics such as these must come from data on program participation. However, this would tend to limit wage record studies to populations of individuals for whom there is data from program participation. The problem is that without good information about the composition of study populations, observed trends of employment and earnings in the labor market may be due more to changes in the composition of the labor force than to any other factors. Likewise, differences in the outcomes of programs may be due more to differences in the populations being served than the effectiveness of the *programs.* Particularly where wage record data are being used in strategies of performance management, not knowing the composition of populations could result in the creation of large incentives for creaming. In many situations of

38 Ibid., figure 39.

39 Ibid., figure 14.

performance management, these characteristics may not be known.

Full-time and Part-time Employment. Another source of error in wage record studies is lack of means for distinguishing between individuals who are employed full-time and part-time. The wage records for the graduates of individual programs typically will include a mix of people who worked only a few weeks within a quarter (and often only 1 or 2 weeks) compared to others who worked the full 13 weeks. As a result, there can be large differences in the estimates of average earnings from one quarter to the next for a given program due to *very small* fluctuations in the proportions of individuals who are working part-time and full-time. In labor market studies, this problem is normally handled by restricting the data used to compute earnings only to fill-time workers. This problem has occurred in FETPIP. In order to obtain stable estimates of earnings, FETPIP's policy is to include only full-time workers, and only workers who are making more than the minimum wage fill-time. Florida is able to do this because of the fact that weeks worked is a data element in their unemployment insurance system. Only seven other states collect this information, so they are the only ones currently able to reduce this large source of variability in earnings estimates in this way. 40

Accuracy of Wage Records Compared to Surveys. While many people have asserted that wage records are "more accurate" than survey data, there have been few efforts to rigorously compare the results of studies using the two different sources. One study where this comparison has been made is the New Jersey Unemployment Insurance Re-employment Demonstration Project. This research and demonstration project is testing whether displaced workers can be successfully identified early in their unemployment spells and provided with alternative forms of job search assistance and training to accelerate their return to work. Comparable data were collected on quarterly earnings from interviews and the New Jersey unemployment insurance system. The study showed that estimates of the subsequent quarterly earnings of displaced workers were $344 dollars *higher* in the first quarter after becoming unemployed using data from the unemployment insurance than they were from interviews, $130 *lower* in the second quarter, and $374 *lower* in the fourth quarter. These figures were about 10 percent of quarterly earnings. Investigation of the reasons for these differences showed that the higher figures from the unemployment insurance data in the first quarter were due to severance payments received after first becoming unemployed, while the lower figures in the later quarters were due to growing numbers of the study population holding jobs out-of-state. Dropping the people who moved out of state from the study population and subtracting the severance payments eliminates all differences in the estimates of earnings between the interview and wage record data.

One other direct comparison of the accuracy of unemployment insurance wage record data with data from other sources is the National Study of JTPA. In this study, estimates of the earnings of the low income individuals studied differed greatly between wage record information and survey data. Among the four population groups considered, the average earnings of individuals were 35 to 80 percent higher using the survey data compared to the wage record data. No attempt was made to obtain out of state wage records but the employment rate for JTPA sites near state borders were not systematically different from other sites. Other possible sources of the discrepancies were also investigated but no conclusions could be drawn about the reasons for the observed differences. [41]

40 NCEP, op. cit., table 1. Hours worked are collected in one other state.

41 U.S. Department of Labor, Employment and Training Administration, "Comparison of Earnings, Employment, and Impact Estimates Based on Data From the First Follow-Up Survey and From State Unemployment Insurance Agencies." In *The National JTPA Study: Title II-A Impacts on Earnings and Employment at 18 Months,* Research Report

Differences Among Population Groups. 130th of these studies may be correct--that is, for displaced workers wage records may be accurate, while for low income, disadvantaged populations they may not be. One possible reason for the differences found in these two studies is that low income, disadvantaged people apparently have more earnings from sources that are not reported to the unemployment insurance system than do displaced workers. These earnings could be from casual work, the underground economy, or tips from reported jobs. [42] The results of these two studies point out that systematic differences in the accuracy of wage records may exist among different population groups. This provides a clear example of the importance of avoiding statements that wage record data are more or less accurate than data from other sources, until more studies can be done to find out what the differences are and how the data can be responsibly used.

It is important to point out that both studies described above were major research efforts in which substantial investments were made in questionnaire development and the careful collection data from respondents. Comparisons of the accuracy of wage record data with the collection of earnings data obtained from interviews conducted in the context of state or local program operations could result in very different conclusions.

Accuracy of SESA Recordkeeping. Finally, another potential source of error in wage record studies is inaccuracy in the actual wage records collected and kept by SESA agencies. The extent of error is currently not known because no large-scale audits of the wage records in state systems have ever been done. The first major audit of wage records is being carried out by the BLS as part of their planning effort. BLS is concentrating their audit on the accuracy of the Social Security numbers and SIC codes in the wage records of all SESA agencies in the Philadelphia region of the Department of Labor. Names and Social Security numbers will be checked by the Social Security Administration. The Social Security number error rates for the first two states checked are seven percent each.[43]

A problem for BLS in developing a NWRD is that the quality of wage record data may vary significantly among the states. This means that the inclusion of any states in the NWRD without carefully auditing their entire system of wage records could contaminate the national data.

Conclusion. More experience with wage record systems and studies of that experience is urgently needed so that guidelines and rules of thumb can be developed regarding the design of studies and wage record systems in which trust can be placed in the results. High quality research on the factors affecting the validity of studies done with wage record data should have a very high priority in future developments.

Studies of the relationships between data quality and system design are just as important for consumer information and local uses of wage record information as they are for purposes of state level program evaluation and national policy. Even for local uses, bad data may be worse than no data at all.

POSSIBILITIES FOR THE DEVELOPMENT OF WAGE RECORD INFORMATION SYSTEMS

The submission of BLS's plan for the creation of a NWRD in June 1994 will set the stage for considering what additional steps, if any, could be taken to encourage or support state capabilities for using wage record information for policy research, program evaluation, or consumer information. The BLS plan is important because it could become a driving force in the development of wage record information systems in the states.

Series, *93-C,* Washington, DC: U.S. Superintendent of Documents, 1993, Appendix E.

42 Ibid., p. 346.

43 Private conversation with Brian MacDonald.

Design choices made by BLS could affect the usefulness of the NWRD and incentives for improving the quality and usefulness of studies done at the state level with wage record information.

. For example, one major choice facing BLS is whether a NWRD should be based on only a sample of individuals in the working population of the U.S. or the entire universe of people covered by unemployment compensation. The sample could be defined for the U.S. as a whole, regions of the United States, within states, by industry, or in many different ways.

. A national database consisting of a sample of wage records could be perfectly adequate for purposes of nationally oriented policy-oriented research on issues of trends in the labor market and broadly defined program participation, but most likely would not be very useful for purposes of program evaluation and monitoring at the state level. Most program evaluation at the state level is likely to involve comparison among specific institutions and programs within those institutions. A national database containing only a sample of individuals, even if the sample is selected within states, would not provide sufficient numbers of individuals for many institutions and programs within a state to be able to allow the conduct of such program evaluation studies or to monitor the outcomes of programs.

. On the other hand, creating a national wage record database containing employment and earnings records for all workers in the U.S. will be an expensive proposition for the states and the federal government, and require dealing effectively with important problems of maintaining confidentiality and privacy. At the same time, the availability of wage record data from a national database could eliminate many of the needs for negotiating state-by-state data sharing agreements, which could significantly reduce their costs of constructing user systems. These reduced costs of obtaining multi-state data could help to stimulate both major expansion of user systems and improvements in the quality of studies.

Option #1: The development of wage record information systems has reached the point where many states have experimented with user systems and have actually begun to use or are poised to begin using wage record systems for purposes of program evaluation and policy analysis. (There are few examples of consumer information so far.) This progress coupled with the potential impetus for expansion following from BLS's submission of a plan for a national database calls for **a deliberate review by the Congress and the consideration of other steps that may be needed to improve the quality and usefulness of studies done at the state level using wage record data and uses of the data for purposes of performance management. These** other steps are important to consider even if plans for the NWRD do not go forward.

Option #2. **The broad review that is needed should include but go beyond questions of the NWRD to issues of how the federal government can be most helpful to state user agencies in all program areas.** Many states and state user agencies have invested substantial amounts of time, effort, and dollars in data-sharing agreements and other elements of capacity for using wage records that could be greatly helped or seriously impeded by new federal initiatives. The review should seek input from the states and federal program offices in all areas, as well as from individuals with the background and experience necessary to raise issues of improving the quality and usefulness of wage record information systems.

Useful advice might be obtained from the:

- *National Occupational Information Coordinating Committee,* which provides a

forum for 10 federal agencies on issues of occupational information. Many of these agencies have strong interests in program followup and career development information. NOICC is also responsible for coordinating the State Occupational Information Coordinating Committee (SOICC) network of state occupational information systems.

. *National Commission on Employment Policy,* which has important responsibilities for helping to coordinate and support the development of national policy on employment and training across federal agencies.

. *Office of Vocational and Adult Education,* which is the federal agency responsible for the implementation of performance standards under the Carl D. Perkins Vocational and Applied Technology Education Act.

• *National Center for Research in Vocational Education,* which is funded under the Perkins Act.

Other organizations in other program areas that could be asked to comment on the plan include: (a) program offices of the federal government in areas other than vocational education, (b) a selection of state user agencies, and (c) the National Governor's Association.

The reviewers should be asked to comment on:

. The implication of the BLS design and plan for a NWRD for the development of user systems at the state level for purposes of: (a) conducting policy-oriented research, (b) performance management, and/or (c) providing consumer information.

. Steps that should be taken beyond the BLS plan to strengthen the capacity of *state user agencies* for organizing wage record information systems and using the data well.

. Actions to improve the quality of studies done with wage record information by state user agencies (and researchers).

. Specific steps to encourage the use of wage record information systems at the *local* level by policy makers in vocational education and other program areas for understanding how programs are working, and developing their own policies and strategies.

BLS is also likely to propose strategies for providing technical assistance to the state offices where access to the NWRD will be provided, if a decision to proceed with such a national database is made.

. The reviewers should be asked specifically to comment on the needs for technical assistance to state user agencies for purposes of policy making or performance management, and improving the quality of studies. [One possibility is that this technical assistance could be provided through federal support for the sharing of information among the states themselves rather than through more direct efforts.]

Option #3: A small **program of cooperative demonstration projects could be supported to encourage the development of regional consortia on using wage record information for policy development, performance management, and/or consumer information.**

. The projects should concentrate on the development of data sharing agreements, the investigation of issues of data quality, demonstration of how program evaluation results can actually be used to improve programs at the local level, and demonstration of the kinds of broader questions of program strategy and dynamics of the labor market that can be

effectively addressed using wage record data.

. The project should also be encouraged to experiment with ways of supporting the local use of wage record information.

. Each consortium should be required to produce a report summarizing the results of their efforts that would be publicly available and distributed.

. Funding in the neighborhood of $100,000 apiece should be provided for these consortia.

The clear purpose of this cooperative program should be to foster the regional development of state-level agencies as effective users of wage record information.

Option #4: **Federal agencies could be directed to conduct systematic research on the sound use of wage record information in policy studies and systems of performance management.**

. More studies like the current 16-state demonstration in JTPA funded by the Employment and Training Administration to compare the impacts of switching to wage records from survey methods are needed.

. A similar demonstration and systematic analysis of the use of wage record information for performance management in comparison to followup surveys is needed.

. As in the New Jersey Unemployment Insurance Re-employment Demonstration Project, federal studies should be encouraged to utilize wage record information in major evaluations of state or federal programs where wage records can be rigorously compared with other sources of earnings and employment information.

OTHER REFERENCES

Bishop, John, "Policy Evaluation and Archived Wage Record Data: Limitations of Existing Data Sets, " Washington, DC: Northeast-Midwest Institute, 1989.

Crosslin, Robert L., "The Feasibility of a National Wage Record Data Base: A Conference Synopsis and Proposed Next Steps, " Washington, DC: Northeast-Midwest Institute, 1989.

King, Christopher T., "Employment and Earnings Dynamics in the 1990s, " Policy Issue Needs and Potential Uses of Nationally Archived UI Wage Records, " Washington, DC: Northeast-Midwest Institute, 1989.

Office of the Governor, State of Florida, "Occupational Identifier Project: Legislative Report," State Job Training Coordinating Council and Council on Vocational Education, Apr. 29, 1988.

Pfeiffer, Jay, "Student Followup Using Automated Record Linkage Techniques: Lessons From Florida's Education and Training Placement Information Program (FETPIP), " unpublished paper, Tallahassee, FL: State Department of Education.

Stevens, David, "The School to Work Transition of High School and Community College Vocational Program Completers: 1990-1992, " draft report prepared for the National Assessment of Vocational Education, Baltimore, MD: University of Baltimore, Nov. 30, 1993.

U.S. Department of Labor, "Exploring Databases for Performance Management and Evaluation, " *Evaluation Forum,* Issue 9, summer 1993, pp. 55-59.